

SPECIAL COMMUNICATION

Strength of Recommendation Taxonomy (SORT): A Patient-Centered Approach to Grading Evidence in the Medical Literature

Mark H. Ebell, MD, MS, Jay Siwek, MD, Barry D. Weiss, MD,
Steven H. Woolf, MD, MPH, Jeffrey Susman, MD, Bernard Ewigman, MD, MPH, and
Marjorie Bowman, MD, MPA

A large number of taxonomies are used to rate the quality of an individual study and the strength of a recommendation based on a body of evidence. We have developed a new grading scale that will be used by several family medicine and primary care journals (required or optional), with the goal of allowing readers to learn one taxonomy that will apply to many sources of evidence. Our scale is called the Strength of Recommendation Taxonomy. It addresses the quality, quantity, and consistency of evidence and allows authors to rate individual studies or bodies of evidence. The taxonomy is built around the information mastery framework, which emphasizes the use of patient-oriented outcomes that measure changes in morbidity or mortality. An A-level recommendation is based on consistent and good quality patient-oriented evidence; a B-level recommendation is based on inconsistent or limited quality patient-oriented evidence; and a C-level recommendation is based on consensus, usual practice, opinion, disease-oriented evidence, or case series for studies of diagnosis, treatment, prevention, or screening. Levels of evidence from 1 to 3 for individual studies also are defined. We hope that consistent use of this taxonomy will improve the ability of authors and readers to communicate about the translation of research into practice. (J Am Board Fam Pract 2004;17:59–67.)

Review articles (or overviews) are highly valued by physicians as a way to keep up to date with the medical literature. Sometimes, though, these articles are based more on the authors' personal experience, or anecdotes, or incomplete surveys of the literature than on a comprehensive collection of the best available evidence. As a result, there is an ongoing effort in the medical publishing field to

improve the quality of review articles through the use of more explicit grading of the strength of evidence on which recommendations are based.^{1–4}

Several journals, including *American Family Physician* and *Journal of Family Practice*, have adopted evidence-grading scales that are used in some of the articles published in those journals. Other organizations and publications have also developed evidence-grading scales. The diversity of these scales can be confusing for readers. More than 100 grading scales are in use by various medical publications.⁵ A level B recommendation in one journal may not mean the same thing as a level B recommendation in another. Even within journals, different evidence-grading scales sometimes are used in different articles within the same issue of a journal. Journal readers do not have the time, energy, or interest to interpret multiple grading scales, and more complex scales are difficult to integrate into daily practice.

Therefore the editors of the US family medicine and primary care journals (ie, *American Family Phy-*

Submitted, revised 20 November 2003.

From the Michigan State University College of Human Medicine, East Lansing (MHE), Georgetown University Medical Center, Washington, DC (JS), University of Arizona College of Medicine, Tucson (BDW), Virginia Commonwealth University School of Medicine, Richmond (SHW), University of Cincinnati College of Medicine, Cincinnati, Ohio (JS), University of Chicago, Pritzker School of Medicine, Chicago, Illinois (BE), and University of Pennsylvania Health System, Philadelphia (MAB). Address correspondence to Mark Ebell, MD, MS, 300 Snapfinger Dr., Athens, GA 30605 (e-mail: ebell@msu.edu).

Simultaneously published in print and online by *American Family Physician*, *Journal of Family Practice*, *Journal of the American Board of Family Practice*, and online by *Family Practice Inquiries Network*. Copyright © 2004 American Family Physician, a publication of the American Academy of Family Physicians. All rights reserved.

sician, *Family Medicine*, *Journal of Family Practice*, *Journal of the American Board of Family Practice*, and *BMJ-USA*) and the Family Practice Inquiries Network (FPIN) came together to develop a unified taxonomy for the strength of recommendations based on a body of evidence. The new taxonomy should include the following attributes: (1) be uniform in most family medicine journals and electronic databases; (2) allow authors to evaluate the strength of recommendation of a body of evidence; (3) allow authors to rate the level of evidence for an individual study; (4) be comprehensive and allow authors to evaluate studies of screening, diagnosis, therapy, prevention, and prognosis; (5) be easy to use and not too time-consuming for authors, reviewers, and editors who may be content experts but not experts in critical appraisal or clinical epidemiology; and (6) be straightforward enough that primary care physicians can readily integrate the recommendations into daily practice.

Definitions

A number of relevant terms must be defined for clarification.

Disease-Oriented Outcomes

These outcomes include intermediate, histopathologic, physiologic, or surrogate results (ie, blood sugar, blood pressure, flow rate, coronary plaque thickness) that may or may not reflect improvements in patient outcomes.

Patient-Oriented Outcomes

These are outcomes that matter to patients and help them live longer or better lives, including reduced morbidity, reduced mortality, symptom improvement, improved quality of life, or lower cost.

Level of Evidence

The validity of an individual study is based on an assessment of its study design. According to some methodologies,⁶ levels of evidence can refer not only to individual studies but also to the quality of evidence from multiple studies about a specific question or the quality of evidence supporting a clinical intervention. For purposes of maintaining simplicity and consistency in this proposal, we use the term level of evidence to refer to individual studies.

Strength of Recommendation

The strength (or grade) of a recommendation for clinical practice is based on a body of evidence (typically more than one study). This approach takes into account the level of evidence of individual studies, the type of outcomes measured by these studies (patient-oriented or disease-oriented), the number, consistency, and coherence of the evidence as a whole, and the relationship between benefits, harms, and costs.

Practice Guideline (Evidence-Based)

These guidelines are recommendations for practice that involve a comprehensive search of the literature, an evaluation of the quality of individual studies, and recommendations that are graded to reflect the quality of the supporting evidence. All search, critical appraisal, and grading methods should be described explicitly and be replicable by similarly skilled authors.

Practice Guideline (Consensus)

Consensus guidelines are recommendations for practice based on expert opinions that typically do not include a systematic search, an assessment of the quality of individual studies, or a system to label the strength of recommendations explicitly.

Research Evidence

This evidence is presented in publications of original research, involving collection of original data or the systematic review of other original research publications. It does not include editorials, opinion pieces, or review articles (other than systematic reviews or meta-analyses).

Review Article

A nonsystematic overview of a topic is a review article. In most cases, it is not based on an exhaustive, structured review of the literature and does not evaluate the quality of included studies systematically.

Systematic Reviews and Meta-Analyses

A systematic review is a critical assessment of existing evidence that addresses a focused clinical question, includes a comprehensive literature search, appraises the quality of studies, and reports results in a systematic manner. If the studies report comparable quantitative data and have a low degree

of variation in their findings, a meta-analysis can be performed to derive a summary estimate of effect.

Existing Strength-of-Evidence Scales

In March 2002, the Agency for Healthcare Research and Quality (AHRQ) published a report that summarized the state of the art in methods of rating the strength of evidence.⁵ The report identified a large number of systems for rating the quality of individual studies: 20 for systematic reviews, 49 for randomized controlled trials, 19 for observational studies, and 18 for diagnostic test studies. It also identified 40 scales that graded the strength of a body of evidence consisting of one or more studies.

The authors of the AHRQ report proposed that any system for grading the strength of evidence should consider 3 key elements: quality, quantity, and consistency. Quality is the extent to which the identified studies minimize the opportunity for bias and is synonymous with the concept of validity. Quantity is the number of studies and subjects included in those studies. Consistency is the extent to which findings are similar between different studies on the same topic. Only 7 of the 40 systems identified and addressed all 3 of these key elements.^{6–11}

Strength of Recommendation Taxonomy (SORT)

The authors of this article represent the major family medicine journals in the United States and a large family practice academic consortium. Our process began with a series of electronic mail exchanges, was developed during a meeting of the editors, and continued through another series of electronic mail exchanges.

We decided that our taxonomy for rating the strength of a recommendation should address the 3 key elements identified in the AHRQ report: quality, quantity, and consistency of evidence. We also were committed to creating a grading scale that could be applied by authors with varying degrees of expertise in evidence-based medicine and clinical epidemiology and interpreted by physicians with little or no formal training in these areas. We believed that the taxonomy should address the issue of patient-oriented evidence versus disease-oriented evidence explicitly and be consistent with the information mastery framework proposed by Slawson and Shaughnessy.²

After considering these criteria and reviewing the existing taxonomies for grading the strength of a recommendation, we decided that a new taxonomy was needed to reflect the needs of our specialty. Existing grading scales were focused on a particular kind of study (ie, prevention or treatment), were too complex, or did not take into account the type of outcome.

Our proposed taxonomy is called the Strength of Recommendations Taxonomy (SORT). It is shown in Figure 1. The taxonomy includes ratings of A, B, or C for the strength of recommendation for a body of evidence. The table in the center of Figure 1 explains whether a body of evidence represents good or limited-quality evidence and whether evidence is consistent or inconsistent. The quality of individual studies is rated 1, 2, or 3; numbers are used to distinguish ratings of individual studies from the letters A, B, and C used to evaluate the strength of a recommendation based on a body of evidence. Figure 2 provides information about how to determine the strength of recommendation for management recommendations, and Figure 3 explains how to determine the level of evidence for an individual study. These 2 algorithms should be helpful to authors preparing manuscripts for submission to family medicine journals. The algorithms are to be considered general guidelines, and special circumstances may dictate assignment of a different strength of recommendation (eg, a single, large, well-designed study in a diverse population may warrant an A-level recommendation).

Recommendations based only on improvements in surrogate or disease-oriented outcomes are always categorized as level C, because improvements in disease-oriented outcomes are not always associated with improvements in patient-oriented outcomes, as exemplified by several well-known findings from the medical literature. For example, doxazosin lowers blood pressure in black patients—a seemingly beneficial outcome—but it also increases mortality rates.¹² Similarly, encainide and flecainide reduce the incidence of arrhythmias after acute myocardial infarction, but they also increase mortality rates.¹³ Finasteride improves urinary flow rates, but it does not significantly improve urinary tract symptoms in patients with benign prostatic hypertrophy,¹⁴ whereas arthroscopic surgery for osteoarthritis of the knee improves the appearance of cartilage but does not reduce pain or improve joint function.¹⁵ Additional examples of clinical sit-

In general, only key recommendations for readers require a grade of the “Strength of Recommendation”. Recommendations should be based on the highest quality evidence available. For example, Vitamin E was found in some cohort studies (Level 2 study quality) to have a benefit for cardiovascular protection, but good-quality randomized trials (Level 1) have not confirmed this effect. It is therefore preferable to base clinical recommendations in a manuscript on the level 1 studies.

Strength of Recommendation	Definition
A	Recommendation based on consistent and good quality patient-oriented evidence *
B	Recommendation based on inconsistent or limited quality patient-oriented evidence *
C	Recommendation based on consensus, usual practice, opinion, disease-oriented evidence, * and case series for studies of diagnosis, treatment, prevention, or screening.

Use the table below to determine whether a study measuring patient-oriented outcomes is of good or limited quality, and whether the results are consistent or inconsistent between studies:

Study Quality	Type of study		
	Diagnosis	Treatment / Prevention/ Screening	Prognosis
Level 1 Good quality patient-oriented evidence	<ul style="list-style-type: none"> Validated clinical decision rule Systematic Review (SR)/meta-analysis of high quality studies High quality diagnostic cohort study ** 	<ul style="list-style-type: none"> SR/meta-analysis of RCTs with consistent findings High quality individual randomized controlled trial (RCT) + All or none study ++ 	<ul style="list-style-type: none"> SR/meta-analysis of good quality cohort studies Prospective cohort study with good follow-up
Level 2 Limited quality patient-oriented evidence	<ul style="list-style-type: none"> Unvalidated clinical decision rule SR/meta-analysis of lower quality studies or studies with inconsistent findings Lower quality diagnostic cohort study or diagnostic case-control study ** 	<ul style="list-style-type: none"> SR/meta-analysis of lower quality clinical trials or of studies with inconsistent findings Lower quality clinical trial + Cohort study Case-control study 	<ul style="list-style-type: none"> SR/meta-analysis of lower quality cohort studies or with inconsistent results Retrospective cohort study or prospective cohort study with poor follow-up Case-control study Case series
Level 3 Other evidence	Consensus guidelines, extrapolations from bench research, usual practice, opinion, disease-oriented evidence (intermediate or physiologic outcomes only), and case series for studies of diagnosis, treatment, prevention, or screening.		

Consistency	Consistency Across Studies
	Consistent
Inconsistent	<ul style="list-style-type: none"> Considerable variation among study findings and lack of coherence or If high quality and up-to-date systematic reviews or meta-analyses exist, they do not find consistent evidence in favor of the recommendation

* Patient-oriented evidence measures outcomes that matter to patients: morbidity, mortality, symptom improvement, cost reduction, quality of life. Disease-oriented evidence measures intermediate, physiologic, or surrogate endpoints that may or may not reflect improvements in patient outcomes (i.e. blood pressure, blood chemistry, physiological function, and pathological findings)

** High quality diagnostic cohort study: cohort design, adequate size, adequate spectrum of patients, blinding, and a consistent, well-defined reference standard

+ High quality RCT: allocation concealed, blinding if possible, intention-to-treat analysis, adequate statistical power, adequate follow-up (>80%).

++ An all-or-none study is one where the treatment causes a dramatic change in outcomes, such as antibiotics for meningitis or surgery for appendicitis, which precludes study in a controlled trial.

Figure 1. The Strength of Recommendation Taxonomy (SORT). SR, systematic review; RCT, randomized controlled trial.

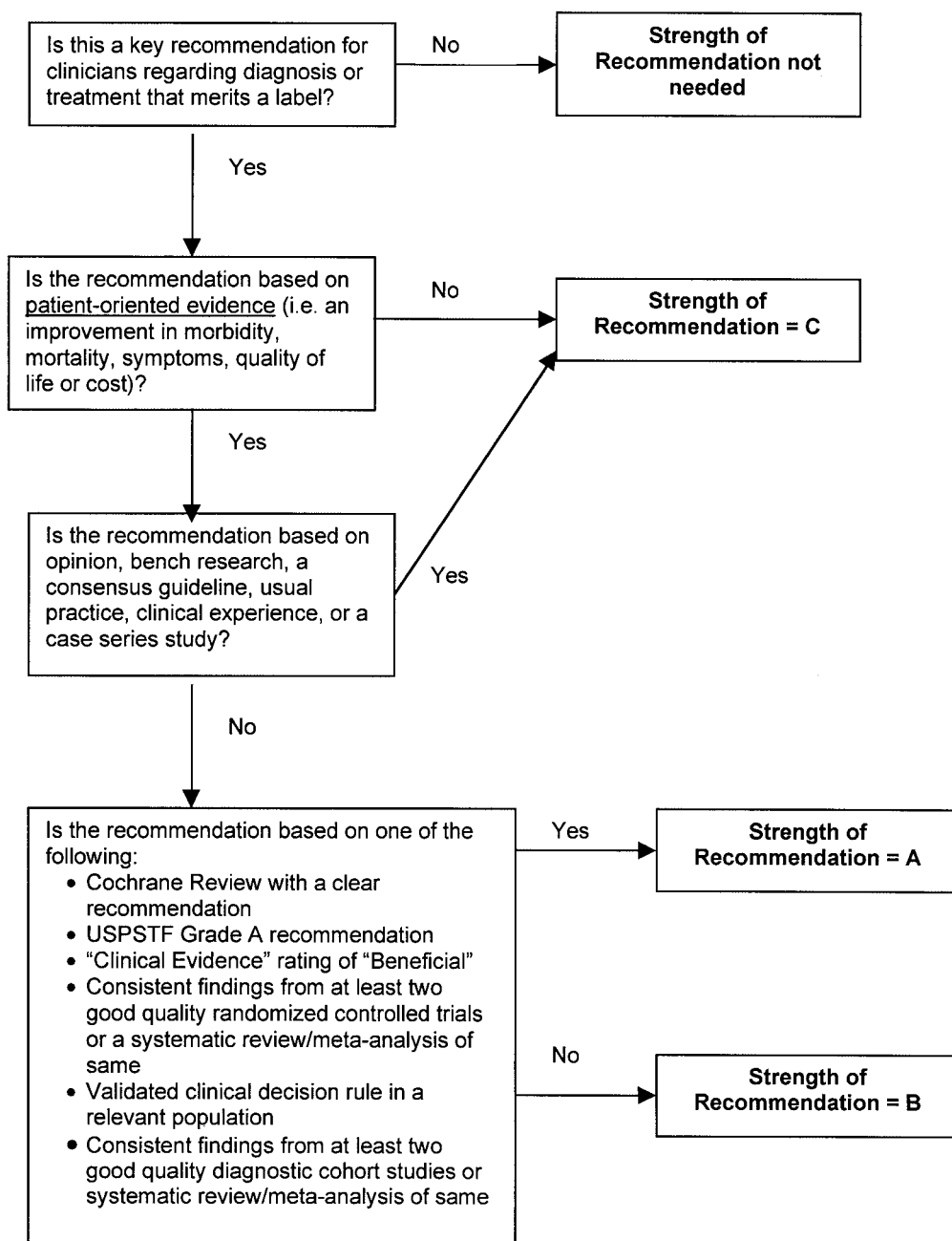


Figure 2. Algorithm for determining the strength of a recommendation based on a body of evidence (applies to clinical recommendations regarding diagnosis, treatment, prevention, or screening). Although this algorithm provides a general guideline, authors and editors may adjust the strength of recommendation based on the benefits, harms, and costs of the intervention being recommended. USPSTF, US Preventive Services Task Force.

uations where disease-oriented evidence disagrees with patient-oriented evidence are shown in Table 1.^{12–24} Examples of how to apply the taxonomy are given in Table 2.

We believe there are several advantages to our proposed taxonomy. It is straightforward and comprehensive, is easily applied by authors and physi-

cians, and explicitly addresses the issue of patient-oriented versus disease-oriented evidence. The latter attribute distinguishes SORT from most other evidence grading scales. These strengths also create some limitations. Some clinicians may be concerned that the taxonomy is not as detailed in its assessment of study designs as others, such as that

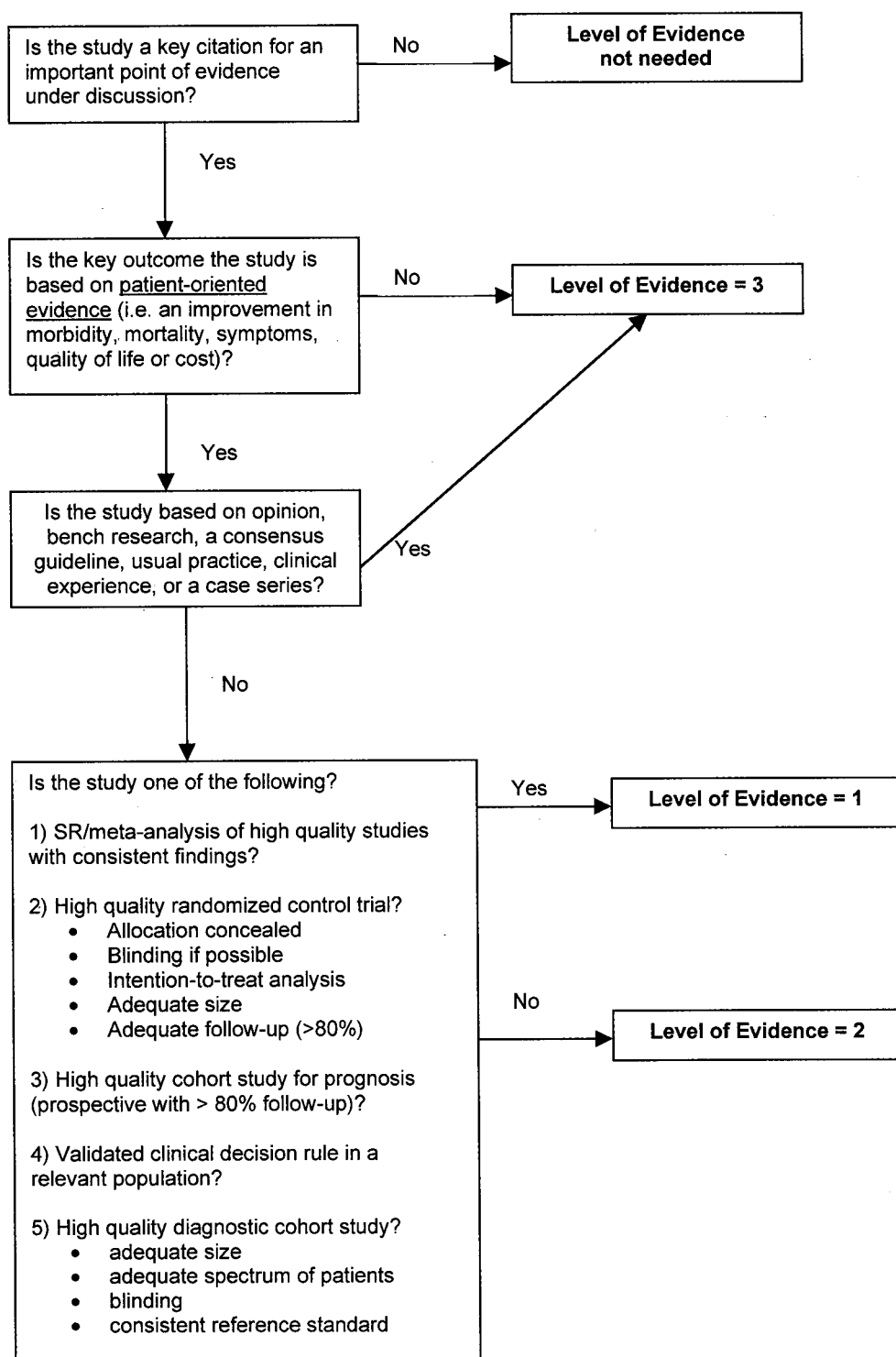


Figure 3. Algorithm for determining the level of evidence for an individual study.

of the Centre for Evidence-Based Medicine (CEBM).²⁵ However, the primary difference between the 2 taxonomies is that the CEBM version distinguishes between good and poor observational studies whereas the SORT version does not. We

concluded that the advantages of a system that provides the physician with a clear recommendation that is strong (A), moderate (B), or weak (C) in its support of a particular intervention outweighs the theoretic benefit of distinguishing between

Table 1. Examples of Inconsistency between Disease-Oriented and Patient-Oriented Outcomes

Disease Or Condition	Disease-Oriented Outcome	Patient-Oriented Outcome
Doxazosin for blood pressure ¹²	Reduces blood pressure	Increases mortality in blacks
Lidocaine for arrhythmia after acute myocardial infarction ¹³	Suppresses arrhythmias	Increases mortality
Finasteride for benign prostatic hypertrophy ¹⁴	Improved urinary flow rate	No clinically important change in symptom scores
Sleeping infants on their stomach or side ¹⁶	Knowledge of anatomy and physiology suggests that this will decrease the risk of aspiration	Increased risk of sudden infant death syndrome
Vitamin E for heart disease ¹⁷	Reduces levels of free radicals	No change in mortality
Histamine antagonists and proton-pump inhibitors for nonulcer dyspepsia ¹⁸	Significantly reduce gastric pH levels	Little or no improvement in symptoms in patients with nongastroesophageal reflux disease, nonulcer dyspepsia
Arthroscopic surgery for osteoarthritis of the knee ¹⁵	Improved appearance of cartilage after debridement	No change in function or symptoms at 1 year
Hormone therapy ¹⁹	Reduced low-density lipoprotein cholesterol, increased high-density lipoprotein cholesterol	No decrease in cardiovascular or all-cause mortality and an increase in cardiovascular events in women older than 60 years (Women's Health Initiative) with combined hormone therapy
Insulin therapy in type 2 diabetes mellitus ²⁰	Keeps blood sugar below 120 mg/dL (6.7 mmol/L)	Does not reduce overall mortality
Sodium fluoride for fracture prevention ²¹	Increases bone density	Does not reduce fracture rate
Lidocaine prophylaxis after acute myocardial infarction ²²	Suppresses arrhythmias	Increases mortality
Clofibrate for hyperlipidemia ²³	Reduces lipids	Does not reduce mortality
β-blockers for heart failure ²⁴	Reduce cardiac output	Reduce mortality in moderate to severe disease

lower quality and higher quality observational studies, particularly because there is no objective evidence that the latter distinction carries important differences in clinical recommendations.

Any publication applying SORT (or any other evidence-based taxonomy) should describe carefully the search process that preceded the assignment of a SORT rating. For example, authors could perform a comprehensive search of MEDLINE and the gray literature, a comprehensive

search of MEDLINE alone, or a more focused search of MEDLINE plus secondary evidence-based sources of information.

Walkovers: Creating Linkages with SORT

Some organizations, such as the CEBM,²⁵ the Cochrane Collaboration,⁷ and the US Preventive Services Task Force (USPSTF),⁶ have developed their own grading scales for the strength of recom-

Table 2. Examples of How to Apply the SORT in Practice

Example 1: Although a number of observational studies (level of evidence—2) suggested a cardiovascular benefit from vitamin E, a large, well-designed, randomized trial with a diverse patient population (level of evidence—1) showed the opposite. The strength of recommendation against routine, long-term use of vitamin E to prevent heart disease, based on the best available evidence, should be A.
Example 2: A Cochrane review finds 7 clinical trials that are consistent in their support of a mechanical intervention for low back pain, but the trials were poorly designed (ie, unblinded, nonrandomized, or with allocation to groups un concealed). In this case, the strength of recommendation in favor of these mechanical interventions is B (consistent but lower quality clinical trials).
Example 3: A meta-analysis finds 9 high-quality clinical trials of the use of a new drug in the treatment of pulmonary fibrosis. Two of the studies find harm, 2 find no benefit, and 5 show some benefit. The strength of recommendation in favor of this drug would be B (inconsistent results of good-quality, randomized controlled trials).
Example 4: A new drug increases the forced expiratory volume in 1 second (FEV ₁) and peak flow rate in patients with an acute asthma exacerbation. Data on symptom improvement is lacking. The strength of recommendation in favor of using this drug is C (disease-oriented evidence only).

Table 3. Suggested Walkovers between Taxonomies for Assessing the Strength of a Recommendation Based on a Body of Evidence

SORT	CEBM	BMJ's Clinical Evidence
A. Recommendation based on consistent and good quality patient-oriented evidence	A. Consistent level 1 studies	Beneficial
B. Recommendation based on inconsistent or limited-quality patient-oriented evidence	B. Consistent level 2 or 3 studies or extrapolations from level 1 studies	Likely to be beneficial Likely to be ineffective or harmful (recommendation against)
	C. Level 4 studies or extrapolations from level 2 or 3 studies	Unlikely to be beneficial (recommendation against)
C. Recommendation based on consensus, usual practice, disease-oriented evidence, case series for studies of treatment or screening, and/or opinion	D. Level 5 evidence or troublingly inconsistent or inconclusive studies of any level	Unknown effectiveness

SORT, Strength Of Evidence Taxonomy; CEBM, Centre for Evidence-Based Medicine; BMJ, BMJ Publishing Group.

Table 4. Suggested Walkover between the CEBM and the SORT Taxonomies for Assessing the Level of Evidence of an Individual Study

SORT	CEBM	
	Treatment/Screening	Other Categories
Level 1	Levels 1a to 1c	Levels 1a to 1c
Level 2	Level 2 or 3	Levels 2 to 4
Level 3	Level 4 or 5 and any study that measures intermediate or surrogate outcomes	Level 5 and any study that measures intermediate or surrogate outcomes

CEBM, Centre for Evidence-Based Medicine; SORT, Strength of Recommendation Taxonomy.

mentations based on a body of evidence and are unlikely to abandon them. Other organizations, such as the FPIN,²⁶ publish their work in a variety of settings and must be able to move between taxonomies. We have developed a set of optional walkovers that suggest how authors, editors, and readers might move from one taxonomy to another. Walkovers for the CEBM and USPSTF taxonomies are shown in Table 3.

Many authors and experts in evidence-based medicine use the "Level of Evidence" taxonomy from the CEBM to rate the quality of individual studies.²⁵ A walkover from the 5-level CEBM scale to the simpler 3-level SORT scale for individual studies is shown in Table 4.

Final Comment

The SORT is a comprehensive taxonomy for evaluating the strength of a recommendation based on a body of evidence and the quality of an individual study. If applied consistently by authors and editors in the family medicine literature, it has the potential to make it easier for physicians to apply

the results of research in their practice through the information mastery approach and to incorporate evidence-based medicine into their patient care.

Like any such grading scale, it is a work in progress. As we learn more about biases in study design, and as the authors and readers who use the taxonomy become more sophisticated about principles of information mastery, evidence-based medicine, and critical appraisal, it is likely to evolve. We remain open to suggestions from the primary care community for refining and improving SORT.

We thank Lee Green, MD, MPH, John Epling, MD, Kurt Stange, MD, PhD, and Margaret Gourlay, MD, for helpful comments on the manuscript.

References

1. Anonymous. Evidence-based medicine. A new approach to teaching the practice of medicine. Evidence-Based Medicine Working Group. *JAMA* 1992;268:2420–5.
2. Slawson DC, Shaughnessy AF, Bennett JH. Becoming a medical information master: feeling good about

- not knowing everything. *J Fam Pract* 1994;38:505–13.
3. Shaughnessy AF, Slawson DC, Bennett JH. Becoming an information master: a guidebook to the medical information jungle. *J Fam Pract* 1994;39:489–99.
 4. Siwek J, Gourlay ML, Slawson DC, Shaughnessy AF. How to write an evidence-based clinical review article. *Am Fam Physician* 2002;65:251–8.
 5. Systems to rate the strength of scientific evidence. Evidence report/technology assessment: number 47. AHRQ publication no. 02-E015. Rockville (MD): Agency for Healthcare Research and Quality; 2002. Available at: URL: .
 6. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001;20(3 Suppl):21–35.
 7. Clarke M, Oxman AD. Cochrane reviewer's handbook 4.0. The Cochrane Collaboration; 2003. Available at: URL: <http://www.cochrane.org/resources/handbook/>.
 8. Gyorkos TW, Tannenbaum TN, Abrahamowicz M, et al. An approach to the development of practice guidelines for community health interventions. *Can J Public Health* 1994;85(Suppl 1):S8–13.
 9. Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based guide to community preventive services—methods. *Am J Prev Med* 2000;18(1 Suppl):35–43.
 10. Greer N, Mosser G, Logan G, Halaas GW. A practical approach to evidence grading. *Jt Comm J Qual Improv* 2000;26:700–12.
 11. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. *JAMA* 2000;284:1290–6.
 12. Anonymous. Major cardiovascular events in hypertensive patients randomized to doxazosin vs chlorthalidone: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALL-HAT). *JAMA* 2000;283:1967–75.
 13. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardia Arrhythmia Suppression Trial. *N Engl J Med* 1991;324:781–8.
 14. Lepor H, Williford WO, Barry MJ, et al. The efficacy of terazosin, finasteride, or both in benign prostatic hyperplasia. Veterans Affairs Cooperative Studies Benign Prostatic Hyperplasia Study Group. *N Engl J Med* 1996;335:533–9.
 15. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;347:81–8.
 16. Dwyer T, Ponsonby AL. Sudden infant death syndrome: after the “back to sleep” campaign. *BMJ* 1996;313:180–1.
 17. Yusuf S, Dagenais G, Pogue J, Bosch J, Sleight P. Vitamin E supplementation and cardiovascular events in high-risk patients. *N Engl J Med* 2000;342:154–60.
 18. Moayyedi P, Soo S, Deeks J, Delaney B, Innes M, Forman D. Pharmacological interventions for non-ulcer dyspepsia. *Cochrane Database Syst Rev* 2003; (1):CD001960.
 19. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321–33.
 20. Anonymous. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group [published erratum appears in *Lancet* 1999;354:602]. *Lancet* 1998;352:837–53.
 21. Meunier PJ, Sebert JL, Reginster JY, et al. Fluoride salts are no better at preventing new vertebral fractures than calcium-vitamin D in postmenopausal osteoporosis: the FAVOStudy. *Osteoporos Int* 1998;8:4–12.
 22. MacMahon S, Collins R, Peto R, Koster RW, Yusuf S. Effects of prophylactic lidocaine in suspected acute myocardial infarction. An overview of results from the randomized, controlled trials. *JAMA* 1988; 260:1910–6.
 23. Grumbach K. How effective is drug treatment of hypercholesterolemia? A guided tour of the major clinical trials for the primary care physician. *J Am Board Fam Pract* 1991;4:437–45.
 24. Heidenreich PA, Lee TT, Massie BM. Effect of beta-blockade on mortality in patients with heart failure: a meta-analysis of randomized clinical trials. *J Am Coll Cardiol* 1997;30:27–34.
 25. Centre for Evidence-Based Medicine. Levels of evidence and grades of recommendation. Available at: URL: http://www.cebm.net/levels_of_evidence.asp.
 26. Family Practice Inquiries Network (FPIN). Available at: URL: <http://www.fpin.org>.