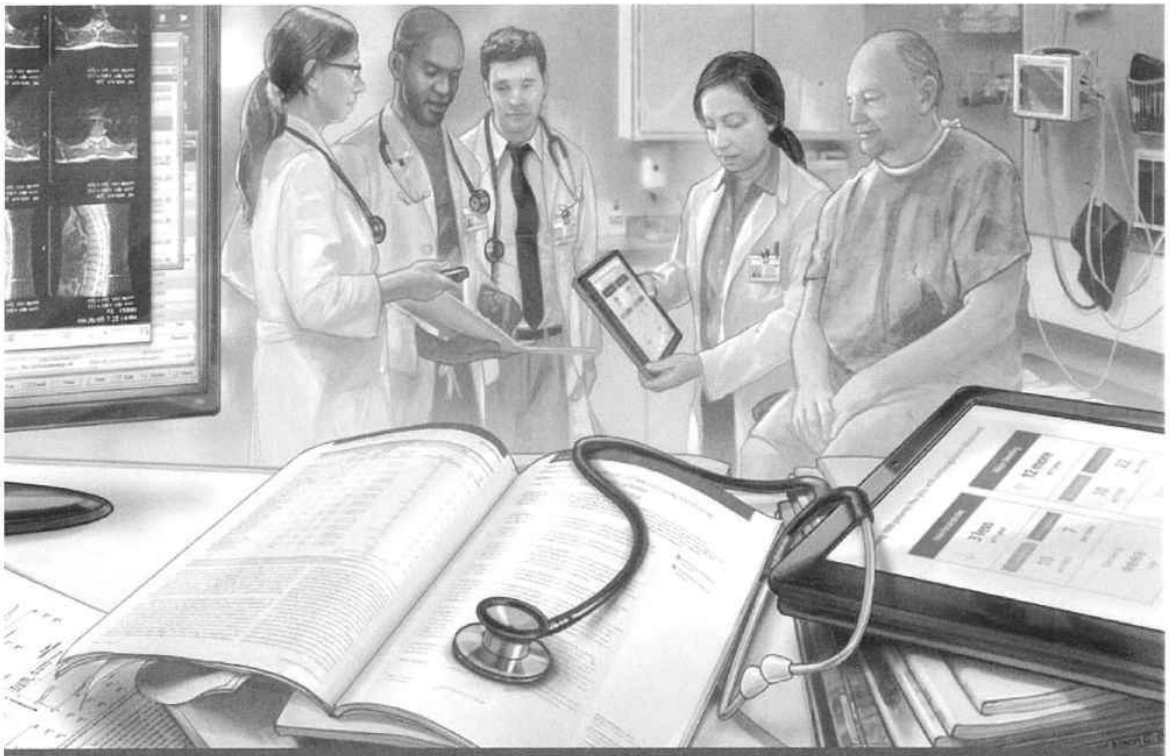
EXHIBIT 45



3rd EDITION

Users' Guides to the Medical Literature

ESSENTIALS OF EVIDENCE-BASED CLINICAL PRACTICE

Gordon Guyatt, MD
Drummond Rennie, MD
Maureen O. Meade, MD
Deborah J. Cook, MD

Mc Graw Hill Education

JAMAevidence®

Users' Guides to the Medical Literature

Notice

Medicine is an ever-changing science. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy are required. The authors and the publisher of this work have checked with sources believed to be reliable in their efforts to provide information that is complete and generally in accord with the standards accepted at the time of publication. However, in view of the possibility of human error or changes in medical sciences, neither the authors nor the publisher nor any other party who has been involved in the preparation or publication of this work warrants that the information contained herein is in every respect accurate or complete, and they disclaim all responsibility for any errors or omissions or for the results obtained from use of the information contained in this work. Readers are encouraged to confirm the information contained herein with other sources. For example and in particular, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this work is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs.

Users' Guides to the Medical Literature

ESSENTIALS OF EVIDENCE-BASED CLINICAL PRACTICE

3rd EDITION

Editors

Gordon Guyatt, MD, MSc
Departments of Clinical Epidemiology & Biostatistics and Medicine
Faculty of Health Sciences
McMaster University
Hamilton, Ontario, Canada

Drummond Rennie, MD
Former Deputy Editor, *JAMA*
Chicago, Illinois
Philip R. Lee Institute for Health Policy Studies
University of California, San Francisco
San Francisco, California, USA

Maureen O. Meade, MD, FRCPC, MSc
Departments of Medicine and Clinical Epidemiology & Biostatistics
Faculty of Health Sciences
McMaster University
Hamilton, Ontario, Canada

Deborah J. Cook, MD, MSc
Departments of Medicine and Clinical Epidemiology & Biostatistics
Faculty of Health Sciences
McMaster University
Hamilton, Ontario, Canada

Mc Graw Hill Education

New York Chicago San Francisco Athens
London Madrid Mexico City Milan
New Delhi Singapore Sydney Toronto

JN
The JAMA Network

JAMA evidence
Using Evidence to Improve Care

ISBN: 978-0-07-180872-9

MHID: 0-07-180872-8

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-179415-2, MHID: 0-07-183314-5.

eBook conversion by codeMantra
Version 1.0

To our students, in many countries, whose interest, passion, and probing questions made possible the development of the methods we use to communicate the concepts of evidence-based medicine.

GG, MOM, and DJC

To Deb, who has watched over and tended me while I have watched over and tended this wonderful group, with gratitude for her love and her good humor.

DR

This page intentionally left blank

CONTENTS

Prognosis

Summarizing the Evidence

Moving From Evidence to Action

This page intentionally left blank

CONTRIBUTORS

Thomas Agoritsas, MD, Dr Med
Health Information Research Unit
Department of Clinical Epidemiology & Biostatistics
McMaster University
Hamilton, Ontario, Canada

Elie A. Akl, MD
Department of Medicine
American University of Beirut
Riad-El-Solh, Beirut, Lebanon

Paul Elias Alexander
Department of Clinical Epidemiology & Biostatistics
Health Research Methodology Graduate Program
McMaster University
Hamilton, Ontario, Canada

Waleed Alhazzani, MD, FRCPC, MSc
Departments of Medicine and Clinical Epidemiology & Biostatistics
McMaster University
Hamilton, Ontario, Canada

Pablo Alonso-Coello, MD
Hospital de la Santa Creu i Sant Pau
Barcelona, Spain

Shannon M. Bates, MDCM, MSc, FRCP(c)
Department of Medicine
McMaster University
Hamilton, Ontario, Canada

Heiner C. Bucher, MPH
Basel Institute for Clinical Epidemiology and Biostatistics
University Hospital Basel
Basel, Switzerland

Alonso Carrasco-Labra, DDS, Msc, PhD(c)
Department of Clinical Epidemiology & Biostatistics
McMaster University
Hamilton, Ontario, Canada
Evidence-Based Dentistry Unit
Universidad de Chile
Santiago, Chile

Deborah J. Cook, MD, FRCPC, MSc, OC
Departments of Clinical Epidemiology & Biostatistics and Medicine
McMaster University
Hamilton, Ontario, Canada

PJ Devereaux, MD, PhD, FRCPC
Departments of Clinical Epidemiology & Biostatistics and Medicine
McMaster University
Hamilton, Ontario, Canada

Glyn Elwyn, MD, MSc, FRCGP, PhD
The Dartmouth Centre for Health Care Delivery Science
Hanover, New Hampshire, USA

Toshi A. Furukawa, MD, PhD
Departments of Health Promotion and Human Behavior and Clinical Epidemiology
Kyoto University Graduate School of Medicine
Kyoto, Japan

Gordon Guyatt, MD, MSc, FRCPC, OC
Departments of Clinical Epidemiology & Biostatistics and Medicine
McMaster University
Hamilton, Ontario, Canada

Alfred Theodore (Ted) Haines, MD, CCFP, MSc, DOHS, FRCPC
Departments of Clinical Epidemiology & Biostatistics and Family Medicine
McMaster University
Chedoke-McMaster Hospitals
LAMP Community Health Centre
Occupational Health Clinic for Ontario Workers
Hamilton, Ontario, Canada

Rose Hatala, MD, MSc
University of British Columbia
Vancouver, British Columbia, Canada

R. Brian Haynes, MD, PhD
Departments of Clinical Epidemiology & Biostatistics and Medicine
McMaster University
Hamilton, Ontario, Canada

Robert Hayward, MD
Owogo Inc.
Centre for Health Evidence
Department of Medicine
University of Alberta
Edmonton, Alberta, Canada

John P. A. Ioannidis, MD, DSc
Departments of Medicine, Health Research and Policy, and Statistics
Stanford Prevention Research Center
Meta-Research Innovation Center
Stanford University
Stanford, California, USA

Cynthia A. Jackevicius, BScPhm, PharmD, MSc, BCPS, FCSHP
Western University of Health Sciences
Pomona, California, USA
Institute for Clinical Evaluative Sciences
Institute for Health Policy, Management and Evaluation
University of Toronto
Toronto, Ontario, Canada
Veterans Affairs Greater Los Angeles Healthcare System
Los Angeles, California, USA

Roman Jaeschke, MD, MSc, FRCPC
Department of Medicine
St. Joseph's Healthcare
Hamilton, Ontario, Canada

Mitchell Levine, MD, MSc
Department of Clinical Epidemiology & Biostatistics
McMaster University
Centre for Evaluation of Medicines
St. Joseph's Healthcare
Hamilton, Ontario, Canada

Braden Manns, MD
Departments of Medicine & Community Health Sciences
University of Calgary
Calgary, Alberta, Canada

K. Ann McKibbon, MLS, PhD, FMLA
Department of Clinical Epidemiology & Biostatistics
McMaster University
Hamilton, Ontario, Canada

Maureen O. Meade, MD, MSc, FRCPC
Departments of Clinical Epidemiology & Biostatistics and Medicine
McMaster University
Hamilton, Ontario, Canada

Edward J. Mills, PhD, MSc, MSt
Global Evaluative Sciences
Vancouver, British Columbia, Canada

Victor M. Montori, MD, MSc
Knowledge and Evaluation Research Unit
Mayo Clinic
Rochester, Minnesota, USA

Sohail M. Mulla, MSc
Department of Clinical Epidemiology & Biostatistics
Health Research Methodology Graduate Program
McMaster University
Hamilton, Ontario, Canada

M. Hassan Murad, MD, MPH
Division of Preventive Medicine
Mayo Clinic
Rochester, Minnesota, USA

Reem A. Mustafa, MD
Department of Medicine
University of Missouri-Kansas City
Overland Park, Kansas, USA

Ignacio Neumann, MD, MSc
Department of Internal Medicine
Pontificia Universidad Católica de Chile
Santiago, Chile
Department of Clinical Epidemiology & Biostatistics
McMaster University
Hamilton, Ontario, Canada

Vlado Perkovic, MBBS, PhD, FASN, FRACP
George Institute for Global Health Australia, Medicine
University of Sydney
Sydney, New South Wales, Australia

Kameshwar Prasad, MD, DM, MMSc
Department of Neurology
Neurosciences Centre

All India Institute of Medical Sciences
New Delhi, India

Milo A. Puhan, MD, PhD
Department of Epidemiology and Public Health
Epidemiology, Biostatistics and Prevention Institute
University of Zurich
Zurich, Switzerland

Adrienne G. Randolph, MD, MSc
Department of Anaesthesia
Harvard Medical School
Department of Anesthesia, Perioperative and Pain Medicine
Boston Children's Hospital
Boston, Massachusetts, USA

W. Scott Richardson, MD
Department of Medicine
Georgia Regents University-University of Georgia Medical Partnership
Athens, Georgia, USA

David M. Rind, MD
Department of Medicine
Harvard Medical School
Editorial and Evidence-Based Medicine, UpToDate
Wolters Kluwer Health
Waltham, Massachusetts, USA

Bram Rochwerg, BSc, MD
Department of Medicine
McMaster University
Hamilton, Ontario, Canada

Nancy Santesso, BSc(Hon), MLIS, PhD(c)
Department of Clinical Epidemiology & Biostatistics
McMaster University,
Hamilton, Ontario, Canada

Holger J. Schünemann, MD, PhD, MSc, FRCPC
Departments of Clinical Epidemiology & Biostatistics and Medicine
McMaster University
Hamilton, Ontario, Canada

Ian A. Scott, MBBS, FRACP, MHA, MEd
Department of Internal Medicine and Clinical Epidemiology
Princess Alexandra Hospital
Department of Medicine
University of Queensland
Brisbane, Queensland, Australia

Sadeesh Srinathan, MD, MSc
University of Manitoba Health Sciences Centre
Winnipeg, Manitoba, Canada

Sharon E. Straus, MSc, MD, FRCPC
Department of Medicine
Division of Geriatric Medicine
University of Toronto
Li Ka Shing Knowledge Institute
St. Michael's Hospital
Toronto, Ontario, Canada

Kristian Thorlund, MSc, PhD
Department of Clinical Epidemiology & Biostatistics
McMaster University
Hamilton, Ontario, Canada

Per Olav Vandvik, MD, PhD
Department of Medicine
University of Oslo
Norwegian Knowledge Centre for the Health Services
Oslo, Norway

Michael Walsh, MD, PhD
Departments of Medicine and Clinical Epidemiology & Biostatistics
Population Health Research Institute
Hamilton Health Sciences and McMaster University
Division of Nephrology
St. Joseph's Hospital
Hamilton, Ontario, Canada

Stephen D. Walter, PhD, FRSC
Department of Clinical Epidemiology & Biostatistics
McMaster University
Hamilton, Ontario, Canada

Mark C. Wilson, MD, MPH
Department of Internal Medicine
Graduate Medical Education
Carver College of Medicine
University of Iowa Hospitals and Clinics
Iowa City, Iowa, USA

Peter Wyer, MD
Columbia University Medical Center
New York, New York, USA

John J. You, MD, MSc
Departments of Medicine and Clinical Epidemiology & Biostatistics
McMaster University
Hamilton, Ontario, Canada

This page intentionally left blank

FOREWORD

When I was attending school in wartime Britain, staples of the curriculum, along with cold baths, mathematics, boiled cabbage, and long cross-country runs, were Latin and French. It was obvious that Latin was a theoretical exercise—the Romans were dead, after all. However, although France was clearly visible just across the Channel, for years it was either occupied or inaccessible, so learning the French language seemed just as impractical and theoretical an exercise. It was unthinkable to me and my teachers that I would ever put it to practical use—that French was a language to be spoken.

This is the relationship too many practitioners have with the medical literature—clearly visible but utterly inaccessible. We recognize that practice should be based on discoveries announced in the medical journals. But we also recognize that every few years the literature doubles in size, and every year we seem to have less time to weigh it,[1] so every day the task of taming the literature becomes more hopeless. The translation of those hundreds of thousands of articles into everyday practice appears to be an obscure task left to others, and as the literature becomes more inaccessible, so does the idea that the literature has any utility for a particular patient become more fanciful.

This book, now in its third edition, is intended to change all that. It is designed to make the clinician fluent in the language of the medical literature in all its forms. To free the clinician from practicing medicine by rote, by guesswork, and by their variably integrated experience. To put a stop to clinicians being ambushed by drug company representatives, or by their patients, telling them of new therapies the clinicians are unable to evaluate. To end their dependence on out-of-date authority. To enable the practitioner to work from the patient and use the literature as a tool to solve the patient's problems. To provide the clinician access to what is relevant and the ability to assess its validity and whether it applies to a specific patient. In other

words, to put the clinician in charge of the single most powerful resource in medicine.

The Users' Guides Series in *JAMA*

I have left it to Gordon Guyatt, MD, MSc, the moving force, principal editor, and most prolific coauthor of the Users' Guides to the Medical Literature series in *JAMA*, to describe the history of this series and of this book in the accompanying preface. But where did *JAMA* come into this story?

In the late 1980s, at the invitation of my friend David Sackett, MD, I visited his department at McMaster University to discuss a venture with *JAMA*—a series that examined the evidence behind the clinical history and examination. After these discussions, a series of articles and systematic reviews was developed and, with the enthusiastic support of then *JAMA* Editor in Chief George Lundberg, MD, *JAMA* began publishing The Rational Clinical Examination series in 1992.[2] By that time, I had formed an excellent working relationship with the brilliant group at McMaster. Like their leader, Sackett, they tended to be iconoclastic, expert at working together and forming alliances with new and talented workers, and intellectually exacting. Like their leader, they delivered on their promises.

So, when I heard that they were thinking of updating the wonderful little series of Readers' Guides published in 1981 in the *Canadian Medical Association Journal* (*CMAJ*), I took advantage of this working relationship to urge them to update and expand the series for *JAMA*. Together with Sackett, and first with Andy Oxman, MD, and then with Gordon Guyatt taking the lead (when Oxman left to take a position in Oslo), the Users' Guides to the Medical Literature series was born. We began publishing articles in the series in *JAMA* in 1993.[3]

At the start, we thought we might have 8 or 10 articles, but the response from readers was so enthusiastic and the variety of types of article in the literature so great that ever since I have found myself receiving, sending for review, and editing new

articles for the series. Just before the first edition of this book was published in 2002, Gordon Guyatt and I closed this series at 25, appearing as 33 separate journal articles.

The passage of years during the preparation of the original *JAMA* series and the publication of the first edition of this book had a particularly useful result. Some subjects that were scarcely discussed in the major medical journals in the early 1990s but that had burgeoned years later could receive the attention that had become their due. For instance, in 2000, *JAMA* published 2 Users' Guides[4,5] on how readers should approach reports of qualitative research in health care. To take another example, systematic reviews and meta-analyses, given a huge boost by the activities of the Cochrane Collaboration, had become prominent features of the literature, and as Gordon Guyatt points out in his preface, the change in emphasis in the Users' Guides to preappraised resources continues.

The Book

From the start, readers kept urging us to put the series together as a book. That had been our intention right from the start, but each new article delayed its implementation. How fortunate! When the original Readers' Guides appeared in the *CMAJ* in 1981, Gordon Guyatt's phrase "evidence-based medicine" had never been coined, and only a tiny proportion of health care workers possessed computers. The Internet did not exist and electronic publication was only a dream. In 1992, the Web—for practical purposes—had scarcely been invented, the dot-com bubble had not appeared, let alone burst, and the health care professions were only beginning to become computer literate. But at the end of the 1990s, when Guyatt and I approached my colleagues at *JAMA* with the idea of publishing not merely the standard printed book but also Web-based and CD-ROM formats of the book, they were immediately receptive. Putting the latter part into practice has been the notable achievement of Rob Hayward, MD, of the Centre for Health Evidence of the University of Alberta.

The science and art of evidence-based medicine, which this book does so much to reinforce, has developed remarkably during the past 25 years, and this is reflected in every page of this book. Encouraged by the immediate success of the first and second editions of the *Users' Guides to the Medical Literature*, Gordon Guyatt and the Evidence-Based Medicine Working Group have once again brought each chapter up to date for this third edition. They have also added 6 completely new chapters: Evidence-Based Medicine and the Theory of Knowledge, How to Use a Noninferiority Trial, How to Use an Article About Quality Improvement, How to Use an Article About Genetic Association, Understanding and Applying the Results of a Systematic Review and Meta-analysis, and Network Meta-analysis. Some of these chapters appear in the larger Manual version of this book.

An updated Web version of the *Users' Guides to the Medical Literature* will accompany the new edition. As part of the online educational resource, JAMAevidence, the *Users' Guides to the Medical Literature* online is intertwined with the online edition of *The Rational Clinical Examination: Evidence-Based Clinical Diagnosis*. Together they serve as the cornerstones of a comprehensive online educational resource for teaching and learning evidence-based medicine. Interactive calculators and worksheets provide practical complements to the content, and downloadable PowerPoint presentations serve as invaluable resources for instructors. Finally, podcast presentations bring the foremost minds behind evidence-based medicine to medical students, residents, and faculty around the world.

Once again, I thank Gordon Guyatt for being an inspired author, a master organizer, and a wonderful teacher, colleague, and friend. I know personally and greatly admire a good number of his colleagues in the Evidence-Based Medicine Working Group, but it would be invidious to name them, given the huge collective effort this has entailed. This is an enterprise that came about only because of the strenuous efforts of many individuals. On the *JAMA* side, I must thank Annette Flanagin, RN, MA, a wonderfully efficient, creative, and diplomatic colleague

at *JAMA*. All of this was coordinated and kept up to schedule by the energy and meticulous efficiency of Kate Pezalla, MA. My colleague, Edward Livingston, MD, a surgeon and a perceptive critic, is taking over the *Users' Guides to the Medical Literature* series at *JAMA*, and I am confident it will prosper in his hands. In addition, I acknowledge the efforts of our partners at McGraw-Hill Education—James Shanahan, Scott Grillo, Michael Crumsho, and Robert Pancotti.

Finally, I thank my friends Cathy DeAngelis, MD, MPH, and her successor, Howard Bauchner, MD, MPH, former and current Editors in Chief of The JAMA Network, for their strong backing of me, my colleagues, and this project. Howard inherited this project. Once I found out that his immediate and enthusiastic acceptance of it was based on his regular use of early articles in the Users' Guides series, any concern about its reception vanished. Indeed, Howard was the instigator of Evidence-Based Medicine—An Oral History,[2,3] a video series of personal views on the birth and early growth of evidence-based medicine that has helped put the Users' Guides into perspective. Howard's infectious good spirits and sharp intelligence bode well for further editions of this book.

Drummond Rennie, MD
University of California, San Francisco

References

1. Durack DT. The weight of medical knowledge. *N Engl J Med*. 1978;298(14):773-775.
2. Smith R, Rennie D. Evidence-based medicine—an oral history. *JAMA*. 2014;311(4):365-367.
3. Evidence-based medicine—an oral history website. http://ebm.jamanetwork.com. Accessed August 17, 2014.

This page intentionally left blank

PREFACE

Evidence-based medicine (EBM)—as a concept with that particular moniker—is now almost 25 years old. Looking back, periods of infancy, childhood, adolescence,[1] and now a mature adulthood are evident.[2] This third edition of the *Users' Guides to the Medical Literature* firmly establishes the maturity of the EBM movement.

The first articulation of the world view that was to become EBM appeared in 1981 when a group of clinical epidemiologists at McMaster University, led by David Sackett, MD, published the first of a series of articles that advised clinicians on how to read clinical journals.[3] Although a huge step forward, the series had its limitations. After teaching what they then called *critical appraisal* for a number of years, the group became increasingly aware of both the necessity and the challenges of going beyond reading the literature in a browsing mode and instead using research studies to solve patient management problems on a day-to-day basis.

In 1990, I assumed the position of residency director of the Internal Medicine Program at McMaster. Through Dave Sackett's leadership, critical appraisal had evolved into a philosophy of medical practice based on knowledge and understanding of the medical literature supporting each clinical decision. We believed that this represented a fundamentally different style of practice and required a term that would capture this difference.

My mission as residency director was to train physicians who would practice this new approach to medicine. In the spring of 1990, I presented our plans for changing the program to the members of the Department of Medicine, many of whom were unsympathetic. The term suggested to describe the new approach was *scientific medicine*. Those already hostile were incensed at the implication that they had previously been "unscientific." My second try at a name for our philosophy of medical practice,

evidence-based medicine, became extremely popular in a very short time. To use the current vernacular, it went viral.[4]

After that fateful Department of Medicine meeting at McMaster, the term *EBM* first appeared in the autumn of 1990 in an information document for residents entering, or considering application to, the residency program. The relevant passage follows:

> Residents are taught to develop an attitude of "enlightened scepticism" towards the application of diagnostic, therapeutic, and prognostic technologies in their day-to-day management of patients. This approach . . . has been called "evidence-based medicine." . . . The goal is to be aware of the evidence on which one's practice is based, the soundness of the evidence, and the strength of inference the evidence permits. The strategy employed requires a clear delineation of the relevant question(s); a thorough search of the literature relating to the questions; a critical appraisal of the evidence and its applicability to the clinical situation; a balanced application of the conclusions to the clinical problem.

The first published appearance of the term was in the American College of Physicians' *Journal Club* in 1991.[5] Meanwhile, our group of enthusiastic evidence-based medical educators at McMaster were refining our practice and teaching of EBM. Believing that we were on to something important, we linked up with a larger group of academic physicians, largely from the United States, to form the first Evidence-Based Medicine Working Group and published an article in *JAMA* that defined and expanded on the description of EBM, labeling it as a "paradigm shift."[6]

This working group then addressed the task of producing a new set of articles, the successor to the Readers' Guides, to present a more practical approach to applying the medical literature to clinical practice. With the unflagging support and wise counsel of *JAMA* Deputy Editor Drummond Rennie, MD, the Evidence-Based Medicine Working Group created a 25-part series called the *Users' Guides to the Medical Literature*, published in *JAMA* between 1993 and 2000.[7] The series continues

to be published in *JAMA*, with articles that address new concepts and applications.

The first edition of the *Users' Guides to the Medical Literature* was a direct descendant of the *JAMA* series. By the time of the book's publication in 2002, EBM had already undergone its first fundamental evolution, the realization that evidence was never sufficient for clinical decision making. Rather, management decisions always involve trade-offs between desirable and undesirable consequences and thus require value and preference judgments. Indeed, in the first edition of the *Users' Guide to the Medical Literature*, the first principle of EBM was presented as Clinical Decision Making: Evidence Is Never Enough, joining the previously articulated principle of a hierarchy of evidence.

It did not take long for people to realize that the principles of EBM were equally applicable for other health care workers, including nurses, dentists, orthodontists, physiotherapists, occupational therapists, chiropractors, and podiatrists. Thus, terms such as *evidence-based health care* and *evidence-based practice* are appropriate to cover the full range of clinical applications of the evidence-based approach to patient care. Because our *Users' Guides* are directed primarily at physicians, we have continued with the term *EBM*.

The second edition incorporated 2 new EBM developments in EBM thinking. First, we had realized that only a few clinicians would become skilled at critically appraising original journal articles and that preappraised evidence would be crucial for evidence-based clinical practice. Second, our knowledge of how best to ensure that clinical decisions were consistent with patient values and preferences was rudimentary and would require extensive study.

This third edition of the *Users' Guides to the Medical Literature* builds on these realizations, most substantially in the revised guide to finding the evidence. The emphasis is now on preappraised resources and particularly on the successor to medical texts: electronic publications that produce updated

evidence summaries as the data appear and provide evidence-based recommendations for practice.

Awareness of the importance of preappraised evidence and evidence-based recommendations is reflected in other changes in the third edition. We have added a fundamental principle to the hierarchy of evidence and the necessity for value and preference judgments: that optimal clinical decision making requires systematic summaries of the best available evidence.

This principle has led to a fundamental revision of the *Users' Guide* to systematic reviews, which now explicitly includes the meta-analyses and acknowledges 2 core considerations. The first is how well the systematic review and meta-analysis were conducted. The second, inspired by the contributions of the GRADE (Grading of Recommendations Assessment, Development and Evaluation) Working Group,[8] demands an assessment of the confidence that one can place in the estimates of effect emerging from the review and meta-analysis. However well done the review, if the primary evidence on which it is based warrants little confidence, inferences from the review will inevitably be very limited.

The third edition of the *Users' Guides to the Medical Literature* incorporates the lessons we have learned in more than 20 years of teaching the concepts of EBM to students with a wide variety of backgrounds, prior preparation, clinical interest, and geographic location. Indeed, among our many blessings is the opportunity to travel the world, helping to teach at EBM workshops. Participating in workshops in Thailand, Saudi Arabia, Egypt, Pakistan, Oman, Kuwait, Singapore, the Philippines, Japan, India, Peru, Chile, Brazil, Germany, Spain, France, Belgium, Norway, the United States, Canada, and Switzerland—the list goes on—provides us with an opportunity to try out and refine our teaching approaches with students who have a tremendous heterogeneity of backgrounds and perspectives. At each of these workshops, the local EBM teachers share their own experiences,

struggles, accomplishments, and EBM teaching tips that we can add to our repertoire.

We are grateful for the extraordinary privilege of sharing, in the form of the third edition of *Users' Guides to the Medical Literature*, what we have learned.

Gordon Guyatt, MD, MSc
McMaster University

References

1. Daly J. *Evidence-based Medicine and the Search for a Science of Clinical Care*. Berkeley, CA: Milbank Memorial Fund and University of California Press; 2005.
2. Smith R, Rennie D. Evidence-based medicine—an oral history. *JAMA*. 2014;311(4):365-367.
3. Department of Clinical Epidemiology & Biostatistics, McMaster University. How to read clinical journals, I: why to read them and how to start reading them critically. *Can Med Assoc J*. 1981;124(5):555-558.
4. Evidence-based medicine—an oral history website. http://ebm.jamanetwork.com. Accessed August 17, 2014.
5. Guyatt G. Evidence-based medicine. *ACP J Club (Ann Intern Med)*. 1991;114(suppl 2):A-16.
6. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992;268(17):2420-2425.
7. Guyatt GH, Rennie D. Users' guides to the medical literature. *JAMA*. 1993;270(17):2096-2097.
8. Guyatt GH, Oxman AD, Vist GE, et al; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.

This page intentionally left blank

JAMAevidence
Using Evidence to Improve Care

1

How to Use the Medical Literature—and This Book—to Improve Your Patient Care

Gordon Guyatt and Maureen O. Meade

IN THIS CHAPTER

The objective of this book is to help you make efficient use of the published literature in guiding your patient care. What does the published literature comprise? Our definition is broad. You may find *evidence*[a] in a wide variety of sources, including original journal articles, *reviews* and *synopses* of *primary studies*, *clinical practice guidelines*, and traditional and innovative medical textbooks. Increasingly, clinicians can most easily access many of these sources through the Internet.

THE STRUCTURE OF THE *USERS' GUIDES TO THE MEDICAL LITERATURE*: THE FOUNDATIONS

This book is not like a novel that you read from beginning to end. Indeed, the *Users' Guides* are designed so that each part is largely self-contained. Thus, we anticipate that clinicians may be selective in their reading of the core content chapters and will certainly be selective when they move beyond the essentials. On the first reading, you may choose only a few advanced areas that interest you. If, as you use the medical literature, you find the need to expand your understanding of, for instance, studies addressing *screening* tests or the use of *surrogate outcomes*, you can consult the relevant chapters to familiarize or reacquaint yourself with the issues. You may also find the glossary a useful reminder of the formal definitions of terms used herein. Finally, we rely heavily on examples to make our points. You will find examples identified by their blue background.

The Essentials version of this book comprises 18 chapters in 7 sections: The Foundations, Therapy, Harm, Diagnosis, Prognosis, Summarizing the Evidence, and Moving From Evidence to Action (Box 1-1). A larger Manual version of this book includes additional chapters in each section.

[a]The italicization, here and in every other chapter, represents the first occurrence in the chapter of a word defined in the glossary.

BOX 1-1

Sections of This Book

The Foundations
Therapy
Harm
Diagnosis
Prognosis
Summarizing the Evidence
Moving From Evidence to Action

The first section of this book introduces the foundations of *evidence-based practice*. Two chapters in this section, What Is Evidence-Based Medicine? and Evidence-Based Medicine and the Theory of Knowledge, present the 3 guiding principles of *evidence-based medicine* (EBM), and place EBM in the context of a humanistic approach to medical practice. The subsequent chapters in this section deal with defining your clinical question, locating the best evidence to address that question, and distinguishing *bias* from *random error* (a key principle of critical appraisal).

Clinicians are primarily interested in making accurate diagnoses and selecting optimal treatments for their patients. They also must avoid exposing patients to *harm* and offer patients prognostic information. Thus, chapters in 4 sections of this book (Therapy, Harm, Diagnosis, and Prognosis) begin by outlining what every medical student, intern and resident, and practicing physician and other clinicians will need to know to use articles that present primary data that address these 4 principal issues in providing patient care.

Increasingly, we have become aware that individual studies are often unrepresentative of all relevant studies (ie, showing larger or smaller *treatment effects* than *pooled estimates* of all relevant

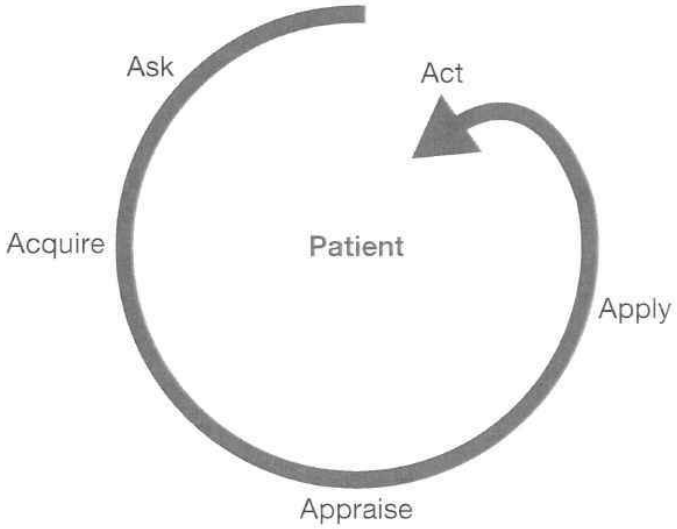
studies), imprecise, or limited in their applicability—so much so that, since the previous edition of this book, we have added the need for systematic summaries of all relevant studies as a core principle of EBM. This has major implications for clinicians looking to use the literature to provide optimal patient care. Efficient and optimally effective evidence-based practice dictates bypassing the critical assessment of primary studies and, if they are available, moving straight to the evaluation of rigorous *systematic reviews*. Even more efficient than using a systematic review is moving directly to an evidence-based recommendation. Ideally, management recommendations—summarized in clinical practice guidelines or *decision analyses*—will incorporate the best evidence and make explicit the value judgments used in moving from evidence to recommendations for action. Unfortunately, many clinical practice guidelines sometimes provide recommendations that are inconsistent with the best evidence or with typical patient *values and preferences*. The last 2 sections of the book, Summarizing the Evidence and Moving From Evidence to Action, provide clinicians with guides for using systematic reviews (with and without *meta-analyses*) and recommendations to optimize their patient care.

Our approach to addressing diagnosis, therapy, harm, and *prognosis* begins when the clinician faces a clinical question (Figure 1-1). Having identified the problem, the clinician then formulates a structured clinical question (the "Ask," Figure 1-1) (see Chapter 3, What Is the Question?) and continues with finding the best relevant evidence (the "Acquire," Figure 1-1) (see Chapter 4, Finding Current Best Evidence).

Many chapters of this book include an example of a search for the best evidence. These searches were accurate at the time they were done, but you are unlikely to get exactly the same results if you replicate the searches now. The reasons for this include additions to the literature and occasional structural changes in databases. Thus, you should view the searches as illustrations of searching principles, rather than as currently definitive searches that address the clinical question. Having identified the best evidence, the clinician then proceeds

FIGURE 1-1

Using the Medical Literature to Provide Optimal Patient Care



through the next 3 steps in evaluating that evidence: appraisal, considering how to apply the results, and acting (Figure 1-1). The appraisal includes 2 questions, "How serious is the *risk of bias*?" and "What are the results?" The first question, "How serious is the risk of bias?" deals with the extent to which the results represent an unbiased estimate of the truth. In the first 2 editions of this book, we referred to risk of bias as *validity* and used the question, "Are the results valid?" We have made this change because "risk of bias" is a more explicit and transparent term. In Chapter 7, How to Use a Noninferiority Trial, limitations of study design related to these topics include issues beyond risk of bias. Therefore, in Chapter 7, we continue to use the term validity and the question "Are the results valid?" to capture the risk of bias and these additional issues.

The second question in the appraisal step is, "What are the results?" For issues of therapy or harm, this will involve assessing the magnitude and precision of the impact of the intervention (a treatment or possible harmful exposure) (see Chapter 6,

Therapy [Randomized Trials]; Chapter 7, How to Use a Noninferiority Trial; Chapter 8, Does Treatment Lower Risk? Understanding the Results; Chapter 9, Confidence Intervals: Was the Single Study or Meta-analysis Large Enough? and Chapter 10, Harm [Observational Studies]). For issues of diagnosis, this will involve generating *pretest probabilities* and then *posttest probabilities* on the basis of test results (see Chapter 11, The Process of Diagnosis, and Chapter 12, Diagnostic Tests). For issues of prognosis, this will involve determining the likelihood of events occurring over time and the precision of those estimates (see Chapter 13, Prognosis).

Once we understand the results, we move to dealing with applicability (Figure 1-1) and ask ourselves the third question: "How can I apply these results to patient care?" This question has 2 parts. First, can you *generalize* (or, to put it another way, particularize) the results to your patient? For instance, your confidence in estimates of treatment effect decreases if your patient is too dissimilar from those who participated in the *trial* or trials. Second, what is the significance of the results for your patient? Have the investigators measured all *patient-important outcomes*? What is the tradeoff among the benefits, *risks*, and *burdens* of alternative management strategies?

Often, you will find a systematic review that, if it is done well and includes a meta-analysis (see Chapter 14, The Process of a Systematic Review and Meta-analysis), will have conducted the search and risk of bias appraisals and, further, summarized the results and suggested the confidence you can place in estimates (see Chapter 15, Understanding and Applying the Results of a Systematic Review and Meta-analysis). In addition, you often will find a recommendation that, if developed rigorously, is based on trustworthy systematic reviews of the evidence and explicitly considers patient values and preferences (see Chapter 17, How to Use a Patient Management Recommendation: Clinical Practice Guidelines and Decision Analyses) and provides guidance on the issue of applying the results to your patient. In our discussions of systematic reviews and guidelines, we introduce the

GRADE (*Grading of Recommendations Assessment, Development and Evaluation*) approach to summarizing evidence and developing recommendations, an approach that we believe represents a major advance in EBM (see Chapter 15, Understanding and Applying the Results of a Systematic Review and Meta-analysis).

The final step in using the evidence is action (Figure 1-1). Often, this will involve shared decision making with your patients (see Chapter 18, Decision Making and the Patient), a key part of the EBM process.

We have kept the initial chapters of each part of this book simple and succinct. From an instructor's point of view, these core chapters constitute a curriculum for a short course in using the literature for medical students, resident physicians, or students of other health professions. They also are appropriate for a continuing education program for practicing physicians and other clinicians.

ADVANCED TOPICS

Moving beyond the foundations, the advanced topics in this book will interest clinicians who want to practice EBM at a more sophisticated level. They are organized according to the core issues addressed in the sections on Therapy, Harm, Diagnosis, and Prognosis.

The presentations of advanced topics will deepen your understanding of study methods, statistical issues, and use of the numbers that emerge from medical research. We wrote the advanced chapters mindful of an additional audience: those who teach evidence-based practice. Many advanced entries read like guidelines for an interactive discussion with a group of learners in a tutorial or on the ward. That is natural enough because the material was generated in such small-group settings. Indeed, the Evidence-Based Medicine Working Group has produced materials that specifically discuss the challenges that arise when these concepts are presented in small-group

settings, including a series of 5 articles published in the *Canadian Medical Association Journal*[1] and another 5 articles in the *Journal of General Internal Medicine*.[2]

Experience on the wards and in outpatient clinics, and with the first 2 editions of the *Users' Guides to the Medical Literature*, has taught us that this approach is well suited to the needs of any clinician who is eager to achieve an evidence-based practice.

References

1. Wyer PC, Keitz S, Hatala R, et al. Tips for learning and teaching evidence-based medicine: introduction to the series. *CMAJ*. 2004;171(4):347-348.
2. Kennedy CC, Jaeschke R, Keitz S, et al; Evidence-Based Medicine Teaching Tips Working Group. Tips for teachers of evidence-based medicine: adjusting for prognostic imbalances (confounding variables) in studies on therapy or harm. *J Gen Intern Med*. 2008;23(3):337-343.

JAMAevidence
Using Evidence to Improve Care

2

What Is Evidence-Based Medicine?

Gordon Guyatt, Roman Jaeschke, Mark C. Wilson, Victor M. Montori, and W. Scott Richardson

IN THIS CHAPTER

9

Evidence-based medicine (EBM) involves conscientiously working with patients to help them resolve (sometimes) or cope with (often) problems related to their physical, mental, and social health. The EBM approach necessitates awareness and understanding of clinical research *evidence*. For those involved in making health care decisions, EBM encompasses creating implementation strategies to ensure practice evidence that is well grounded in best evidence research summaries.

At the core of EBM is a care and respect for patients who will suffer if clinicians fall prey to muddled clinical reasoning and to neglect or misunderstanding of research findings. Practitioners of EBM strive for a clear and comprehensive understanding of the evidence underlying their clinical care and work with each patient to ensure that chosen courses of action are in that patient's best interest. Practicing EBM requires clinicians to understand how uncertainty about clinical research evidence intersects with an individual patient's predicament and preferences. In this chapter, we outline how EBM proposes to achieve these goals and, in so doing, define the nature of EBM.

THREE FUNDAMENTAL PRINCIPLES OF EBM

Conceptually, EBM involves 3 fundamental principles. First, optimal clinical decision making requires awareness of the best available evidence, which ideally will come from systematic summaries of that evidence. Second, EBM provides guidance to decide whether evidence is more or less trustworthy—that is, how confident can we be of the properties of diagnostic tests, of our patients' *prognosis*, or of the impact of our therapeutic options? Third, evidence alone is never sufficient to make a clinical decision. Decision makers must always trade off the benefits and *risks*, *burden*, and costs associated with alternative management strategies and, in doing so, consider their patients' unique predicament and *values and preferences*.[1]

Best Evidence Summaries

In 1992, Antman et al[2] published an article that compared the recommendations of experts for management of patients with myocardial infarction to the evidence that was available at the time the recommendations were made. Figures 2-1 and 2-2 summarize their results in *forest plots*. Both are cumulative *meta-analyses*: the first of thrombolytic therapy for myocardial infarction and the second for lidocaine antiarrhythmic therapy. In both cases, the line in the center represents an *odds ratio* of 1.0 (treatment is neither beneficial nor harmful). As in any forest plot, the dots represent the best estimates of treatment effect (often from individual studies; in this case from the totality of accumulated evidence), and the associated lines represent the 95% *confidence intervals* (CIs).

The "Patients" column presents the total number of patients enrolled in all *randomized clinical trials* (RCTs) conducted to the date specified in the "Year" column—the reason we call it a cumulative meta-analysis. In both figures, early on, with relatively few patients, the CIs are wide, but they progressively narrow as new trials were reported.

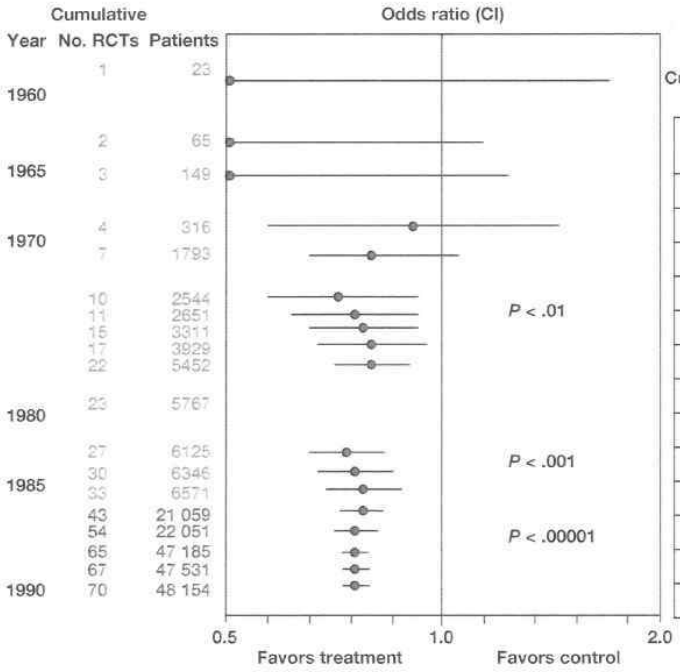
For the thrombolytic example, by 10 trials and approximately 2500 patients, it appears that thrombolytic therapy reduces mortality, but the CIs are still wide enough to permit residual uncertainty. By 30 trials and more than 6000 patients, the reduction in odds of death of approximately 25% seems secure.

Despite this apparently definitive result, additional trials that enrolled 40 000 patients—half of whom did not receive the benefits of life-prolonging thrombolytic therapy—were conducted. Why was this necessary?

The right side of each figure, which presents the guidance expressed in then-current reviews and textbooks as the data were accumulating, provides the answer to this question. Until approximately a decade after the answer was in, there was considerable disagreement among experts, with many recommending against, or not mentioning, thrombolytic therapy. To the detriment of patients who did not receive thrombolytic

FIGURE 2-1

Thrombolytic Therapy in Acute Myocardial Infarction



Current Review and Textbook Recommendations

Yes	No	Not mentioned
		21
		5
	1	10
	1	2
	2	8
		7
		8
1		12
1	8	4
1	7	3
7	2	1
23		1
7		

Abbreviations: CI, confidence interval; RCTs, randomized clinical trials.

This is a cumulative meta-analysis of thrombolytic therapy for myocardial infarction. The line down the center, the odds ratio, equals 1.0. The dots represent best estimates, and the lines around the dots are 95% CIs. The numbers on the left side of the figure are trials and patient totals across trials.

Early on, the CIs are very wide. By 10 trials, it appears therapy reduces mortality, but the effect is still uncertain. By 30 trials, the effect seems secure. However, 40 000 more patients were enrolled after the answer was in. Why?

The right side of the figure displays current reviews and textbook recommendations as data accumulated. Recommendations are in favor ("Yes"), against ("No"), or "Not mentioned." Two key points: (1) at the same time, experts disagreed, and (2) it took 10 years for experts to catch up with evidence.

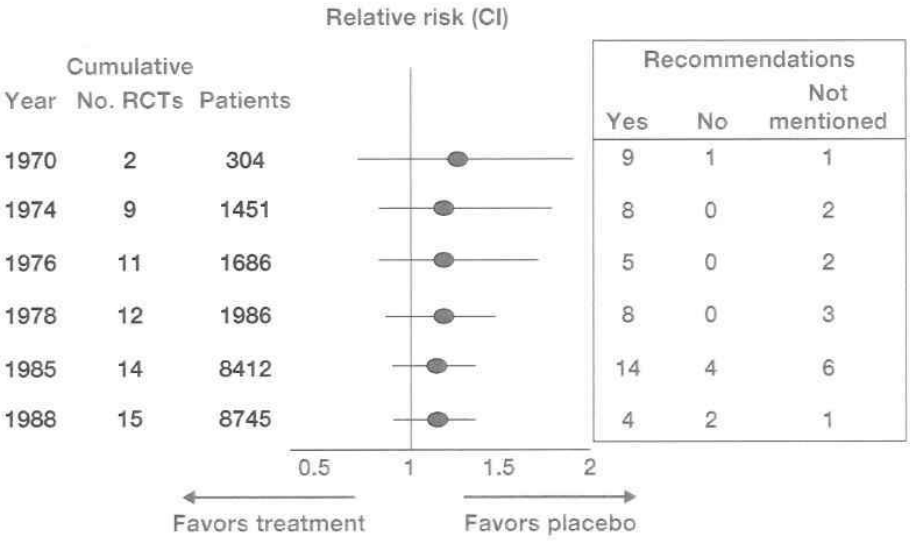
Reproduced from Antman et al.[2]

therapy during this period, it took a decade for the experts to catch up with the evidence.

Figure 2-2 tells a perhaps even more disturbing story. This cumulative meta-analysis reveals that there was never any RCT

FIGURE 2-2

Prophylactic Lidocaine in Acute Myocardial Infarction



Abbreviations: CI, confidence interval; RCTs, randomized clinical trials.

This figure shows a cumulative meta-analysis of the effect of prophylactic lidocaine in preventing death from myocardial infarction. In this case, there is never any evidence of benefit. Ultimately, harm is not proved, but there clearly is no benefit. Most experts, however, were recommending therapy despite RCT evidence. Also, as in Figure 2-1, there was a lot of disagreement among experts.

Reproduced from Antman et al.[2]

evidence that suggested a lower mortality with prophylactic lidocaine after myocardial infarction—indeed, *point estimates* suggested an increase in death rate. Nevertheless, although we once again see widespread disagreement among the experts, most texts and reviews were recommending prophylactic lidocaine during the 2 decades during which the RCT evidence was accumulating.

Why the expert disagreement, the lag behind the evidence, and the recommendations inconsistent with the evidence? These stories come from the era before *systematic reviews* and meta-analyses were emerging in the late 1980s. If the evidence

summaries presented in the forest plots had been available to the experts, they would have grasped the benefits of thrombolytic therapy far earlier than they did and abandoned prophylactic lidocaine far earlier. Indeed, following EBM principles that limit reliance on biologic rationale and place far more emphasis on empirical evidence, the experts may never have started using lidocaine.

Rational clinical decisions require systematic summaries of the best available evidence. Without such summaries, clinicians—expert or otherwise—will be unduly influenced by their own preconceptions and by unrepresentative and often lower-quality evidence. This, the first principle of EBM, immediately raises another question: "How does one recognize the best evidence?"

Guides to Confidence in Estimates

Summaries of the best evidence for diagnosis, prognosis, or treatment present evidence, respectively, for how to interpret test results, predict patients' likely fate, or understand the impact of alternative management strategies. Sometimes, such evidence is trustworthy—we have high confidence in estimates of test properties, patients' prognosis, or treatment effects. At other times, limitations in evidence leave us uncertain. Evidence-based medicine provides guidance to distinguish between these situations and the range of confidence between them.

Historically, EBM answered the question, "What is the best evidence?" with *hierarchies of evidence*, the most prominent of which was the hierarchy related to evidence that supported therapeutic interventions (Figure 2-3). Issues of diagnosis or prognosis require different hierarchies. For studies of the accuracy of diagnostic tests, the top of the hierarchy includes studies that enrolled patients about whom clinicians had diagnostic uncertainty and that undertook a *blind* comparison between the candidate test and a *criterion standard* (see Chapter 12, Diagnostic Tests, and Chapter 13, Prognosis). For prognosis, prospective observational studies accurately documenting *exposures* and outcomes and

FIGURE 2-3

Hierarchy of Evidence

Because we would like to optimally individualize patient care, n-of-1 randomized clinical trials are at the top of the hierarchy of study designs, followed by conventional randomized trials. Next in the hierarchy are observational studies; we should try to find studies that focus on outcomes important to the patient. Next, if there are no clinical studies available, we may look at basic scientific research, although caution must be used in extrapolating the results to the clinical setting. Clinical experience is at the bottom of the hierarchy, either your own or that of colleagues or experts.

following up all patients during relevant periods would sit atop the hierarchy.

Returning to the hierarchy of therapy, noting the limitations of human intuition,[3] EBM places the unsystematic observations of individual clinicians lowest on the hierarchy. Noting that predictions based on physiologic experiments are often right but sometimes disastrously wrong, EBM places such experiments at the next step up in the hierarchy. Observational studies

that measure the apparent impact on *patient-important outcomes* and RCTs constitute the next 2 steps up the hierarchy of evidence.

All of the sources of evidence mentioned thus far involve generalizations from groups of patients to an individual, and all are limited in this regard. The same strategies that minimize *bias* in conventional therapeutic trials that involve multiple patients, however, can guard against misleading results in studies that involve single patients.[4] In the *n-of-1 RCT*, a patient and clinician are blind to whether that patient is receiving active or *placebo* medication. The patient makes quantitative ratings of troublesome symptoms during each period, and the n-of-1 RCT continues until both the patient and the clinician conclude that the patient is or is not obtaining benefit from the target intervention. An n-of-1 RCT can provide definitive evidence of treatment effectiveness in individual patients[5,6] and is thus at the top of the evidence hierarchy. Unfortunately, n-of-1 RCTs are restricted to chronic conditions with treatments that act and cease acting quickly and are subject to considerable logistic challenges. We therefore must usually rely on studies of other patients to make inferences regarding our patient.

This hierarchy is far from absolute, and a more sophisticated framework has emerged for judging confidence in estimates of effect. Table 2-1 summarizes that framework, formulated by the *GRADE* (*Grading of Recommendations Assessment, Development and Evaluation*) Working Group, originally to provide an approach to the development of *clinical practice guidelines*.[7,8] The GRADE approach involves rating our confidence in estimates of the effects of health care interventions (also referred to as quality of evidence) as high, moderate, low, or very low. Consistent with the previous hierarchy approach, in the GRADE guidance, RCTs begin as high confidence and observational studies begin as low confidence. We lose confidence in a body of RCT evidence, however, if studies have major problems in design and execution (*risk of bias*); results are imprecise, inconsistent, or indirect (eg, the population of

TABLE 2-1

Confidence Assessment Criteria[8]

Study Design	Confidence in Estimates	Lower If[a]	Higher If[a]
Randomized trial	High	Risk of bias –1 Serious –2 Very serious	Large effect +1 Large +2 Very large Dose response +1 Evidence of a gradient
	Moderate	Inconsistency –1 Serious –2 Very serious	
Observational study	Low	Indirectness –1 Serious –2 Very serious	
	Very low	Imprecision –1 Serious –2 Very serious	
		Publication bias –1 Likely –2 Very likely	

[a]Minus and plus signs refer, respectively, to rating down and rating up confidence in estimates. The 1 refers to rating down or up by 1 level (eg, from high to moderate or moderate to high), and the 2 refers to rating down or up by 2 levels (eg, high to low or low to high).

interest differs from the population studied); or we have a high suspicion of publication bias (see Chapter 15, Understanding and Applying the Results of a Systematic Review and Meta-analysis). When a body of RCT evidence suffers from a number of these limitations, the confidence in estimates may be low or even very low.

Similarly, if treatment effects are sufficiently large and consistent, the GRADE approach allows for moderate or even high confidence ratings from carefully conducted observational studies.

For example, observational studies have allowed extremely strong inferences about the efficacy of insulin in diabetic ketoacidosis or that of hip replacement in patients with debilitating hip osteoarthritis.

The EBM approach implies a clear course of action for clinicians addressing patient problems. They should seek the highest-quality evidence available to guide their clinical decisions. This approach makes it clear that any claim that there is no evidence for the effect of a particular treatment is a non sequitur. The available evidence may warrant very low confidence—it may be the unsystematic observation of a single clinician or physiologic studies that point to mechanisms of action that are only indirectly related—but there is always evidence.

Evidence Is Never Enough to Drive Clinical Decision Making

First, picture a woman with chronic pain from terminal cancer. She has come to terms with her condition, resolved her affairs, said her good-byes, and wishes to receive only palliative care. She develops severe pneumococcal pneumonia. Evidence that antibiotic therapy reduces morbidity and mortality from pneumococcal pneumonia warrants high confidence. This evidence does not, however, dictate that this patient should receive antibiotics. Her values—emerging from her comorbidities, social setting, and beliefs—are such that she would prefer to forgo treatment.

Now picture a second patient, an 85-year-old man with severe dementia who is mute and incontinent, is without family or friends, and spends his days in apparent discomfort. This man develops pneumococcal pneumonia. Although many clinicians would argue that those responsible for his decision making should elect not to administer antibiotic therapy, others would suggest that they should. Again, evidence of treatment effectiveness does not automatically imply that treatment should be administered.

Finally, picture a third patient, a healthy 30-year-old mother of 2 children who develops pneumococcal pneumonia. No clinician would doubt the wisdom of administering antibiotic therapy to this patient. This does not mean, however, that an underlying value judgment has been unnecessary. Rather, our values are sufficiently concordant, and the benefits so overwhelm the risk of treatment that the underlying value judgment is unapparent.

By values and preferences, we mean the collection of goals, expectations, predispositions, and beliefs that individuals have for certain decisions and their potential outcomes. The explicit enumeration and balancing of benefits and risks that are central to EBM bring the underlying value judgments involved in making management decisions into bold relief.

Acknowledging that values play a role in every important patient care decision highlights our limited understanding of how to ensure that decisions are consistent with individual and, where appropriate, societal values. As we discuss further in the final section of this chapter, developing efficient processes for helping patients and clinicians work together toward optimal decisions consistent with patient values and preferences remains a frontier for EBM.

Next, we comment on additional skills that clinicians must master for optimal patient care and the relation of those skills to EBM.

CLINICAL SKILLS, HUMANISM, AND EBM

In summarizing the skills and attributes necessary for *evidence-based practice*, Box 2-1 highlights how EBM complements traditional aspects of clinical expertise. One of us, an intensive care specialist, developed a lesion on his lip shortly before an important presentation. He was concerned and, wondering whether he should take acyclovir, proceeded to spend the next 30 minutes searching for and evaluating the highest-quality evidence. When he began to discuss his remaining uncertainty

BOX 2-1

Knowledge and Skills Necessary for Optimal Evidence-Based Practice

- Diagnostic expertise
- In-depth background knowledge
- Effective searching skills
- Effective critical appraisal skills
- Ability to define and understand benefits and risks of alternatives
- In-depth physiologic understanding that allows application of evidence to the individual
- Sensitivity and communication skills required for full understanding of patient context
- Ability to elicit and understand patient values and preferences and work with patients in shared decision making

with his partner, an experienced dentist, she cut short the discussion by exclaiming, "But, my dear, that isn't herpes!"

This story illustrates the necessity of obtaining the correct diagnosis before seeking and applying research evidence regarding optimal treatment. After making the diagnosis, the clinician relies on experience and background knowledge to define the relevant management options. Having identified those options, the clinician can search for, evaluate, and apply the best evidence regarding patient management.

In applying evidence, clinicians rely on their expertise to define features that affect the applicability of the results to the individual patient. The clinician must judge the extent to which differences in treatment (for instance, local surgical expertise or the possibility of patient *nonadherence*) or patient characteristics (such as age, comorbidity, or the patient's

personal circumstances) may affect estimates of benefit and risk that come from the published literature.

We note that some of these skills—the sensitivity to the patient's unique predicament and the communication skills necessary for shared decision making—are often not typically associated with EBM. We believe they are, in fact, at the core of EBM. Understanding the patient's personal circumstances is of particular importance and requires advanced clinical skills, including listening skills and compassion. For some patients, incorporation of patient values for major decisions will mean a full enumeration of the possible benefits, risks, and inconveniences associated with alternative management strategies. For some patients and problems, this discussion should involve the patient's family. For other problems—the discussion of *screening* with prostate-specific antigen with older male patients, for instance—attempts to involve family members might violate cultural norms.

Some patients are uncomfortable with an explicit discussion of benefits and risk and object to clinicians placing what they perceive as excessive responsibility for decision making on their shoulders. In such cases, it is the physician's responsibility to develop insight to ensure that choices will be consistent with the patient's values and preferences while remaining sensitive to the patient's preferred role in decision making.

ADDITIONAL CHALLENGES FOR EBM

Busy clinicians—particularly those early in their development of the skills needed for *evidence-based practice*—will find that they often perceive time limitations as the biggest challenge to evidence-based practice. This perception may arise from having inadequate access to various evidence-based resources. Fortunately, a tremendous array of sophisticated evidence-based information is now available for clinicians working in high-income countries, and the pace of innovation remains extremely rapid (see Chapter 4, Finding Current Best Evidence).

Access to preprocessed information cannot, however, address other skills required for efficient evidence-based practice. These skills include formulating focused clinical questions, matching prioritized questions to the most appropriate resources, assessing confidence in estimates, and understanding how to apply results to clinical decision making. Although these skills take time to learn, the reward in terms of efficient and effective practice can more than compensate.

Another challenge for evidence-based practice is ensuring that management strategies are consistent with patients' values and preferences. In a time-constrained environment, how can we ensure that patients' involvement in decision making has the form and extent that they desire and that the outcome reflects their needs and desires? Evidence-based medicine leaders are now making progress in addressing these challenges.[9,10]

This book deals primarily with decision making at the level of the individual patient. Evidence-based approaches can also inform health care policy making, day-to-day decisions in public health, and systems-level decisions, such as those facing hospital managers. In each of these areas, EBM can support the appropriate goal of gaining the greatest health benefit from limited resources.

In the policy arena, dealing with differing values poses even more challenges than in the arena of individual patient care. Should we restrict ourselves to alternative resource allocation within a fixed pool of health care resources, or should we consider expanding health care services at the cost, for instance, of higher tax rates for individuals or corporations? How should we deal with the large body of observational studies that suggest that social and economic factors may have a larger influence on the health of populations than health care provision? How should we deal with the tension between what may be best for a person and what may be optimal for the society of which that person is a member? The debate about such issues is at the core of evidence-based policy making in health care; it also has implications for decision making at the individual patient level.

References

1. Napodano R. *Values in Medical Practice.* New York, NY: Humana Sciences Press; 1986.
2. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *JAMA*. 1992;268(2): 240-248.
3. Nisbett R, Ross L. *Human Inference.* Englewood Cliffs, NJ: Prentice-Hall; 1980.
4. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy—randomized trials in individual patients. *N Engl J Med*. 1986;314(14):889-892.
5. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness: our three-year experience. *Ann Intern Med*. 1990;112(4):293-299.
6. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2-year period. *JAMA*. 1993;270(22):2708-2712.
7. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008;336(7651):995-998.
8. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines, 3: rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-406.
9. Montori VM, Guyatt GH. Progress in evidence-based medicine. *JAMA*. 2008;300(15):1814-1816.
10. Stiggelbout AM, Van der Weijden T, De Wit MP, et al. Shared decision making: really putting patients at the centre of healthcare. *BMJ*. 2012;344:e256.

This page intentionally left blank

JAMAevidence
Using Evidence to Improve Care

3 What Is the Question?

Gordon Guyatt, Maureen O. Meade, Thomas Agoritsas, W. Scott Richardson, and Roman Jaeschke

IN THIS CHAPTER

THREE WAYS TO USE THE MEDICAL LITERATURE

Consider a medical student, early in her training, seeing a patient with newly diagnosed type 2 diabetes mellitus. She will ask questions such as the following: "What is type 2 diabetes mellitus?" "Why does this patient have polyuria?" "Why does this patient have numbness and pain in his legs?" "What treatment options are available?" These questions address normal human physiology and the pathophysiology associated with a medical condition.

Traditional medical textbooks, whether in print or online, that describe underlying pathophysiology or epidemiology of a disorder provide an excellent resource for addressing these *background questions*. In contrast, the sorts of *foreground questions* that experienced clinicians usually ask require different resources. Formulating a question is a critical and generally unappreciated skill for *evidence-based practice*. The following ways to use the medical literature provide opportunities to practice that skill.

Staying Alert to Important New Evidence

A general internist is checking e-mails on a smartphone while riding public transit to work. While screening a weekly e-mail alert from EvidenceUpdates (http://plus.mcmaster.ca/EvidenceUpdates, Figure 3-1), the internist sees an article titled, Cardiovascular Effects of Intensive Lifestyle Intervention in Type 2 Diabetes,[1] recently published and rated by internist colleagues as newsworthy and highly relevant for practice.

This internist is in the process of addressing a question that clinicians at all stages of training and career development are constantly posing: "What important new evidence should I know to optimally treat patients?" Clinicians traditionally addressed this question by attending rounds and conferences

FIGURE 3-1

Example of E-mail Alert From EvidenceUpdates

EvidenceUPDATES
FROM THE BMJ EVIDENCE CENTRE

Dear Dr. Agoritsas:

New articles: colleagues in your discipline have identified the following article(s) as being of interest:

Article Title	Discipline	Relevance	Newsworthiness
Cardiovascular Effects of Intensive Lifestyle Intervention in Type 2 Diabetes. N Engl J Med	General Practice(GP)/Family Practice(FP)	7	6
	General Internal Medicine-Primary Care(US)	7	6
Comparison of different regimens of proton pump inhibitors for acute peptic ulcer bleeding. Cochrane Database Syst Rev	Hospital Doctor/Hospitalists	6	6
	Internal Medicine	6	6
Effects of Combined Application of Muscle Relaxants and Celecoxib Administration After Total Knee Arthroplasty (TKA) on Early Recovery: A Randomized, Double-Blind, Controlled Study. J Arthroplasty	Hospital Doctor/Hospitalists	6	4
	Internal Medicine	6	4
Corticosteroids for acute bacterial meningitis. Cochrane Database Syst Rev	Hospital Doctor/Hospitalists	6	4
	Internal Medicine	6	4

and by subscribing to target medical journals in which articles relevant to their practice appear. They kept up-to-date by skimming the table of contents and reading relevant articles.

This traditional approach to what we might call the browsing mode of using the medical literature has major limitations of inefficiency and its resulting frustration. Many screened articles may prove of little relevance or newsworthiness or fail to meet the critical appraisal criteria that are presented in this book. To make matters worse, the volume of research is markedly increasing,[2] and relevant studies appear in a large variety of journals.[3] *Evidence-based medicine* offers solutions to these problems.

The most efficient strategy for ensuring you are aware of recent developments relevant to your practice is to subscribe to e-mail alerting systems, such as EvidenceUpdates, used by the internist in this example. This free service has research staff screening approximately 45 000 articles per year in more than 125 clinical journals for methodologic quality and a worldwide panel of practicing physicians rating them for clinical relevance and newsworthiness.[4] You can tailor alerting systems to your information needs (clinical disciplines and frequency of alerts)

and identify the 20 to 50 articles per year that will influence your practice.[5] Several other free or subscription-based alerting systems are available, both for a wide scope of disciplines (eg, NEJM Journal Watch, http://www.jwatch.org) and for specific subspecialties (eg, OrthoEvidence, http://www.myorthoevidence.com).

An alternative to alerting systems are *secondary evidence-based journals*. For example, in internal and general medicine, *ACP Journal Club* (http://acpjc.acponline.org) publishes *synopses* of articles that meet criteria of both high clinical relevance and methodologic quality. We describe such secondary journals in more detail in Chapter 4, Finding Current Best Evidence. If you prefer browsing to receiving alerts, such preappraised sources of *evidence* may increase your efficiency.

Some specialties (primary care and mental health care) and subspecialties (cardiology, oncology, and obstetrics and gynecology) already have specialty-devoted secondary journals; others do not. The New York Academy of Medicine keeps a current list of available secondary journals in many health care disciplines (http://www.nyam.org/fellows-members/ebhc/eb_publications.html). If your specialty does not yet have its own journal, you can apply your own relevance and methodologic screening criteria to articles in your target specialty or subspecialty journals. When you have learned the skills, you will be surprised at the small proportion of studies to which you need attend and the efficiency with which you can identify them.

Problem Solving

Experienced clinicians managing a patient with type 2 diabetes mellitus will ask questions such as "In patients with new-onset type 2 diabetes mellitus, which clinical features or test results predict the development of diabetic complications?" "In patients with type 2 diabetes mellitus requiring drug therapy, does starting with metformin treatment yield improved diabetes control and reduce long-term complications better than other initial treatments?" Here, clinicians are defining specific

questions raised in caring for patients and then consulting the literature to resolve these questions.

Asking Background and Foreground Questions

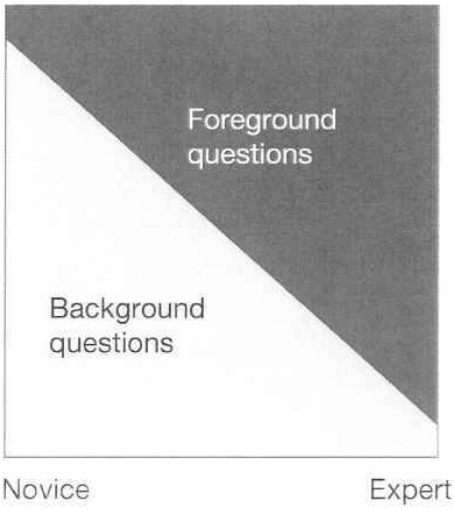
One can think of the first set of questions, those of the medical student, as background questions and of the browsing and problem-solving sets as foreground questions. In most situations, you need to understand the background thoroughly before it makes sense to address foreground issues.

Experienced clinicians may occasionally require background information when a new condition or medical *syndrome* (eg, Middle East respiratory syndrome coronavirus), a new diagnostic test (eg, molecular diagnosis), or a new treatment modality (eg, dipeptidyl peptidase 4 inhibitors) appears in their clinical arena.

Figure 3-2 represents the evolution of the questions we ask as we progress from being novices posing background questions to experts posing foreground questions. This book explores how clinicians can use the medical literature to solve their foreground questions.

FIGURE 3-2

Background and Foreground Questions



CLARIFYING YOUR QUESTION

The Structure: Patients, Exposures, Outcome

Clinical questions often spring to mind in a form that makes finding answers in the medical literature a challenge. Dissecting the question into its component parts to facilitate finding the best evidence is a fundamental skill. One can divide questions of therapy or *harm* into 4 parts following the *PICO* framework: patients or population, intervention(s) or exposure(s), comparator, and outcome (Box 3-1). For questions of *prognosis*, you can use 1 of 2 alternative structures. One has only 3 elements: patients, exposure (time), and outcome. An alternative focuses on patient-related factors, such as age and sex, that can modify prognosis: patients, exposure (eg, older age or male), comparison (eg, younger age or female), and outcome. For diagnostic tests, the structure we suggest is patients, exposure (test), and outcome (*criterion standard*).[6]

BOX 3-1

Framing Clinical Questions: PICO

Patients or Population: Who are the relevant patients?

Intervention(s) or Exposure(s): For example, diagnostic tests, foods, drugs, surgical procedures, time, or risk factors. What are the management strategies we are interested in comparing or the potentially harmful exposures about which we are concerned?

Comparator: For issues of therapy, prevention, or harm, there will always be both an experimental intervention or putative harmful exposure and a control, alternative, or comparison intervention.

Outcome: What are the patient-relevant consequences of the exposures in which we are interested? We may also be interested in the consequences to society, including cost or resource use. It may also be important to specify the period of interest.

Five Types of Foreground Clinical Questions

In addition to clarifying the population, intervention or exposure, and outcome, it is productive to label the nature of the question that you are asking. There are 5 fundamental types of clinical questions:

1. Therapy: determining the effect of interventions on *patient-important outcomes* (*symptoms*, function, morbidity, mortality, and costs)
2. Harm: ascertaining the effects of potentially harmful agents (including therapies from the first type of question) on patient-important outcomes
3. *Differential diagnosis*: in patients with a particular clinical presentation, establishing the frequency of the underlying disorders
4. Diagnosis: establishing the *power* of a test to differentiate between those with and without a *target condition* or disease
5. Prognosis: estimating a patient's future course

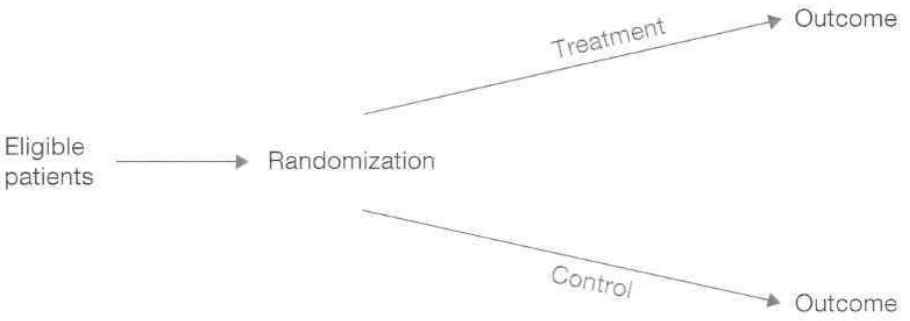
Finding a Suitably Designed Study for Your Question Type

You need to correctly identify the category of study because, to answer your question, you must find an appropriately designed study. If you look for a *randomized trial* to inform the properties of a diagnostic test, you will not find the answer you seek. We will now review the study designs associated with the 5 major types of questions.

To answer questions about a therapeutic issue, we seek studies in which a process analogous to flipping a coin determines participants' receipt of an *experimental treatment* or a control or standard treatment: a randomized trial (see Chapter 6, Therapy [Randomized Trials]). Once investigators allocate participants to treatment or *control groups*, they follow them forward in time to determine whether they have, for instance, a stroke or

FIGURE 3-3

Structure of Randomized Trials



myocardial infarction—what we call the outcome of interest (Figure 3-3). When randomized trials are not available, we look to *observational studies* in which—rather than randomization—clinician or patient preference, or happenstance, determines whether patients receive an intervention or alternative (see Chapter 5, Why Study Results Mislead: Bias and Random Error).

Ideally, we would also look to randomized trials to address issues of harm. For most potentially harmful exposures, however, randomly allocating patients is neither practical nor ethical. For instance, one cannot suggest to potential study participants that an investigator will decide by the flip of a coin whether or not they smoke during the next 20 years. For exposures such as smoking, the best one can do is identify observational studies (often subclassified as *cohort* or *case-control studies*) that provide less trustworthy evidence than randomized trials (see Chapter 10, Harm [Observational Studies]).

Figure 3-4 depicts a common observational study design in which patients with and without the exposures of interest are followed forward in time to determine whether they experience the outcome of interest. For smoking, an important outcome would likely be the development of cancer.

For sorting out differential diagnosis, we need a different study design (Figure 3-5). Here, investigators collect a group of patients with a similar presentation (eg, painless jaundice,

FIGURE 3-4

Structure of Observational Cohort Studies

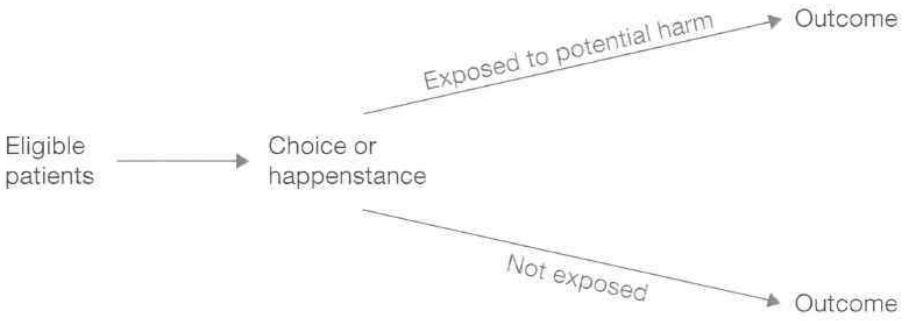


FIGURE 3-5

Structure of Studies of Differential Diagnosis

syncope, or headache), conduct an extensive battery of tests, and, if necessary, follow patients forward in time. Ultimately, for each patient the investigators hope to establish the underlying cause of the symptoms and *signs* with which the patient presented.

Establishing the performance of a diagnostic test (ie, the test's properties or operating characteristics) requires a slightly different design (Figure 3-6). In diagnostic test studies, investigators identify a group of patients among whom they suspect a disease or condition of interest exists (such as tuberculosis, lung cancer, or iron deficiency anemia), which we call the target condition. These patients undergo the new diagnostic test and a *reference standard* (also referred to as *gold standard* or criterion standard). Investigators evaluate the diagnostic test by comparing its classification of patients with that of the reference standard (Figure 3-6).

A final type of study examines a patient's prognosis and may identify factors that modify that prognosis. Here, investigators identify patients who belong to a particular group (such as pregnant women, patients undergoing surgery, or patients with cancer) with or without factors that may modify their prognosis (such as age or *comorbidity*). The exposure here is time, and investigators follow up patients to determine whether they experience the *target outcome*, such as an adverse obstetric or neonatal event at the end of a pregnancy, a myocardial infarction after surgery, or survival in cancer (Figure 3-7).

FIGURE 3-6

Structure of Studies of Diagnostic Test Properties

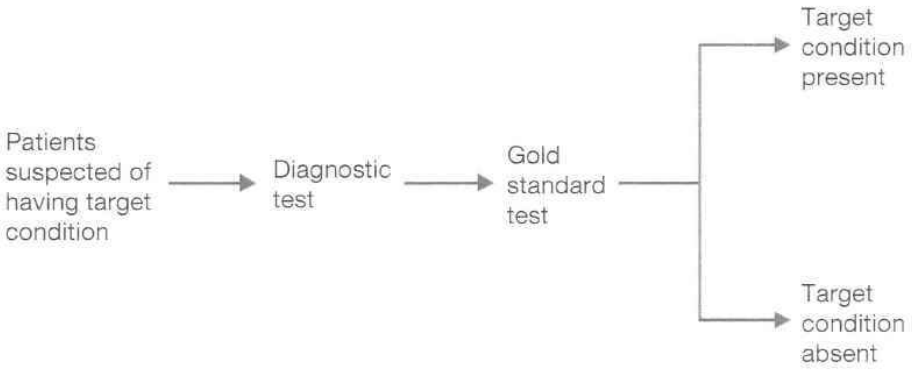
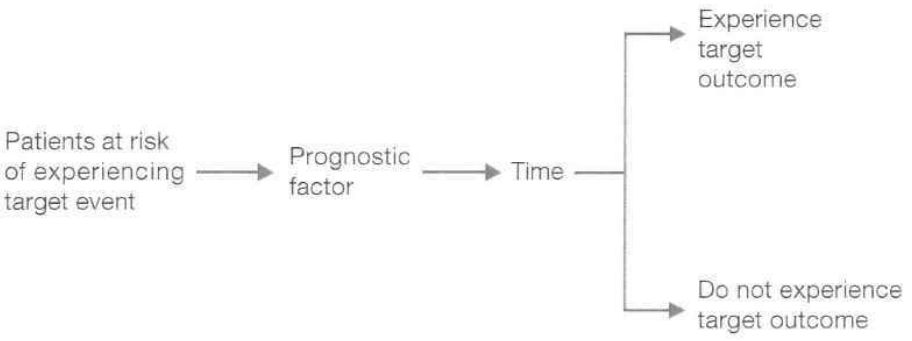


FIGURE 3-7

Structure of Studies of Prognosis



Three Examples of Question Clarification

We will now provide examples of the transformation of unstructured clinical questions into the structured questions that facilitate the use of the medical literature.

Example 1: Diabetes and Target Blood Pressure

A 55-year-old white woman presents with type 2 diabetes mellitus and hypertension. Her glycemic control is excellent with metformin, and she has no history of complications. To manage her hypertension, she takes a small daily dose of a thiazide diuretic. During a 6-month period, her blood pressure is near 155/88 mm Hg.

Initial Question: When treating hypertension, at what target blood pressure should we aim?

Digging Deeper: One limitation of this formulation of the question is that it fails to specify the population in adequate detail. The benefits of tight control of blood pressure may differ among patients with diabetes vs those without diabetes, in type 1 vs type 2 diabetes, and among patients with and without diabetic complications.

The detail in which we specify the patient population is a double-edged sword. On the one hand, being very specific (middle-aged women with uncomplicated type 2 diabetes) will ensure that the answer we get is applicable to our patient. We may, however, fail to find any studies that restrict themselves to this population. The solution is to start with a specific patient population but be ready to remove specifications to find a relevant article. In this case, we may be ready to remove the "female," "middle-aged," "uncomplicated," and "type 2," in that order. If we suspect that the optimal target blood pressure may be similar among patients

with and without diabetes, and if it proves absolutely necessary, we might remove "diabetes" from the question.

The order in which we remove the patient specifications depends on how likely it is that those characteristics will influence response to treatment. We suggest removing "female" first because we think it likely that optimal target blood pressure will be similar in men and women. Similarly, younger, middle-aged, and elderly individuals are likely to have the same optimal targets (although here we are not quite so sure). As our doubts about the same optimal targets across populations becomes progressively greater (uncomplicated vs complicated diabetes, type 1 vs type 2, or patients with diabetes vs those without), we become increasingly reluctant to remove the particular patient characteristic from the question.

We may wish to specify that we are interested in the addition of a specific antihypertensive agent. Alternatively, the intervention of interest may be any antihypertensive treatment. Furthermore, a key part of the intervention will be the target for blood pressure control. For instance, we might be interested in knowing whether it makes any difference if our target diastolic blood pressure is less than 80 mm Hg vs less than 90 mm Hg. Another limitation of the initial question formulation is that it fails to specify the criteria (the outcomes of interest) by which we will judge the appropriate target for our hypertensive treatment.

Improved (Searchable) Question: A Question About Therapy

- *Patients:* Patients with hypertension and type 2 diabetes without diabetic complications.
- *Intervention/Exposure:* Any antihypertensive agent that aims at a target diastolic blood pressure of 90 mm Hg.

- *Comparator:* Target diastolic blood pressure of 80 mm Hg.
- *Outcomes:* Stroke, myocardial infarction, cardiovascular death, and total mortality.

Example 2: Transient Loss of Consciousness

A previously well, although a heavy drinker, 55-year-old man presents to the emergency department with an episode of transient loss of consciousness. On the evening of presentation, he had his usual 5 beers and started to climb the stairs at bedtime. The next thing he remembers is being woken by his son, who found him lying near the bottom of the stairs. The patient took about a minute to regain consciousness and remained confused for another 2 minutes. His son did not witness any shaking, and there had not been any incontinence. Physical examination findings were unremarkable; the electrocardiogram revealed a sinus rhythm with a rate of 80/min and no abnormalities. Glucose, sodium, and other laboratory results were normal, and a blood alcohol test result was negative.

Initial Question: How extensively should I investigate this patient?

Digging Deeper: The initial question gives us little idea of where to look in the literature for an answer. As it turns out, there are a host of questions that could be helpful in choosing an optimal investigational strategy. We could, for instance, pose a question of differential diagnosis: If we knew the distribution of ultimate diagnoses in such patients, we could choose to investigate the more common and omit investigations targeted at remote possibilities.

Other information that would help us would be the properties of individual diagnostic tests. If an electroencephalogram were extremely accurate for diagnosing a

seizure or a 24-hour Holter monitor for diagnosing arrhythmia, we would be far more inclined to order these tests than if they missed patients with the underlying problems or falsely identified patients as not having the problems.

Alternatively, we could ask a question of prognosis. If patients had benign prognoses, we might be much less eager to investigate extensively than if patients tended to have poor outcomes. Finally, the ultimate answer to how intensively we should investigate might come from a randomized trial in which patients similar to this man were allocated to more vs less intensive investigation.

Improved (Searchable) Questions: A Question About Differential Diagnosis

- *Patients:* Middle-aged patients presenting with transient loss of consciousness.
- *Intervention/Exposure:* Thorough investigation and *follow-up* for common and less common diagnoses.
- *Comparator:* Minimal investigation and follow-up.
- *Outcomes:* Frequency of underlying disorders, such as vasovagal syncope, seizure, arrhythmia, and transient ischemic attack.

A Question About Diagnosis

- *Patients:* Middle-aged patients presenting with transient loss of consciousness.
- *Intervention/Exposure:* Electroencephalogram.
- *Outcomes:* Reference standard investigation (probably long-term follow-up).

A Question About Prognosis

- *Patients:* Middle-aged patients presenting with transient loss of consciousness.

- *Exposure/Comparison:* Time.
- *Outcomes:* Morbidity (complicated arrhythmias or seizures, strokes, or serious accidents) and mortality in the year after presentation.

A Question About Diagnostic Impact

You can think of this also as a question of therapy; the principles of critical appraisal are the same.

- *Patients:* Middle-aged patients presenting with loss of consciousness.
- *Intervention/Exposure:* Comprehensive investigation.
- *Comparator:* Minimal investigation.
- *Outcomes:* Morbidity and mortality in the year after presentation.

Example 3: Squamous Cell Carcinoma

A 60-year-old man with a 40-pack-year smoking history presents with hemoptysis. A chest radiograph shows a parenchymal mass with a normal mediastinum, and a fine-needle aspiration and biopsy of the mass reveals non–small cell carcinoma. Aside from hemoptysis, the patient is asymptomatic, and the physical examination results are normal.

Initial Question: What investigations should we undertake before deciding whether to offer this patient surgery?

Digging Deeper: The key defining features of this patient are his non–small cell carcinoma and the fact that his medical history, physical examination, and chest radiograph indicate no evidence of intrathoracic or extrathoracic metastatic disease. Alternative investigational strategies address 2 issues: Does the patient have occult mediastinal disease, and does he have occult extrathoracic metastatic

disease? Investigational strategies for addressing the possibility of occult mediastinal disease include undertaking a mediastinoscopy or performing computed tomography (CT) of the chest and proceeding according to the results of this investigation. Investigational strategies for extrathoracic disease include brain and abdominal CT and bone scanning. Positron emission tomography–CT (PET-CT) represents an alternative approach for both intrathoracic and extrathoracic disease.

What outcomes are we trying to influence in our choice of investigational approach? We would like to prolong the patient's life, but the extent of his underlying tumor is likely to be the major determinant of survival, and our investigations cannot change that. We wish to detect occult mediastinal metastases if they are present because, if the cancer has spread, resectional surgery is unlikely to benefit the patient. Thus, in the presence of mediastinal metastatic disease, patients will usually receive palliative approaches and avoid an unnecessary thoracotomy.

We could frame our structured clinical question in 2 ways. One would be asking about the usefulness of the PET-CT scan for identifying metastatic disease. More definitive would be to ask a question of diagnostic impact, analogous to a therapy question: What investigational strategy would yield superior patient-important outcomes?

Improved (Searchable) Questions: A Question About Diagnosis

- *Patients:* Newly diagnosed non–small cell lung cancer with no evidence of extrapulmonary metastases.
- *Intervention*: PET-CT scan of the chest.
- *Outcome:* Mediastinal spread at mediastinoscopy.

A Question About Diagnostic Impact (Therapy)

- *Patients:* Newly diagnosed non–small cell lung cancer with no evidence of extrapulmonary metastases.
- *Intervention:* PET-CT.
- *Comparator*: Alternative diagnostic strategies.
- *Outcome:* Unnecessary thoracotomy.

CONCLUSION: DEFINING THE QUESTION

Constructing a searchable and answerable question that allows you to use the medical literature to solve problems is no simple matter. It requires a detailed understanding of the clinical issues involved in patient management. The 3 examples in this chapter illustrate that each patient encounter may trigger a number of clinical questions and that you must give careful thought to what you really want to know. Bearing the structure of the question in mind—patient or population, intervention or exposure, outcome, and, for therapy or harm questions, comparison—is helpful in arriving at an answerable question. Identifying the type of questions—therapy, harm, differential diagnosis, diagnosis, and prognosis—will not only ensure you choose the right question structure but also ensure that you are looking for a study with an appropriate design.

Careful definition of the question will provide another benefit: you will be less likely to be misled by a study that addresses a question related to that in which you are interested, but with 1 or more important differences. For instance, making sure that the study compares experimental treatment to current optimal care may highlight the limitations of trials that use a *placebo* comparator rather than an alternative

active agent. Specifying that you are interested in patient-important outcomes (such as long bone fractures) identifies the limitations of studies that focus on *substitute* or *surrogate end points* (such as bone density). Specifying that you are primarily interested in avoiding progression to dialysis will make you appropriately wary of a *composite end point* of progression to dialysis or doubling of serum creatinine level. You will not reject such studies out of hand, but the careful definition of the question will help you to critically apply the results to your patient care.

A final crucial benefit from careful consideration of the question is that it sets the stage for efficient and effective literature searching to identify and retrieve the current best evidence (see Chapter 4, Finding Current Best Evidence). Specifying a structured question and identifying an appropriate study design to answer it will allow you to select and use searching resources efficiently and thus enhance your evidence-based practice.

References

1. Wing RR, Bolin P, Brancati FL, et al; Look AHEAD Research Group. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med*. 2013;369(2):145-154.
2. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010;7(9):e1000326.
3. McKibbon KA, Wilczynski NL, Haynes RB. What do evidence-based secondary journals tell us about the publication of clinically important articles in primary healthcare journals? *BMC Med*. 2004;2:33.
4. Haynes RB, Cotoi C, Holland J, et al; McMaster Premium Literature Service (PLUS) Project. Second-order peer review of the medical literature for clinical practitioners. *JAMA*. 2006;295(15):1801-1808.
5. Haynes RB. ACP Journal Club: the best new evidence for patient care. *ACP J Club*. 2008;148(3):2.
6. Agoritsas T, Merglen A, Courvoisier DS, et al. Sensitivity and predictive value of 15 PubMed search strategies to answer clinical questions rated against full systematic reviews. *J Med Internet Res*. 2012;14(3):e85.

JAMAevidence
Using Evidence to Improve Care

4 Finding Current Best Evidence

Thomas Agoritsas, Per Olav Vandvik, Ignacio Neumann, Bram Rochwerg, Roman Jaeschke, Robert Hayward, Gordon Guyatt, and K. Ann McKibbon

IN THIS CHAPTER

(continued on following page)

INTRODUCTION

Searching for Evidence: A Clinical Skill

Searching for current best evidence in the medical literature has become a central skill in clinical practice.[1,2] On average, clinicians have 5 to 8 questions about individual patients per daily shift[3-5] and regularly use online *evidence-based medicine* (EBM) resources to answer them.[6-9] Some now even consider that "the use of search engines is as essential as the stethoscope."[10]

However, because of the increasing volume of new literature and speed of new research, finding useful evidence efficiently remains challenging. Approximately 2000 new articles are indexed in PubMed every day,[10] and although few of them directly inform clinical practice, as many as 75 are *randomized clinical trials* and 11 are *systematic reviews*.[11] These numbers explain why searching in PubMed is not the most efficient way to look for evidence-based answers. For example, when typing "stroke prevention in atrial fibrillation" in PubMed, you will see that current best evidence is literally lost in an output of almost 4000 citations, with a mix of trials, reviews, guidelines, and editorials that are impossible to screen for relevance during your daily practice.

Fortunately, numerous EBM resources now provide shorter and more efficient paths. These resources select, process, and organize the evidence; some, however, are more trustworthy than others. This chapter will help you navigate through existing EBM resources, distinguish the trustworthy from the less trustworthy, and maximize your chances of quickly finding answers based on current best evidence.

Start by Clarifying the Question

As we have seen in Chapter 3, What Is the Question? framing the question appropriately is an important prerequisite to any search. An initial distinction to make is whether you are asking a *background question* (eg, definition or pathophysiology

of a syndrome or mechanism of a treatment modality) or a *foreground question* (eg, targeted questions of therapy, *harm*, diagnosis, or *prognosis* that provide the evidentiary basis for decision making). Although some EBM resources also answer background questions, this chapter, and the *Users' Guides to the Medical Literature* overall, focuses on efficiently finding answers to foreground questions.

Foreground questions often arise in a form that does not facilitate finding an answer (see Chapter 3, What Is the Question?). A first step is to translate and structure the question into its components, using the *PICO* framework, which accounts for the patient or population, the intervention or exposure, the comparator, and the outcomes (see Chapter 3, Box 3-1). When framing your question, remember to consider all *patient-important outcomes*. Doing so will guide you in selecting the body of evidence that adequately addresses your patient's dilemma between benefits and harms that matter to your patient's decision.

Structuring the question will not only clarify what you are looking for but also help you formulate relevant search terms and combine them into search strategies, adapted to each type of EBM resource. We explore, toward the end of this chapter (see Translating a Question Into Search Terms), how the issues of question formulation and choice of search strategies become particularly crucial when evidence is harder to find using preappraised resources and you need to search in larger databases, such as PubMed. Finally, clarifying your question will help you search for appropriate study designs (see Chapter 3, What Is the Question?) and select corresponding search filters (eg, Clinical Queries) to reduce the number of citations in search outputs and enhance your chances of finding the best relevant evidence.

Searching the Medical Literature Is Sometimes Futile

Consider the following clinical question: "In patients with pulmonary embolism, to what extent do those with pulmonary

infarction have a poorer *health outcome* than those without pulmonary infarction?"

Before beginning your search to answer this question, you should think about how investigators would differentiate between those with and without infarction. Because there is no definitive method, short of autopsy, to make this differentiation, our literature search is doomed before we begin.

This example illustrates that the medical literature will not help you when no feasible study design or measurement tools exist that investigators could use to resolve an issue. Your search also will be futile if no one has conducted and published the necessary study. Before embarking on a search, carefully consider whether the yield is likely to be worth the time expended.

HOW EVIDENCE IS PROCESSED AND ORGANIZED INTO EBM RESOURCES

Evidence-based medicine resources are rapidly evolving and provide innovative solutions to deal with the production, summary, and appraisal of the evidence.[1] Numerous EBM resources are currently available. To clearly see how to navigate across available resources, we offer 3 classification systems: (1) *hierarchy of evidence* in primary studies, (2) level of processing of the evidence, and (3) categories of EBM resources (Figure 4-1). Together, these 3 classification systems describe the flow of evidence from primary studies to existing EBM resources.

Hierarchy of Evidence

At the level of *primary studies*, our first classification relates to the hierarchy of evidence (Figure 4-1, left box). For each type of question, EBM suggests a hierarchy of research designs to minimize the *risk of bias*. For questions regarding therapy or harm, well-conducted randomized clinical trials are superior

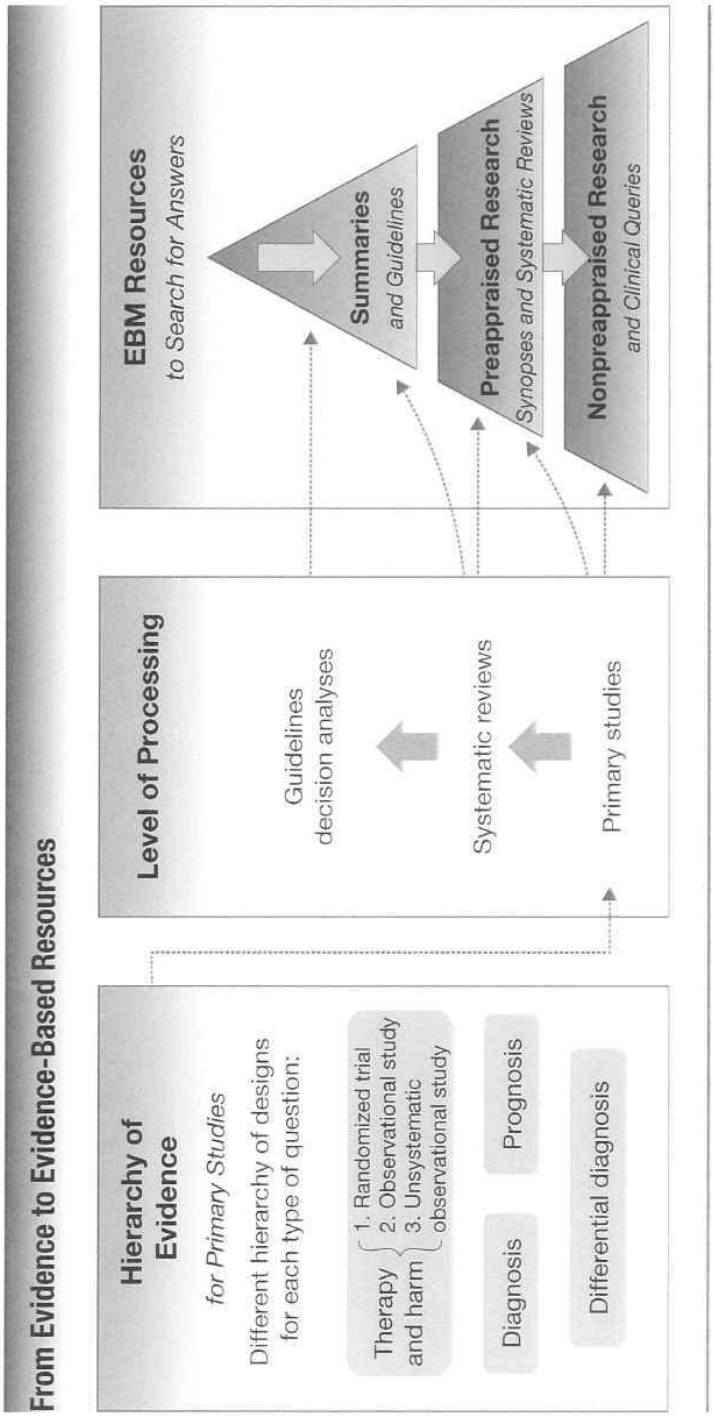

From Evidence to Evidence-Based Resources
Hierarchy of Evidence
for Primary Studies
Different hierarchy of designs for each type of question:
Therapy and harm
1. Randomized trial
2. Observational study
3. Unsystematic observational study
Diagnosis
Prognosis
Differential diagnosis
Level of Processing
Guidelines decision analyses
Systematic reviews
Primary studies
EBM Resources
to Search for Answers
Summaries
and Guidelines
Preappraised Research
Synopses and Systematic Reviews
Nonpreappraised Research
and Clinical Queries

FIGURE 4-1

to *observational studies*, which are superior to unsystematic clinical observations. Questions of diagnostic test properties, *differential diagnosis*, or prognosis require different hierarchies of study design (see Chapter 2, What Is Evidence-Based Medicine?).

Furthermore, within each type of design, some studies provide evidence of higher quality than others. The ideal EBM resource should facilitate access to studies with the most appropriate design and lowest risk of bias.

Levels of Processing

A second classification refers to the level of processing of the evidence (Figure 4-1, middle box). Primary studies can stand alone or be processed into systematic reviews. On the basis of clear eligibility criteria, authors of a systematic review conduct a comprehensive search for all primary studies, critically appraise their quality, and, when it is considered appropriate, provide a summary estimate of effects across studies. Well-conducted systematic reviews are far more useful than single primary studies because they represent the entire body of relevant evidence (see Chapter 14, The Process of a Systematic Review and Meta-analysis). Searching for systematic reviews instead of primary studies will save you substantial time and effort.

A further level of processing is to move from evidence (ideally systematic reviews) to recommendations for practice, as in *clinical practice guidelines* (see Moving From Evidence to Action). Providing recommendations requires judging the relative desirability of alternative courses of action. Therefore, this level of processing requires looking at the entire body of evidence, integrating and appraising the evidence from systematic reviews for each patient-important outcome, taking into account patient *values and preferences*, and being mindful of resource considerations. *Decision analyses* (see Chapter 17, How to Use a Patient Management Recommendation: Clinical Practice Guidelines and Decision Analyses) and health technology assessment

reports also may provide a similar level of processing of the evidence. As for primary studies, some guidelines are more trustworthy than others, and the ideal EBM resources should provide access to the more trustworthy ones.

Pyramid of EBM Resources

Although the 2 previous classifications—the hierarchy of evidence and level of processing—help you decide what type of evidence is likely to answer your question, they do not inform you of where to search for the evidence. For example, you may wonder where to search for high-quality systematic reviews. Should you start your search in the Cochrane Library, use review filters in PubMed, or look in the reference list of an online summary such as UpToDate? To make that choice, you need to understand how evidence is organized into a third classification: the *pyramid of EBM resources* (Figure 4-1, right box). From a practical perspective, resources can be viewed in 3 broad categories: summaries and guidelines, preappraised research, and nonpreappraised research.

Table 4-1 outlines these categories of EBM resources. Box 4-1 and the subsequent paragraph provide a fuller description of each category with examples of resources.

You can navigate efficiently across these different types of resources, as well as search all 3 categories simultaneously, using *federated search engines*, such as ACCESSSS (http://plus.mcmaster.ca/accessss), Trip (http://www.tripdatabase.com), Sum Search (http://sumsearch.org), or Epistemonikos (http://www.epistemonikos.org). Before we describe these search engines in detail, we will look at general criteria that will help clinicians choose which EBM resources to select given their question and which to avoid.

To complement resources that help you answer clinical questions, additional resources can link the evidence with your daily practice, such as *clinical decision support systems*[15] or context-specific access to online resources within electronic

TABLE 4-1

Categories of EBM Resources

Category	Layers[a]	Description	Examples
Summaries and guidelines	Online summary resources Databases of clinical practice guidelines	Summary of the body of evidence at a topic-level (not limited to a question, intervention, or outcome) Often with actionable recommendations for clinical decision making Regularly updated	UpToDate DynaMed Clinical Evidence Best Practice US National Guidelines Clearinghouse
Preappraised research	Synopses of systematic reviews Systematic reviews Synopses of studies	Structured abstracts or 1-page summaries of selected systematic reviews or studies Various degrees of preappraisal – Selection according to methodologic criteria – Clinicians' ratings – Clinicians' comments – Experts' structured appraisal Continuously updated Source of evidence alerts	ACP Journal Club McMaster *PLUS* DARE Cochrane Evidence Updates

(Continued)

TABLE 4-1

Categories of EBM Resources *(Continued)*

Category	Layers[a]	Description	Examples
Nonpreappraised research	Filtered studies	All primary studies with no preappraisal	PubMed (MEDLINE) CINAHL CENTRAL
	Unfiltered studies	Automatic filtering of databases for specific study designs or clinical content	Filters: Clinical Queries in PubMed
Federated searches	All layers of resources searched at once	Search engines that retrieve evidence from summaries and preappraised and nonpreappraised research, and organize the results accordingly	ACCESSSS Trip SumSearch Epistimonikos

Abbreviations: ACCESSSS, ACCess to Evidence-based Summaries, Synopses, Systematic Reviews and Studies; CENTRAL, Cochrane Central Register of Controlled Trials; CINAHL, Cumulative Index to Nursing and Allied Health Literature; DARE, Database of Abstracts of Reviews of Effects.

[a]These layers correspond to the 6-S pyramid from Haynes et al.[1,2]

BOX 4-1

Overview of EBM Resources

1. Summaries and guidelines.

Summaries are regularly updated online resources that aim to integrate the body of evidence at a topic level for several related questions. For example, a topic such as "treatment of type 2 diabetes mellitus in the elderly patient" will usually summarize evidence for drug therapy, strategies to control glycemic levels and avoid hypoglycemia, and lifestyle modification and the reduction of cardiovascular risk. These summaries often provide actionable recommendations for practice. Current examples widely used by clinicians include UpToDate (http://www.uptodate.com), DynaMed (https://dynamed.ebscohost.com), and Best Practice (http://bestpractice.bmj.com).

Guidelines follow a similar approach, usually focused on a specific topic or disease (eg, "antithrombotic therapy and prevention of thrombosis"[12]). Even more than summaries, guidelines are focused on providing recommendations for optimal patient management. Searching for available guidelines is more challenging because they are scattered across specialty journals and organization websites. A useful resource to search for guidelines is the US National Guideline Clearinghouse (http://www.guideline.gov), which includes guidelines from many countries.

2. Preappraised research.

When summaries or guidelines do not provide a satisfactory answer (eg, they provide an answer that is apparently not based on current best evidence or do not provide an answer at all), you must look directly at research findings, first from systematic reviews and then, if necessary, from primary studies. Many resources can prevent the unpleasant experience of searching the whole medical literature (at the risk of getting lost) or having to screen and read articles as PDFs. These resources select only

systematic reviews and studies that meet defined methodologic criteria and provide synopses—a 1-page structured abstract or description of reviews or studies. The degree and quality of preappraisal vary across resources. Some provide clinicians' ratings or short comments on relevance or newsworthiness, whereas others include a structured appraisal from experts. An example of the former is McMaster *PLUS* (Premium LiteratUre Service[13,14]; http://plus.mcmaster.ca/evidenceupdates), and examples of the latter are *ACP Journal Club* (http://acpjc.acponline.org) and DARE (Database of Abstracts of Reviews of Effects; www.crd.york.ac.uk/crdweb). You can access preappraised research in 2 complementary ways: by searching these specific databases for a given question and, for some of them, by subscribing to an e-mail alerting system. Personalized alerts are an efficient way to remain up-to-date on important new research in your area of interest (see, for example, BMJ EvidenceUpdates; http://plus.mcmaster.ca/evidenceupdates).

3. Nonpreappraised research.

Only when other sources have failed to provide an answer should you search for primary studies in the larger databases, such as MEDLINE (http://www.ncbi.nlm.nih.gov/pubmed) or CINAHL (http://www.cinahl.com). Because these databases include millions of articles, using them efficiently requires more advanced searching skills. Limiting your search with filters, such as Clinical Queries (http://www.ncbi.nlm.nih.gov/pubmed/clinical), provides a useful way to reduce the number of abstracts you need to review to identify the best evidence to address your clinical question.

medical records.[16] However, although some clinical decision support systems have the potential to improve processes of care or patient outcomes,[17] most cover only a limited range of clinical problems, are not necessarily based on current best evidence, and are often "homebuilt" so that their use is questionable.[1]

THREE CRITERIA FOR CHOOSING AN EBM RESOURCE

All EBM resources are not equally trustworthy, and none provide answers to all questions. Efficient searching involves choosing the appropriate resources for your clinical question—in much the same way you choose diagnostic tests appropriate for your patient's symptoms. Table 4-2 offers an initial guideline for making resource choices.

Based on Current Best Evidence

Many online summaries and guideline websites promote themselves as "evidence-based," but few have explicit links to research

TABLE 4-2

Selection Criteria for Choosing or Evaluating EBM Resources

Criterion	Description of Criterion
On the basis of current best evidence	How strong is the commitment to evidence to support inference?
	Does it have citations to references to all evidence summaries and recommendations?
	Is the process for keeping it up-to-date transparent and trustworthy?
	Is the quality of the evidence assessed?
	Is the strength of recommendations reported?
	Are numerical effect estimates reported for patient important outcomes?
Coverage and specificity	Does the resource cover my discipline and specific area of practice adequately?
	Does it cover questions of the type I am asking (eg, therapy, diagnosis, prognosis, harm)?
Availability and access	Is it readily available in all locations in which I would use it?
	Can I easily afford it?

findings. To judge the strength of the commitment to evidence to support inference, check whether you can distinguish statements that are based on high-quality vs low-quality evidence. If you cannot make this distinction, dismiss the resource altogether. Resources should provide citations to references to relevant research findings. Currency is important, and a simple way to judge whether the evidence is up-to-date is to look for the date of the most recent reference cited: if it is more than 2 years old, it is possible that future studies lead to a different conclusion.[1,18,19] Generally, the process for keeping a resource up-to-date should be transparent and trustworthy. A date stamp should accompany each summarized topic or piece of evidence (eg, "This topic last updated: Sep 17, 2013"), along with access to the explicit mechanism used to screen for related new findings. An opaque process should raise a red flag that the evidence may be partial, biased, or already outdated.

A summary or guideline should use a rating system to assess the risk of bias of cited studies and the quality of reviews. Resources that provide recommendations should be based on the entire body of existing evidence, ideally summarized in systematic reviews, and provide the benefits and harms of available options. The resources also should use an appropriate system to grade strength of recommendations and provide explicit judgments concerning underlying values and preferences (see Moving From Evidence to Action). Finally, to be actionable, the recommendations should report numerical effect estimates for patient-important outcomes to support clinical judgment and shared decision making at the point of care. For example, the ninth edition of the Antithrombotic Therapy and Prevention of Thrombosis guideline issued a weak recommendation for aspirin for primary prevention of cardiovascular events in people older than 50 years, based on moderate confidence in estimates of effect (grade 2B).[20] The authors provide numerical estimates: for example, in people at moderate risk of cardiovascular events, prophylactic aspirin resulted in 19 fewer myocardial infarctions per 1000 (from 26 fewer to 12 fewer) but 16 more major extracranial bleeds per 1000 (from 7 more to 20 more).

Coverage and Specificity

An ideal resource will cover most of the questions relevant to your practice—and not much more. However, few, if any, resources are sufficient as such a one-stop shop for the evidence you need,[18] and resources from the 3 levels of the pyramid of EBM resources are often complementary. The higher you look in the pyramid, the more time it takes for the resource developers to process and summarize the evidence at a topic level, making these resources potentially out of date. To be comprehensive in your searching, you will need to look for preappraised research for more recent evidence. Conversely, the lower you look in the pyramid, the larger, and often less specific, the resource. Thus, preappraised research limited to your area of practice, such as collections of synopses designed to help you keep up with information on the latest developments in a specific field or specialty—eg, *Evidence-Based Mental Health* (http://ebmh.bmj.com) or *Evidence-Based Nursing* (http://ebn.bmj.com)—may serve your needs efficiently.

The type of question also will affect your choice of a specific resource. For example, resources that focus on management issues informed mainly by randomized clinical trials, such as the Cochrane Database of Systematic Reviews, may not be ideal to answer questions of harm or rare adverse events. Similarly, background questions are more likely to be answered by summaries (eg, UpToDate or DynaMed) than preappraised research (eg, systematic reviews or synopses). For example, if you have background questions about the Middle East respiratory syndrome coronavirus, both UpToDate and DynaMed have a dedicated entry on the topic that summarizes its case definition and the incidence of recent clusters.

Availability and Access

The most trustworthy and efficient resources are frequently expensive, particularly those at the top of the pyramid of EBM resources. For example, an individual subscription to an online

summary often costs more than $250 annually. To establish your information resource regimen, you can map the EBM resources that are accessible to you through your university, school, or clinical institution and check whether they meet your information needs. Academic clinicians typically have access to the resources of their academic institution or hospital libraries, including the full texts of many studies and reviews.

Clinicians in private practice in high-income countries may have access to some resources through their professional associations but otherwise may be burdened by the cost of subscriptions. Some countries have national libraries that centralize access to many resources. Often, the institutional choice of resources is not made by practicing clinicians and may be guided by financial constraints. If an important resource is not available, make the case for it to your librarian (and suggest which other resources are less useful in practice).[1] If your institution is not willing to pay a license, consider subscribing individually. Health professionals in lower-income countries may have institutional access to information resources through the World Health Organization's Health InterNetwork Access to Research Initiative (http://www.who.int/hinari/en) or other organizations but otherwise face even greater financial obstacles to information resources. Additional strategies include seeking open-access journals, writing to authors for a reprint or e-print of their article, and contacting colleagues in academic centers who have access to more extensive library facilities.

Preappraised resources are sometimes expensive as well, and therefore we further describe how searching federated search engines, such as ACCESSSS or Trip, can give you an overview of the clinical content of various resources to help you make subscription decisions.

Free e-mail systems, such as BMJ EvidenceUpdates (http://plus.mcmaster.ca/evidenceupdates), can alert you to important new findings, although access to full texts will vary according to your institutional or personal licenses. An increasing number of full-text articles are accessible through PubMed or Google

Scholar or directly via open-access journals (eg, *CMAJ*, PLOS journals, and BioMed Central journals; see http://www.doaj.org for a directory of open-access journals). Many other journals provide free access to full-text articles 6 to 12 months after publication (eg, *BMJ*, *JAMA*, and *Mayo Clinic Proceedings*) or a portion of their content at the time of publication. However, focusing on free full-text articles and free Internet resources may give a partial and potentially biased view of current best evidence.[21]

Finally, ask your institution or professional organization how to access EBM resources at the point of care and obtain proxy server permission or remote access at home (eg, using a VPN connection). This will give you direct access to evidence on your smartphone and tablets and considerably enhance your *evidence-based practice*.

USING THE PYRAMID OF EBM RESOURCES TO ANSWER YOUR QUESTIONS

Numerous EBM resources are available, including many providers of summaries at the top level of the pyramid. Each has a different clinical scope, as well as different methodologic and editorial processes. No single portal lists them all, but many can be found through the New York Academy of Medicine (http://www.nyam.org/fellows-members/ebhc/eb_resources.html) or the Cochrane Collaboration (http://www.cochrane.org/about-us/webliography-evidence-based-health-care-resources) websites.

It is beyond the scope of this chapter to discuss the pros and cons of each resource. Instead, we will focus on how to navigate across the pyramid of EBM resources and discuss how these resources can complement each other. We provide examples of resources to illustrate important aspects both from research on evidence retrieval and from our own practice but do not aim to be comprehensive or prescriptive on which resource to use.

Summaries and Guidelines

Start your searches by using resources at the top of the pyramid for summaries and guidelines that address your question. These resources can provide a comprehensive view of the body of evidence at a topic level. Imagine, for example, that you are looking for antithrombotic therapies most appropriate for prevention of stroke in patients with atrial fibrillation. Available options include aspirin; other antiplatelet agents, such as clopidogrel; a combination of aspirin plus other antiaggregants; warfarin; or new anticoagulants, such as direct thrombin inhibitors or factor Xa inhibitors. To fully address your question from lower levels of the pyramid, you would need to retrieve, read, and integrate several systematic reviews or trials that cover all of the relevant comparisons and important outcomes. Summaries and guidelines aim to integrate this body of evidence and also often provide actionable recommendations for practice.

Table 4-3 lists examples of 10 widely used online summaries and their corresponding URLs. A recent analytical survey compared them on 3 aspects: the timeliness of updates, coverage of clinical topics, and quality of processing and reporting of the evidence.[19] At the time of this assessment (2011), the mean time since update ranged from 3.5 months (DynaMed) to 29 months (First Consult), and the percentage of clinical topics covered ranged from 25% (Clinical Evidence) to 83% (UpToDate). Quality substantially varied across the resources. For example, despite its limited coverage, the authors rated Clinical Evidence as the highest-quality resource. Because EBM resources continuously evolve, these numbers may be outdated but illustrate that online summaries can be complementary. Summaries also differ on their methods and commitment to providing actionable recommendations (eg, UpToDate now formulates recommendations using the *GRADE* [*Grading of Recommendations Assessment, Development and Evaluation*] framework, whereas

TABLE 4-3

Rank Orderings of 10 Online Summaries Compared on 3 Aspects[19]

Summary Resource	URL	Updates	Coverage, No. (%)	Quality
DynaMed	https://dynamed.ebscohost.com	1	3 (70)	2
UpToDate	http://www.uptodate.com	5	1 (83)	2
Micromedex	http://www.micromedex.com	2	8 (47)	2
Best Practice	http://bestpractice.bmj.com	3	4 (63)	7
Essential Evidence Plus	http://www.essentialevidenceplus.com	7	7 (48)	2
First Consult	http://www.firstconsult.com	9	5 (60)	2
Medscape Reference	http://reference.medscape.com	6	2 (82)	9
Clinical Evidence	http://clinicalevidence.bmj.com	8	10 (25)	1
ACP PIER	http://acpjc.acponline.org	4	9 (33)	7
PEPID	http://www.pepidonline.com	NA	6 (58)	10

Abbreviation: NA, data not available.

Reproduced with permission from the *Journal of Clinical Epidemiology*.[19]

Clinical Evidence focuses more on the summary of evidence, also using GRADE) and their editorial style (eg, structured bullet points in DynaMed and Best Practice vs textbook-like structured chapters in UpToDate).

Unlike summaries, most guidelines are scattered across journals or websites from individual countries or health organizations. One of the most comprehensive portals to search for guidelines is the US National Guideline Clearinghouse (http://www.guideline.gov). It includes the full text of many US guidelines and thousands of international guidelines. Searching is easy, although initial retrievals are often relatively large. Other international guidelines can be found through the UK National Institute for Health and Care Excellence (https://www.evidence.nhs.uk) or the Guideline International Network (http://www.g-i-n.net/library/international-guidelines-library).

Perhaps even more than other types of preappraised evidence, practice guidelines are extremely variable in their trustworthiness.[22,23] When you conduct your search, look for guidelines that are transparent in how they process the evidence and formulate recommendations (see Chapter 17, How to Use a Patient Management Recommendation: Clinical Practice Guidelines and Decision Analyses). The US National Guideline Clearinghouse website also allows side-by-side comparisons of the guideline process and components for guidelines on the same topic.

Finally, the top of the EBM pyramid also includes decision analyses, which process a body of evidence in a similar way to guidelines, map out the options with outcomes and probabilities, and help you judge the benefits and harms of different treatment options for a specific patient (see Chapter 17, How to Use a Patient Management Recommendation: Clinical Practice Guidelines and Decision Analyses). These decision analyses often can be found in stand-alone studies, *economic evaluation* reports, and health technology assessment reports. An efficient way to locate decision analyses is through the Centre for Reviews and Dissemination at the UK University of York (http://www.crd.york.ac.uk/crdweb) by selecting the search filters "HTA" and "NHS EED" (for economic evaluation).

Preappraised Research

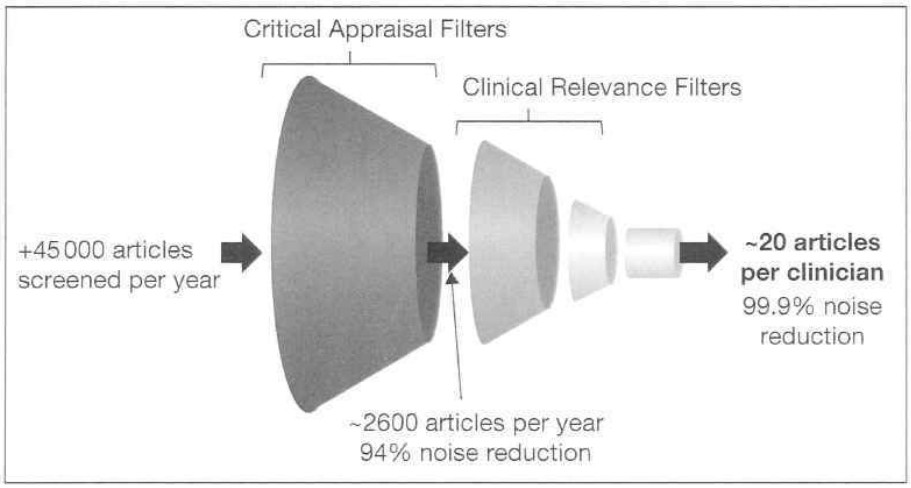
If you do not find a satisfactory answer in summaries or guidelines, either because your question is not covered or because you have reasons to doubt what you found, you may need to look for preappraised research. You also might search preappraised research to look for more recent evidence published since the summary or guideline was last updated.[24] You might wonder how often this additional searching would be worth the trouble. A recent study of the quality of online summaries found that, on average, new high-quality evidence providing potentially different conclusions than existing summaries was available for approximately 52% of the topics evaluated in UpToDate, 60% in Best Practice, and 23% in DynaMed.[18] This potential discrepancy between newly published evidence and existing recommendations would occur more frequently, and likely with greater adverse consequences, for most clinical practice guidelines, which are usually updated every 2 to 8 years.[25]

Consider, for example, the question of whether cardiac resynchronization therapy (CRT) reduces mortality in patients with heart failure and a narrow QRS complex. An initial search in mid-September 2013 in DynaMed or UpToDate provided an excellent summary of available evidence on the efficacy of CRT according to the degree of heart failure and the QRS duration but did not yet identify a more recent trial published in the *New England Journal of Medicine*.[26] This trial found that CRT did not reduce the composite rate of death or hospitalization for heart failure and actually may increase mortality. This important new evidence will of course be included in subsequent updates, but this process typically takes a couple of months to up to 29 months, depending on the online summary.[19]

A quick and efficient way to find preappraised research is to search specific databases, which include only studies and reviews that are more likely to be methodologically sound and clinically relevant. Figure 4-2 shows a typical example of this improved selection process from McMaster *PLUS* (Premium

FIGURE 4-2

Example of Preappraised Research: McMaster PLUS



Reproduced with permission of the Health Information Research Unit, McMaster University.

LiteratUre Service), a large database created by the McMaster Health Knowledge Refinery (http://hiru.mcmaster.ca/hiru/HIRU_McMaster_PLUS_Projects.aspx). The selection process used is as follows: trained research staff continually critically appraise more than 45 000 articles per year, from more than 125 empirically selected, high-quality clinical journals, and identify studies and systematic reviews that meet prespecified methodologic standards. For example, studies of prevention or therapy must have random allocation, a follow-up rate of at least 80%, and at least 1 patient-important outcome. These selected articles are then rated for relevance and newsworthiness by frontline clinicians from around the globe.[27] McMaster *PLUS* is thus a continuously updated database of more than 32 000 highly selective articles (with approximately 3300 added every year) that also feeds several other EBM resources and journals (eg, *ACP Journal Club*, Clinical Evidence, and DynaMed). A simple way to access McMaster *PLUS* is through the free search

engine of BMJ EvidenceUpdates (http://plus.mcmaster.ca/EvidenceUpdates/QuickSearch.aspx) or through the McMaster search engine, ACCESSSS, which we discuss further below (see Searching All Levels of the Pyramid at the Same Time). McMaster *PLUS* also has distinct databases for nursing (http://plus.mcmaster.ca/np) and rehabilitation studies (http://plus.mcmaster.ca/rehab).

In a further level of preappraisal, the more clinically relevant studies and systematic reviews are selected to become *synopses* (<1% of the initial selection). These synopses are usually a 1-page, structured summary of the research findings, along with a brief commentary from an expert in the field. You can find various types of synopses in specialized evidence-based *secondary evidence-based journals.* Figure 4-3 shows an example of a synopsis of a systematic review from *ACP Journal Club* (http://acpjc.acponline.org) on the impact of eplerenone on mortality compared with other aldosterone antagonists in heart failure. The abstract summarizes salient elements of the methods and results and an expert provides a commentary. This appraisal is not always systematic or as thorough as a full critical appraisal, but it usually provides the gist of the strengths and weaknesses of a study. Similar resources include *Evidence-Based Medicine* (http://ebm.bmj.com), *Evidence-Based Mental Health* (http://ebmh.bmj.com), *Evidence-based Oncology* (www.sciencedirect.com/science/journal/13634054), or POEMs (Patient-Oriented Evidence that Matters) (www.essentialevidenceplus.com/content/poems). The New York Academy of Medicine keeps a current list of specialized EBM journals in many health care disciplines (www.nyam.org/fellows-members/ebhc/eb_publications.html).

When searching preappraised research, make synopses of systematic reviews your first priority because they summarize the body of evidence on a question. In addition to evidence-based journals, you can find synopses of systematic reviews in DARE (Database of Abstracts of Reviews of Effects) (http://www.cochrane.org/editorial-and-publishing-policy-resource/database-abstracts-reviews-effects-dare). If no synopses answer

FIGURE 4-3

Example of Synopsis of a Systematic Review From *ACP Journal Club*

Therapeutics

Review: Eplerenone is not more effective for reducing mortality than other aldosterone antagonists

Chatterjee S, Moeller C, Shah N, et al. Eplerenone is not superior to older and less expensive aldosterone antagonists. Am J Med. 2012. 125:817-25.

Clinical impact ratings: ★★★★★☆☆ ★★★★★☆☆

Question
In patients with left ventricular (LV) dysfunction, what is the relative efficacy of eplerenone and other aldosterone antagonists (AAs)?

Review scope
Included studies compared eplerenone or other AAs with control (placebo, angiotensin-converting enzyme inhibitor, angiotensin-receptor blocker, or β-blocker) in patients > 18 years of age with symptomatic or asymptomatic LV dysfunction, had ≥ 8 weeks of follow-up, and reported ≥ 1 outcome of interest. Studies comparing AAs with each other were excluded. Outcomes were all-cause mortality, cardiovascular (CV) mortality, gynecomastia [per trial definition in individual studies]*, and hyperkalemia [serum potassium > 5.5 mEq/L]*.

Review methods
MEDLINE, EMBASE/Excerpta Medica, CINAHL, and Cochrane Central Register of Controlled Trials (all to Jul 2011); reference lists; and reviews were searched for randomized controlled trials (RCTs). 16 RCTs (*n* = 12 505, mean age 55 to 69 y, 54% to 87% men) met selection criteria. 4 RCTs included patients after acute myocardial infarction LV dysfunction, and 12 included patients with heart failure. Study drugs were spironolactone (10 RCTs), canrenone (3 RCTs), and eplerenone (3 RCTs). Risk for bias (Cochrane criteria) was low for 8 RCTs, intermediate for 7, and high for 1.

Main results
Eplerenone and other AAs reduced all-cause mortality and CV mortality compared with no AA (Table). Eplerenone increased risk for hyperkalemia, and other AAs increased risk for gynecomastia, compared with no AA (Table). Based on an indirect comparison, other AAs reduced mortality more than eplerenone (P = 0.009).

Eplerenone or other AAs vs control in patients with left ventricular dysfunction†

Outcomes	Number of trials (*n*)	Weighted event rates		At 2 to 24 mo	
		Eplerenone	Control‡	RRR (95% CI)	NNT (CI)
All-cause mortality	2 (9369)	14%	16%	15% (7 to 22)	41 (27 to 86)
CV mortality	2 (9369)	12%	14%	17% (8 to 25)	42 (29 to 88)
Gynecomastia	2 (5061)	0.49%	0.46%	26% (–27 to 57)	NS
				RRI (CI)	NNH (CI)
Hyperkalemia	3 (9489)	6.1%	3.8%	72% (19 to 147)	37 (19 to 140)
		Other AAs§	Control‡	RRR (95% CI)	NNT (CI)
All-cause mortality	12 (3569)	19%	25%	26% (17 to 34)	16 (12 to 24)
CV mortality	4 (2553)	26%	34%	25% (16 to 33)	12 (9 to 18)
				RRI (CI)	NNH (CI)
Gynecomastia	6 (3279)	5.4%	0.86%	526% (238 to 1057)	23 (11 to 49)
Hyperkalemia	10 (3342)	8.1%	4.5%	80% (–17 to 293)	NS

AA = aldosterone antagonist; CV = cardiovascular; NS = not significant; other abbreviations defined in Glossary. Weighted event rates, RRR, RRI, NNT, NNH, and CI calculated from crude event rates and risk ratios in article using a random-effects model.

†Vasale, angiotensin-converting enzyme inhibitor, angiotensin-receptor blocker, or β-blocker.

‡Other AAs were spironolactone or canrenone.

Conclusion
Based on an indirect comparison, eplerenone is not more effective for reducing mortality in adults with left ventricular dysfunction than other aldosterone antagonists.

**Information provided by author.*

Source of funding: No external funding.

For correspondence: Dr. S. Chatterjee, Maimonides Medical Center, Brooklyn, NY, USA. E-mail saurachatterjeemd@gmail.com.

Commentary
In their thorough review of the use of AAs in systolic heart failure, Chatterjee and colleagues conclude that data are insufficient to recommend eplerenone over spironolactone. Only 3 large outcome trials actually address the issue: RALES, assessing spironolactone (1), and EPHESUS (2) and EMPHASIS-HF (3), assessing eplerenone. Although the populations evaluated in each study were quite different, the relative reductions in mortality were similar (25%, 14%, and 19%, respectively). Indirect comparisons of drug efficacy across clinical trials with different patient populations and study protocols are challenging. Without head-to-head trials of AAs, we should not draw conclusions about their relative efficacy.

Chatterjee and colleagues confirm that spironolactone increases risk for gynecomastia. Hyperkalemia is a known adverse effect of any AA, although potassium increases were "not clinically important" in RALES (1). After RALES was published, however, there was a marked increase in the number of spironolactone prescriptions, with an increase in hyperkalemia and associated mortality (4). Gynecomastia can be distressing to male patients, but hyperkalemia may be fatal to either sex.

A strict, evidence-based practitioner would base drug and dosage selection on the clinical trial most closely matching a patient's presentation. While waiting for a definitive head-to-head trial—noting that benefits seem similar in the studied populations—I start with the less expensive spironolactone, switching to eplerenone if troublesome sexual adverse effects develop (while closely monitoring potassium!).

Ellis Lauer, MD, FACC
Mid Valley Cardiology, New York University School of Medicine
Kingston, New York, USA

References

1. **Pitt B, Zannad F, Remme WJ, et al.** The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. N Engl J Med. 1999;341:709-17.
2. **Pitt B, Remme W, Zannad F, et al; Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study Investigators.** Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. N Engl J Med. 2003;348:1309-21.
3. **Zannad F, McMurray JJ, Krum H, et al; EMPHASIS-HF Study Group.** Eplerenone in patients with systolic heart failure and mild symptoms. N Engl J Med. 2011;364:11-21.
4. **Juurlink DN, Mamdani MM, Lee DS, et al.** Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. N Engl J Med. 2004;351:543-51.

Reproduced with permission of *ACP Journal Club*.[26]

your question, move to a direct search for other systematic reviews. A useful resource is the Cochrane Library (http://www.thecochranelibrary.com).

Regardless of the resources you use, remember that preappraisal and the collection of these synopses can only increase the likelihood of finding sound evidence efficiently. It does not guarantee it. You should also apply your own critical appraisal to the research findings that are summarized, as explained throughout the *Users' Guides to the Medical Literature.*

Alerts to Important New Evidence

In addition to building continuously updated databases of preappraised research, an increasing number of resources offer e-mail *alerting services*. To make the volume of new evidence manageable, these alerts are usually tailored to your information needs when you register (eg, clinical disciplines, quality choices, and frequency of alerts).

For example, the whole process leading to McMaster *PLUS*, including clinicians' ratings for relevance and newsworthiness, results in up to a 99.9% noise (non–clinically relevant) reduction and produces a manageable stream of approximately 20 to 50 key articles per year in a clinical area that may influence your practice (Figure 4-2).[28] You can receive these alerts by subscribing to BMJ EvidenceUpdates or ACCESSSS. Several other free or fee-based alerting systems are available for both a wide scope of disciplines (eg, NEJM Journal Watch, http://www.jwatch.org) and specific subspecialties (eg, OrthoEvidence, http://www.myorthoevidence.com). When using any of these alerting resources, check whether their process of selecting and appraising the evidence is explicit, trustworthy, and meeting your needs.

Nonpreappraised Research

Only when summaries, guidelines, and preappraised research have failed to provide an answer should you search among the

tens of millions of nonpreappraised research articles. They are stored in many different databases (the ones usually searched in systematic reviews), such as PubMed's MEDLINE, EMBASE, CINAHL, or Web of Science. These databases can be accessed directly or through different search engines. Some search engine companies, such as Ovid (http://www.ovid.com), are designed to facilitate complex search strategies, such as those done by medical librarians or authors of systematic reviews. For clinical purposes, PubMed is the most popular search engine, providing free access to the entire MEDLINE database (http://www.ncbi.nlm.nih.gov/pubmed).

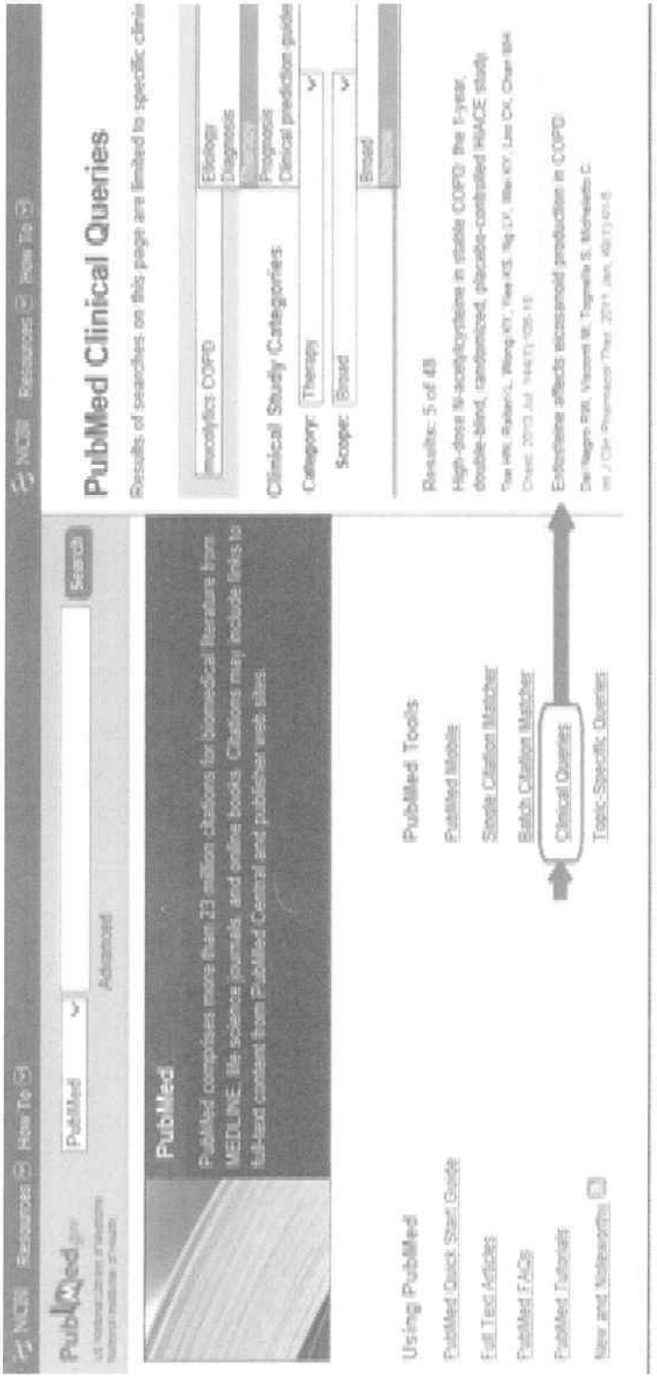
Consider, for example, the question of whether statins can prevent dementia. Summaries and preappraised research provide limited or selected evidence to answer that question. Because of its volume, searching PubMed to find relevant evidence requires more advanced searching skills, particularly in the choice and combination of search terms. Simple searches typically yield large outputs with few easily identified relevant studies in the first pages.

To limit irrelevant studies in the outputs, use methodologic filters, such as Clinical Queries. As shown in Figure 4-4, instead of typing your search terms on the main page of PubMed, select Clinical Queries or go directly to http://www.ncbi.nlm.nih.gov/pubmed/clinical. Empirically validated "methods" search terms are added to your search, according to your type of question. For example, Table 4-4 lists the filters used for questions of therapy that facilitate the retrieval of randomized clinical trials.[29] Two filters are available for each search category, 1 broad (sensitive) and 1 narrow (specific), the latter being more adapted to clinical practice. Use of a filter will increase the proportion of relevant studies from approximately 2% to 30% in the first 2 pages of PubMed's output (first 40 citations).[2] Similar filters are available for questions of diagnosis, etiology, prognosis, and *clinical prediction rules*.

Table 4-5 lists similar broad and narrow filters to find systematic reviews from PubMed.[30] In contrast with Clinical Queries,

FIGURE 4-4

Clinical Queries in PubMed: Accessing From Main Page and Choosing of Filter (Category and Scope)



Reproduced with permission of the US National Library of Medicine and PubMed.

TABLE 4-4

Clinical Queries "Therapy" Filter: Performance and Strategy Used[a]

	Sensitivity, %	Specificity, %	PubMed Equivalent
Broad filter	99	70	((clinical[Title/Abstract] AND trial[Title/Abstract]) OR clinical trials[MeSH Terms] OR clinical trial[Publication Type] OR random*[Title/Abstract] OR random allocation[MeSH Terms] OR therapeutic use[MeSH Subheading])
Narrow filter	93	97	(randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract]))

Abbreviation: MeSH, medical subject headings.

[a]Reproduced with permission of the US National Library of Medicine and PubMed.

these filters are not implemented in PubMed; the search strategy needs to be copied and pasted right after your search. Going back to our example of the search phrase "statins for the prevention of dementia," an unfiltered search retrieves hundreds of citations that cannot be reliably screened in clinical practice. When adding the narrow filter of Table 4-5 to your search, the output shrinks to 19 citations (in October 2013), and a quick review will identify 6 systematic reviews, including 1 Cochrane Review, updated in 2009, and the most recent review, published in *Mayo Clinic Proceedings* in September 2013, Statins and Cognition: A Systematic Review and Meta-analysis of Short- and Long-term Cognitive Effects. The University of York keeps a comprehensive list of available filters and the publications that describe

TABLE 4-5

Filters to Retrieve Systematic Reviews From PubMed[a30]

	Sensitivity, %	Specificity, %	PubMed Equivalent
Broad filter	99.9	52	search*[Title/Abstract] OR meta analysis[Publication Type] OR meta analysis[Title/Abstract] OR meta analysis[MeSH Terms] OR review[Publication Type] OR diagnosis[MeSH Subheading] OR associated[Title/Abstract]
Narrow filter	71	99	MEDLINE[Title/Abstract] OR (systematic[Title/Abstract] AND review[Title/Abstract]) OR meta analysis[Publication Type]

Abbreviation: MeSH, medical subject headings.

[a]These filters are not implemented in PubMed; the search strategy needs to be copied and pasted right after the search to optimally filter systematic reviews. Reproduced with permission from the *BMJ*.

their development and validations. For example, in addition to the ones we have already discussed, you will find filters for adverse events, economic evaluation, observational studies, and even qualitative studies (https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home/search-filters-by-design).

Another useful database for clinical practice is the Cochrane Controlled Trials Registry, the largest electronic compilation of controlled trials, built from MEDLINE, EMBASE, and other sources, including hand searches of most major health care journals. Because it includes only trials, this registry is the fastest, most reliable method of determining whether a controlled trial has been published on any topic. You can search the registry

in the Cochrane Library's advanced search function (http://onlinelibrary.wiley.com/cochranelibrary/search; select "Search Limits," then "Trials"). However, to access the full text of articles, you will need a subscription to the Cochrane Library or several Ovid Evidence-Based Medicine Review packages of databases (http://www.ovid.com/site/catalog/DataBase/904.jsp).

Searching All Levels of the Pyramid at the Same Time

At this point, you may wonder if you can search across all levels of the pyramid of resources, instead of having sequential searches in different resources to get the current best evidence. Federated search engines do this easily. One of the most comprehensive and transparent federated resources is ACCESSSS (http://plus.mcmaster.ca/accessss). Typing a single question in ACCESSSS will run parallel searches in major resources from each level of the pyramid, from summaries to all types of preappraised research and all Clinical Queries filters in PubMed. Table 4-6 presents the resources searched by ACCESSSS. Results are given in 1 page organized by level in the pyramid of EBM resources, with the most relevant and useful for clinical practice on the top (see Figure 4-5). Subscribing to ACCESSSS is free, although access to the full text of some resources will depend on institutional or personal subscriptions. To directly link your own subscriptions to all features of ACCESSSS, you can ask to add your institution to its list.

Other interesting and free federated searches that similarly search multiple resources at more or less each level of the pyramid are available. Instead of looking into summaries at the top, Trip (http://www.tripdatabase.com) has an algorithm to retrieve clinical practice guidelines, classified by country, along with many sources of synopses and other preappraised and nonpreappraised research. Its navigation is easy, and additional interesting features include the ability to structure your search with PICO (patient, intervention, comparator, outcome) and tailor your search to issues in developing countries. SumSearch (http://sumsearch.org) shares similar

TABLE 4-6

Example of a Federated Search: EBM Resources Searched in Parallel in ACCESSSS[a]

Summaries	DynaMed
	UpToDate
	Best Practice
	ACP PIER
Preappraised research	
Synopses of systematic reviews	*ACP Journal Club* DARE
Systematic reviews	McMaster *PLUS* (including Cochrane reviews)
Synopses of studies	McMaster *PLUS*
Nonpreappraised research	
Filtered studies	Clinical Queries in PubMed
Unfiltered studies	PubMed (MEDLINE)

Abbreviations: ACCESSSS, ACCess to Evidence-based Summaries, Synopses, Systematic Reviews and Studies; DARE, Database of Abstracts of Reviews of Effects; EBM, evidence-based medicine.

[a]Reproduced with permission of the Health Information Research Unit, McMaster University.

features, particularly for the retrieval of practice guidelines, but it organizes output according to level of processing (original studies, systematic reviews, and guidelines; Figure 4-1, middle box). SumSearch is linked to alerts from NEJM JournalWatch (http://www.jwatch.org). Finally, Epistemonikos (http://www.epistemonikos.org) is innovative both in simultaneously searching multiple resources and in indexing and interlinking relevant evidence. For example, Epistemonikos connects systematic reviews and their included studies and thus allows clustering of systematic reviews based on the primary studies they have in common. Epistemonikos is also unique in offering an appreciable multilingual user interface, multilingual search, and translation of abstracts in more than 9 languages.

FIGURE 4-5

Output of a Federated Search in ACCESSSS

Reproduced with permission of the Health Information Research Unit, McMaster University.

When to Use Google

Google (http://www.google.com) has brought a revolution in the way we search the Internet. Its powerful algorithm retrieves answers to any type of question. Many factors seem to influence its output, including the relevance to your query but also

the number of times a specific website has been previously accessed or cited, the computer IP and server from which you conduct your search, your nationality, and possibly other financial and nonfinancial interests. Because of its lack of transparency, Google is not a reliable way to filter current best evidence from unsubstantiated or nonscientifically supervised sources. When searching the Web, be aware that you are not searching defined databases but rather surfing the constantly shifting seas of electronic communications. The material you need that is supported by evidence may not float to the surface at any particular time.

On the other hand, "Googling" can be useful for defined purposes. It is often the fastest way to answer general background questions, often through multilingual resources such as Wikipedia (http://www.wikipedia.org), or research new topics, conditions, or treatments that have attracted media attention before being included in any EBM resources (eg, at the time of viral outbreaks around the globe, you may have wondered what Middle East respiratory syndrome coronavirus is). Google also can help you refine the wording of your search terms by rapidly finding 1 relevant citation. For example, you might want to learn whether incretins are associated with pancreatic cancer, but you are unclear about the different types of incretins. By searching Google and Wikipedia, you will rapidly remember how to spell (or copy and paste) dipeptidyl peptidase 4 inhibitor or glucagon-like peptide 1 analogs. Finally, Google can be a surprisingly powerful tool to search for uncommon patterns of symptoms and findings by simply typing them together as a query. These uncommon combinations would usually retrieve little or no information in most medical databases. Google can sometimes find the rare citation that would give you a clue about that syndrome.

A better alternative to Google for answering foreground questions is Google Scholar, which applies Google algorithms to scholarly literature (http://www.google.com/scholar). Although Google Scholar's search algorithms are not transparent,

comparisons have found Google Scholar to be comparable to other databases,[31] and an analysis has found increasing evidence that Google Scholar retrieves twice as many relevant articles as PubMed, with almost 3 times greater access to free full-text articles,[32] as well as access to conference abstracts that might be useful for rare topics. Google Scholar has a complex searching system, and the help feature provides useful guidance in refining your searches (http://scholar.google.com/intl/en/scholar/help.html).

TRANSLATING A QUESTION INTO SEARCH TERMS

How to Choose and Combine Search Terms

Table 4-7 illustrates how you can break down a question into its PICO components and corresponding search terms. You next choose and combine search terms into a variety of search strategies, adapted to each resources. One advantage of searching the top EBM resources is that you can keep searches simple because the databases are highly selective and relatively small. One or 2 search terms for the population or problem and for your intervention or exposure will find most relevant topics. For example, if you are interested in the impact of mucolytics on patients with chronic obstructive pulmonary disease (COPD) who are stable, simply searching with the terms "COPD mucolytic" in summaries (eg, UpToDate) and preappraised research (eg, DARE) will usually suffice. Being too specific in your search can cause you to lose important information. In contrast, searching nonpreappraised research (eg, PubMed) usually requires more specific and structured searches.

To find the evidence you need in large databases, your search terms should closely relate to the components of your PICO question (see Chapter 3, What Is the Question?). For some components, the corresponding search terms are straightforward.

TABLE 4-7

Combining Search Terms Into Different Search Strategies

PICO Components		Potential Search Terms
P	Patients with stable chronic bronchitis	COPD OR (chronic bronchitis)
I	Any mucolytic agent	Mucolytic
C	Placebo (and current best care)	Placebo
O	Number of exacerbation, mortality	Exacerbation OR mortality
Level of the Pyramid		**Examples of Search Strategies[a]**
Summaries and preappraised research		Chronic bronchitis mucolytic
		COPD mucolytic
Nonpreappraised research		COPD mucolytic exacerbation
		(COPD OR (chronic bronchitis)) AND mucolytic
		(COPD OR (chronic bronchitis)) AND mucolytic AND exacerbation
		(COPD OR (chronic bronchitis)) AND mucolytic AND (exacerbation OR mortality)

Abbreviation: COPD, chronic obstructive pulmonary disease; PICO, patient or population, intervention or exposure, comparator, and outcome.

[a]OR and AND are Boolean operators in these searches.

For example, if your population is patients with diabetes, you may simply use "diabetes" or "diabetic." Other components of PICO may prove more challenging, such as "antithyroid drug therapy" as an intervention. Indeed, you might choose "antithyroid" as a single term or consider combining several drugs, such as "carbimazole OR propylthiouracil OR methimazole." Notice that the latter example combines search terms with "OR" in capital letters to signify this is a *Boolean operator*: the search

will retrieve studies for either of these treatments. In contrast, adding no operator actually corresponds to linking search terms with "AND." For example, typing "neuraminidase inhibitors" is equivalent to typing "neuraminidase AND inhibitors" and will retrieve only studies that include both terms, instead of all studies that include any type of inhibitor.

Efficient wording of search terms is based in part on your familiarity with the topic but is also based on trial and error. The *Medical Subject Headings* (MeSH) Thesaurus (http://www.nlm.nih.gov/mesh/MBrowser.html) can help you find words generally used by indexers for a given medical concept. A quick Google search often can give you a sense of appropriate wording in a faster way. If you are surprised that a search yields little relevant evidence, ask yourself if you misspelled a term or were too specific (eg, adding too many words that will automatically be linked with "AND"). Definitions also can differ. For example, in MeSH, "ventilation" refers to "supplying a building or house, their rooms and corridors, with fresh air." "Pulmonary ventilation" is the preferred term for clinicians because it indicates "the total volume of gas inspired or expired per unit of time, usually measured in liters per minute."

Broad vs Narrow Searches

Table 4-8 indicates how to refine your search. If you initially found little evidence, you can broaden your search (eg, increase its sensitivity) by adding synonyms for each concept or using truncated terms (eg, diabet* will retrieve diabetes, diabetic, and many other similar terms with different endings). Conversely, if your initial search retrieved too many citations to be screened, you can narrow your search (eg, increase its specificity) by linking more PICO components with "AND" or by adding limits and methodologic filters (eg, narrow Clinical Queries; http://www.ncbi.nlm.nih.gov/pubmed/clinical). More sophisticated approaches include entering PICO components sequentially

TABLE 4-8

Refining the Search Strategy[1,19,30,34]

Ways to Increase Sensitivity	Ways to Increase Specificity
Many search terms for a similar PICO component, linked with "OR"	More PICO concepts linked with "AND":
	(P) AND (I) AND (C) AND (O)
Truncated terms, wildcards (eg, diabet*, wom?n)	Use of NOT to exclude irrelevant terms
Synonyms (pressure sore, decubitus ulcer)	Use of NOT as Boolean operator
Variant spelling (tumour, tumor)	Limits (date, age group, etc)
Explosion of MeSH terms	Methodologic filters (Clinical Queries)
Use of PubMed "Related citations" or bibliography of relevant articles	Content filters (topic or disease specific)

Abbreviation: MeSH, medical subject headings; PICO, patient or population, intervention or exposure, comparator, and outcome.

according to their importance to obtain a manageable number of articles in large databases, such as PubMed.[33]

Finding Related Articles

When your PubMed search seems laborious, a useful trick is to find at least 1 potentially relevant article to your question and use the "Related citations" feature, as highlighted in Figure 4-6. It will automatically look for other articles that are similar in their titles, abstracts, and index terms. You then can screen the new output and select "Related citations" for each potentially relevant article you find. To keep track of potentially relevant citations, send them to the PubMed clipboard as you screen, and they will be labeled as items in the clipboard (Figure 4-6). This strategy may help you gather relevant articles rapidly in a snowball sampling.

FIGURE 4-6

Features in PubMed: Related Citations and Clipboard

A, Link to "Related citations" from a relevant article. B, Dialogue box allowing user to send relevant articles to the clipboard. C, After having sent an article to the clipboard, it is labeled so in the output.

Getting Help

Finally, because of the complexity and interconnections of medical databases, some searches simply require the help of information specialists. In anticipation of such cases in your clinical practice, befriend your medical librarians. They can be a great resource to help answer difficult questions or those that require elaborate search strategies.

BOX 4-2

Tips to Help Improve Searching Skills

With the pyramid of EBM resources in mind, map the EBM resources that are accessible to you through your affiliations or personal subscriptions.

Choose which resources you would like to explore next, according to your information needs and the criteria described in this chapter.

Bookmark these resources in the browsers of all of your devices—your desktop computer, smartphone, or tablet. Find out if you can get remote access from your institution and implement it so that access is automatic.

Subscribe to an e-mail alerting system for newly published evidence that is transparent and trustworthy.

Train yourself on questions that are familiar to you and compare EBM resources.

Keep track of your questions. It can enhance your learning and help you reflect back on your evidence-based practice.

Finally, always keep the patient perspective. This will help you focus on the appropriate body of evidence that informs all patient important outcomes, instead of being driven by the evidence that is first presented to you.

CONCLUSION: IMPROVING YOUR SEARCHING SKILLS IN DAILY PRACTICE

Box 4-2 presents a few practical tips to help you improve your searching skills in daily practice. Because of the continuous flow of new research findings of variable quality, finding current best evidence is challenging. However, this process has been greatly facilitated by the development of numerous EBM resources that can provide fast answers at the point of care. No resource is sufficient for all information needs, and you will need to use several in combination to find current best evidence. This chapter provides guidance on how to navigate across the pyramid of resources efficiently, ideally by using federated search engines.

References

1. Straus SE. *Evidence-Based Medicine: How to Practice and Teach EBM*. 4th ed. New York, NY: Elsevier/Churchill Livingstone; 2011.
2. Agoritsas T, Merglen A, Courvoisier DS, et al. Sensitivity and predictive value of 15 PubMed search strategies to answer clinical questions rated against full systematic reviews. *J Med Internet Res*. 2012;14(3):e85.
3. Green ML, Ciampi MA, Ellis PJ. Residents' medical information needs in clinic: are they being met? *Am J Med*. 2000;109(3):218-223.
4. González-González AI, Dawes M, Sánchez-Mateos J, et al. Information needs and information-seeking behavior of primary care physicians. *Ann Fam Med*. 2007;5(4):345-352.
5. Graber MA, Randles BD, Ely JW, Monnahan J. Answering clinical questions in the ED. *Am J Emerg Med*. 2008;26(2):144-147.
6. Hoogendam A, Stalenhoef AF, Robbé PF, Overbeke AJ. Answers to questions posed during daily patient care are more likely to be answered by UpToDate than PubMed. *J Med Internet Res*. 2008;10(4):e29.
7. Hoogendam A, Stalenhoef AF, Robbé PF, Overbeke AJ. Analysis of queries sent to PubMed at the point of care: observation of search behaviour in a medical teaching hospital. *BMC Med Inform Decis Mak*. 2008;8:42.
8. Thiele RH, Poiro NC, Scalzo DC, Nemergut EC. Speed, accuracy, and confidence in Google, Ovid, PubMed, and UpToDate: results of a randomised trial. *Postgrad Med J*. 2010;86(1018):459-465.

9. McKibbon KA, Fridsma DB. Effectiveness of clinician-selected electronic information resources for answering primary care physicians' information needs. *J Am Med Inform Assoc*. 2006;13(6):653-659.
10. Glasziou P, Burls A, Gilbert R. Evidence based medicine and the medical curriculum. *BMJ*. 2008;337:a1253.
11. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010;7(9):e1000326.
12. Guyatt GH, Akl EA, Crowther M, Schunemann HJ, Gutterman DD, Zelman Lewis S. Introduction to the ninth edition: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(2 suppl):48S-52S.
13. Haynes RB, Cotoi C, Holland J, et al; McMaster Premium Literature Service (PLUS) Project. Second-order peer review of the medical literature for clinical practitioners. *JAMA*. 2006;295(15):1801-1808.
14. Holland J, Haynes RB; McMaster PLUS Team Health Information Research Unit. McMaster Premium Literature Service (PLUS): an evidence-based medicine information service delivered on the Web. *AMIA Annu Symp Proc*. 2005;2005:340-344.
15. Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*. 2005;293(10):1223-1238.
16. Del Fiol G, Curtis C, Cimino JJ, et al. Disseminating context-specific access to online knowledge resources within electronic health record systems. *Stud Health Technol Inform*. 2013;192:672-676.
17. Roshanov PS, Fernandes N, Wilczynski JM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. *BMJ*. 2013;346:f657.
18. Jeffery R, Navarro T, Lokker C, Haynes RB, Wilczynski NL, Farjou G. How current are leading evidence-based medical textbooks? an analytic survey of four online textbooks. *J Med Internet Res*. 2012;14(6):e175.
19. Prorok JC, Iserman EC, Wilczynski NL, Haynes RB. The quality, breadth, and timeliness of content updating vary substantially for 10 online medical texts: an analytic survey. *J Clin Epidemiol*. 2012;65(12):1289-1295.
20. Vandvik PO, Lincoff AM, Gore JM, et al. Primary and secondary prevention of cardiovascular disease: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(2 suppl):e637S-e668S.
21. Wentz R. Visibility of research: FUTON bias. *Lancet*. 2002;360(9341):1256.
22. Kung J, Miller RR, Mackowiak PA. Failure of clinical practice guidelines to meet institute of medicine standards: two more decades of little, if any, progress. *Arch Intern Med*. 2012;172(21):1628-1633.
23. Vandvik PO, Brandt L, Alonso-Coello P, et al. Creating clinical practice guidelines we can trust, use, and share: a new era is imminent. *Chest*. 2013;144(2):381-389.

24. Banzi R, Cinquini M, Liberati A, et al. Speed of updating online evidence based point of care summaries: prospective cohort analysis. *BMJ*. 2011;343:d5856.
25. Martínez García L, Arévalo-Rodríguez I, Solà I, Haynes RB, Vandvik PO, Alonso-Coello P; Updating Guidelines Working Group. Strategies for monitoring and updating clinical practice guidelines: a systematic review. *Implement Sci*. 2012;7:109.
26. Ruschitzka F, Abraham WT, Singh JP, et al; EchoCRT Study Group. Cardiac-resynchronization therapy in heart failure with a narrow QRS complex. *N Engl J Med*. 2013;369(15):1395-1405.
27. Haynes RB, Holland J, Cotoi C, et al. McMaster PLUS: a cluster randomized clinical trial of an intervention to accelerate clinical use of evidence-based information from digital libraries. *J Am Med Inform Assoc*. 2006;13(6):593-600.
28. Lader E. Review: Eplerenone is not more effective for reducing mortality than other aldosterone antagonists. *Ann Intern Med*. 2012;157:JC6-10.
29. Haynes RB. ACP Journal Club: the best new evidence for patient care. *ACP J Club*. 2008;148(3):2.
30. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR; Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ*. 2005;330(7501):1179.
31. Montori VM, Wilczynski NL, Morgan D, Haynes RB; Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ*. 2005;330(7482):68.
32. Kulkarni AV, Aziz B, Shams I, Busse JW. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA*. 2009;302(10):1092-1096.
33. Shariff SZ, Bejaimal SA, Sontrop JM, et al. Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *J Med Internet Res*. 2013;15(8):e164.
34. Dans AL, Dans LF, Silvestre MAA. Literature searches. In: Dans AL, Dans LF, Silvestre MAA, eds. *Painless Evidence-Based Medicine.* Chichester, England: John Wiley & Sons; 2008:115-136.
35. DiCenso A, Bayley L, Haynes RB. ACP Journal Club. Editorial: Accessing preappraised evidence: fine-tuning the 5S model into a 6S model. *Ann Intern Med*. 2009;151(6):JC3-JC2, JC3-JC3.

JAMA evidence
Using Evidence to Improve Care

5

Why Study Results Mislead: Bias and Random Error

Gordon Guyatt, Roman Jaeschke, and Maureen O. Meade

IN THIS CHAPTER

Our clinical questions have correct answers that correspond to an underlying reality or truth. For instance, there is a true underlying magnitude of the impact of β-blockers on mortality in patients with heart failure, the impact of inhaled corticosteroids on exacerbations in patients with asthma, the impact of reamed vs unreamed nailing of tibial fractures, the *prognosis* of patients with hip osteoarthritis, and the diagnostic properties of a pregnancy test. Research studies attempt to estimate that underlying truth. Unfortunately, however, we will never know the exact truth. Studies may be flawed in their design or conduct and introduce *systematic error* (or *bias*). Even if a study could be perfectly designed and executed, the estimated *treatment effect* may miss the mark because of *random error*. The next section explains why.

RANDOM ERROR

Consider a perfectly balanced coin. Every time we flip the coin, the *probability* of it landing with its head up or tail up is equal—50%. Assume, however, that we as investigators do not know that the coin is perfectly balanced—in fact, we have no idea how well balanced it is, and we would like to find out. We can state our question formally: What is the true underlying probability of a resulting head or tail on any given coin flip? Our first experiment addressing this question is a series of 10 coin flips; the result: 8 heads and 2 tails. What are we to conclude? Taking our result at face value, we infer that the coin is very unbalanced (ie, biased in such a way that it yields heads more often than tails) and that the probability of heads on any given flip is 80%.

Few would be happy with this conclusion. The reason for our discomfort is that we know that the world is not constructed so that a perfectly balanced coin will always yield 5 heads and 5 tails in any given set of 10 coin flips. Rather, the result is subject to the play of chance, otherwise

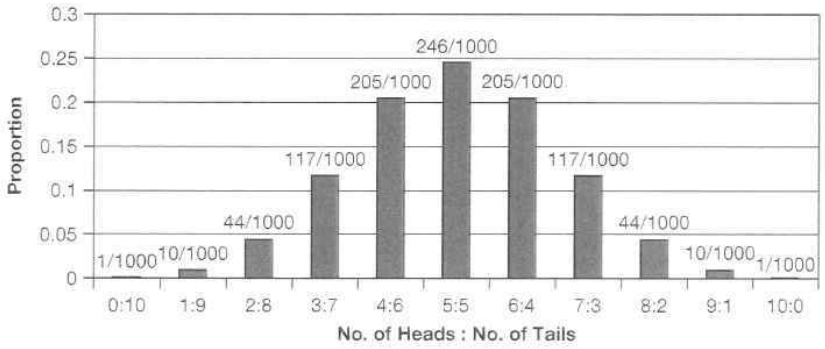
known as random error. Some of the time, 10 flips of a perfectly balanced coin will yield 8 heads. On occasion, 9 of 10 flips will turn up heads. On rare occasions, we will find heads on all 10 flips. Figure 5-1 shows the actual distribution of heads and tails in repeated series of coin flips.

What if the 10 coin flips yield 5 heads and 5 tails? Our awareness of the play of chance leaves us uncertain that the coin is balanced: a series of 10 coin flips of a very biased coin (a true probability of heads of 0.8, for instance) could, by chance, yield 5 heads and 5 tails.

Let us say that a funding agency, intrigued by the results of our first small experiment, provides us with resources to conduct a larger study. This time, we increase the sample size of our experiment markedly, conducting a series of 1000 coin flips. If we end up with 500 heads and 500 tails, are we ready to conclude that we are dealing with a true coin? We are much more confident but still not completely sure. The reason for our remaining uncertainty is that we know that, were the true underlying probability of heads 51%, we would sometimes see 1000 coin flips yield the result we have just observed.

FIGURE 5-1

Theoretical Distribution of Results of an Infinite Number of Repetitions of 10 Flips of an Unbiased Coin



We can apply the above logic to the results of studies that address questions of prognosis, diagnosis, and *harm*, and also to *randomized clinical trials* (RCTs) that address treatment issues. For instance, an RCT finds that 10 of 100 treated patients die during treatment, as do 20 of 100 control patients. Does treatment really reduce the death rate by 50%? Maybe, but awareness of chance will leave us with some degree of uncertainty about the magnitude of the *treatment effect*—and perhaps about whether treatment helps at all.

In a study of congestive heart failure, 228 of 1320 patients (17%) with moderate to severe heart failure allocated to receive *placebo* died, as did 156 of 1327 (12%) allocated to receive bisoprolol.[1] Although the true underlying reduction in the *relative risk* of dying is likely to be in the vicinity of the 32% suggested by the study, we must acknowledge that appreciable uncertainty remains about the true magnitude of the effect (see Chapter 9, Confidence Intervals: Was the Single Study or Meta-analysis Large Enough?).

We have now addressed the question with which we started: "Why is it that no matter how powerful and well designed a study, we will never be sure of the truth?" The answer is that chance is directionless, and it is equally likely, for instance, to overestimate or underestimate treatment effects.

BIAS

Bias is the term we use for the other reason study results may be misleading. In contrast to random error, bias leads to systematic deviations (ie, the error has direction) from the underlying truth. In studies of prognosis, bias leads us to falsely optimistic or pessimistic conclusions about a patient's fate. In studies of diagnosis, bias leads us to an overly optimistic (usually) or

pessimistic assessment of a test's value in differentiating between those with and without a target condition. In treatment or harm studies, bias leads to either an underestimate or an overestimate of the underlying benefit or harm (Box 5-1).

Bias may intrude as a result of differences, other than the *experimental intervention*, between patients in treatment and *control groups* at the time they enter a study. At the start of a study, each patient, if left untreated, is destined to do well—or poorly. To do poorly means to have an adverse event (eg, a stroke) during the study. We often refer to the adverse event that is the focus of a study as the *target outcome* or *target event*. Bias will result if treated and control patients differ in their prognosis (ie, their likelihood of experiencing the target outcome) at the start of the study. For instance, if patients in the control group have more severe atherosclerosis or are older than their counterparts, their destiny will be to have a greater proportion of adverse events than those in the intervention or treatment group, and the results of the study will be biased in favor of the treatment group; that is, the study will yield a larger treatment

BOX 5-1

How Can a Study of an Intervention (Treatment) Be Biased?

Intervention and control groups may be different at the start

Example: patients in control group are sicker or older

Intervention and control groups may, independent of the experimental treatment, become different as the study proceeds

Example: patients in the intervention group receive effective additional medication

Intervention and control groups may differ, independent of treatment, at the end

Example: more sick patients lost to follow-up in the intervention group

effect than would be obtained were the study groups prognostically similar at baseline.

Even if patients in the intervention and control groups begin the study with the same prognosis, the result may still be biased. This will occur if, for instance, effective interventions are differentially administered to treatment and control groups. For instance, in a study of a novel agent for the *prevention* of complications of atherosclerosis, the intervention group might receive more intensive statin therapy than the control group.

Finally, patients may begin prognostically similar, and stay prognostically similar, but the study may end with a biased result. This could occur if, for example, the study loses patients to *follow-up* (see Chapter 6, Therapy [Randomized Trials]) or because a study is *stopped early* because of an apparent large treatment effect.

STRATEGIES FOR REDUCING THE RISK OF BIAS

This book teaches you how to recognize *risk of bias* not only in studies that address issues of therapy and harm but also in studies of prognosis and diagnosis. In studies of prognosis, investigators can reduce bias by enrolling a representative sample and ensuring they are completely followed up. In studies of diagnosis, investigators can ensure that they have chosen an appropriate *criterion* or *gold standard* for diagnosis and that those interpreting test results are unaware of the gold standard findings. In the remainder of this chapter, however, we focus on issues of therapy and harm.

We have noted that bias arises from differences in *prognostic factors* in treatment and control groups at the start of a study or from differences in prognosis that arise as a study proceeds. What can investigators do to reduce these biases? Table 5-1 summarizes the available strategies to reduce biases in RCTs and *observational studies*.

TABLE 5-1

Ways of Reducing Bias in Studies of Therapy and Harm		
Source of Bias	**Therapy: Strategy for Reducing Bias**	**Harm: Strategy for Reducing Bias**
Differences Observed at the Start of the Study		
Treatment and control patients differ in prognosis	Randomization	Statistical adjustment for prognostic factors in the analysis of data
	Randomization with stratification	Matching
Differences That Arise as the Study Proceeds		
Placebo effects	Blinding of patients	Choice of outcomes (such as mortality) less subject to placebo effects
Cointervention	Blinding of caregivers	Documentation of treatment differences and statistical adjustment
Bias in assessment of outcome	Blinding of assessors of outcome	Choice of outcomes (such as mortality) less subject to observer bias

(Continued)

TABLE 5-1

Ways of Reducing Bias in Studies of Therapy and Harm *(Continued)*

Source of Bias	Therapy: Strategy for Reducing Bias	Harm: Strategy for Reducing Bias
Differences at the Completion of the Study		
Loss to follow-up	Ensuring complete follow-up	Ensuring complete follow-up
Stopping study early because of large effect	Completing study as initially planned by sample size calculation	Not applicable
Omitting patients who did not receive assigned treatment	Including all patients for whom data are available in the arm to which they were randomized	Not applicable

When studying new treatments, investigators can implement a large of number of strategies to limit the risk of bias. They can reduce the likelihood of differences in the prognostic features in treated and untreated patients at baseline by *randomly allocating* patients to the 2 groups. They can balance placebo effects by administering identical-appearing but biologically inert treatments—placebos—to patients in the control group. *Blinding* clinicians to whether patients are receiving active or placebo therapy can eliminate the risk of important *cointerventions*, and blinding outcome assessors minimizes bias in the assessment of event rates.

Investigators studying either treatment effects or harm using observational study designs have far less control over the risk of bias. They must be content to compare patients whose *exposure* is determined by their choice or circumstances, and they can address potential differences in patients' fate only by statistical adjustment for known prognostic factors. Blinding is impossible, so their best defense against placebo effects and bias in outcome assessment is to choose *end points*, such as death, that are less subject to these biases. Investigators who address both sets of questions can reduce bias by minimizing loss to follow-up (Table 5-1).

Note that when investigators choose observational study designs to study treatment issues, clinicians must apply the risk of bias criteria developed primarily for questions of harm. Similarly, if the potentially harmful exposure is a drug with beneficial effects, investigators may be able to randomize patients to intervention and control groups. In this case, clinicians can apply the risk of bias criteria designed primarily for therapy questions. Whether for issues of therapy or harm, the strength of inference from RCTs will almost invariably be greater than the strength of inference from observational studies.

Reference

1. CIBIS-II Investigators and Committees. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet*. 1999;353(9146):9-13.

Official journal of the American College of Medi

Open

SYSTEMATIC REVIEW | Genetics inMedicine

A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States

Nanibaa' A. Garrison, PhD[1,2], Nila A. Sathe, MA, MLIS[3,4], Armand H. Matheny Antommaria, MD, PhD[5], Ingrid A. Holm, MD, MPH[6,7], Saskia C. Sanderson, PhD[8], Maureen E. Smith, MS, CGC[9], Melissa L. McPheeters, PhD, MPH[3,4] and Ellen W. Clayton, MD, JD[1,2,4,10]

Purpose: In 2011, an Advanced Notice of Proposed Rulemaking proposed that de-identified human data and specimens be included in biobanks only if patients provide consent. The National Institutes of Health Genomic Data Sharing policy went into effect in 2015, requiring broad consent from almost all research participants.

Methods: We conducted a systematic literature review of attitudes toward biobanking, broad consent, and data sharing. Bibliographic databases included MEDLINE, Web of Science, EthxWeb, and GenETHX. Study screening was conducted using DistillerSR.

Results: The final 48 studies included surveys ($n = 23$), focus groups ($n = 8$), mixed methods ($n = 14$), interviews ($n = 1$), and consent form analyses ($n = 2$). Study quality was characterized as good ($n = 19$), fair ($n = 27$), and poor ($n = 2$). Although many participants objected, broad consent was often preferred over tiered or study-specific consent, particularly when broad consent was the only option, samples were de-identified, logistics of biobanks were communicated, and privacy was addressed. Willingness for data to be shared was high, but it was lower among individuals from under-represented minorities, individuals with privacy and confidentiality concerns, and when pharmaceutical companies had access to data.

Conclusions: Additional research is needed to understand factors affecting willingness to give broad consent for biobank research and data sharing in order to address concerns to enhance acceptability.

Genet Med advance online publication 19 November 2015

Key Words: biobank; broad consent; data sharing; systematic review; tiered consent

Vast amounts of genomic and phenotypic data are needed for many types of research. Frequently, data must be aggregated from several sites to achieve the necessary sample size. These data are often placed in biobanks or biorepositories, which may exist at both the site(s) of collection and in aggregated or centralized sites, such as the database of Genotypes and Phenotypes. These data, which often were collected for one purpose—whether for clinical use or a specific research project—frequently can be studied for other research. These facts raise two distinct, but related, questions. The first is under what conditions data can and should be repurposed for other research in order to increase what can be learned from them. The second is whether data can and should be shared with other investigators in academic institutions, the government, and the commercial sector.

Currently, regulations for the protection of research participants and the Health Information Technology for Economic and Clinical Health Act amendments to the Health Insurance Portability and Accessibility Act Privacy Rule[1] permit the sharing and repurposing of data under certain conditions without the consent of the individual from whom the data were obtained. However, the regulatory landscape is rapidly changing. The Advanced Notice of Proposed Rule Making (ANPRM) issued in 2011 by the Office of Human Research Protections deemed all biospecimens "identifiable per se" and so would require that individuals sign a "standard, brief general consent form" that would provide to participants an opportunity to say no to all future research.[2] In 2014 the National Institutes of Health (NIH) required that investigators obtain broad consent for research and data sharing as a condition of funding for genomics research, with very few exceptions.[3]

Nonetheless, questions remain about the ethical and practical desirability and acceptability of broad consent for research and data sharing. Approaches to obtain permission for use of genomic samples and data include no consent, opt-out, opt-in, case-by-case, tiered or categorical,[4] and broad or blanket consent. Many have argued that blanket consent for unanticipated future research uses is unethical[5] or unworkable,[6] whereas others argue that such consent is acceptable as long as additional protections are in place,[7] especially since broad data sharing

[1]Center for Biomedical Ethics and Society, Vanderbilt University Medical Center, Nashville, Tennessee, USA; [2]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee, USA; [3]Vanderbilt Evidence-Based Practice Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, Tennessee, USA; [4]Department of Health Policy, Vanderbilt University Medical Center, Nashville, Tennessee, USA; [5]Ethics Center, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; [6]Division of Genetics and Genomics and The Manton Center for Orphan Diseases Research, Boston Children's Hospital, Boston, Massachusetts, USA; [7]Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA; [8]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York; [9]Department of Medicine, Northwestern University, Chicago, Illinois; [10]School of Law, Vanderbilt University, Nashville, Tennessee, USA. Correspondence: Nanibaa' A. Garrison (nanibaa.garrison@seattlechildrens.org)

Submitted 3 May 2015; accepted 1 September 2015; advance online publication 19 November 2015. doi:10.1038/gim.2015.138

promotes discovery related to health and disease. Debates have also addressed what sort of control, if any, individuals ought to have over the sharing of data obtained from them, with a similar array of options.[6,8,9] Each option has proponents who present ethical, legal, and social arguments for their positions, often citing studies of public opinion.[10,11] This raises the question of what impact public opinion should have on the development of public policy in this arena.

In 2013, the NIH asked the Consent, Education, Regulation, and Consultation (CERC) Working Group of the Electronic Medical Records and Genomics (eMERGE) Network to conduct a population-based survey of public opinion about the acceptability of both broad consent for research and wide data sharing. To inform the development of this survey and to synthesize the existing literature, we conducted a systematic literature review of empirical research that has been conducted on these topics, the results and policy implications of which are reported here.

MATERIALS AND METHODS

Definitions

We defined "broad consent" as a process in which participants agree prospectively to have their samples, genomic data, and health information retained for use in any future research deemed appropriate by a biobank and/or relevant oversight bodies. Studies of broad consent may use an opt-in or an opt-out model. "Categorical consent," by contrast, is a process in which participants agree prospectively to future use of their samples and data for particular types of research, usually by categories of disease (e.g., cardiac diseases, diabetes). "Data sharing" refers to the transfer of biospecimens with their associated genotypic and/or phenotypic information, data derived from biospecimens, and/or health information to researchers at institutions that are not directly affiliated with the biobanks or to other biorepositories.

Literature search strategy

We systematically searched the literature on broad consent and data sharing for biobank research using the following databases: MEDLINE via the PubMed interface, Web of Science, National Reference Center for Bioethics Literature databases (EthxWeb, GenETHX), and Dissertation Abstracts International. Search strategies used subject heading terms appropriate for each database and key words relevant to biobanking, consent, and data sharing (**Supplementary Table S1** online). Searches were limited to the literature published since 1990 to capture current views about biobanking. We also manually searched the reference lists of included studies and of recent narrative and systematic reviews addressing the topic. Our initial searches were done between October and December 2013 and were updated in March 2015. All citations were imported into DistillerSR systematic review software.

Two reviewers (N.A.S. and a colleague) initially screened titles and abstracts, and two investigators (N.A.G. and E.W.C.) reviewed the full text of the included articles. Articles were included if they reported empirical data with sufficient detail to enable use and aggregation of the data and results about individuals in the United States regarding one or more of the following: participant perceptions of broad consent or data sharing for biobank research, preferences for different consent models for biobank research, information about people's opinions about participating in biobank research, or providing broad consent for biobank research. Disagreements between reviewers were resolved by discussion that included a third reviewer (A.H.M.A.) to reach consensus.

Data extraction and analysis

We identified and screened a total of 3,205 citations and abstracts through the electronic database searches and manual review of articles and bibliographies (**Figure 1**). After reviewing titles and abstracts, we excluded 2,714 studies that did not meet our criteria. We assessed the full text of the 491 remaining studies and excluded another 440 articles because they (i) did not address biobanking, consent, or data sharing ($n = 403$); (ii) were not conducted in the United States ($n = 206$); or (iii) were not obtainable ($n = 1$). Fifty-one publications comprising 48 unique cohorts met our inclusion criteria.

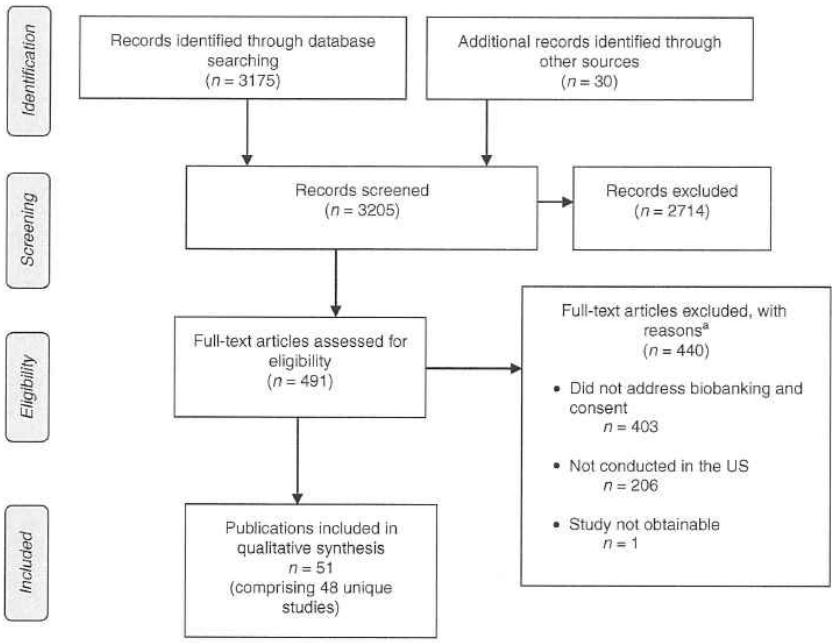
Two investigators (N.A.G. and E.W.C.) assessed the quality of studies using questions adapted from published criteria for the quality assessment of survey and focus group studies.[12–14] Scoring criteria fell into the following broad domains: (i) description of the methods, (ii) participant recruitment from a representative pool and response rates, (iii) appropriateness of objective study questions, and (iv) data analysis lending to reproducible results. Articles that adequately defined criteria in all four domains were rated as "good." Articles containing information that had adequate descriptions of the methods but did not fulfill the criteria for all of the other domains received a rating of "fair." Articles that failed to adequately define their methods, thus preventing an evaluation of representativeness, bias, or reproducibility, received a rating of "poor." Each study was evaluated based on published and Web-accessible information. The questions used in the quality review are contained in **Supplementary Table S2** online. Two investigators (N.A.G. and E.W.C.) also characterized the studies as conducted in urban, rural, or combined settings. The reviewers independently assessed each article and resolved disagreements via discussion to reach consensus.

Data were extracted into summary tables (**Supplementary Table S3** online) by outlining the study population and biobank focus, methods, quality assessment, urban/rural residency, and key outcomes related to consent and data sharing. We report the relevant findings based on the terminology, percentages, and number of significant digits as presented in the publications. We qualitatively analyzed results of studies using summary tables and descriptive synthesis. The heterogeneity of study methods and populations precluded performing a meta-analysis.

RESULTS

Article selection

A total of 51 publications comprising 48 studies were included in this review.[15–65] Most studies involved surveys ($n = 23$), followed



[a]Numbers do not tally as studies could be excluded for multiple reasons

Figure 1 Disposition of studies identified for this review.

by focus groups ($n = 8$), mixed methods ($n = 14$), interviews ($n = 1$), and analyses of consent forms ($n = 2$) (**Supplementary Table S3** online). Two publications used a mixed-methods approach that included qualitative studies that informed the development and implementation of a survey.[35,45] Nineteen studies were of good quality, 27 of fair quality, and 2 of poor quality. Regardless of the assigned quality score, we included all studies in this review. Roughly one-third of the studies ($n = 20$) were written and published after the Office of Human Research Protections issued the ANPRM in July 2011.[46–65] The number of studies published per year from 2008 to 2014 ranged from five to seven, with no notable difference after the ANPRM was issued. Some of the studies published after 2011 mention the ANPRM.[46,56,59,65] Although we examined studies published since 1990, no studies that met our inclusion criteria were published before 2001.

Participant demographics

Studies included a total of 35,969 individuals. Race and/or ethnicity were available for 78.8% of the participants (Table 1). Of these, just over half (51.3%) of participants identified as white, and 13.6% were African American and 6.3% were Hispanic/Latino. Native-American, Alaska Native, Native Hawaiian, and Pacific Islander participants made up 2.2% of the sample. Representation of Asian participants was particularly low at 1.4%. Details for gender were available for 93.3% of participants. Women made up 54.2% of the total sample.

Many studies did not report other demographic data. Only 21 studies reported socioeconomic status, and 43 reported

Table 1 Sociodemographic characteristics

Demographic	*N*	%
Race/ethnicity		
White/Caucasian	18,467	51.3
African American	4,876	13.6
Hispanic/Latino	2,275	6.3
Native American/Alaska Native/Native Hawaiian/Pacific Islander	800	2.2
Asian	515	1.4
Other	1,423	4.0
Missing data	7,622	21.2
Total[a]	35,978	100.0
Gender		
Female	19,491	54.2
Male	14,075	39.1
Missing data	2,403	6.7
Total	35,969	100.0
Other factors		
Socioeconomic Status	21	
Education	43	
Location		
Urban	28	
Rural	2	
Urban + rural	9	
Nationwide	9	

[a]Some participants reported mixed heritage.

educational level. Twenty-eight studies were conducted primarily in urban settings, two were conducted in rural settings, nine were conducted in both urban and rural settings, and nine studies were conducted nationwide.

Studies of broad consent

We identified 48 unique studies that focused on different approaches to obtaining consent.[15–36,38,39,41,42,44–47,49–53,56–65] Three papers each reported two unique studies;[35,42,45] other studies were reported in multiple papers.[19,21,34,35,48,49,59]

Willingness to provide broad consent. Investigators used a variety of approaches to ascertain support for broad consent. Some analyzed the actual choice that participants made when enrolling in research. For example, a retrospective analysis of signed informed consent forms found that 87.1% of 1,298 research participants at the NIH authorized all future research.[20] In a different large national study, the National Health and Nutrition Examination Survey (NHANES), 84.8% of 4,480 overall participants recruited in 1999 and 2000 agreed to DNA specimen collection for inclusion in a national repository for genetic research.[18]

Many studies asked participants hypothetical questions about their willingness to provide broad consent for research. In Indiana, 88.4% of 273 cancer patients agreed that they would be "willing to permit their tissue sample to be used in research on any condition."[24] After time for deliberation, 85% of 40 focus group participants in North Carolina reported that they would agree to have blood and information stored indefinitely in a biorepository for future research.[25] Similarly, 78% of 49 focus group participants in Chicago were interested in participating in a biobank, and the majority stated they would give broad consent.[37] Of 30 patients who were interviewed at a Hawaiian cancer center, 77% endorsed broad consent.[57] One representative nationwide survey found that 68% of 1,593 respondents were willing to give broad consent for research, although their enthusiasm waned if they had a moral objection to certain types of studies for which their samples might be used.[65] Two studies examined patients' willingness to participate in biobanks managed by Kaiser Permanente: 69% of 500 Kaiser patients in the Northwest[30] and 69% of 203 in Colorado[54] agreed to participate in a biobank. In a focus group study in Boston, patients with breast cancer were generally positive about having their samples used for secondary studies that were not planned at the time they gave consent.[22] One older survey deserves special comment. Scott et al.[39] reported the results of a 1998 survey of blood donors that asked about their views regarding storage and use of the blood for research. Of the 49,775 respondents, 60.3% said that "testing stored blood for *any* research" was acceptable with the donor's permission, and 35.5% would not require permission for research use. These studies reported substantial acceptance for broad consent.

Asking participants for their preference among different types of consent—broad, study-by-study, or categorical consent—revealed more mixed support for broad consent. For example, 47% of 931 veterans preferred to give broad consent over other types of consent for all research approved by an oversight board.[33] After adjusting for missing data, a national survey of 4,569 adults found that 52% preferred broad consent, whereas 48% preferred study-by-study consent.[59] In a survey of 751 Iowans, 42% preferred broad consent and 29% favored study-specific consent, compared with 25% who favored categorical consent.[45] In another study of 315 cancer patients at two hospitals in Atlanta, 92 and 97% were willing to allow their samples to be used for research on other diseases; when asked to specify a preference, 56% preferred one-time broad consent and 11% preferred study-by-study consent over no consent or no preference.[23] In a 2001 nationwide survey, 43% of 2,621 participants were willing to donate blood for genetic research and to allow it to be stored for future research.[16] Similarly, only 39.3% of 30 patients who had already donated samples preferred broad consent over consent for specific studies.[41] By contrast, 77.7% of 1,276 people recruited through a crowd-sourced Internet marketplace were willing to donate to biobanks, even after receiving disclosures about potentially objectionable research; however, 40.8% of participants still felt that specific consent was necessary, even if it might inhibit research progress.[58] A similar nationwide survey of 1,599 individuals conducted through a probability-based online panel of adults found a wide range of opinions, with broad consent and real-time study-by-study consent considered the "worst" of five options.[65]

Several studies showed that participants preferred to give informed consent for each study rather than a broad consent, with preferences ranging from 42 to 72%: 42% of a national sample of 4,700 US adults[34,35] (which rose to 48% after adjusting for missing data[59]), 43 to 50% of 931 veterans nationwide,[33,50] and 60.7% of adults recruited in New York.[41] Of 393 parents, 72% reported that they would want to consent each time to allow their child's dried bloodspots to be used for research.[56] In focus groups of 92 Native Hawaiians, respondents repeatedly expressed desire to re-consent, although some stated that they would be content if they trusted the researcher or the biobank's governance.[63] In one study of 273 Jewish individuals, 60–75% believed that consent should be required regardless of whether the DNA was collected in a research or clinical setting.[15]

In a focus group study of 178 Alaska Native participants, some indicated a preference to have consent options for a variety of specimen uses, storage duration, and destruction of the sample at the completion of the study.[48,49] In the same group, some wanted re-contact each time, whereas others felt that a one-time consent was appropriate for new studies. In Chicago, 239 postpartum women were asked about their willingness to enroll their children into a pediatric biobank: 48% of women would enroll their child, but 24% would not; of the latter, 82% of the participants were African American.[28] In another focus group study, 11 of 15 participants preferred tiered consent over other methods to exert the greatest level of control regarding how they wanted their data to be shared; however, participants who were willing to provide broad consent also appreciated the option to opt in or opt out of DNA data sharing.[27]

Preferences for opt-out or opt-in. Some studies reported that most respondents favored an opt-in approach,[15,26,29,45,56] whereas others found that opt-out was acceptable or even preferred by the majority.[23,38,42,44,53] A majority of participants—67% of 751 survey respondents and 63% of 57 focus group participants—who were asked about biobank participation in Iowa preferred opt-in, whereas 18% of survey respondents and 25% of focus group participants in the same study preferred opt-out.[45] In a study of 451 nonactive military veterans, 82% thought it would be acceptable for the proposed Million Veterans biobank to use an opt-in approach, and 75% thought that an opt-out approach was acceptable; 80% said that they would take part if the biobank were opt-in as opposed to 69% who would participate if it were an opt-out approach.[50] When asked to choose which option they would prefer, 29% of respondents chose the opt-in method, 14% chose opt-out, 50% said either would be acceptable, and 7% would not want to participate.

In some cases, biobank participants were re-contacted to inquire about their thoughts regarding proposed changes to the biobank in which they participated. Thirty-two biobank participants who attended focus groups in Wisconsin regarding proposed minimal-risk protocol changes were comfortable with using an opt-out model for future studies because of the initial broad consent given at the beginning of the study and their trust in the institution.[44] A study of 365 participants who were re-contacted about their ongoing participation in a biobank in Seattle showed that 55% thought that opt-out would be acceptable, compared with 40% who thought it would be unacceptable.[38]

Similarly, several studies explored perspectives on the acceptability of an opt-out biobank at Vanderbilt University. First, 91% of 1,003 participants surveyed in the community thought leftover blood and tissues should be used for anonymous medical research under an opt-out model; these preferences varied by population, with 76% of African Americans supporting this model compared with 93% of whites.[29] In later studies of community members, approval rates for the opt-out biobank were generally high (around 90% or more) in all demographic groups surveyed, including university employees, adult cohorts, and parents of pediatric patients.[42,53]

Three studies explored community perspectives on using newborn screening blood spots for research through the Michigan BioTrust for Health program. First, 77% of 393 parents agreed that parents should be able to opt out of having their child's blood stored for research.[56] Second, 87 participants were asked to indicate a preference: 55% preferred an opt-out model, 29% preferred to opt-in, and 16% felt that either option was acceptable.[47] Finally, 39% of 856 college students reported that they would give broad consent to research with their newborn blood spots, whereas 39% would want to give consent for each use for research.[60] In a nationwide telephone survey regarding the use of samples collected from newborns, 46% of 1,186 adults believed that researchers should re-consent participants when they turn 18 years old.[31]

Identifiability of samples influences the acceptability of broad consent. Some studies examined the differences in participants' willingness to provide broad consent for samples that were de-identified or anonymous as compared with identifiable. Respondents generally preferred to give consent if their samples were identifiable. In two studies involving 429 primarily Native Hawaiian participants, 78% of Native Hawaiians and 66% of whites indicated that they would require consent for research if the specimens were identifiable and collected in the clinical setting.[19] For genetics research, 81% Native Hawaiians and 78% of whites indicated that they would require consent if the specimens were identifiable.[21] In a US-wide telephone survey, 81% of 1,193 respondents stated that they would want to be informed about research being done with their sample if it were identifiable; additionally, 57% said they would require permission to use their samples if they were identifiable.[26] De-identification tended to allay concerns. For example, 65.8% of 504 adults who participated in a telephone survey across the United States reported that they would require consent for samples collected in the clinic if they were identifiable, compared with 27.3% who reported they would require consent if samples were anonymized.[17] In the research setting, fewer people thought consent was required for identifiable (29.0%) or anonymized (12.1%) samples.[17] In a study utilizing a hypothetical biobank scenario, 43% of 565 government and medical employees in New Mexico indicated that they would donate their sample for future genetic testing if it could not be traced to them.[32] Not all studies found that people were worried about identifiability. In one survey of 144 clinicians, 86% said that they would donate a DNA sample to a hypothetical biobank in New York regardless of whether it was linked to or unlinked from their identity.[36] In the study in New Mexico, 36% of 565 respondents found it acceptable for broadly consented samples to be used by their local university, even if the samples were linked to them.[32]

Factors associated with views about broad consent. Few studies reported the correlation between demographic variables and respondents' opinions. Characteristics associated with favoring broad consent included being male,[30,34,59] white/Caucasian,[30,34] older,[30,50] and more affluent.[30,50] By contrast, Asians,[34] black non-Hispanics,[33,50] African Americans,[20] and others[19,21,50] (who represented 14.3% of the total) were less likely than whites to believe that research without explicit permission was acceptable. One study of consent forms showed that 75.0% of African Americans gave broad consent compared with 88.4% of whites ($P = 0.002$).[20] Similarly, in the NHANES data, 78.7% of African Americans and 87.1% of whites consented to genetics research.[18]

A few studies looked at other factors that correlated with preferring broad consent. One study reported that participants who were significantly more likely to prefer broad consent also believed that participating would "make me feel like I was contributing to society" (odds ratio = 1.85; $P = 0.001$), that the study would accelerate medical treatments and cures

(odds ratio = 2.20; $P = 0.001$), and that participating in the cohort study would be easy (odds ratio = 1.59; $P < 0.001$).[59] Other investigators reported that the large majority (97.7%) of respondents said "yes" or "maybe" to the idea that it is a "gift" to society when an individual takes part in medical research.[46] Many other studies cited the benefit of research to improve health as a reason to favor broad consent.

Studies of data sharing

We identified 23 studies of data sharing.[23–25,27,28,30,33,34,40–44,46,48,51,52,54,55,59,61,62,64] The earliest publications about participants' preferences on data sharing date to 2006. Most studies of data sharing were conducted with studies of consent preferences; however, six studies were conducted with the primary goal of eliciting preferences on data sharing.[37,40,43,48,54,55]

Willingness to share with other researchers. Participants were generally willing to have their samples and information shared with other academic institutions. Willingness to share data with academic and medical researchers was acceptable for 92% of 4,659 US adults generally,[34] and 80% of 931 US veterans specifically.[33] More than 70% of 100 young adults in Baltimore who were enrolled in a longitudinal study of prevention were willing to share results arising from their DNA.[62] "Nearly three fourths" of 40 community members in a focus group study in North Carolina were comfortable with academic researchers having access to their samples.[25] Many of 79 focus group participants in Seattle endorsed the value of sharing, agreed that sharing locally and with close collaborators was acceptable, and were comfortable with nonprofit and public-interest organizations using data from their samples.[40] In one focus group study of 48 primarily white and female participants in Iowa, the majority cited positive reasons for donating their samples to help and to contribute to advancements in research, and that data sharing would not affect their decision to enroll in a biobank.[44] In another focus group study of 100 African Americans in North Carolina, many recognized the benefits of data sharing but wanted the potential risks to be disclosed, and some wanted the data to be restricted.[43] In another study of patients with inflammatory bowel diseases, 97.3% of 92 respondents were comfortable with sharing their biological sample with investigators in the United States, but 23.8% were uncomfortable with sharing with investigators outside the United States.[64]

Willingness to share in national databases or federal repositories. Some participants expressed concern over sharing their data and information with federal repositories. In one study, 18.5% of 4,050 Vanderbilt University faculty and staff were more likely to want to participate in their institution's biobank if the de-identified data were deposited into a national database; however, 12.1% were less likely to want to participate.[42] In many studies, the location of the repository was often important. In two large nationwide surveys, 80% of 4,659 adults were willing to have their data shared with government researchers; however, 75% of the same sample also were concerned about "the government having [their] samples and information."[34] Similarly, another study found that 71% of 931 veterans were willing to grant database access to government researchers, but half were concerned about "the government having [their] samples and information."[33]

Other studies have shown that some people are concerned about government involvement in maintaining databases containing biomedical information. More than half of the 40 participants in a focus group study of North Carolina community members were concerned about government researchers having access to their institution's biorepository.[25] Despite concerns, 61% of 203 Kaiser patients in Colorado would still provide a sample even if the data would be submitted to a government database,[54] and 82% of 500 Kaiser patients in Oregon agreed to have their information posted in a US government database.[30] In a large metropolitan area in southwest Florida, some of 95 focus group participants believed that biospecimens were already being collected from leftover tissue, and others suspected that tissues were already being shared with researchers in other countries who lack "'strict laws' governing research."[52] In a focus group study of 178 Alaska Native participants, some cited mistrust of the government and police having access to their samples and wanted transparency from the researchers about how their samples were used.[48]

Willingness to share with commercial enterprises. The majority of participants were willing to share with pharmaceutical company researchers, but the percentage was generally less than the percentage willing to share with academic researchers. Seventy-five percent of 4,659 US adults,[34] 54% of 931 veterans,[33] 55.2% of 1,599 adults responding to a nationwide survey,[65] and 75.1% of members of the Crohn's and Colitis Foundation of America Partners cohort were willing to share with pharmaceutical company researchers.[64] Focus group participants in Florida voiced concern about providing blanket consent because they would not benefit financially from any resulting discoveries.[52]

Factors associated with views about data sharing. With the exception of gender, few demographic data (e.g., about race/ethnicity, socioeconomic status, education, and urban/rural residency) were available. Even when demographic information was obtained, investigators did not always report how these variables correlated with respondents' opinions. Therefore, it was largely not possible to draw meaningful conclusions about the associations between sociodemographic factors and views on data sharing.

The willingness of patients with cancer to share seemed to be shaped by their devotion to the institution at which they were receiving care. For example, patients with cancer in Indiana who agreed to participate in a biobank were less likely to be willing to allow their tissue samples to be used by researchers who were not affiliated with the local researchers (89.7%), compared with 96.3% who were willing to share with local university

researchers ($P < 0.01$).[24] Half of 100 patients with breast cancer at MD Anderson Cancer Center preferred to allow only their physician (24%) or other researchers at their hospital (26%) to use their de-identified genetic data for research; fewer patients were willing to share their de-identified data with any cancer researcher (25%) or any researcher (18%).[61] In another study, 95% of 315 patients with cancer in Atlanta were willing to allow researchers to share samples with other local researchers, but only 85% and 92% of participants at two different sites were willing to have their samples shared elsewhere in the United States ($P < 0.05$).[23]

DISCUSSION AND CONCLUSION

In 2013, NIH funded the eMERGE consortium to perform a broad population-based survey to assess public opinion about broad consent for research and data sharing. This systematic literature review, which ultimately contained 48 studies involving 35,969 participants, was conducted to identify gaps and issues that needed to be addressed in this survey.

The most notable finding is that many people do not favor broad consent for either research itself or for research and subsequent wide data sharing. While the majority often expressed support for broad consent when that was the only choice offered, only a minority of respondents favored broad consent when other options, such as tiered or study-by-study consent, were offered. Furthermore, earlier studies focused on the importance of obtaining consent for research, whereas later studies focused on the preferences for different consent options. Willingness to give broad consent increased if data were de-identified. While individuals were generally willing for data or biospecimens to be shared with other academic researchers, individuals were less willing for their data to be shared in federal databases or with commercial enterprises. These findings differ from recent assertions that the public generally supports broad consent.[66,67]

What is equally striking are the large gaps in what is known about factors that affect people's decisions. Gender is the only demographic for which there is essentially complete information. Yet while a few studies generally found that men were more likely to support broad consent, most investigators did not examine the impact of gender on attitudes. Although data about race/ethnicity are incomplete, it seems that minorities often have more concerns about broad consent, although existing evidence suggests that these concerns can be ameliorated in some cases by discussion and education. Much less is known about the impact of sociodemographic factors—such as socioeconomic status, education, and whether people live in urban or rural environments—on attitudes toward broad consent and data sharing. Building on these findings, the eMERGE CERC survey developed a sampling strategy, experimental study design, and survey questions to ascertain more uniformly the views of individuals throughout society in order to identify and address concerns.

This study had several limitations. First, we used broad search terms to capture the existing literature on broad consent and data sharing. The literature addressing these concepts is not well indexed. Thus, while we used multiple approaches (e.g., searching multiple sources, reviewing reference lists, and searching the unpublished, "gray" literature, such as dissertations and reports) to comprehensively identify studies, we may not have identified all salient research. We excluded commentaries and one dissertation from which data could not be extracted. Second, we adapted existing metrics of quality scores to our study. For many studies, we were unable to ascertain the appropriateness of study questions or an analysis plan, thus limiting our ability to thoroughly assess the quality of the studies. Third, the studies that have been conducted to date have a number of limitations, which in turn limit the generalizability of this literature review. Several methodologies were used across studies, often in ways that limit direct comparability. Many of the surveys focus what people say they think, rather what they actually do, even though opinions may differ from action. Definitions of consent were not always consistent and have changed over time, which not only limits our ability to compare studies but also may affect our evaluation of older studies given today's ethical standards for biobanking governance. However, all studies were sufficiently focused on broad consent for research or for data sharing to permit some comparison. Most of the surveys heavily oversampled whites, whereas the qualitative studies disproportionately involved minority participants. Studies that incorporated an educational component may have influenced respondents compared with those studies that did not involve education around biobanking practices. This review also was limited to the United States, which is warranted given the different policy preferences in other countries.

The ultimate goal of this literature review and the eMERGE CERC survey is to obtain a more comprehensive understanding of public opinion about broad consent for data sharing and use. The studies included here typically noted a general acceptance for broad consent and endorsement of data sharing, but with notable privacy and governance concerns, especially by minority participants. The policy question will be what to do if some people, particularly from certain demographics, express a desire for more granular control over the use of data obtained from them in light of the policy trend toward requiring individual consent for broad data use and sharing. At a minimum, it suggests the need to engage those who are skeptical, even if it is decided that the public good of research to improve health outweighs honoring individual objections in some cases or the risk that some people will choose not to participate.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

ACKNOWLEDGMENTS

The eMERGE Network was initiated and funded by National Human Genome Research Institute through the following grants: U01HG006828 (Cincinnati Children's Hospital Medical Center/Boston Children's Hospital), U01HG006830 (Children's Hospital of Philadelphia), U01HG006389 (Essentia Institute of Rural Health),

U01HG006382 (Geisinger Clinic), U01HG006375 (Group Health Cooperative), U01HG006379 (Mayo Clinic), U01HG006380 (Icahn School of Medicine at Mount Sinai), U01HG006388 (Northwestern University), U01HG006378 (Vanderbilt University Medical Center), U01HG006385 (Vanderbilt University Medical Center serving as the Coordinating Center), and the eMERGE Consent, Education, Regulation, and Consultation (CERC) working group supplement grant. The authors thank Rachel R. Walden, previously at Vanderbilt's Eskind Biomedical Library, for conducting the electronic searches and assisting with the initial screening of titles and abstracts.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

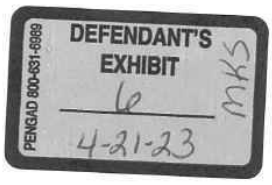
1. HITECH Act amendments to the HIPAA Privacy Rule. 45 C.F.R. §164.508 (c) (2015).
2. Office of Human Research Protections. *Advanced Notice of Proposed Rule Making, 76 Fed.* Reg. 44512–31. 26 July 2011. http://www.gpo.gov/fdsys/pkg/FR-2011-07-26/html/2011–18792.htm. Accessed 17 April, 2015.
3. National Institutes of Health. *Genomic Data Sharing Policy, 79 Fed.* Reg. 51345–54. 28 August 2014. https://www.federalregister.gov/articles/2014/08/28/2014–20385/final-nih-genomic-data-sharing-policy. Accessed 17 April, 2015.
4. National Cancer Institute/National Action Plan on Breast Cancer. *NCI/NAPBC Model Informed Consent Documents*. http://oprs.usc.edu/files/2013/01/modelconsenthandouts.pdf. Accessed 17 April 2015.
5. Tomlinson T. Respecting donors to biobank research. *Hastings Cent Rep* 2013;43:41–47.
6. Stein DT, Terry SF. Reforming biobank consent policy: a necessary move away from broad consent toward dynamic consent. *Genet Test Mol Biomarkers* 2013;17:855–856.
7. Hofmann B. Broadening consent–and diluting ethics? *J Med Ethics* 2009;35:125–129.
8. Williams H, Spencer K, Sanders C, et al. Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. *JMIR Med Inform* 2015;3:e3.
9. *National Institutes of Health. NIH, Lacks family reach understanding to share genomic data of HeLa cells*. Press release, 7 August 2013. http://www.nih.gov/news/health/aug2013/nih-07.htm. Accessed 17 April 2015.
10. Pereira S, Gibbs RA, McGuire AL. Open access data sharing in genomic research. *Genes (Basel)* 2014;5:739–747.
11. Master Z, Nelson E, Murdoch B, Caulfield T. Biobanks, consent and claims of consensus. *Nat Methods* 2012;9:885–888.
12. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–357.
13. Kelley K, Clark B, Brown V, Sitzia J. Good practice in the conduct and reporting of survey research. *Int J Qual Health Care* 2003;15:261–266.
14. Huston P. Reporting on surveys: information for authors and peer reviewers. *CMAJ* 1996;154:1695–1704.
15. Schwartz MD, Rothenberg K, Joseph L, Benkendorf J, Lerman C. Consent to the use of stored DNA for genetics research: a survey of attitudes in the Jewish population. *Am J Med Genet* 2001;98:336–342.
16. Wang SS, Fridinger F, Sheedy KM, Khoury MJ. Public attitudes regarding the donation and storage of blood specimens for genetic research. *Community Genet* 2001;4:18–26.
17. Wendler D, Emanuel E. The debate over research on stored biological samples: what do sources think? *Arch Intern Med* 2002;162:1457–1462.
18. McQuillan GM, Porter KS, Agelli M, Kington R. Consent for genetic research in a general population: the NHANES experience. *Genet Med* 2003;5:35–42.
19. Fong M, Braun KL, Chang RM. Native Hawaiian preferences for informed consent and disclosure of results from research using stored biological specimens. *Pac Health Dialog* 2004;11:154–159.
20. Chen DT, Rosenstein DL, Muthappan P, et al. Research with stored biological samples: what do research participants want? *Arch Intern Med* 2005;165:652–655.
21. Fong M, Braun KL, Chang RM. Native Hawaiian preferences for informed consent and disclosure of results from genetic research. *J Cancer Educ* 2006;21(suppl 1):S47–S52.
22. Kaphingst KA, Janoff JM, Harris LN, Emmons KM. Views of female breast cancer patients who donated biologic samples regarding storage and use of samples for genetic research. *Clin Genet* 2006;69:393–398.
23. Pentz RD, Billot L, Wendler D. Research on stored biological samples: views of African American and White American cancer patients. *Am J Med Genet A* 2006;140:733–739.
24. Helft PR, Champion VL, Eckles R, Johnson CS, Meslin EM. Cancer patients' attitudes toward future research uses of stored human biological materials. *J Empir Res Hum Res Ethics* 2007;2:15–22.
25. Beskow LM, Dean E. Informed consent for biorepositories: assessing prospective participants' understanding and opinions. *Cancer Epidemiol Biomarkers Prev* 2008;17:1440–1451.
26. Hull SC, Sharp RR, Botkin JR, et al. Patients' views on identifiability of samples and informed consent for genetic research. *Am J Bioeth* 2008;8:62–70.
27. McGuire AL, Hamilton JA, Lunstroth R, McCullough LB, Goldman A. DNA data sharing: research participants' perspectives. *Genet Med* 2008;10:46–53.
28. Neidich AB, Joseph JW, Ober C, Ross LF. Empirical data about women's attitudes towards a hypothetical pediatric biobank. *Am J Med Genet A* 2008;146A:297–304.
29. Pulley JM, Brace MM, Bernard GR, Masys DR. Attitudes and perceptions of patients towards methods of establishing a DNA biobank. *Cell Tissue Bank* 2008;9:55–65.
30. Goddard KA, Smith KS, Chen C, McMullen C, Johnson C. Biobank recruitment: motivations for nonparticipation. *Biopreserv Biobank* 2009;7:119–121.
31. Goldenberg AJ, Hull SC, Botkin JR, Wilfond BS. Pediatric biobanks: approaching informed consent for continuing research after children grow up. *J Pediatr* 2009;155:578–583.
32. Hoop JG, Roberts LW, Hammond KA. Genetic testing of stored biological samples: views of 570 U.S. workers. *Genet Test Mol Biomarkers* 2009;13:331–337.
33. Kaufman D, Murphy J, Erby L, Hudson K, Scott J. Veterans' attitudes regarding a database for genomic research. *Genet Med* 2009;11:329–337.
34. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public opinion about the importance of privacy in biobank research. *Am J Hum Genet* 2009;85:643–654.
35. Murphy J, Scott J, Kaufman D, Geller G, LeRoy L, Hudson K. Public perspectives on informed consent for biobanking. *Am J Public Health* 2009;99:2128–2134.
36. Kern MH. Clinician attitudes and intentions toward participation in a DNA biobank. Thesis, Case Western Reserve University, 2010.
37. Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME. Public and biobank participant attitudes toward genetic research participation and data sharing. *Public Health Genomics* 2010;13:368–377.
38. Ludman EJ, Fullerton SM, Spangler L, et al. Glad you asked: participants' opinions of re-consent for dbGap data submission. *J Empir Res Hum Res Ethics* 2010;5:9–16.
39. Scott EA, Schlumpf KS, Mathew SM, et al. Biospecimen repositories: are blood donors willing to participate? *Transfusion* 2010;50(9):1943–1950.
40. Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W. Genomic research and wide data sharing: views of prospective participants. *Genet Med* 2010;12:486–495.
41. Valle-Mansilla JI, Ruiz-Canela M, Sulmasy DP. Patients' attitudes to informed consent for genomic research with donated samples. *Cancer Invest* 2010;28:726–734.
42. Brothers KB, Morrison DR, Clayton EW. Two large-scale surveys on community attitudes toward an opt-out biobank. *Am J Med Genet A* 2011;155A:2982–2990.
43. Haga SB, O'Daniel J. Public perspectives regarding data-sharing practices in genomics research. *Public Health Genomics* 2011;14:319–324.
44. McCarty CA, Garber A, Reeser JC, Fost NC; Personalized Medicine Research Project Community Advisory Group and Ethics and Security Advisory Board. Study newsletters, community and ethics advisory boards, and focus group discussions provide ongoing feedback for a large biobank. *Am J Med Genet A*. 2011;155A:737–741.
45. Simon CM, L'heureux J, Murray JC, et al. Active choice but not too active: public perspectives on biobank consent models. *Genet Med* 2011;13:821–831.
46. Conley JM, Mitchell R, Cadigan RJ, Davis AM, Dobson AW, Gladden RQ. A trade secret model for genomic biobanking. *J Law Med Ethics* 2012;40:612–629.
47. Duquette D, Langbo C, Bach J, Kleyn M. Michigan BioTrust for Health: public support for using residual dried blood spot samples for health research. *Public Health Genomics* 2012;15:146–155.

48. Hiratsuka V, Brown J, Dillard D. Views of biobanking research among Alaska native people: the role of community context. *Prog Community Health Partnersh* 2012;6:131–139.
49. Hiratsuka VY, Brown JK, Hoeft TJ, Dillard DA. Alaska native people's perceptions, understandings, and expectations for research involving biological specimens. *Int J Circumpolar Health* 2012;71:18642.
50. Kaufman D, Bollinger J, Dvoskin R, Scott J. Preferences for opt-in and opt-out enrollment and consent models in biobank research: a national survey of veterans administration patients. *Genet Med* 2012;14:787–794.
51. Lemke AA, Halverson C, Ross LF. Biobank participation and returning research results: perspectives from a deliberative engagement in South Side Chicago. *Am J Med Genet A* 2012;158A:1029–1037.
52. Luque JS, Quinn GP, Montel-Ishino FA, et al.; Tampa Bay Community Cancer Network Partners. Formative research on perceptions of biobanking: what community members think. *J Cancer Educ* 2012;27:91–99.
53. Brothers KB, Westbrook MJ, Wright MF, et al. Patient awareness and approval for an opt-out genomic biorepository. *Per Med* 2013;10. doi:10.2217/pme.13.34
54. Rahm AK, Wrenn M, Carroll NM, Feigelson HS. Biobanking for research: a survey of patient population attitudes and understanding. *J Community Genet* 2013;4:445–450.
55. Ridgeway JL, Han LC, Olson JE, et al. Potential bias in the bank: what distinguishes refusers, nonresponders and participants in a clinic-based biobank? *Public Health Genomics* 2013;16:118–126.
56. Thiel DB, Platt T, Platt J, King SB, Kardia SL. Community perspectives on public health biobanking: an analysis of community meetings on the Michigan BioTrust for Health. *J Community Genet* 2014;5:125–138.
57. Braun KL, Tsark JU, Powers A, et al. Cancer patient perceptions about biobanking and preferred timing of consent. *Biopreserv Biobank* 2014;12:106–112.
58. Gornick MC, Ryan KA, Kim SY. Impact of non-welfare interests on willingness to donate to biobanks: an experimental survey. *J Empir Res Hum Res Ethics* 2014;9:22–33.
59. Platt J, Bollinger J, Dvoskin R, Kardia SL, Kaufman D. Public preferences regarding informed consent models for participation in population-based genomic research. *Genet Med* 2014;16:11–18.
60. Platt T, Platt J, Thiel DB, Fisher N, Kardia SL. 'Cool! and creepy': engaging with college student stakeholders in Michigan's biobank. *J Community Genet* 2014;5:349–362.
61. Rogith D, Yusuf RA, Hovick SR, et al. Attitudes regarding privacy of genomic information in personalized cancer therapy. *J Am Med Inform Assoc* 2014;21(e2):e320–e325.
62. Storr CL, Or F, Eaton WW, Ialongo N. Genetic research participation in a young adult community sample. *J Community Genet* 2014;5:363–375.
63. Tauali I M, Davis EL, Braun KL, et al. Native Hawaiian views on biobanking. *J Cancer Educ* 2014;29:570–576.
64. Long MD, Cadigan RJ, Cook SF, et al. Perceptions of patients with inflammatory bowel diseases on biobanking. *Inflamm Bowel Dis* 2015;21:132–138.
65. Tomlinson T, De Vries R, Ryan K, Kim HM, Lehpamer N, Kim SY. Moral concerns and the willingness to donate to a research biobank. *JAMA* 2015;313:417–419.
66. Grady C, Eckstein L, Berkman B, et al. Broad consent for research with biological samples: workshop conclusions. *Am J Bioeth* 2015;15:34–42.
67. Truven Health Analytics. *Health Poll: Data Privacy*. Ann Arbor, MI. November 2014.

Jennifer E. deSante-Berthau, Michelle L. McGowan, and Armand H. Matheny Antommaria, "Systematic Review of Typologies Used to Characterize Clinical Ethics Consultations," *The Journal of Clinical Ethics* 29, no. 4 (Winter 2018): 291-304.

Systematic Review of Typologies Used to Characterize Clinical Ethics Consultations

Jennifer E. deSante-Bertkau, Michelle L. McGowan, and Armand H. Matheny Antommaria



ABSTRACT

Introduction

Classifying the ethical issues in clinical ethics consultations is important to clinical practice and scholarship. We conducted a systematic review to characterize the typologies used to analyze clinical ethics consultations.

Methods

We identified empirical studies of clinical ethics consultation that reported types of ethical issues using PubMed. We screened these articles based on their titles and abstracts, and then by a review of their full text. We extracted study characteristics and typologies and coded the typologies.

Results

We reviewed 428 articles; 30 of the articles fulfilled our inclusion criteria. We identified 27 unique typologies. Each typology contained five to 47 categories (mean = 18). The most common categories were do-not-attempt-resuscitation orders (19 typologies, 70 percent), capacity (18 typologies, 67 percent), withholding (18 typologies, 67 percent), withdrawing (17 typologies, 63 percent), and surrogate or proxy (16 typologies, 59 percent). Only seven (26 percent) of the typologies contained all five of the most common categories.

The typologies we used to characterize clinical ethics consultations exhibit significant heterogeneity and several conceptual limitations. A common typology is needed whose development may require multi-institutional collaboration and could be facilitated by professional organizations.

Jennifer E. deSante-Bertkau, MD, MBE, is an Assistant Professor in the Ethics Center and Division of Hospital Medicine, Cincinnati Children's Hospital Medical Center and the Department of Pediatrics at the University of Cincinnati College of Medicine in Cincinnati, Ohio. *Jennifer.desante@cchmc.org*

Michelle L. McGowan, PhD, is a Research Associate Professor in the Ethics Center and Division of General and Community Pediatrics, Cincinnati Children's Hospital Medical Center, the Department of Pediatrics at the University of Cincinnati College of Medicine, and the Department of Women's, Gender, and Sexuality Studies at the University of Cincinnati College of Arts & Sciences.

Armand H. Matheny Antommaria, MD, PhD, is a Professor and Director of the Ethics Center at Cincinnati Children's Hospital Medical Center and Professor in the Department of Pediatrics at the University of Cincinnati College of Medicine.

INTRODUCTION

Hospital ethics committees and clinical ethics consultation proliferated between 1983 and 2007. One survey published in 1983 found that approximately 1 percent of all hospitals in the United States had ethics committees that could become involved in decisions regarding individual patients.[1] By 2007, 81 percent of general hospitals and 100 percent of hospitals with more than 400 beds had established ethics consultation services.[2] During this time, there was substantial discussion of the goals of consultations, the competencies of consultants, and the evaluation of consultation processes.[3] Observational studies described the process of consultation including the frequency of consultation requests, the characteristics of requestors and patients, and the types of ethical issues encountered. Some studies examined potential associations between consultation and mortality and/or length of stay, and/or reported the satisfaction of requestors. Few intervention trials, however, have been conducted, and trials that have been conducted show variable results.[4]

Observational studies of consultations have a variety of potential benefits. Studies of individual institutions can be used to evaluate trends in consultation requests, plan educational programs, identify systems issues, and evaluate changes in response to interventions. They may provide data that can be used to justify institutional support for clinical ethics consultation services. Comparisons between institutions also may be beneficial. Such comparisons allow descriptions of variation between different types of institutions, including those associated with different clinical ethics consultation methods.

A typical example of an observational study is Johnson, Church, Metzger, and Baker's analysis of ethics consultations conducted at St. Jude Children's Research Hospital from 2000 to 2011.[5] St. Jude is a 78-bed pediatric hospital that specializes in the treatment of children with cancer, human immunodeficiency virus infection, blood disorders, and primary immunodeficiencies. The authors reported descriptive demographic data, the primary reason for consult requests, outcomes, and involvement with external services, for example, palliative care and child protective services. They compared their results with other recently published studies.

Several commentators have identified methodological issues regarding this type of study. Antommaria argued for the need for a common list of reasons to advance scholarship on clinical ethics consultation[6] and Henriksen Hellyer and colleagues argued that "one of the most challenging aspects of interpreting ethics consultation practices across settings . . . is a nonstandard classification of consult types or 'reason for consult.' "[7] Gilliam, McDougall, and Delany proposed their own alternative typology of categories.[8] Given the debate over the appropriate development and content of consultation typologies, we conducted a systematic review of the literature to describe the typologies used to characterize clinical ethics consultations.

METHODS

Systematic Review

Inclusion criteria for the systematic review were (1) empirical studies of clinical ethics consultation that (2) categorized the ethical issues that prompted or were identified in the consultation, and (3) provided data on the number of consultations performed. Studies of the ethical issues encountered by healthcare professionals, causes of moral distress, and ethical issues in research were excluded. Articles that described each individual consultation, but did not categorize them, and articles that described selected consultations were also excluded. A search strategy was developed with the assistance of two medical librarians (Alison Kissling and Martina Darragh). It included indexed terms and text words to capture concepts related to empirical studies, clinical ethics consultation, and categories or types. (The full search strategy is available from the corresponding author.) The search was limited to the U.S. National Library of Medicine's PubMed, as this database indexes the journals most likely to publish such studies. The search was limited to articles published in English since 1980. Our review was registered with the international prospective registry of systematic reviews, PROSPERO.[9]

Two of the authors (JdB and AHMA) independently screened the titles and abstracts of the articles, and then independently reviewed the full text of the articles. The authors identified additional articles from the references and independently reviewed the full text of those articles. Articles that either author believed might be relevant based on review of the title and abstract or the references advanced to re-

view of the full text and disagreements following review of the full text were resolved by consensus.

Data Extraction and Coding

Two of the authors (JdB and AHMA) extracted the name of the institution from the articles that were included in the study, as well as the type of institution, the study population, the data collection period, the number of consultations, the primary outcome measure used, and the method used to derive the typology. These two authors characterized the primary outcome measure used in each article as either the reason(s) for consultation or the ethical issue(s) identified during the process of the consultation.

These authors characterized the primary outcome measure in the selected articles by whether the measure was identified prospectively or retrospectively, and whether the measure was defined by the requestor of the consultation, the ethics consultant, or the author(s) of the article. For each article, the authors categorized the method used to derive the typology as deductive, inductive, or both. By *deductive,* the authors mean that the typology used in each article was based on *a priori* categories or a review of the literature, and by *inductive,* the authors mean that the typology used emerged from a review of the consultations. Typologies were extracted from the articles by one of the authors (JdB) and reviewed for accuracy by another individual (Jennifer Longbottom). Some articles presented multiple typologies, for example, typologies of "primary consult activity" and "organizational issues," but only one typology of ethical issues was extracted per article.[10] Disagreements regarding the assignment of typology were resolved by consensus between the authors.

Using inductive and deductive reasoning, the authors developed a coding scheme that would allow comparison across studies.[11] Some codes were narrowly defined based on their common use in the ethics literature, for example, "best interest." Other codes were created to combine categories from different studies that were felt to represent similar ethical issues. For example, "durable powers of attorney for healthcare" and "living wills" were included in the code "advanced care planning." Some codes were gathered into clusters based on their relation to each other. For instance, the distinct codes for the different types of interpersonal conflict were combined into a cluster, or grouping that we called "conflict." The typologies were not exhaustively coded to the level of codes that only appeared a small number of times. For example, "community considerations,"[12] "guns in the home of home care patients,"[13] and "initiation of an individual attempt to cure"[14] each appeared in only one typology and were not coded. The resulting code schema was then utilized for thematic analysis of the typologies.

After reviewing the initial subsets of the typologies to refine the codes and coding rules, the authors reviewed and discussed how to apply the codes to enhance intercoder reliability.[15] The typologies were independently coded by all three of the authors using ATLAS.ti 8.0 qualitative software. Any discrepancies were discussed and resolved by consensus.

Figure 1. Data extraction process

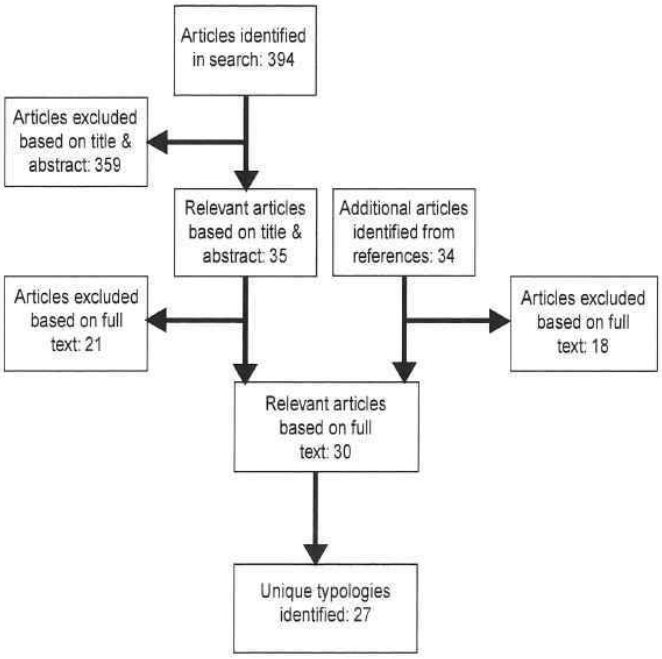


TABLE 1. Characteristics of studies

Article authors	Institution	Institution type	Study population	Data collection period	Duration of data collection (months)	Number of consultations	Total number of reasons or issues	Primary outcome measure	Derivation of typology	Number of categories
Boissy, Ford, Edgell, and Furlan	Cleveland Clinic, Cleveland, Ohio	Academic	Neurology inpatient, neurology step down, neuro-intensive care units. Nonneurological diagnoses excluded	1998-2004	72	49	49	Reasons for consultation identified by authors	Inductive	20
Bruce, Smith, Hizlan, and Sharp	Cleveland Clinic Cleveland, Ohio	Academic	--	Jan. 2007-Dec. 2008	24	478	NS	Ethical issues identified retrospectively by consultant	--	8
Forde and Vandvik	National Hospital Oslo, Norway	Academic	--	1996-2002	72	31	100	Reasons for consultation identified retrospectively by consultant	--	13
Fukuyama, Asai, Itai, and Bito	Clinical Ethics Support and Education Project, Japan[1]	--	--	Oct. 2006-Dec. 2007	15	25	25	"Consultation request classification" identified prospectively by consultant	--	13
Henriksen Hellyer et al.	Mayo Clinic, Rochester, Minn.	Academic	< 18 years old, or ≥ 18 years old, never competent, and treated in pediatrics	May 1995-June 2014	241	64	64	Reason for consultation identified retrospectively by consultant	Deductive	47
Johnson, Church, Metzger, and Baker	St. Jude's Children's Research Hospital, Memphis, Tenn.	Children's specialty	--	May 2000-Dec. 2011	140	53	79	Ethical issues identified retrospectively by authors	Both	27
Johnson, Lesandrini, and Rozycki	Grady Memorial Hospital, Atlanta, Ga.	Academic	Trauma patients	Jan. 2000-Dec. 2010	132	108	108	Ethical issues identified retrospectively by consultants	Deductive, based on Fox, Myers, and Pearlman[5]	28
La Puma (1987)	University of Chicago Hospitals	Academic	--	July 1985-June 1986	12	27	27	Ethical issues identified retrospectively by consultants	--	14
La Puma et al. (1988)	University of Chicago Hospitals and Clinics	Academic	--	July 1986-June 1987	12	51	138[6]	Reasons for consultation identified prospectively by requestor and consultant	Inductive	13
La Puma, Stocking, Darling, and Siegler	Lutheran General Hospital, Park Ridge, Ill.	Community	--	Jan. 1988-Dec. 1989	24	104	313[7]	Reason for consultation identified by requestor and consultant	--	18

McDougall and Notini	Royal Children's Hospital, Melbourne Vic., Australia	Children's	--	May 2005-Dec. 2014	116	184	184	Reason for consultation identified by requestor and consultant	Both	9
Moeller et al.	Akron City Hospital Akron, Ohio	Community	--	July 1997-March 2009	152	100	195	Reason for consultation identified retrospectively	--	9
Nilson, Acres, Tamerin, and Fins	New York-Presbyterian Healthcare System[2]	Healthcare system	--	2002 (30 days)	1	53	79	Reasons for consultation identified retrospectively by consultant	Both	9
Opel et al.	Seattle Children's Hospital, Seattle, Wash.	Children's	--	1996-2006	132	71	71	Reason for consultation identified retrospectively by authors	Deductive	12
Orr and Moon	Loma Linda University Medical Center, Loma Linda, Calif.	Academic	--	Aug. 1990-July 1991	12	46	144	Ethical issues identified retrospectively by requestor	Deductive, based on La Puma et al.[8]	19
Orr and Perkin	Loma Linda University Medical Center, Loma Linda, Calif.	Academic	≤ 18 years old	Aug. 1990-July 1992	24	64	179	Ethical issues identified retrospectively by consultant	--	21
Perkins and Saathoff	University of Texas Health Science Center, San Antonio, Tex.	Academic	--	Jan. 1984-June 1985	18	44	97	Ethical issues identified retrospectively by consultant	--	15
Ramsauer and Frewer	Erlangen University Hospital, Erlangen, Germany	Academic	Pediatric	2002-2008	84	16	NS	Reason for consultation and "Consultation Contents" identified retrospectively by authors	--	8
Romano et al.	Columbia University Medical Center, New York, N.Y.	Academic	Intensive care unit	Aug. 2006-July 2007	12	168	198	Ethical issue identified retrospectively by authors	Deductive, used Swetz, Crowley, Hook, and Mueller[9]	
Schenkenberg	Salt Lake VA Medical Center, Salt Lake City, Utah	Veterans Administration	--	Jan. 1987-April 1996	124	160	249	Ethical issues	--	29
Shuman et al. (Sept. 2013)	National Cancer Institute designated comprehensive cancer centers[3]	Specialty	Adult oncology	2007-2011	48	208	NS	Reason for consultation identified prospectively	Deductive, based on Nilson, Tamerin, and Fins[10]	8

Table 1. Continued next page.

TABLE 1. Characteristics of studies, *continued*

Article authors	Institution	Institution type	Study population	Data collection period	Duration of data collection (months)	Number of consultations	Total number of reasons or issues	Primary outcome measure	Derivation of typology	Number of categories
Shuman et al. (Nov 2013)	Memorial Sloan-Kettering Cancer Center, New York, N.Y.	Specialty	Head and neck cancer	2007-2011	60	14	14	Reason for consultation identified prospectively	Deductive, based on Nilson, Tamerin, and Fins[11]	13
Streuli et al.	Zurich University Children's Hospital, Zurich, Switzerland	Children's	--	Jan. 2006-Dec. 2010	60	95	95	Ethical issues	--	11
Swetz, Crowley, Hook, and Mueller (June 2007)	Mayo Clinic, Rochester, Minn.	Academic	--	April 1995-Dec. 2005	129	255	1,181	Reason for consultation identified prospectively	Deductive	19
Swetz, Crowley, Hook, and Mueller (Dec. 2007)	Mayo Clinic, Rochester, Minn.	Academic	Neurological diagnosis	April 1995-Dec. 2005	120	47	129	Reason for consultation identified prospectively	Deductive, used Swetz, Crowly, Hook, and Mueller[12]	11
Tapper, Vercler, Cruze, and Sexson	Atlanta, Ga.[4]	Academic	--	2004-2006	36	285	7,598	Ethical issues identified retrospectively by consultant	--	32
Thomas et al.	Cleveland Clinic, Cleveland, Ohio	Academic	< 18 years of age. Presurgical neurologic consultations excluded	Jan. 2005-July 2013	106	102	261	Ethical issues identified by consultant	Inductive	29
Voigt et al.	Memorial Sloan-Kettering Cancer Center, New York, N.Y.	Specialty	Medical-surgical intensive care unit	Sept. 2007-Dec. 2011	52	53	53	Reasons for consultation identified retrospectively by authors	Deductive, used Nilson, Acres, Tamerin and Fins[13]	8
Wasson et al.	Loyola University Medical Center, Chicago, Ill.	Academic	--	2008-2013	60	156	53	Ethical issues identified retrospectively by authors	Both	40
Yen and Schneiderman	San Diego Children's Hospital and Health Center, San Diego, Calif.	Children's	--	Sept. 1990-April 1995	68	23	34	Ethical issues identified by consultant	--	5

NOTES

See the appendix for an alphabetical list of these articles. The column "Number of consultations" is the number of consultations analyzed, but may not be the total number of consultations during this time period. Some studies excluded some consultations based on type and/or completeness of records. The column "Primary outcome measure" is reported in the following format: reason for consultation or ethical issue identified prospectively or retrospectively by the requestor;

consultant, or authors; missing components were not specified in the article. NS = not specified.

1. Japanese medical institutions
2. Two large medical centers in Manhattan, three community teaching hospitals in Brooklyn and Queens, and two community teaching hospitals in northern New Jersey.
3. One freestanding urban cancer center in a Northeastern metropolis, and one cancer center integrated within a large academic health system in a small Midwestern city.
4. A large urban public teaching hospital.
5. E. Fox, S. Myers, and R.A. Pearlman, "Ethics Consultation in United States Hospitals: A National Survey," *American Journal of Bioethics* 7, no. 2 (February 2007): 13-25.
6. This is the total number of reasons identified by the consulting physician.
7. Ibid.
8. J. La Puma et al., "An Ethics Consultation Service in a Teaching Hospital: Utilization and Evaluation," *Journal of the American Medical Association* 260, no. 6 (12 August 1988): 808-11.
9. K.M. Swetz, M.E. Crowley, C. Hook, and P.S. Mueller, "Report of 255 Clinical Ethics Consultations and Review of the Literature," *Mayo Clinic Proceedings* 82, no. 6 (June 2007): 686-91.
10. E.G. Nilson, C.A. Acres, N.G. Tamerin, and J.J. Fins, "Clinical Ethics and the Quality Initiative: A Pilot Study for the Empirical Evaluation of Ethics Case Consultation," *American Journal of Medical Quality* 23, no. 5 (September-October 2008): 356-64.
11. Ibid.
12. Swetz, Crowley, Hook, and Mueller, "Report of 255 Clinical Ethics Consultations," see note 9 above.
13. Nilson, Acres, Tamerin, and Fins, "Clinical Ethics and the Quality Initiative," see note 10 above.

RESULTS

Study Characteristics

The literature search, performed on 30 December 2016, identified 394 articles. Of these, 359 were excluded based on a review of the titles and abstracts, and 21 were excluded based on a review of the full text. An additional 34 articles were identified from references, and 18 of these were excluded based on a review of their full text. For example, articles that reported on clinician focus groups,[16] a survey of ethics committee chairs,[17] and published professional codes[18] were excluded. Of the 394 articles, 30 met our inclusion criteria and are included in this review. (See the appendix for a bibliography of the 30 articles.) Figure 1 depicts the data-extraction process.

Studies were conducted in the U.S. and a number of other countries including Australia,[19] Germany,[20] Japan,[21] Norway,[22] and Switzerland.[23] (See table 1.) The majority of studies were conducted at academic medical centers (n = 17). Cleveland Clinic,[24] Loma Linda University Medical Center,[25] Mayo Clinic,[26] and University of Chicago Hospitals[27] were the subject of multiple reports. A number of studies were conducted at children's hospitals or focused on pediatric patients (n = 9). Other sites included community hospitals,[28] health systems,[29] specialty hospitals,[30] and Veterans Affairs hospitals.[31] Several studies focused on specific patient populations including neurology patients,[32] trauma patients,[33] and patients with cancer.[34]

The earliest study was published in 1987.[35] The duration of data collection ranged from one month[36] to 241 months.[37] Some of the studies excluded consultations for a variety of reasons, and the resulting number of consultations ranged from 14[38] to 478.[39]

The studies described their primary outcomes in a variety of ways; 14 typologies described the reasons that triggered the consultation; 13 described the issues identified during the consultation. Of the 30 articles, 11 studies reported one ethical issue per consultation, 16 reported one or more issues, and three did not specify the number of issues. The articles differed regarding whether the issues were identified prospectively or retrospectively, or by the requestor, the consultant, or the investigator.

Content of the Typologies

The 30 articles that met our inclusion criteria included 27 unique typologies (a table of all of the typologies is available from the corresponding author). Three articles utilized previously published typologies: Swetz, Crowley, Hook, and colleagues[40] utilized their previously developed typology;[41] Romano, Wahlander, Lang, and colleagues[42] utilized a typology developed by Swetz, Crowley, Hook, and colleagues;[43] and Voigt, Rajendram, Shuman, and colleagues[44] utilized a typology developed by Nilson, Acres, Tamerin, and Fins.[45] While four articles reported that they utilized existing typologies, the categories included were not identical to the previously published typology, and they were included as distinct typologies.[46]

TABLE 2. Contents of typologies

Code	Numer of typologies that include code	% of typologies that include code	Henriksen Hellyer et al.	Moeller et al.	Nilson, Acres, Tamerin, and Fins	Orr and Moon	Shuman et al. (Nov. 2013)	Swetz, Crowley, Hook, and Mueller	Wasson et al.	Boissy, Ford, Edgell, and Furlan	Johnson, Church, Meltzer, and Baker
Number of the 5 most common codes	--	--	5	5	5	5	5	5	5	4	4
DNAR orders	19	70	X	X	X	X	X	X	X	X	X
Capacity	18	67	X	X	X	X	X	X	X	X	X
Withholding	18	67	X	X	X	X	X	X	X		X
Withdrawing	17	63	X	X	X	X	X	X	X	X	X
Surrogate or proxy	16	59	X	X	X	X	X	X	X	X	
Futility	15	56	X	X	X		X	X	X	X	X
Conflict cluster	15	56	X	X		X		X		X	X
Not otherwise specified	8	30				X				X	
Between patient/family and team	9	33	X	X						X	X
Family conflict	3	11						X			X
Within family	5	19								X	X
Within team	8	30						X		X	X
Life-sustaining treatment	14	52		X	X		X		X	X	X
Professionalism cluster	14	52	X			X	X	X	X		X
Not otherwise specified	8	30	X					X	X		X
Truth-telling	10	37	X			X	X				X
Boundaries	4	15	X								
Conflict of interest	2	7									X
Refusing treatment	14	52			X	X	X		X	X	X
Legal	13	48	X			X		X	X	X	
Resources	13	48	X			X	X	X			X
Advanced care planning	12	44	X		X		X	X	X	X	X
Autonomy	12	44	X			X		X	X	X	X
Medical subspecialty cluster	11	41						X	X	X	
Reproductive health	6	22						X	X		
Psychiatry	5	19						X		X	
Other	5	19							X		
Culture	10	37	X					X	X		X
Discharge	10	37				X	X	X	X		
Informed consent	10	37	X		X		X		X		
Privacy and confidentiality	10	37	X				X	X			X
Specific interventions	10	37				X	X		X	X	
Goals of care	9	33	X	X				X	X		X
Research	9	33	X					X			X
Death	8	30	X						X	X	
Decision making	8	30	X						X	X	X
End-of-life care	8	30						X	X		X
Communication	7	26	X							X	
Palliative care	7	26					X		X		X
Permission and assent	7	26	X						X		X
Quality of life	7	26						X	X		
Difficult patients	5	19	X								
Nonadherence	5	19	X	X							X
"Other"	5	19		X				X		X	
Best interest	3	11							X		
Demanding	3	11				X					X
Hastening death	3	11	X								
Justice	3	11	X								X
TOTAL number of codes			27	11	10	14	16	23	26	20	28

NOTES: Clusters do not count towards the total. See the appendix for a bibliography of these articles.

La Puma, Stocking, Darling, and Siegler	La Puma et al.	La Puma	Orr and Perkin	Perkins and Saathoff	Shuman et al. (Sept. 2013)	Tapper, Vercler, Cruze, and Sexson	Johnson, Lesandrini, and Rozycki	Schenkenberg	Steuli et al.	Forde and Vandvik	Fukuyama, Asai, Itai, and Bito	Opel et al.	Thomas et al.	Yen and and Schneiderman	Bruce, Smith, Hizlan, and Sharp	McDougall and Notini	Ramsauer and Frewer
4	4	4	4	4	4	4	3	3	3	2	2	1	1	1	0	0	0
X	X	X	X	X	X	X	X	X	X								
X	X	X	X	X			X	X			X		X				
X	X	X	X	X	X	X			X	X				X			
X	X	X			X	X			X	X	X						
			X	X	X	X	X	X				X					
			X		X	X	X		X		X						
X	X		X			X		X	X				X	X		X	X
X	X		X											X			X
			X			X		X	X							X	
									X								
			X			X										X	
			X			X		X	X				X				
		X		X	X	X		X			X		X			X	
X	X		X			X			X					X			
X	X	X				X	X		X				X				
X	X				X		X				X	X					
	X	X				X	X		X		X	X					
X	X					X		X	X	X		X	X			X	
X										X	X		X			X	
	X							X									
						X			X	X		X	X			X	
X			X			X	X	X			X						
	X		X					X					X	X			
X					X	X	X	X			X	X	X				
X		X				X	X	X			X						
			X	X		X		X									X
						X		X									X
X		X	X			X		X				X	X				
X		X				X	X	X									
							X		X		X		X				
		X	X				X				X		X				
						X			X			X	X				
		X				X				X		X	X				
				X									X				
X			X			X						X	X			X	
							X		X		X		X		X		
						X						X	X				
X	X																
											X						X
				X													
	X						X										
											X						X
18	14	13	16	10	9	26	18	17	12	5	14	9	20	4	6	6	6

The studies developed their typologies using a variety of methods. They characterized the consultations deductively based on *a priori* categories or a review of the literature ($n = 8$), inductively based on categories developed from a qualitative analysis of the cases ($n = 2$), or based on a combination of both approaches ($n = 3$). Some studies categorized their consultations based on a published coding catalog or typology.[47] Almost half of the studies (14) did not state how they developed their categories. Only two studies included examples of their categories[48] and only one included a code book with definitions.[49]

All but one of the studies presented their typologies in a table, figure, or box.[50] Of these, 11 typologies were divided into major headings and subcategories. The number of categories in each typology ranged from five to 47 (mean = 18).

We created 45 codes based on the concepts that appeared in the published typologies (see table 2). The most commonly used codes were "DNAR orders" (19 typologies, 70 percent), "capacity" (18 typologies, 67 percent), "withholding" (18 typologies, 67 percent), "withdrawing" (17 typologies, 63 percent), and "surrogate or proxy" (16 typologies, 59 percent). Seven (26 percent) of the typologies contained all five of the most frequently appearing codes. None of the typologies contained all 10 of the most frequently used codes; two typologies contained nine of the 10 most frequently used codes.[51] One typology contained none of the 10 most frequently used codes.[52]

Some codes were related to ethical principles—for example, "autonomy" and "justice"—or ethical issues—for example, "DNAR [do-not-attempt-resuscitation] orders," "capacity," "surrogate or proxy," "advance care planning," "informed consent," and "privacy and confidentiality." Other codes referred to decision making in general, for example, "decision making" and "goals of care," decision making dynamics, for example, "withholding," "withdrawing," "refusing," and "demanding," or types of interventions, for example, "DNAR orders," "life-sustaining treatment," "end-of-life care," and "palliative care," without specifying specific ethical issues. Finally, some codes identified sources of ethical norms without specifying particular ethical issues, for example, "legal" and "culture," which included religion, specific religious groups, and spirituality.

DISCUSSION

Our systematic review identified 30 articles containing 27 unique typologies of the reasons for or ethical issues identified in clinical ethics consultations. The studies varied in type of institution, duration of data collection, number of consultations, primary outcome measure, and number of categories and typology. The number of categories in each typology ranged from five to 47 (mean = 18). The most commonly used codes were "DNAR orders," "capacity," "withholding," "withdrawing," and "surrogate or proxy." Only seven of the 27 (26 percent) contained all five of the most common codes.

While evaluation of the reasons for clinical ethics consultations has generated a substantial body of literature, this literature has a number of limitations. First, the studies utilized a variety of primary outcome measures. It may be beneficial to identify the benefits and detriments of focusing on the perspectives of the requestor, the consultant, or the investigator as well as the benefits and detriments of prospective and retrospective coding. Furthermore, some studies identified a single ethical issue per consultation, and others multiple ethical issues per consultation. It was not clear how the former studies identified the most important issue.

Second, 13 studies did not specify how they developed their typologies, and those that did specify used a variety of methods. Ideally, one might use both deductive and inductive approaches, draw on ethical theory and the published literature, as well as analyses of the consultations themselves. Some of the typologies did not include important ethical concepts, for example, Moeller and colleagues did not include "privacy and confidentiality."[53] Collaboration will be required to overcome the limitations of inductive analyses of consultations from one institution; for example, augmenting the experience of institutions that do not provide obstetric or pediatric care with those that do.

Third, there was no standard typology; the existing typologies were significantly heterogeneous. There was no consensus on even the most frequent codes. For example, "DNAR orders," the most frequent code, did not appear in almost one-third of the studies. This lack of uniformity made it difficult to compare institutional experiences; for example, how did the reasons for consultations differ between types of institutions, institutions in different geographic re-

gions, or methods of clinical ethics consultation.[54]

Fourth, with one exception, the studies did not provide a code book with clear definitions or examples of their categories.[55] This made coding and interpretation of some of the categories more subjective. For example, "family conflict," which appeared in three typologies, is ambiguous; it is unclear whether it referred to conflict between the family and the medical staff and/or conflict within the family. Additionally, typologies included categories that were difficult to distinguish from each other, for example, separate categories for "constrained decision making" and "threatened autonomy."[56] A lack of well-defined categories made it difficult for institutions that wished to utilize an existing typology to apply it consistently. Additional research is needed to establish the reliability of different raters applying a typology's categories.

Fifth, many of the typologies included multiple, conceptually distinct topics in a single typology.[57] For example, some typologies included categories for "conflict and/or types of conflict." Consultations may be the result of either dilemmas or conflicts. There may be uncertainty about the ethical issue, or interpersonal conflict regarding the ethical issue, but the dynamic is separate from the ethical issue itself. Other typologies included categories regarding particular types of treatment, for example, "DNAR orders," "end-of-life care," and "palliative care." It was unclear whether these categories identified specific ethical issues or clinical scenarios. Finally, some typologies included categories that were coded as "culture" or "legal." These categories generally identified a source of ethical norms rather than the ethical issue; they generally modified rather than characterized the ethical issue. It would have been clearer for typologies to treat these topics as separate issues rather than include them in a single typology. For example, Johnson, Church, Metzger, and Baker[58] distinguished the primary reason for a consult request and involvement of external services (chaplaincy, palliative care, legal, and child protective services); Henriksen Hellyer and colleagues[59] distinguished "evidence of interpersonal conflict," "interpersonal conflict type," "primary reason for consult," "legal involvement," and "consult and end of live [*sic*]".

These limitations suggest the need for a uniform typology. Such a typology would have a number of benefits: it could support clinical practice, scholarship, and professionalization. Data on the frequency of different ethical issues could, for example, inform the development of specified content for a certification examination for clinical ethics consultants. The development and adoption of a uniform typology would be facilitated by collaboration among a variety of institutions. This would provide a diversity of perspectives in developing a typology and promote investment in utilizing the resulting product. Professional organizations could play a crucial role in funding and coordinating the development process.

This systematic review has several limitations. It only retrieved published typologies and did not include typologies that were not published in the scholarly literature. Our utilization of a single database, PubMed, and language, English, may have inadvertently excluded some published clinical ethics typologies. Our review also only retrieved typologies that were published in particular types of articles. The review may, therefore, have omitted some typologies. It, nonetheless, resulted in relatively large listing of typologies. Our preconceptions or biases may have inadvertently influenced our coding of the data. The heterogeneity of the typologies prevents a meta-analysis of the results of the studies.

Our systematic literature review of typologies of clinical ethics consultation identified 30 articles and 27 unique typologies. The studies varied in terms of institution type, geographic location, time frame, and number of consultations. The studies used different primary outcome measures. The typologies differed from one another in number and types of categories, which made comparisons between the studies difficult. This suggests the need for a uniform typology with clear definitions to advance practice and scholarship within the field. We believe that such a typology will provide a common language and framework to categorize consultations and compare consultation patterns.

ACKNOWLEDGMENTS

The authors wish to thank Alison Kissling, MLIS, Cincinnati Children's Hospital Medical Center, and Martina Darragh, MLS, the Kennedy Institute of Ethics, for their assistance in developing the search strategy and conducting the research, and Jennifer Longbottom for verifying the abstraction of data.

NOTES

1. S.J. Youngner et al.,"A National Survey of Hospital Ethics Committees," *Critical Care Medicine* 11, no. 11 (November 1983): 902-05.

2. E. Fox, S. Myers, and R.A. Pearlman, "Ethics Consultation in United States Hospitals: A National Survey," *American Journal of Bioethics* 7, no. 2 (February 2007): 13-25.

3. M.P. Aulisio, R.M. Arnold, and S.J. Youngner, "Health Care Ethics Consultation: Nature, Goals, and Competencies: A Position Paper from the Society for Health and Human Values-Society for Bioethics Consultation Task Force on Standards for Bioethics Consultation," *Annuals of Internal Medicine* 133, no. 1 (4 July 2000): 59-69; J.C. Fletcher and M. Siegler, "What Are the Goals of Ethics Consultation? A Consensus Statement," *The Journal of Clinical Ethics* 7, no. 2 (Summer 1996): 122-6.

4. L.J. Schneiderman et al.,"Effect of Ethics Consultations on Nonbeneficial Life-Sustaining Treatments in the Intensive Care Setting: A Randomized Controlled Trial," *Journal of the American Medical Association* 290, no. 9 (3 September 2003): 1166-72; SUPPORT, "A Controlled Trial to Improve Care for Seriously Ill Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT): The SUPPORT Principal Investigators," *Journal of the American Medical Association* 274, no. 20 (22-29 November 1995): 1591-8.

5. L.M. Johnson, C.L. Church, M. Metzger, and J.N. Baker, "Ethics Consultation in Pediatrics: Long-Term Experience from a Pediatric Oncology Center," *American Journal of Bioethics* 15, no. 5 (2015): 3-17.

6. A.H. Antommaria, "Characterizing Clinical Ethics Consultations: The Need for a Standardized Typology of Cases," *American Journal of Bioethics* 15, no. 5 (2015): 18-20.

7. J. Henriksen Hellyer et al.,"Pediatric Clinical Ethics Consultations at an Academic Medical Center: Does One Size Fit All?" *American Journal of Bioethics* 15, no. 5 (2015): 20-4.

8. L. Gillam, R. McDougall, and C. Delany, "Making Meaning from Experience: A Working Typology for Pediatrics Ethics Consultations," *American Journal of Bioethics* 15, no. 5 (2015): 24-6.

9. "Prospero is an international prospective register of systematic reviews. . . . relevant to health and social care, welfare, public health, education, crime, justice, and international development, where there is a health related outcome." *https://www.crd.york.ac.uk/prospero/.*

10. D.J. Opel et al.,"Characterisation of Organisational Issues in Paediatric Clinical Ethics Consultation: A Qualitative Study," *Journal of Medical Ethics* 35, no. 8 (August 2009): 477-82.

11. S. Merriam, *Qualitative Research: A Guide to Design and Implementation* (San Francisco, Calif.: Jossey-Bass, 2009).

12. Opel et al.,"Characterisation of Organisational Issues," see note 10 above.

13. T. Schenkenberg, "Salt Lake City VA Medical Center's First 150 Ethics Committee Case Consultations: What We Have Learned (So Far)," *HEC Forum* 9, no. 2 (June 1997): 147-58.

14. T. Ramsauer and A. Frewer, "Clinical Ethics Committees and Pediatrics: An Evaluation of Case Consultations,"*Diametros* 22 (2009): 90-104.

15. D. Hruschka et al., "Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research," *Field Methods* 16, no. 3 (2004): 307-31.

16. N. Cobanoglu and L. Algier, "A Qualitative Analysis of Ethical Problems Experienced by Physicians and Nurses in Intensive Care Units in Turkey," *Nursing Ethics* 11, no. 5 (September 2004): 444-58.

17. G. McGee et al., "Successes and Failures of Hospital Ethics Committees: A National Survey of Ethics Committee Chairs," *Cambridge Quarterly of Healthcare Ethics* 11, no. 1 (Winter 2002): 87-93.

18. L.S. Robins, C.H. Braddock, and K.A. Fryer-Edwards, "Using the American Board of Internal Medicine's 'Elements of Professionalism' for Undergraduate Ethics Education," *Academic Medicine* 77, no. 6 (June 2002): 523-31.

19. R.J. McDougall and L. Notini, "What Kinds of Cases Do Paediatricians Refer to Clinical Ethics? Insights from 184 Case Referrals at an Australian Paediatric Hospital," *Journal of Medical Ethics* 42, no. 9 (September 2016): 586-91.

20. Ramsauer and Frewer, "Clinical Ethics Committees and Pediatrics," see note 14 above.

21. M. Fukuyama, A. Asai, K. Itai, and S. Bito, "A Report on Small Team Clinical Ethics Consultation Programmes in Japan," *Journal of Medical Ethics* 34, no. 12 (December 2008): 858-62.

22. R. Forde and I.H. Vandvik, "Clinical Ethics, Information, and Communication: Review of 31 Cases from a Clinical Ethics Committee," *Journal of Medical Ethics* 31, no. 2 (February 2005): 73-7.

23. J.C. Streuli et al., "Five-Year Experience of Clinical Ethics Consultations in a Pediatric Teaching Hospital," *European Journal of Pediatrics* 173, no. 5 (May 2014): 629-36.

24. A.R. Boissy, P.J. Ford, R.C. Edgell, and A.J. Furlan, "Ethics Consultations in Stroke and Neurological Disease: A 7-Year Retrospective Review," *Neurocritical Care* 9, no. 3 (2008): 394-9; C.R. Bruce, M.L. Smith, S. Hizlan, and R.R. Sharp, "A Systematic Review of Activities at a High-Volume Ethics Consultation Service," *The Journal of Clinical Ethics* 22, no. 2 (Summer 2011): 151-64; S.M. Thomas et al.,"Not Just Little Adults: A Review of 102 Paediatric Ethics Consultations," *Acta Paediatrica* 104, no. 5 (May 2015): 529-34.

25. R.D. Orr and E. Moon, "Effectiveness of an Ethics Consultation Service," *Journal of Family Practice* 36, no. 1 (January 1993): 49-53; R.D. Orr and R.M. Perkin, "Clinical Ethics Consultations with Children," *The Journal of Clinical Ethics* 5, no. 4 (Winter 1994): 323-28.

26. Henriksen Hellyer et al.,"Pediatric Clinical Ethics Consultations," see note 7 above; K.M. Swetz, M.E. Crowley, C. Hook, and P.S. Mueller, "Ethics Consultations and Patients with Neurological Diseases: Author Reply," *Mayo Clinic Proceedings* 82, no. 12 (December 2007): 1577-8; K.M. Swetz, M.E. Crowley, C. Hook, and P.S. Mueller, "Report of 255 Clinical Ethics Consultations and Review of the Literature," *Mayo Clinic Proceedings* 82, no. 6 (June 2007): 686-91.

27. J. La Puma et al., "An Ethics Consultation Service in a Teaching Hospital: Utilization and Evaluation," *Journal of the American Medical Association* 260, no. 6 (12 August 1988): 808-11; J. La Puma, C.B. Stocking, C.M. Darling, and M. Siegler, "Community Hospital Ethics Consultation: Evaluation and Comparison with a University Hospital Service," *American Journal of Medicine* 92, no. 4 (April 1992): 346-51.

28. J. La Puma, "Consultations in Clinical Ethics—Issues and Questions in 27 Cases," *Western Journal of Medicine* 146, no. 5 (May 1987): 633-7.

29. E.G. Nilson, C.A. Acres, N.G. Tamerin, and J.J. Fins, "Clinical Ethics and the Quality Initiative: A Pilot Study for the Empirical Evaluation of Ethics Case Consultation," *American Journal of Medical Quality* 23, no. 5 (September-October 2008): 356-64.

30. Johnson, Church, Metzger, and Baker, "Ethics Consultation in Pediatrics," see note 5 above.

31. Schenkenberg, "Salt Lake City VA," see note 13 above.

32. Boissy, Ford, Edgell, and Furlan, "Ethics Consultations in Stroke and Neurological Disease," see note 24 above; Swetz, Crowley, Hook, and Mueller, "Ethics Consultations," see note 26 above.

33. L.S. Johnson, J. Lesandrini, and G.S. Rozycki, "Use of the Medical Ethics Consultation Service in a Busy Level I Trauma Center: Impact on Decision-Making and Patient Care," *American Journal of Surgery* 78, no. 7 (July 2012): 735-40.

34. A.G. Shuman et al., "Clinical Ethics Consultation in Patients with Head and Neck Cancer," *Head & Neck* 35, no. 11 (November 2013): 1647-51; A.G. Shuman et al., "Clinical Ethics Consultation in Oncology," *Journal of Oncology Practice* 9, no. 5 (September 2013): 240-5.

35. La Puma, "Consultations in Clinical Ethics," see note 28 above.

36. Nilson, Acres, Tamerin, and Fins, "Clinical Ethics and the Quality Initiative," see note 29 above.

37. Henriksen Hellyer et al.,"Pediatric Clinical Ethics Consultations," see note 7 above.

38. Shuman et al.,"Clinical Ethics Consultation," see note 34 above.

39. Bruce, Smith, Hizlan, and Sharp, "A Systematic Review," see note 24 above.

40. Swetz, Crowley, Hook, and Mueller, "Ethics Consultations," see note 26 above.

41. Swetz, Crowley, Hook, and Mueller, "Report of 255," see note 26 above.

42. M.E. Romano et al.,"Mandatory Ethics Consultation Policy," *Mayo Clinic Proceedings* 84, no. 7 (July 2009): 581-5.

43. Swetz, Crowley, Hook, and Mueller, "Report of 255," see note 26 above.

44. L.P. Voigt et al.,"Characteristics and Outcomes of Ethics Consultations in an Oncologic Intensive Care Unit," *Journal of Intensive Care Medicine* 30, no. 7 (October 2015): 436-42.

45. Nilson, Acres, Tamerin, and Fins, "Clinical Ethics and the Quality Initiative," see note 29 above.

46. La Puma, Stocking, Darling, and Siegler, "Community Hospital Ethics Consultation," see note 27 above; La Puma et al., "An Ethics Consultation Service in a Teaching Hospital," see note 27 above; Shuman et al., "Clinical Ethics Consultation," see note 34 above; Shuman et al., "Clinical Ethics Consultation in Oncology," see note 34 above.

47. Johnson, Lesandrini, and Rozycki, "Use of the Medical Ethics Consultation," see note 33 above; La Puma et al., "An Ethics Consultation Service in a Teaching Hospital," see note 27 above; Shuman et al., "Clinical Ethics Consultation," see note 34 above; Shuman et al., "Clinical Ethics Consultation in Oncology," see note 34 above.

48. Johnson, Church, Metzger, and Baker, "Ethics Consultation in Pediatrics," see note 5 above; McDougall and Notini, "What Kinds of Cases Do Paediatricians Refer?" see note 19 above.

49. Johnson, Lesandrini, and Rozycki, "Use of the Medical Ethics Consultation," see note 33 above.

50. Nilson, Acres, Tamerin, and Fins, "Clinical Ethics and the Quality Initiative," see note 29 above.

51. Shuman et al., "Clinical Ethics Consultation in Oncology," see note 34 above; K. Wasson et al., "What Ethical Issues Really Arise in Practice at an Academic Medical Center? A Quantitative and Qualitative Analysis of Clinical Ethics Consultations from 2008 to 2013," *HEC Forum* 28, no. 3 (September 2016): 217-28.

52. Bruce, Smith, Hizlan, and Sharp, "A Systematic Review," see note 24 above.

53. J.R. Moeller et al., "Functions and Outcomes of a Clinical Medical Ethics Committee: A Review of 100 Consults," *HEC Forum* 24, no. 2 (June 2012): 99-114.

54. Antommaria, "Characterizing Clinical Ethics Consultations," see note 6 above; Henriksen Hellyer et al., "Pediatric Clinical Ethics Consultations," see note 7 above.

55. Johnson, Lesandrini, and Rozycki, "Use of the Medical Ethics Consultation," see note 33 above.

56. Johnson, Church, Metzger, and Baker, "Ethics Consultation in Pediatrics," see note 5 above.

57. Antommaria, "Characterizing Clinical Ethics Consultations," see note 6 above; Henriksen Hellyer et al.,"Pediatric Clinical Ethics Consultations," see note 7 above.

58. Johnson, Church, Metzger, and Baker, "Ethics Consultation in Pediatrics," see note 5 above.

59. Henriksen Hellyer et al.,"Pediatric Clinical Ethics Consultations," see note 7 above.

APPENDIX

Boissy, A.R., P.J. Ford, R.C. Edgell, and A.J. Furlan. "Ethics Consultations in Stroke and Neurological Disease: A 7-Year Retrospective Review." *Neurocritical Care* 9, no. 3 (2008): 394-9.

Bruce, C.R., M.L. Smith, S. Hizlan, and R.R. Sharp. "A Systematic Review of Activities at a High-Volume Ethics Consultation Service." *The Journal of Clinical Ethics* 22, no. 2 (Summer 2011): 151-64.

Forde, R., and I.H. Vandvik. "Clinical Ethics, Information, and Communication: Review of 31 Cases from a Clinical Ethics Committee." *Journal of Medical Ethics* 31, no. 2 (February 2005): 73-7.

Fukuyama, M., A. Asai, K. Itai, and S. Bito. "A Report on Small Team Clinical Ethics Consultation Programmes in Japan." *Journal of Medical Ethics* 34, no. 12 (December 2008): 858-62.

Henriksen Hellyer, J. et al. "Pediatric Clinical Ethics Consultations at an Academic Medical Center: Does One Size Fit All?" *American Journal of Bioethics* 15, no. 5 (2015): 20-4.

Johnson, L.M., C.L. Church, M. Metzger, and J.N. Baker. "Ethics Consultation in Pediatrics: Long-Term Experience from a Pediatric Oncology Center." *American Journal of Bioethics* 15, no. 5 (2015): 3-17.

Johnson, L.S., J. Lesandrini, and G.S. Rozycki. "Use of the Medical Ethics Consultation Service in a Busy Level I Trauma Center: Impact on Decision-Making and Patient Care." *American Journal of Surgery* 78 no. 7 (July 2012): 735-40.

La Puma, J. "Consultations in Clinical Ethics—Issues and Questions in 27 Cases." *Western Journal of Medicine* 146, no. 5 (May 1987): 633-7.

La Puma, J. et al. "An Ethics Consultation Service in a Teaching Hospital: Utilization and Evaluation." *Journal of the American Medical Association* 260, no. 6 (12 August 1988): 808-11.

La Puma, J., C.B. Stocking, C.M. Darling, and M. Siegler. "Community Hospital Ethics Consultation: Evaluation and Comparison with a University Hospital Service." *American Journal of Medicine* 92, no. 4 (April 1992): 346-51.

McDougall, R.J., and L. Notini. "What Kinds of Cases Do Paediatricians Refer to Clinical Ethics? Insights from 184 Case Referrals at an Australian Paediatric Hospital." *Journal of Medical Ethics* 42, no. 9 (September 2016): 586-91.

Moeller, J.R. et al. "Functions and Outcomes of a Clinical Medical Ethics Committee: A Review of 100 Consults." *HEC Forum* 24, no. 2 (June 2012): 99-114.

Nilson, E.G., C.A. Acres, N.G. Tamerin, and J.J. Fins. "Clinical Ethics and the Quality Initiative: A Pilot Study for the Empirical Evaluation of Ethics Case Consultation." *American Journal of Medical Quality* 23, no. 5 (September-October 2008): 356-64.

Opel, D.J. et al. "Characterisation of Organisational Issues in Paediatric Clinical Ethics Consultation: A Qualitative Study." *Journal of Medical Ethics* 35, no. 8 (August 2009): 477-82.

Orr, R.D., and E. Moon. "Effectiveness of an Ethics Consultation Service." *Journal of Family Practice* 36, no. 1 (January 1993): 49-53.

Orr, R.D., and R.M. Perkin. "Clinical Ethics Consultations with Children." *The Journal of Clinical Ethics* 5, no. 4 (Winter 1994): 323-8.

Perkins, H.S., and B.S. Saathoff. "Impact of Medical Ethics Consultations on Physicians: An Exploratory Study." *American Journal of Medicine* 85, no. 6 (December 1988): 761-65.

Ramsauer, T., and A. Frewer. "Clinical Ethics Committees and Pediatrics: An Evaluation of Case Consultations." *Diametros* 22 (2009): 90-104.

Romano, M.E. et al. "Mandatory Ethics Consultation Policy." *Mayo Clinic Proceedings* 84, no. 7 (July 2009): 581-5.

Schenkenberg, T. "Salt Lake City VA Medical Center's First 150 Ethics Committee Case Consultations: What We Have Learned (So Far)." *HEC Forum* 9, no. 2 (June 1997): 147-58.

Shuman, A.G. et al. "Clinical Ethics Consultation in Oncology." *Journal of Oncology Practice* 9, no. 5 (September 2013): 240-5.

Shuman, A.G. et al. "Clinical Ethics Consultation in Patients with Head and Neck Cancer." *Head & Neck* 35, no. 11 (November 2013): 1647-51.

Streuli, J.C. et al. "Five-Year Experience of Clinical Ethics Consultations in a Pediatric Teaching Hospital." *European Journal of Pediatrics* 173, no. 5 (May 2014): 629-36.

Swetz, K.M., M.E. Crowley, C. Hook, and P.S. Mueller. "Report of 255 Clinical Ethics Consultations and Review of the Literature." *Mayo Clinic Proceedings* 82, no. 6 (June 2007): 686-91.

Swetz, K.M., M.E. Crowley, C. Hook, and P.S. Mueller. "Ethics Consultations and Patients with Neurological Diseases: Author Reply." *Mayo Clinic Proceedings* 82, no. 12 (December 2007): 1577-8.

Tapper, E.B., C.J. Vercler, D. Cruze, and W. Sexson. "Ethics Consultation at a Large Urban Public Teaching Hospital." *Mayo Clinic Proceedings* 85, no. 5 (May 2010): 433-8.

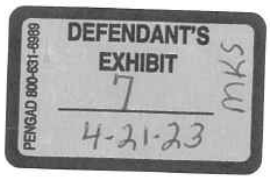
Thomas, S.M. et al. "Not Just Little Adults: A Review of 102 Paediatric Ethics Consultations." *Acta Paediatrica* 104, no. 5 (May 2015): 529-34.

Voigt, L.P. et al. "Characteristics and Outcomes of Ethics Consultations in an Oncologic Intensive Care Unit." *Journal of Intensive Care Medicine* 30, no. 7 (October 2015): 436-42.

Wasson, K. et al. "What Ethical Issues Really Arise in Practice at an Academic Medical Center? A Quantitative and Qualitative Analysis of Clinical Ethics Consultations from 2008 to 2013." *HEC Forum* 28, no. 3 (September 2016): 217-28.

Yen, B.M., and L.J. Schneiderman. "Impact of Pediatric Ethics Consultations on Patients, Families, Social Workers, and Physicians." *Journal of Perinatology* 19, no. 5 (July-August 1999): 373-8.

92



ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 401–406

Journal of Clinical Epidemiology

GRADE guidelines: 3. Rating the quality of evidence

Howard Balshem[a,*], Mark Helfand[a,b], Holger J. Schünemann[c], Andrew D. Oxman[d], Regina Kunz[e], Jan Brozek[c], Gunn E. Vist[d], Yngve Falck-Ytter[f], Joerg Meerpohl[g,h], Susan Norris[i], Gordon H. Guyatt[c]

[a]*Oregon Evidence-based Practice Center, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, USA*
[b]*Portland VA Medical Center, Portland, OR, USA*
[c]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[d]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs plass, 0130 Oslo, Norway*
[e]*Basel Institute of Clinical Epidemiology, University Hospital Basel, Hebelstrasse 10, 4031 Basel, Switzerland*
[f]*Division of Gastroenterology, Case Medical Center and VA, Case Western Reserve University, Cleveland, OH 44106, USA*
[g]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, 79104 Freiburg, Germany*
[h]*Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, University Medical Center Freiburg, 79106 Freiburg, Germany*
[i]*Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239-3098, USA*

Accepted 30 July 2010

Abstract

This article introduces the approach of GRADE to rating quality of evidence. GRADE specifies four categories—high, moderate, low, and very low—that are applied to a body of evidence, not to individual studies. In the context of a systematic review, quality reflects our confidence that the estimates of the effect are correct. In the context of recommendations, quality reflects our confidence that the effect estimates are adequate to support a particular recommendation. Randomized trials begin as high-quality evidence, observational studies as low quality. "Quality" as used in GRADE means more than risk of bias and so may also be compromised by imprecision, inconsistency, indirectness of study results, and publication bias. In addition, several factors can increase our confidence in an estimate of effect. GRADE provides a systematic approach for considering and reporting each of these factors. GRADE separates the process of assessing quality of evidence from the process of making recommendations. Judgments about the strength of a recommendation depend on more than just the quality of evidence.

Keywords: Quality assessment; Body of evidence; Imprecision; Indirectness; Inconsistency; Publication bias

1. Introduction

In the two previous articles in this series, we introduced GRADE; provided an overview of the GRADE process for developing recommendations and the final outputs of that process, the evidence profile, and Summary of Findings table; and described the process for framing questions and identifying outcomes [1,2]. In this third article, we will introduce GRADE's approach to rating the quality of evidence. The goal is to provide a conceptual overview of the approach. A more detailed description, accompanied by examples, will follow in articles dealing with factors that may lead to rating down or rating up the quality of evidence [3–7].

2. What we do not mean by quality of evidence

In discussions of quality of evidence, confusion often arises between evidence and opinion and between quality of evidence and strength of recommendations. We, therefore, begin by explaining what we do not mean by quality of evidence.

3. Opinion is not evidence

In the absence of high-quality evidence, clinicians must look to lower quality evidence to guide their decisions.

The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the *Journal of Clinical Epidemiology* Web site.

* Corresponding author. Oregon Evidence-based Practice Center, OHSU-BICC, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, USA. Tel.: 503-220-8262 x54487; fax: 503-418-3332.
E-mail address: balshemh@ohsu.edu (H. Balshem).

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2010.07.015

Key Points

- GRADE provides a framework for assessing quality that encourages transparency and an explicit accounting of the judgments made.
- GRADE distinguishes between quality assessment conducted as part of a systematic review and that undertaken as part of guideline development.
- The optimal application of GRADE requires systematic review of the impact of alternative management strategies on all patient-important outcomes.
- Information about study limitations, imprecision, inconsistency, indirectness, and publication bias is necessary for decision makers, clinicians, and patients to understand and have confidence in the assessment of quality and estimate of effect size.

Confusion arises when, in such situations, guideline developers classify "expert opinion" as a type of evidence. Developing recommendations always requires the opinion of experts, the basis of which includes experience with patients, an understanding of biology and mechanism, and knowledge and understanding of preclinical and early clinical research as well as of the results of randomized clinical trials and observational studies. Guideline developers should always engage experts to help understand the evidence; they must also uncover and make clear the evidence that underlies the experts' opinions and rate the quality of that evidence, not the opinions that follow from the evidence and its interpretation.

An example illustrates the difference between evidence and expert opinion. Suppose that during attending rounds with medical students and residents, an endocrinologist explains the rationale for tight glycemic control in diabetes. Table 1 shows the two assertions he makes and the evidence he cites to support them. The evidence he cites for opinion 1 is exclusively his personal clinical experience. For opinion 2, he cites his own experience and refers (with no more than a general statement) to evidence from clinical research.

It seems highly plausible that opinion 1 might reasonably be based on careful observation. If patients who complain of fatigue, polyuria, or other symptoms return in a few days saying they are better, initiation of treatment is the likeliest explanation. The phenomenon of a patient who had no complaints returning, a few days later, to say how much better she is would be particularly memorable. Unfortunately, there are many other potential explanations of these observations. The endocrinologist's impression of the extent of patients' reports of benefit may be inaccurate, he may be forgetting many patients who failed to improve, or the apparent improvement in some patients may be because of natural history, placebo effects, leading questions on the part of the clinician, or the patient's desire to please. Without, at the very least, a rigorous and structured approach to data collection, we could consider the endocrinologist's report of his clinical experience (but not the opinion that he arrived at from his interpretation of that experience) as evidence from an uncontrolled case series and classify it as very low quality.

Whereas the implicit study design underlying the evidence for opinion 1 is a before–after study, opinion 2 suggests a parallel group comparison, which in this case has serious problems. If indeed his memory is accurate (patients with tighter control in his practice do achieve better outcomes), the reason may be that their success in controlling their glucose reflects differences in their underlying disease strongly associated with their likelihood of suffering complications. This risk of bias from unrecognized prognostic imbalance, as well as from the uncertainty and imprecision associated with the endocrinologist's memory of the events, would lead us again to classify his observations as very low quality evidence.

4. A particular quality of evidence does not necessarily imply a particular strength of recommendation

A second area of confusion relates to the distinction between assessing the quality of evidence and making a recommendation. Later articles in this series will provide a detailed discussion of GRADE's approach to deciding on the direction and strength of recommendations. We note here the importance of GRADE's explicit separation of the process for assessing the quality of a body of evidence from the process for making recommendations based in part on those assessments. Although higher quality evidence is more likely to be associated with strong recommendations than lower quality evidence, a particular level of quality does not imply a particular strength of recommendation. Sometimes, low or very low quality evidence can lead to a strong recommendation.

For instance, consider the decision to administer aspirin or acetaminophen to children with chicken pox. Observational studies have observed an association between aspirin administration and Reye's syndrome [8–11]. Because aspirin and acetaminophen are similar in their analgesic and antipyretic effects, the low-quality evidence regarding the potential harms of aspirin does not preclude a strong recommendation for acetaminophen.

Similarly, high-quality evidence does not necessarily imply strong recommendations. For example, faced with a first deep venous thrombosis (DVT) with no obvious provoking factor patients must, after the first months of anticoagulation, decide whether to continue taking warfarin long term. High-quality randomized controlled trials show that continuous warfarin will decrease the risk of recurrent thrombosis but at the cost of increased risk of bleeding and inconvenience [12–15]. Because patients with varying values and

Table 1
Expert opinion vs. evidence

Expert opinion	Evidence
Tight control will make a patient feel better	"In my 20 years in practice I have started treatment for newly diagnosed diabetes many times. I almost always see these patients back a week or so after starting treatment, and the great majority say they feel much better than they did before. Even a patient who denied having any complaints or symptoms will come back and say she has more energy, particularly in the afternoons, and will marvel at how much better she feels in general."
Tight control will reduce the long-term risk of developing kidney disease, neuropathy, and blindness	"I institute tight control on every patient—I believe they all deserve the best possible treatment—so I have a lot of experience with this. I have many patients who have been with me for a decade, or even several decades, and who take their medicine faithfully and have great blood sugars. These patients also have very few complications. On the other hand, I have a lot of patients who have terrible control and develop complications early on. Also, there are a lot of studies showing that tight control reduces the risk of complications."

preferences are likely to make different choices, guideline panels addressing whether patients should continue or terminate warfarin may—despite the high-quality evidence—offer a weak recommendation.

5. So what do we mean by "quality of evidence"?

GRADE distinguishes between quality assessment conducted as part of a systematic review and that undertaken in the process of guideline development. We, therefore, provide two definitions of "quality of evidence."

The optimal application of GRADE requires systematic reviews of the impact of alternative management approaches on all patient-important outcomes [1]. In the context of a systematic review, the ratings of the quality of evidence reflect the extent of our confidence that the estimates of the effect are correct. In the context of making recommendations, the quality ratings reflect the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.

The reason for the different definitions is that the conduct of systematic reviews does not include processes required for making rigorous recommendations. In particular, unless the systematic review team includes members who will use the review as part of guideline development, authors of systematic reviews are, generally, not in a position to weigh the trade-offs between the desirable and undesirable consequences of adhering to a recommendation. Relevant stakeholders are in a better position to make these judgments. For example, in the DVT case described earlier, a systematic review might provide reliable estimates of the magnitude of effect and associated confidence intervals (CIs) for symptomatic thromboembolism and bleeding and the mortality associated with both of these events, but the reviewers who wrote it would not be able to provide reliable judgments about whether the benefit of warfarin treatment is worth the risk. Such judgments must also include considerations of values, cost, and pertinent stakeholder input.

On the other hand, a guideline (or a clinician applying the evidence from a systematic review) must assess the quality of the evidence in the context of the decision regarding anticoagulation. In considering this trade-off, a guideline panel must decide whether or not to recommend anticoagulation (and the strength of that recommendation) in light of the effect on the risk of symptomatic thromboembolism, their confidence in the effect estimates, and the corresponding risks and confidence in estimates of serious bleeding. Although the processes for assessing quality are the same, authors of systematic reviews and authors of guidelines will apply the criteria differently. We will highlight this different application of criteria in the fifth article in this series, which addresses the assessment of precision in rating the quality of the evidence [5].

6. Quality in GRADE means more than risk of bias

In the clinical epidemiological literature, when used at all, "quality" commonly refers to a judgment on the internal validity (i.e., risk of bias) of an individual study. To arrive at a rating, reviewers consider features in controlled trials such as randomization, allocation concealment, blinding, and use of intention to treat analysis. In observational studies, they consider appropriate measurement of exposure and outcome as well as appropriate control of confounding; in both controlled trials and observational studies, they consider loss to follow-up and may consider other aspects of design, conduct, and analysis that influence the risk of bias.

GRADE judgments refer not to individual studies but to a body of evidence, and quality, as used in GRADE, means more than risk of bias. A body of evidence (for instance, a number of well-designed and executed trials) may be associated with a low risk of bias, but our confidence in effect estimates may be compromised by a number of other factors (imprecision, inconsistency, indirectness, and publication bias). There are also factors, particularly relevant to observational studies, that may lead to rating up quality, including the magnitude of treatment effect and the presence of a dose–response gradient.

GRADE's specific uses of the terms "quality" and "risk of bias" (labeled "study limitations" in previous GRADE publications) require authors to take care in using these terms when they describe their findings and reasoning in

Table 2
Significance of the four levels of evidence

Quality level	Current definition	Previous definition
High	We are very confident that the true effect lies close to that of the estimate of the effect	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect	Any estimate of effect is very uncertain

the context of a systematic review or guideline. Well-conducted studies may be part of a body of evidence rated low quality because they only provide indirect or imprecise evidence for the question of interest. Although clinical epidemiologists and others have attributed other meanings to the word "quality" (typically risk of bias), we believe the meaning described here corresponds more closely to the common and nontechnical understanding of "quality."

7. GRADE specifies four categories for the quality of a body of evidence

Although the quality of evidence represents a continuum, the GRADE approach results in an assessment of the quality of a body of evidence as high, moderate, low, or very low. Table 2 presents what GRADE means by each of these four categories and contrasts their current definition with the previous definition [16], which focused on the implications of the levels of evidence for future research (the lower the quality, the more likely further research would change our confidence in the estimates, and the estimates themselves). The earlier characterization has been criticized—we believe legitimately—because there are many situations in which we cannot expect higher quality evidence to be forthcoming. We, nevertheless, consider the prior characterization of quality to provide an alternative under circumstances when obtaining new compelling evidence is plausible.

8. Arriving at a quality rating

When we speak of evaluating quality, we are referring to an overall rating for each important outcome across studies. As discussed in the previous article in this series that addressed the framing of the question [2], before assessing the quality of the evidence, systematic reviewers and guideline developers should identify all potential patient-important outcomes, including benefits, harms, and costs. Reviewers will then assess the quality of evidence for each important outcome.

Table 3 summarizes GRADE's approach to rating the quality of evidence, which begins with the study design (trials or observational studies) and then addresses five reasons to possibly rate down the quality of evidence and three to possibly rate up the quality. Subsequent articles in this series will address, in detail, the meaning and use of each of these criteria. Here, we discuss why these criteria, in particular, have been identified as important in assessing the quality of a body of evidence.

Table 3
A summary of GRADE's approach to rating quality of evidence

Study design	Initial quality of a body of evidence	Lower if	Higher if	Quality of a body of evidence
Randomized trials	High ⊕⊕⊕⊕	Risk of Bias −1 Serious −2 Very serious Inconsistency −1 Serious −2 Very serious Indirectness −1 Serious −2 Very serious Imprecision −1 Serious −2 Very serious Publication bias −1 Likely −2 Very likely	Large effect +1 Large +2 Very large Dose response +1 Evidence of a gradient All plausible residual confounding +1 Would reduce a demonstrated effect +1 Would suggest a spurious effect if no effect was observed	High (four plus: ⊕⊕⊕⊕) Moderate (three plus: ⊕⊕⊕◯) Low (two plus: ⊕⊕◯◯) Very low (one plus: ⊕◯◯◯)
Observational studies	Low ⊕⊕◯◯			

9. Rationale for using GRADE's definition of quality

To be useful to decision makers, clinicians, and patients, systematic reviews must provide not only an estimate of effect for each outcome but also the information needed to judge whether these estimates are likely to be correct. What information about the studies in a review affects our confidence that the estimate of an effect is correct?

To answer this question, consider an example. Suppose you are told that a recent Cochrane review reported that, in patients with chronic pain, the number needed to treat (NNT) for clinical success with topical salicylates was 6 (95% CI = 4–13) compared with placebo. What additional information would you seek to help you decide whether to believe this estimate and how to apply it?

The most obvious questions might be the following: how many studies were pooled to get this estimate; how many patients did they include; and how wide were the CIs around the effect estimate? Were they randomized controlled trials? Did the studies have important limitations, such as lack of blinding or large or differential loss to follow-up in the compared groups? The questions thus far relate to GRADE categories of imprecision and risk of bias.

But there are also other important questions. Is there evidence that more studies of this treatment were conducted, but some were inaccessible to the reviewers? If so, how likely is it that the results of the review reflect the overall experience with this treatment? Did the trials have similar or widely varying results? Was the outcome measured at an appropriate time, or were the studies too short in duration to have much relevance? What part of the body was involved in the interventions (and thus, to what part of the body can we confidently apply these results)? These latter questions refer to the GRADE categories of publication bias, inconsistency, and indirectness. Without answers to (or at least information about) these questions, it is not possible to determine how much confidence to attach to the reported NNT and CIs.

GRADE identified its five categories—risk of bias, imprecision, inconsistency, indirectness, and publication bias—because they address nearly all issues that bear on the quality of evidence. For any given question, moreover, information about each of these categories is likely to be essential to judge whether the estimate is likely to be correct. These categories were arrived at through a case-based process by members of GRADE, who identified a broad range of issues and factors related to the assessment of the quality of studies. All potential factors were considered, and through an iterative process of discussion and review, concerns were scrutinized and solutions narrowed by consensus to these five categories.

GRADE's approach to quality implies that every systematic review should provide information about each of the categories (and any other pertinent issues in a particular case). Decision makers, whether they are guideline developers or clinicians, find it difficult to use a systematic review that does not provide this information. Good systematic reviews and clinical practice guidelines have commonly emphasized appraisal of the risk of bias (study limitations) using explicit criteria. Often, however, the focus has been on assessments across outcomes for each study rather than on each important outcome across studies. Assessment of other factors that determine how much confidence can be placed in estimates of effect has often been lacking. Before the adoption of GRADE, standards for reporting systematic reviews have not made clear how this information should be presented. GRADE provides a structure for systematic reviews and clinical practice guidelines to ensure they address the key questions that are pertinent to rating the quality of the evidence for all outcomes relevant to a particular question in a consistent systematic manner.

10. Conclusion

In closing, we caution against a mechanistic approach toward the application of the criteria for rating the quality of the evidence up or down. Although GRADE suggests the initial separate consideration of five categories of reasons for rating down the quality of evidence, and three categories for rating up, with a yes/no decision regarding rating up or down in each case, the final rating of overall evidence quality occurs in a continuum of confidence in the validity, precision, consistency, and applicability of the estimates. Fundamentally, the assessment of evidence quality is a subjective process, and GRADE should not be seen as obviating the need for or minimizing the importance of judgment or as suggesting that quality can be objectively determined.

As we repeatedly stress throughout this series, use of GRADE will not guarantee consistency in assessment, whether of the quality of evidence or of the strength of recommendations. There will be cases in which competent reviewers will have honest and legitimate disagreement about the interpretation of evidence. In such cases, the merit of GRADE is that it provides a framework that guides one through the critical components of this assessment and an approach to analysis and communication that encourages transparency and an explicit accounting of the judgments involved.

References

[1] Guyatt GH, Oxman AD, Kunz R, Vist GE, Brozek J, Norris S, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383–94 [in this issue].

[2] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist GE, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395–400 [in this issue].

[3] Guyatt GH, Oxman AD, Vist GE, Kunz R, Brozek J, Alonso-Coello, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol 2011;64:407–15.

[4] Guyatt GH, Oxman AD, Montori V, Vist GE, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence - publication bias. J Clin Epidemiol. In press.

[5] Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines: 6. Rating the quality of evidence—imprecision (random error). J Clin Epidemiol. In press.

[6] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. J Clin Epidemiol. In press.

[7] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. J Clin Epidemiol. In press.

[8] Waldman RJ, Hall WN, McGee H, Van Amburg G. Aspirin as a risk factor in Reye's syndrome. JAMA 1982;247:3089–94.

[9] Starko KM, Ray CG, Dominguez LB, Stromberg WL, Woodall DF. Reye's syndrome and salicylate use. Pediatrics 1980;66:859–64.

[10] Halpin TJ, Holtzhauer FJ, Campbell RJ, Hall LJ, Correa-Villasenor A, Lanese R, et al. Reye's syndrome and medication use. JAMA 1982;248:687–91.

[11] Hurwitz ES, Barrett MJ, Bregman D, Gunn WJ, Pinsky P, Schonberger LB, et al. Public health service study of Reye's syndrome and medications: report of the main study. JAMA 1987;257:1905–11.

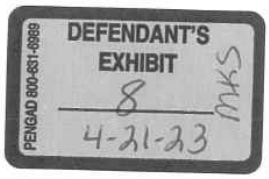
[12] Kearon C, Gent M, Hirsh J, Weitz J, Kovacs MJ, Anderson DR, et al. A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. N Engl J Med 1999;340:901.

[13] Campbell IA, Bentley DP, Prescott RJ, Routledge PA, Shetty HGM, Williamson IJ. Anticoagulation for three versus six months in patients with deep vein thrombosis or pulmonary embolism, or both: randomised trial. BMJ 2007;334:674.

[14] Kearon C, Ginsberg JS, Anderson DR, Kovacs MJ, Wells P, Julian JA, et al. Comparison of 1 month with 3 months of anticoagulation for a first episode of venous thromboembolism associated with a transient risk factor. J Thromb Haemost 2004;2:743–9.

[15] Agnelli G, Prandoni P, Santamaria MG, Bagatella P, Iorio A, Bazzan M, et al. Three months versus one year of oral anticoagulant therapy for idiopathic deep venous thrombosis. N Engl J Med 2001;345:165.

[16] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.



ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 407–415

Journal of Clinical Epidemiology

GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias)

Gordon H. Guyatt[a,*], Andrew D. Oxman[b], Gunn Vist[b], Regina Kunz[c], Jan Brozek[a], Pablo Alonso-Coello[d], Victor Montori[e], Elie A. Akl[f], Ben Djulbegovic[g,h,i], Yngve Falck-Ytter[j], Susan L. Norris[k], John W. Williams Jr.[l], David Atkins[m], Joerg Meerpohl[n,o], Holger J. Schünemann[a]

[a]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[b]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St. Olavs plass, 0130 Oslo, Norway*
[c]*Academy of Swiss Insurance Medicine, University Hospital Basel; and Basel Institute of Clinical Epidemiology and Biostatistics, University Hospital Basel, Hebelstrasse 10, 4031 Basel, Switzerland*
[d]*Iberoamerican Cochrane Center-Servicio de Epidemiología Clínica y Salud Pública and CIBER de Epidemiología y Salud Pública (CIBERESP), Hospital de Sant Pau, Universidad Autónoma de Barcelona, Barcelona 08041, Spain*
[e]*Knowledge and Encounter Research Unit, Mayo Clinic, Rochester, MN, USA*
[f]*Department of Medicine, State University of New York at Buffalo, NY, USA*
[g]*Center for Evidence-based Medicine and Health Outcomes Research, University of South Florida, Tampa, FL 33612, USA*
[h]*Department of Hematology, H. Lee Moffitt Cancer Center & Research Institute, 12901 Bruce B. Downs Boulevard, MDC02, Tampa, FL 33612, USA*
[i]*Department of Health Outcomes and Behavior, H. Lee Moffitt Cancer Center & Research Institute, 12901 Bruce B. Downs Boulevard, MDC02, Tampa, FL 33612, USA*
[j]*Division of Gastroenterology, Case and VA Medical Center, Case Western Reserve University, Cleveland, OH 44106, USA*
[k]*Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239-3098, USA*
[l]*Duke University Medical Center and Durham Veterans Affairs Center for Health Services Research in Primary Care Durham, NC 27705, USA*
[m]*QUERI Program, Office of Research and Development, Department of Veterans Affairs, Washington, DC, USA*
[n]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, 79104 Freiburg, Germany*
[o]*Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, University Medical Center Freiburg, 79106 Freiburg, Germany*

Accepted 30 July 2010

Abstract

In the GRADE approach, randomized trials start as high-quality evidence and observational studies as low-quality evidence, but both can be rated down if most of the relevant evidence comes from studies that suffer from a high risk of bias. Well-established limitations of randomized trials include failure to conceal allocation, failure to blind, loss to follow-up, and failure to appropriately consider the intention-to-treat principle. More recently recognized limitations include stopping early for apparent benefit and selective reporting of outcomes according to the results. Key limitations of observational studies include use of inappropriate controls and failure to adequately adjust for prognostic imbalance. Risk of bias may vary across outcomes (e.g., loss to follow-up may be far less for all-cause mortality than for quality of life), a consideration that many systematic reviews ignore. In deciding whether to rate down for risk of bias—whether for randomized trials or observational studies—authors should not take an approach that averages across studies. Rather, for any individual outcome, when there are some studies with a high risk, and some with a low risk of bias, they should consider including only the studies with a lower risk of bias.

Keywords: GRADE; quality of evidence; risk of bias; confidence in estimates; blinding; concealment

1. Introduction

In three previous articles in our series describing the GRADE system of rating the quality of evidence and grading the strength of recommendations, we have described the process of framing the question and introduced GRADE's approach to rating the quality of evidence. This fourth article deals with one of the five categories of reasons for rating down the quality of evidence, study limitations (risk of bias).

The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the *Journal of Clinical Epidemiology* Web site.

* Corresponding author. CLARITY Research Group, Department of Clinical Epidemiology and Biostatistics, Room 2C12, 1200 Main Street, West Hamilton, Ontario, Canada L8N 3Z5. Tel.: +905-527-4322; fax: +905-523-8781.

E-mail address: guyatt@mcmaster.ca (G.H. Guyatt).

0895-4356/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi: 10.1016/j.jclinepi.2010.07.017

Key points

- In the GRADE approach, both randomized trials (which start as high quality evidence) and observational studies (which start as low quality evidence) can be rated down if relevant evidence comes from studies that suffer from a high risk of bias.
- Risk of bias can differ across outcomes when, for instance, each outcome is informed by a different subset of studies (e.g. mortality from some trials, quality of life from others).
- Current systematic reviews are often limited in their usefulness for guidelines because they rate risk of bias by studies across outcomes rather than by outcome across studies.

2. Rating down quality for risk of bias

Both randomized controlled trials (RCTs) and observational studies may incur additional risk of misleading results if they are flawed in their design or conduct—what other publications refer to as problems with "validity" or "internal validity" and we label "study limitations" or "risk of bias."

3. Study limitations in randomized trials

Readers can refer to many authoritative discussions of the study limitations that often afflict RCTs (Table 1). Two of these discussions are particularly consistent with GRADE's conceptualization, which include a focus on outcome specificity (i.e., the focus of risk of bias is not the individual study but rather the individual outcome, and quality can differ across outcomes in individual trials, or a series of trials [1,2]). We shall highlight three of the criteria in Table 1. The importance of the first of these, stopping early for benefit, has only recently been recognized. Recent evidence has also emerged regarding the second, selective outcome reporting [3,4]. Furthermore, the positioning of selective outcome reporting in taxonomies of bias can be confusing. Some may intuitively think it should be categorized with publication bias, rather than as an issue of risk of bias within individual studies. Finally, we highlight loss to follow-up because it is often misunderstood.

Before we do so, however, we note one additional issue. Recent evidence suggests that bias associated with lack of blinding and lack of concealment may be greater in trials with subjective outcomes [5]. Systematic review authors and guideline developers should consider this evidence when making decisions about rating down quality for risk of bias.

4. Stopping early for benefit

Theoretical consideration [6], simulations [7], and empirical evidence [8] all suggest that trials stopped early for benefit overestimate treatment effects. The most recent empirical work suggests that in the real world, formal stopping rules do not reduce this bias, that it is evident in stopped early trials with less than 500 events and that on average the ratio of relative risks in trials stopped early vs. the best estimate of the truth (trials not stopped early) is 0.71 [9].

Because in most cases the major contributor to the overestimation of treatment effects in trials stopped early for benefit is chance, including stopping early as a source of bias is questionable. Nevertheless, the presence of stopped early trials, particularly when they contribute substantial weight in a meta-analysis, should alert systematic review authors and guideline developers to the possibility of a substantial overestimate of treatment effect. Systematic reviews should provide sensitivity analyses of results including and excluding studies that stopped early for benefit; if estimates differ appreciably, those restricted to the trials that did not stop early should be considered the more credible. When evidence comes

Table 1
Study limitations in randomized trials

1. Lack of allocation concealment
   Those enrolling patients are aware of the group (or period in a crossover trial) to which the next enrolled patient will be allocated (major problem in "pseudo" or "quasi" randomized trials with allocation by day of week, birth date, chart number, etc)
2. Lack of blinding
   Patient, care givers, those recording outcomes, those adjudicating outcomes, or data analysts are aware of the arm to which patients are allocated (or the medication currently being received in a crossover trial)
3. Incomplete accounting of patients and outcome events
   Loss to follow-up and failure to adhere to the intention-to-treat principle in superiority trials; or in noninferiority trials, loss to follow-up, and failure to conduct both analyses considering only those who adhered to treatment, and all patients for whom outcome data are available
4. Selective outcome reporting bias
   Incomplete or absent reporting of some outcomes and not others on the basis of the results
5. Other limitations
   Stopping early for benefit
   Use of unvalidated outcome measures (e.g., patient-reported outcomes)
   Carryover effects in crossover trial
   Recruitment bias in cluster-randomized trials

primarily or exclusively from trials stopped early for benefit, authors should infer that substantial overestimates are likely in trials with fewer than 500 events and that large overestimates are likely in trials with fewer than 200 events [9].

5. Selective outcome reporting

When authors or study sponsors selectively report positive outcomes and analyses within a trial, critics have used the label "selective outcome reporting." Recent evidence suggests that selective outcome reporting, which tends to produce overestimates of the intervention effects, may be widespread [4,10–13].

For example, a systematic review of the effects of testosterone on erection satisfaction in men with low testosterone identified four eligible trials [14]. The largest trial's results were reported only as "not significant" and could not, therefore, contribute to the meta-analysis. Data from the three smaller trials suggested a large treatment effect (1.3 standard deviations, 95% confidence interval 0.2, 2.3). The review authors ultimately obtained the complete data from the larger trial: after including the less impressive results of the large trial, the magnitude of the effect was smaller and no longer statistically significant (0.8 standard deviations, 95% confidence interval −0.05, 1.63) [15].

The Cochrane handbook suggests that definitive evidence that selective reporting has not occurred requires access to a protocol developed before the study was undertaken [2]. Selective reporting is present if authors acknowledge prespecified outcomes that they fail to report or report outcomes incompletely such that they cannot be included in a meta-analysis. One should suspect reporting bias if the study report fails to include results for a key outcome that one would expect to see in such a study or if composite outcomes are presented without the individual component outcomes.

Note that within the GRADE framework, which rates the quality of a body of evidence, suspicion of selective reporting bias in a number of included studies may lead to rating down of quality of the body of evidence. For instance, in the testosterone example above, had the authors not obtained the missing data, they would have considered rating down the body of evidence for the selective reporting bias suspected in the largest study.

6. Loss to follow-up

Historically, methodologists have sometimes suggested arbitrary thresholds for acceptable loss to follow-up (e.g., less than 20%). The significance of particular rates of loss to follow-up, however, varies widely and is dependent on the relation between loss to follow-up and number of events. For instance, loss to follow-up of 5% in both intervention and control groups would entail little threat of bias if event rates were 20% and 40% in intervention and control groups, respectively. If event rates were 2% and 4%, however, concern with 5% loss to follow-up is much greater.

To state this as a general rule, the higher the proportion lost to follow-up in relation to intervention and control event rates, and differences between intervention and control groups, the greater the threat of bias. Even with relatively high rates of loss to follow-up, however, bias will result only if the number lost is imbalanced between groups or the relationship between loss to follow-up and the likelihood of events differs between intervention and control groups. Unfortunately, we never know if the relationship between loss to follow-up and the likelihood of events does or does not differ in intervention and control groups; large loss to follow-up in relation to the number of events always, therefore, raises the issue of a serious threat of bias.

The issue is conceptually identical with continuous outcomes: Was the loss to follow-up such that reasonable assumptions about differences in outcomes among those lost to follow-up in intervention and control groups could change the overall results in an important way? One can test a variety of assumptions about rates of events in those lost to follow-up when the outcome is a binary variable. One can also conduct such sensitivity analyses when the data are continuous, although the statistical modeling is more challenging.

7. Study limitations in observational studies

Systematic reviews of tools to assess the methodological quality of nonrandomized studies have identified more than 200 checklists and instruments [16–19]. Table 2 summarizes key criteria for observational studies that reflect the contents of these checklists. Judgments associated with assessing study limitations in observational studies are often complex; here, we address two key issues that arise in assessing risk of bias.

7.1. Case series: the problem of missing internal controls

Ideally, observational studies will choose contemporaneous comparison groups that, as far as possible, differ from intervention groups only in the decision (typically by

Table 2
Study limitations in observational studies

1. Failure to develop and apply appropriate eligibility criteria (inclusion of control population)
Under- or overmatching in case–control studies
Selection of exposed and unexposed in cohort studies from different populations
2. Flawed measurement of both exposure and outcome
Differences in measurement of exposure (e.g., recall bias in case–control studies)
Differential surveillance for outcome in exposed and unexposed in cohort studies
3. Failure to adequately control confounding
Failure of accurate measurement of all known prognostic factors
Failure to match for prognostic factors and/or lack of adjustment in statistical analysis
4. Incomplete follow-up

patient or clinician) not to use the intervention. Researchers will enroll and observe intervention and comparison group patients in identical ways. This is the prototypical design using what might be called "internal controls"—internal, that is, to the study under conduct.

An alternative approach is to study only patients exposed to the intervention—a design we refer to as a case series (others may use "single group cohort"). To make inferences regarding intervention effects, case series must still refer to results in a comparison group. In many case series, however, the source of comparison group results is implicit or unclear. Such vagueness raises serious questions about the prognostic similarity of intervention and comparison groups and will usually warrant rating down from low- to very low-quality evidence. For instance, in considering the relative impact of low–molecular weight heparin vs. unfractionated heparin in pregnant women, we find systematic reviews of the incidence of bleeding in women receiving the former agent [20,21] but no direct comparisons with the latter.

Thus, case series typically yield very low-quality evidence. There are, however, exceptions. Consider the question of the impact of routine colonoscopy vs. no screening for colon cancer on the rate of perforation associated with colonoscopy. Here, a large series of representative patients undergoing colonoscopy will provide high-quality evidence. When control rates are near zero, case series of representative patients (one might call these cohort studies) can provide high-quality evidence of adverse effects associated with an intervention. One should not confuse these with isolated case reports of associations between exposures and rare adverse outcomes (as have, for instance, been reported with vaccine exposure).

7.2. *Dealing with prognostic imbalance*

Observational studies are at risk of bias because of differences in prognosis in exposed and unexposed populations; to the extent that the two groups come from the same time, place, and population, this risk of bias is diminished. Nevertheless, prognostic imbalance threatens the validity of all observational studies. If the available studies have failed to measure known important prognostic factors, have measured them badly, or have failed to take these factors into account in their analysis (by matching or statistical adjustment), review authors and guideline developers should consider rating down the quality of the evidence from low to very low.

For example, a cohort study using a large administrative database demonstrated an increased risk of cancer-related mortality in diabetic patients using sulfonylureas or insulin relative to metformin [22]. The investigators did not have data available and could, therefore, not adjust for key prognostic variables, including smoking, family history of cancer, occupational exposure, dietary history, and exposure to pollutants. Thus, the study—and others like it that fail to adjust for key prognostic variables—provides only very low-quality evidence of a causal relation between the hypoglycemic agent and cancer deaths.

8. Limitations of GRADE's approach to assessing risk of bias in individual studies

GRADE's approach to assessing risk of bias shares two fundamental limitations with the very large number of alternative approaches. First, empirical evidence supporting the criteria is limited—attempts to show systematic difference between studies that meet and do not meet specific criteria have shown inconsistent results. Second, the relative weight one should put on the criteria remains uncertain.

The GRADE approach is less comprehensive than many systems, emphasizing simplicity and parsimony over completeness. GRADE's approach does not provide a quantitative rating of risk of bias. Although such a rating has advantages, we share with the Cochrane Collaboration methodologists a reluctance to provide a risk of bias score that, by its nature, must make questionable assumptions about the relative extent of bias associated with individual items and fails to consider the context of the individual items.

9. Summarizing study limitations must be outcome specific

Sources of bias may vary in importance across outcomes. Thus, within a single study, one may have higher quality evidence for one outcome than for another. For instance, RCTs of steroids for acute spinal cord injury measured both all-cause mortality and, based on a detailed physical examination, motor function [23–25]. Blinding of outcome assessors is irrelevant for mortality but crucial for motor function. Thus, as in this example, if the outcome assessors in the primary studies on which a guideline panel relies were not blinded, the panel might categorize evidence for all-cause mortality as having no serious study limitations and rate down the evidence for motor function by one level on the basis of serious study limitations.

10. Summarizing risk of bias requires consideration of all relevant evidence

Every study addressing a particular outcome will differ, to some degree, in risk of bias. Review authors and guideline developers must make an overall judgment, considering all the evidence, whether quality of evidence for an outcome warrants rating down on the basis of study limitations.

Table 3 presents the structure of GRADE's approach to study limitations in RCTs. The second column in Table 3 presents the approach as applied to individual studies; the remaining columns refer to the entire body of evidence. Individual trials achieve a low risk of bias when most or all key criteria are met and any violations are not crucial. Studies that suffer from one crucial violation—a violation of crucial importance with regard to a point estimate (in the

Table 3
Summarizing study limitations for randomized trials

Extent of risk of bias	Risk of bias within a study	Risk of bias across studies	Interpretation across studies[a]	Example of summary across studies
No serious limitations, do not downgrade	Low risk of bias for all key criteria (Table 1)	Most information is from studies at low risk of bias	High-quality evidence: the true effect lies close to that of the estimate of the effect	Beta-blockers reduce mortality in patients with heart failure [26]
Serious limitations, rate down one level (i.e., from high to moderate quality)	Crucial limitation for one criterion or some limitations for multiple criteria sufficient to lower ones confidence in the estimate of effect	Most information is from studies at moderate risk of bias	Quality of evidence reduced from high- to moderate-quality evidence: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different	Amodiaquine and SP together likely reduce treatment failures compared with SP alone in patients with malaria [27]
Very serious limitations rate down two levels (i.e., from high to low quality or moderate to very low)	Crucial limitation for one or more criteria sufficient to substantially lower ones confidence in the estimate of effect	Most information is from studies at high risk of bias	Quality of evidence reduced from high- to low-quality evidence: the true effect may be substantially different from the estimate of the effect	Open discectomy may reduce symptoms after 1 yr compared with conservative treatment of lumbar disc prolapse [28]

Abbreviation: SP, sulfadoxine-pyrimethamine.
[a] This interpretation assumes no problems that necessitate rating down because of imprecision, inconsistency, indirectness, and publication bias.

context of a systematic review) or decision (in the context of a guideline)—provide limited-quality evidence. When one or more crucial limitations substantially lower confidence in a point estimate, a body of evidence provides only very limited support for inferences regarding the magnitude of a treatment effect.

Table 3 illustrates that high-quality evidence is available when most studies from a body of evidence meet bias-minimizing criteria. For example, of the 22 trials addressing the impact of beta-blockers on mortality in patients with heart failure, most, probably or certainly, used concealed allocation, all blinded at least some key groups, and follow up of randomized patients was almost complete [26].

GRADE considers a body of evidence of moderate quality when the best evidence comes from individual studies of moderate quality. For instance, we cannot be confident that, in patients with falciparum malaria, amodiaquine and sulfadoxine-pyrimethamine together reduce treatment failures compared with sulfadoxine-pyrimethamine alone because the apparent advantage of sulfadoxine-pyrimethamine was sensitive to assumptions regarding the event rate in those lost to follow-up in two of three studies [27].

Surgery vs. conservative treatment in the management of patients with lumbar disc prolapse provides an example of rating down two levels because of risk of bias in RCTs [28]. We are uncertain of the benefit of open disectomy in reducing symptoms after 1 year or longer because of very serious limitations in one trial of open disectomy compared with conservative treatment without a large number of early crossovers in both comparison groups. That trial suffered from inadequate concealment of allocation and unblinded assessment of outcome by potentially biased raters (surgeons) using unvalidated rating instruments (Table 4).

11. Existing systematic reviews are often limited in summarizing study limitations across studies

To rate overall quality of evidence with respect to an outcome, review authors and guideline developers must

Table 4
Quality assessment for open discectomy vs. conservative treatment (Gibson and Waddell [28])

Quality assessment						
No of patients (studies)	Design	Limitations	Inconsistency	Indirectness	Imprecision	Publication bias
Outcome: poor/bad result at 1yr—surgeon rated						
126 (1)	RCT	Very serious limitations[a]	Not relevant	No serious indirectness	Serious imprecision[b]	Unlikely
Outcome: poor/bad result at 4yr—surgeon rated						
126 (1)	RCT	Very serious limitations[a]	Not relevant	No serious indirectness	Serious imprecision[b]	Unlikely
Outcome: poor/bad result at 10yr—surgeon rated						
126 (1)	RCT	Very serious limitations[a]	Not relevant	No serious indirectness	Serious imprecision[b]	Unlikely

Abbreviation: RCT, randomized controlled trial.
[a] Inadequate concealment of allocation and unblinded unvalidated assessment by the surgeon.
[b] Wide confidence intervals and few events (16 or fewer).

consider and summarize study limitations considering all the evidence from multiple studies. For a guideline developer, using an existing systematic review would be the most efficient way to address this issue.

Unfortunately, systematic reviews usually do not address all important outcomes, typically focusing on benefit and neglecting harm. For instance, one is required to go to separate reviews to assess the impact of beta-blockers on mortality [26] and on quality of life [29]. No systematic review has addressed beta-blocker toxicity in heart failure patients.

Review authors' usual practice of rating the quality of studies across outcomes, rather than separately for each outcome, further limits the usefulness of existing systematic reviews for guideline developers. This approach becomes even more problematic when review authors use summary measures that aggregate across quality criteria (e.g., allocation concealment, blinding, loss to follow-up) to provide a single score. These measures are often limited in that they focus on quality of reporting rather than on the design and conduct of the study [30]. Furthermore, they tend to be unreliable and less closely correlated with outcome than individual quality components [31–33]. These problems arise, at least in part, because calculating a summary score inevitably involves assigning arbitrary weights to different criteria.

Finally, systematic reviews that address individual components of study limitations are often not comprehensive and fail to make transparent the judgments needed to evaluate study limitations. These judgments are often challenging, at least in part, because of inadequate reporting: just because a safeguard against bias is not reported does not mean it was neglected [34,35].

Thus, although systematic reviews are often extremely useful in identifying the relevant primary studies, members of guideline panels or their delegates must often review individual studies if they wish to ensure accurate ratings of study limitations for all relevant outcomes. As review authors increasingly adopt the GRADE approach (and in particular as Cochrane review authors do so in combination with using the Cochrane risk of bias tool), the situation will improve.

12. What to do when there is only one RCT

Many people are uncomfortable designating a single RCT as high-quality evidence. Given the many instances in which the first positive report has not held up under subsequent investigation, this discomfort is warranted. On the other hand, automatically rating down quality when there is a single study is not appropriate. A single, very large, rigorously planned and conducted multicentre RCT may provide high-quality evidence. GRADE suggests especially careful scrutiny of all relevant issues (risk of bias, precision, directness, and publication bias) when only a single RCT addresses a particular question.

13. Moving from Cochrane risk of bias tables in individual studies to rating quality of evidence across studies

Moving from 6 risk of bias criteria for each individual study to a judgment about rating down for quality of evidence for risk of bias across a group of studies addressing a particular outcome presents challenges. We suggest the following principles.

First, in deciding on the overall quality of evidence, one does not average across studies (for instance if some studies have no serious limitations, some serious limitations, and some very serious limitations, one does not automatically rate quality down by one level because of an average rating of serious limitations). Rather, judicious consideration of the contribution of each study, with a general guide to focus on the high-quality studies (as we will illustrate), is warranted.

Second, this judicious consideration requires evaluating the extent to which each trial contributes toward the estimate of magnitude of effect. This contribution will usually reflect study sample size and number of outcome events—larger trials with many events will contribute more, much larger trials with many more events will contribute much more.

Third, one should be conservative in the judgment of rating down. That is, one should be confident that there is substantial risk of bias across most of the body of available evidence before one rates down for risk of bias.

Fourth, the risk of bias should be considered in the context of other limitations. If, for instance, reviewers find themselves in a close-call situation with respect to two quality issues (risk of bias and, say, precision), we suggest rating down for at least one of the two.

Fifth, notwithstanding the first four principles, reviewers will face close-call situations. They should both acknowledge that they are in such a situation, make it explicit why they think this is the case, and make the reasons for their ultimate judgment apparent.

14. Application of principles

A systematic review of flavonoids to treat pain and bleeding associated with hemorrhoids [36], with respect to the primary outcome of persisting symptoms, most trials did not provide sufficient information to determine whether randomization was concealed, the majority violated the intention-to-treat principle and did not provide the data allowing the appropriate analysis (Table 5), and none used a validated symptom measure. On the other hand, most authors described their trials as double blind, and although concealment and blinding are different concepts, blinded trials of drugs are very likely to be concealed [34] (Table 5). Because the questionnaires appeared simple and transparent, and because of the blinding of the studies, we would be hesitant to consider lack of validation introducing a serious risk of bias.

Table 5
Risk of bias for measurement of symptoms in studies of flavonoids in patients with hemorrhoids

Study[c]	Randomization	Allocation concealment	Blinding	Loss to follow-up[a]/IT principle observed or per protocol analysis	Other
Dimitroulopoulos D, 2005	Adequate[b] Computer-generated random numbers[b]	Sealed opaque envelopes[b]	Described as single blind Care givers, patients, and data collectors blinded[b]	3%/protocol	Unvalidated symptom measure
Misra MC, 2000	Adequate Computer-generated random numbers[b]	Adequate Sealed opaque envelopes[b]	Patients and physicians[b] Described as double blind Placebo identical appearance	2%/protocol	Unvalidated symptom measure
Godeberge P, 1994	Adequate[b]	Adequate Sealed opaque envelopes[b]	Patients, physician-investigator, data manager, statistician, and authors blinded	6%/protocol	
Cospite M, 1994	Unclear	Unclear	Unclear Described as double blind	12%/IT	Unvalidated symptom measure
Chauvenet-M, 1994	Unclear	Unclear	Unclear	11%/protocol	Unvalidated symptom measure
Ho Y-H, 2000	Adequate Drawing of sealed opaque envelopes[b]	Adequate Sealed opaque envelopes	All parties blinded[b]	0%/IT	Unvalidated symptom measure
Thanapongsathorn W, 1992	Unclear	Unclear	Unclear Described as double blind	I2%/protocol	Unvalidated symptom measure
Titapant V, 2001	Unclear	Unclear	Unclear Described as double blind Placebo identical appearance	12%/protocol	Unvalidated symptom measure
Wijayanegara H, 1992	Unclear	Unclear	Unclear Described as double blind	3%/protocol	Unvalidated symptom measure
Annoni F, 1986	Unclear	Unclear	Unclear Described as double blind Placebo identical appearance	Uncertain/unclear	Unvalidated symptom measure
Thorp RH, 1970	Unclear	Unclear	Physicians and patients blinded Described as double blind Placebo identical appearance	20%/protocol	Unvalidated symptom measure
Clyne MB, 1967	Bottles numbered consecutively in accordance to random tables	Unclear	Physicians and patients blinded Described as double blind Placebo identical appearance	Uncertain/protocol	Unvalidated symptom measure
Sinnatamby CS, 1973	Unclear	Unclear	Physicians and patients blinded Described as double blind	53%/protocol	Unvalidated symptom measure
Trochet JP, 1992	Randomized by blocks of three (method unclear)	Unclear	Physicians blinded Placebo identical appearance	Uncertain/IT	Unvalidated symptom measure

Abbreviation: IT, intention-to-treat principle observed.
[a] No important differences in rate of loss to follow-up between flavonoid and control groups in any study.
[b] Data provided by authors.
[c] For full citation of the references cited in this table, see Alonso-Coello et al.[36]

Nevertheless, in light of these study limitations, one might consider focusing on the highest quality trials. Substantial precision would, however, be lost (requiring rating down for imprecision), and the quality of the trials did not explain variability in results (i.e., the magnitude of effect was similar in the methodologically stronger and weaker studies). Both considerations argue for basing an estimate on the results of all RCTs.

In our view, this represents a borderline situation in which it would be reasonable either to rate down for risk of bias or not to do so. This illustrates that the great merit of GRADE is not that it ensures consistency of conclusions but that it requires explicit and transparent judgments. Considering these issues in isolation, and following the principles articulated above, however, we would be inclined not to rate down for quality for risk of bias.

The possibility of discrepant judgments between intelligent and well-informed review authors is more than theoretical. A number of RCTs have evaluated the extent to which graduated pressure stockings can prevent deep venous thrombosis (DVT) in airline passengers taking long flights. Cochrane review authors concluded that the studies provided high-quality evidence for DVT prevention [37]. In contrast, a group of thrombosis experts involved in producing a guideline concluded that because of use of an unreliable method of diagnosing DVT, and lack of blinding, the evidence was of

low quality [38]. Even after direct contact and discussion, each group adhered to its own position—and it remains possible that either group is correct.

Three RCTs addressing the impact of 24-hour administration of high-dose corticosteroids on motor function in patients with acute spinal cord injury illustrate another principle of aggregation [23–25]. Although the degree of limitations is in fact a continuum (as Fig. 1 illustrates), GRADE simplifies the process by categorizing these studies—or any other study—as having "no serious limitations," "serious limitations," or "very serious limitations" (as in Table 3).

The first of the three trials (Bracken in Fig. 1), which included 127 patients treated within 8 hours of injury, ensured allocation concealment through central randomization, almost certainly blinded patients, clinicians, and those measuring motor function, and lost 5% of patients to follow-up at 1 year [23]. The flaws in this RCT are sufficiently minor to allow classification as "no serious limitations."

The second trial (Pointillart et al. [25] in Fig. 1) was unlikely to have concealed allocation, did blind those assessing outcome (but not patients or clinicians), and lost only one of 106 patients to follow-up. Here, quality falls in an intermediate range, and classification as either "no serious limitations" or "serious limitations" may be appropriate. The third trial (Otani et al. [24] in Fig. 1), which included 158 patients, almost certainly failed to conceal allocation, used no blinding, and lost 26% of patients to follow-up, many more in the steroid group than the control group. This third trial is probably best classified as having "very serious limitations."

Considering these three RCTs, should one rate down for design and implementation with respect to the motor function outcome? If we considered only the first two trials, the answer would be no. Therefore, the review authors must decide either to exclude the third trial (thereby only including trials with few limitations) or include it based on a judgment that overall there is a low risk of bias (because most of the evidence comes from trials with few limitations) despite the contribution of the trial with very serious limitations to the overall estimate of effect. This example illustrates that averaging across studies will not be the right approach.

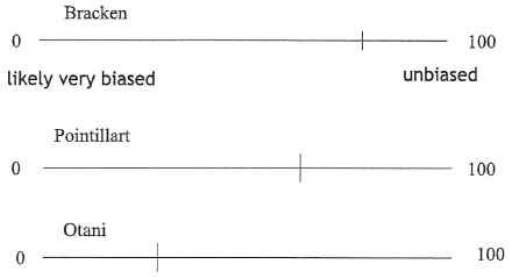


Fig. 1. Validity of three randomized controlled trials addressing the effect of steroids on motor function in acute spinal cord injury.

15. Recording judgments about study limitations

One great merit of GRADE is its lucid categorization of factors that decrease quality of evidence and the resultant transparency of judgments. This transparency, however, requires careful documentation of judgments. Including a risk of bias table that summarizes key criteria used to assess study limitations for each outcome for each study helps ensure transparency.

Table 5 presents an example of such a table. Note that the table focuses on only one outcome, symptoms. Each study will need only one line on such a table if, as in this case, there is only one important outcome or if each quality criterion is the same for every important outcome. Each outcome for which quality criteria differ in important ways will need a separate line. Outcomes may, for instance, differ for blinding (e.g., in surgical trials patients completing questionnaires measuring health-related quality of life may be unblinded, but adjudicators of cause-specific mortality may be blinded) or loss to follow-up (e.g., greater loss to follow-up for quality of life than for all-cause mortality).

Review authors and guideline developers can then summarize their assessments across studies in a "quality assessment" table to fully ensure the transparency of their judgments (Table 4). A footnote provides the reasoning behind the decision to rate down the quality of the evidence from high to low quality on the basis of study limitations (alternatively, one can very briefly summarize the key information in a cell in the table). In this example, there was an additional concern about imprecision, which further decreases the quality of evidence from low to very low. We will describe guidelines for making judgments about imprecision (the risk of random error), in the sixth article in this series.

References

[1] The users' guides to the medical literature: a manual for evidence-based clinical practice. In: Guyatt G, Rennie D, Meade M, Cook D, editors. 2nd ed. New York, NY: McGraw-Hill; 2008.
[2] Higgins JP, Altman D. Assessing the risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions 5.0.1. Chichester, UK: John Wiley & Sons; 2008.
[3] Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. N Engl J Med 2009;361:1963–71.
[4] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA 2009;302:977–84.
[5] Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601–5.
[6] Pocock SJ. When (not) to stop a clinical trial for benefit. JAMA 2005;294:2228–30.
[7] Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. Control Clin Trials 1989;10(4 Suppl): 209S–21S.

[8] Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. JAMA 2005;294:2203–9.

[9] Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA 2010;303:1180–7.

[10] Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. JAMA 2007;297:468–70.

[11] Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. BMJ 2005;330:753.

[12] Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004;291:2457–65.

[13] Chan AW, Krleza-Jerić K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ 2004;171:735–40.

[14] Bolona ER, Uraga MV, Haddad RM, Tracz MJ, Sideras K, Kennedy CC, et al. Testosterone use in men with sexual dysfunction: a systematic review and meta-analysis of randomized placebo-controlled trials. Mayo Clin Proc 2007;82:20–8.

[15] Sinha M, Montori VM. Reporting bias and other biases affecting systematic reviews and meta-analyses: a methodological commentary. Expert Rev Pharmacoecon Outcomes Res 2006;6(1):603–11.

[16] Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:iii–x. 1–173.

[17] West S, King V, Carey TS, Lohr KN, Mckoy N, Sutton SF, et al. Systems to rate the strength of scientific evidence [evidence report/technology assessment no 47]. AHRQ Publication No 02–E016. Rockville, MD: Agency for Healthcare Research and Quality; 2002.

[18] Proposed Evaluation Tools for COMPUS: assessment, November 29, 2005. Ottawa, Canada: Canadian Coordinating Office for Health Technology; 2005.

[19] Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol 2007;36:666–76.

[20] Greer IA, Nelson-Piercy C. Low-molecular-weight heparins for thromboprophylaxis and treatment of venous thromboembolism in pregnancy: a systematic review of safety and efficacy. Blood 2005;106:401–7.

[21] Sanson BJ, Lensing AW, Prins MH, Ginsberg JS, Barkagan ZS, Lavenne-Pardonge E, et al. Safety of low-molecular-weight heparin in pregnancy: a systematic review. Thromb Haemost 1999;81:668–72.

[22] Bowker SL, Majumdar SR, Veugelers P, Johnson JA. Increased cancer-related mortality for patients with type 2 diabetes who use sulfonylureas or insulin. Diabetes Care 2006;29:254–8.

[23] Bracken MB, Shepard MJ, Collins WF Jr, Holford TR, Baskin DS, Eisenberg HM, et al. Methylprednisolone or naloxone treatment after acute spinal cord injury: 1-year follow-up data. Results of the second National Acute Spinal Cord Injury Study. J Neurosurg 1992;76:23–31.

[24] Otani K, Abe H, Kadoya S. Beneficial effect of methylprednisolone sodium succinate in the treatment of acute spinal cord injury. Sekitsui Sekizui 1994;7:633–47.

[25] Pointillart V, Petitjean ME, Wiart L, Vital JM, Lassié P, Thicoipé M, et al. Pharmacological therapy of spinal cord injury during the acute phase. Spinal Cord 2000;38:71–6.

[26] Brophy JM, Joseph L, Rouleau JL. Beta-blockers in congestive heart failure. A Bayesian meta-analysis. Ann Intern Med 2001;134:550–60.

[27] McIntosh H, Jones K. Chloroquine or amodiaquine combined with sulfadoxine-pyrimethamine for treating uncomplicated malaria. Cochrane Database Syst Rev 2005;(4). CD000386.

[28] Gibson J, Waddell G. Surgical interventions for lumbar disc prolapse. Cochrane Database Syst Rev 2007;(2)10.1002/14651858. CD001350.

[29] Dobre D, van Jaarsveld CH, deJongste MJ, Haaijer Ruskamp FM, Ranchor AV. The effect of beta-blocker therapy on quality of life in heart failure patients: a systematic review and meta-analysis. Pharmacoepidemiol Drug Saf 2007;16:152–9.

[30] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995;16:62–73.

[31] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–12.

[32] Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. Control Clin Trials 1990;11:339–52.

[33] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054–60.

[34] Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schünemann HJ, et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol 2004;57:1232–6.

[35] Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, Djulbegovic B, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. BMJ 2004;328:22–4.

[36] Alonso-Coello P, Zhou Q, Martinez-Zapata MJ, Mills E, Heels-Ansdell D, Johanson JF, et al. Meta-analysis of flavonoids for the treatment of haemorrhoids. Br J Surg 2006;93(8):909–20.

[37] Clarke M, Hopewell S, Juszczak E, Eisinga A, Kjeldstrøm M. Compression stockings for preventing deep vein thrombosis in airline passengers. Cochrane Database Syst Rev 2007;(3). CD004002.

[38] Geerts WH, Bergqvist D, Pineo GF, Heit JA, Samama CM, Lassen MR, et al. Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest 2008;133(6 Suppl):381S–453S.