JAMA
The Journal of the American Medical Association

Volume 274(20), 22/29 November 1995, pp 1630-1632

Users' Guides to the Medical Literature: VIII. How to Use Clinical Practice Guidelines: B. What Are the Recommendations and Will They Help You in Caring for Your Patients?

[The Medical Literature]

Wilson Mark C., MD MPH; Hayward, Robert S. A. MD, MPH; Tunis, Sean R. MD, MSc; Bass, Eric B. MD, MPH; Guyatt, Gordon MD, MSc; for the Evidence-Based Medicine Working Group

From the Division of Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, Md (Dr Bass); the Departments of Medicine (Drs Hayward and Guyatt) and Clinical Epidemiology and Biostatistics (Drs Hayward and Guyatt), McMaster University, Hamilton, Ontario; Health Program, Office of Technology Assessment, US Congress, Washington, DC (Dr Tunis); and the Department of Medicine, Bowman Gray School of Medicine of Wake Forest University, Winston-Salem, NC (Dr Wilson).

CLINICAL SCENARIO

At the conclusion of our first article on practice guidelines [1] in this series, we left you examining the full text of a practice guideline [2] that could help you marshal a convincing response to a colleague who disagrees with your approach to hormone replacement therapy (HRT) in postmenopausal women. Later that day, chatting with another colleague, you mention the disagreement. He shrugs, and avows, "It's entirely a matter of personal preference, the evidence doesn't support either of you." You return to the guideline, looking for how particular recommendations may be justified and adapted to your patient's circumstances.

WHAT ARE THE RECOMMENDATIONS?

Are Practical, Clinically Important, Recommendations Made?

Outline

To be useful, recommendations should give practical, unambiguous advice about a specific health problem. For guidelines about managing health conditions, you should determine if the intent is to prevent, screen for, diagnose, treat, or palliate the disorder. For guidelines about the appropriate uses of health interventions, the recommendations should include a definition of the intervention and its optimal role in patient management. In the American College of Physicians (ACP) guideline on HRT, [2] recommendations are divided into general observations that can help the clinician discuss with patients the effects of therapy, and specific management recommendations concerning what should be done in patient evaluation, risk assessment, hormone administration, and follow-up to achieve the outcomes predicted by the available evidence.

To be clinically important, a practice guideline should convince you that the benefits of following the recommendations are worth the expected harms and costs. You should consider both the relative and absolute changes in outcomes. A 25 percent reduction in relative risk of death from a disease is much more compelling if it involves a reduction in the proportion of deaths from 40 of 100 to 30 of 100 (an absolute risk reduction of 10 in 100), than if it involves a reduction in the proportion of deaths from four of 100 to three of 100 (an absolute risk reduction of one in 100). [3]

The ACP guideline cites extensive and consistent observational data to show that unopposed estrogen therapy (ET) reduces the lifetime risk of developing coronary heart disease (CHD) by about 35 percent (for 50-year-old women with no extraordinary CHD risks, about 12 of 100 would be spared CHD in their lifetimes) and hip fractures by about 15 percent (two to three of 100 avoid hip fracture because of ET use). In women who have a uterus and take unopposed ET, the risk of developing endometrial cancer increases up to eightfold (approximately 17 women of 100 who take ET and would not otherwise have developed endometrial cancer will develop the disease) and the risk for breast cancer may increase as much as 25 percent (absolute increase of about three of 100 women). Clearly, the relative increases or decreases in outcomes can be misleading if baseline risks and absolute changes in outcomes are not reported. Addition of progestin maintains hip fracture risk reduction and removes the increased risk of endometrial cancer, but has uncertain effects on risks for breast cancer and cardiovascular disease. Hormone replacement therapy can increase life expectancy by 10 months to 2 years, depending on the presence of risk factors, a gain similar to that achieved by treatment of hypertension. The guideline did not consider personal or societal costs associated with HRT.

How Strong Are the Recommendations?

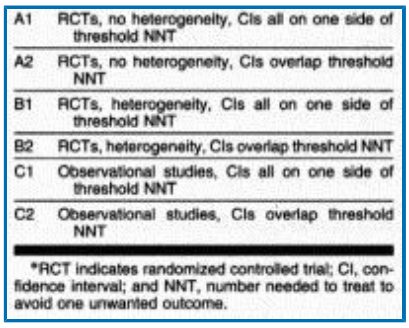
The "strength," "grade," "confidence," or "force" of a recommendation should be informed by multiple considerations: the quality of the investigations that provide the evidence for the recommendations, the magnitude and consistency of positive outcomes relative to negative outcomes (adverse effects, burdens to the patient and the health care system, costs), and the relative value placed on different outcomes. Even in the presence of strong evidence from randomized clinical trials, the effect size of an intervention may be marginal. The intervention may be associated with costs, discomforts, or impracticalities that downgrade the strength of a summary

recommendation about what practicing clinicians should do. It is important to consider this distinction and to scrutinize a guideline document for what, in addition to evidence, determines the wording of actual recommendations. These factors are key to understanding conflicts among guidelines on similar topics from different organizations. [4]

In our first article about using practice guidelines, [1] we pointed out that the best available evidence about the effects of health interventions may come from sources as diverse as, on the one hand, well-conducted randomized trials and, on the other, expert opinion. Thus, users of practice guidelines will find tremendous variability in strength of the evidence linking options and outcomes. Among guidelines developed by different groups about the same health condition or intervention, there should be little variability in estimates of the strength of evidence as long as the supporting overviews considered the same body of literature. [5-7] Here, differences in recommendations probably reflect differences in the relative value placed on various health and economic outcomes. [8] Unfortunately, these considerations are rarely exposed in guideline documents and there is no commonly accepted approach for grading evidence or recommendations. [9-12]

Formal taxonomies of "levels of evidence" and "grades of recommendations" were first popularized by the Canadian Task Force on the Periodic Health Examination, [13] and later revised in cooperation with the United States Preventive Services Task Force. [9] Like previous articles in this series, [14] these guideline developers emphasized that the strongest evidence comes from rigorous randomized controlled trials and weaker evidence from observational studies using cohort or case-control designs. Inferring strength of evidence from study design alone, however, may overlook other determinants of the quality of evidence, such as sample size, recruitment bias, losses to follow-up, unmasked outcome assessment, atypical patient groups, unreproducible interventions, impractical clinical settings, and other threats to internal and external validity. Moreover, results from a single randomized controlled trial with a small sample size are not necessarily more convincing than consistent results with high precision from a large number of high-quality trials of nonrandomized design conducted in a variety of places and times. Recent proposals for summarizing strength of evidence have emphasized the need for overviews to filter out studies with major design flaws, and meta-analyses to consider the precision, magnitude, and heterogeneity of study results. [11] The United States Preventive Services Task Force now supplements its "study design categories" with prose descriptions of flaws in the published evidence. [15]

Another approach to categorizing evidence from multiple studies offers a hierarchy from overviews of observational studies with inconsistent results to overviews of randomized controlled trials with consistent results Table 1. [16] Since inferences about the health effects of interventions are weakened when there are unexplained major differences in effects in different studies, guidelines based on randomized controlled trials are stronger when the results of individual studies are similar, and weaker when major differences between studies (heterogeneity) are present. If the evidence linking interventions and outcomes came from overviews of articles, you could apply the criteria for a valid overview and the schema in the Table 1 to decide on the strength of evidence supporting recommendations.



A1	RCTs, no heterogeneity, CIs all on one side of threshold NNT
A2	RCTs, no heterogeneity, CIs overlap threshold NNT
B1	RCTs, heterogeneity, CIs all on one side of threshold NNT
B2	RCTs, heterogeneity, CIs overlap threshold NNT
C1	Observational studies, CIs all on one side of threshold NNT
C2	Observational studies, CIs overlap threshold NNT

*RCT indicates randomized controlled trial; CI, confidence interval; and NNT, number needed to treat to avoid one unwanted outcome.

Table 1. Grades of Recommendations for a Specified Level of Baseline Risk

[Help with image viewing]
[Email Jumpstart To Image]

This approach is constrained by its focus on only one major outcome (for HRT we are interested in many outcomes), but it exemplifies how the strength of evidence and the strength of recommendations could be integrated on a common scale. It considers study design, heterogeneity, effect size, confidence intervals (CIs) around the effect sizes, and threshold effect sizes over which negative outcomes outweigh the benefits. The threshold effect size presumes value judgments about the relative importance of various outcomes resulting from the health intervention have been applied. In principle, strong recommendations are warranted when the smallest effect compatible with the data (the lower boundary of the CI) is still greater than the threshold below which the negative outcomes outweigh the benefits. (In an upcoming article [16] in this series, we describe this approach to levels of recommendation in much more detail.)

If the guidelines are developed on the basis of observational studies or if the estimate of the treatment effect is imprecise, the user should not expect strong recommendations unless major harms and costs are associated with the intervention or a catastrophic outcome (eg, death) may be prevented by a low-risk, low-cost intervention of probable efficacy. Guideline developers could compensate for weak evidence by testing the effect of their guideline on patient outcomes in a real-world clinical situation. [17] Such a study, if methodologically strong, could enhance the strength of the recommendations in the absence of strong evidence from original studies.

While the ACP HRT guideline does not grade its recommendations, the guideline does cross-reference recommendations to discussions about evidence and effect sizes in the associated overview. Because the guideline is based largely on observational studies, the recommendations are relatively weak, and would be categorized as C1 in the schema in the Table 1.

What is the Impact of Uncertainty Associated With the Evidence and Values Used in the Guidelines?

Guideline developers should consider the possibility that the effect of a management option on an outcome, or the relative value of different outcomes, is much greater, or much less, than their best estimate. We have discussed how to examine this possibility, a process we call sensitivity analysis, in the users' guide for decision analysis. [18] The weaker the evidence linking intervention and outcome, and the greater the possible range of competing values, the greater the need for a sensitivity analysis. For example, the range of plausible estimates of the impact of HRT on breast cancer is very wide, and guideline developers should test how their recommendations would differ across the range of possible effects. When the evidence is of the weakest sort, arising from expert

opinion, sensitivity analysis is essential.

The authors of the HRT guideline acknowledge that the observational design of the studies may introduce bias, and they alert us to areas where the evidence is particularly weak (such as the effect of combined estrogen and progestins on breast cancer). They don't, however, provide a formal sensitivity analysis. Such a sensitivity analysis might have been useful in highlighting the uncertainty of many of the estimates on which the recommendations are based, particularly those relating to life expectancy.

WILL THE RECOMMENDATIONS HELP YOU IN CARING FOR YOUR PATIENTS?

Is the Primary Objective of the Guideline Consistent With Your Objectives?

You should try to anticipate how a guideline will be used. Guidelines may be disseminated to assist physicians with clinical decision making (for example, clinical algorithms and reminders), to enable evaluation of physician practices (eg, utilization review, quality assurance), or to set limits on physician choices (eg, recertification, reimbursement). Guidelines may be directed at different practitioners. Some guidelines about detection and treatment of depression have, for example, aimed to guide primary care providers and others to guide psychiatrists. [19] You should ensure the purpose of the guideline meets the use you intend for it.

Are the Recommendations Applicable to Your Patients?

To be really useful, guidelines should describe interventions well enough for their exact duplication. You must determine whether your patients are the intended target of a particular guideline. If your patients have a different prevalence of disease or risk factors, for instance, the guidelines may not apply.

The flexibility of the guideline may be indicated by patient or practice characteristics that require individualizing recommendations or that justify departures from the recommendations. For example, the American College of Cardiology, the American Heart Association, and the ACP advise against using electrocardiograms to screen asymptomatic adults, but they acknowledge that this advice may not be valid for persons who smoke; are male and of "increased age"; have a family history of coronary artery disease; have hypertension, diabetes, or other cardiovascular risk factors; are sedentary; or whose occupation affects public safety. [20-24] The caveats reflect reluctance to make recommendations in the absence of good evidence. They also exclude groups of patients who, in total, may account for a majority of an internist's patients!

You should look for information that must be obtained from and provided to patients and for patient preferences that should be considered. It is important to consider whether the values assigned (implicitly or explicitly) to outcomes could differ enough from your patients' preferences to change a decision about whether to adopt a recommendation.

When you review the HRT guidelines, you may begin to understand why your colleague in the scenario with which this article began felt that recommendations regarding HRT must be different for every patient. In its HRT guideline, the ACP offers separate recommendations for women at increased risk for CHD, hip fracture and breast cancer, and for women who have had a hysterectomy. These different recommendations reflect the fact that different women are at varying

risk of adverse outcomes, and the impact of HRT on them will therefore differ. The most vivid example is women who have had hysterectomies: since they are not at risk of endometrial cancer, unopposed estrogen is much more likely to be the right treatment choice.

RESOLUTION OF THE SCENARIO

The ACP recommends that all women consider taking preventive hormone therapy, while admitting that no evidence supports strong advice except for some women who are at increased risk for some outcomes. The guidelines suggest that women at increased risk for CHD are likely to achieve longevity gains from HRT, but that conclusion needs to be confirmed by randomized controlled trials. Hormone replacement therapy is likely to decrease the risk of hip, vertebral, and wrist fractures, but, without a progestin, risks for endometrial cancer increase up to eightfold. Women who have had a hysterectomy should take ET alone; others should add a progestin or comply with careful endometrial monitoring. The effect of estrogen on breast cancer appears to be small, but the evidence is weak and many women may not be willing to "take a chance," particularly if they bear low or average risks for CHD. Clinicians should assess risks, estimate benefits and harms, educate patients, and facilitate individualized decision making for all postmenopausal patients.

There is certainly much more to making decisions about HRT than perhaps you or your colleague had at first appreciated. There are many options, multiple outcomes, and significant trade-offs in benefits and harms. A good guideline, based on solid scientific evidence and an explicit process for judging the value of alternative practices, allows you to review, at one sitting, links between multiple options and outcomes. Unfortunately, well-developed and usefully summarized guidelines are still rare in the clinical literature. We hope that more consistent reporting of guideline development methods will prevail, making the guidelines literature more accessible to and useful for prospective guideline users. [25]

We offer special thanks to Deborah Maddock who has provided outstanding administrative support and coordination for the activities of the Evidence-Based Medicine Working Group.

A complete list of members (with affiliations) of the Evidence-Based Medicine Working Group appears in the first article of this series (JAMA. 1993;270:2093-2095). The following members contributed to this article: Deborah Cook, MD, MSc; Brian Haynes, MD, MSc, PhD; Roman Jaeschke, MD, MSc; Andreas Laupacis, MD, MSc; Virginia Moyer, MD, MPH; David Naylor, MD, DPhil; John Philbrick, MD; Scott Richardson, MD; David Sackett, MD, MSc; and Stephen Walter, PhD.

Reprint requests to Room 2C12, McMaster University Health Sciences Centre, 1200 Main St W, Hamilton, Ontario, Canada L8N 3Z5 (Dr Guyatt).

Users' Guides to the Medical Literature section editor: Drummond Rennie, MD, Deputy Editor (West), JAMA.

REFERENCES

1. Hayward RSA, Wilson MC, Tunis SR, Bass EB, Guyatt G, for the Evidence-Based Medicine Working Group. User's guides to the medical literature, VIII: how to use clinical practice guidelines, A: are the recommendations valid? JAMA. 1995;274:570-574. **Ovid Full Text** | **Bibliographic Links** | [Context Link]

2. American College of Physicians. Guidelines for counseling postmenopausal women about preventive

hormone therapy. Ann Intern Med. 1992;117: 1038-1041. **Bibliographic Links** [Context Link]

3. Laupacis A, Naylor CD, Sackett DL. How should the results of clinical trials be presented to clinicians [editorial]? ACP J Club. May/June 1992:A12-A14. Ann Intern Med. Vol 116, suppl 3). [Context Link]

4. Eddy DM. Resolving conflicts in practice policies. JAMA. 1990;264:389-391. **Bibliographic Links** [Context Link]

5. Whelton PK. Reflections on the U.S. Preventive Services Task Force recommendations for screening for hypertension and hypercholesterolemia. J Gen Intern Med. 1990;5:S17-S19. **Bibliographic Links** [Context Link]

6. American College of Physicians. Screening low risk, asymptomatic adults for cardiac risk factors: serum serum cholesterol and triglycerides. In: Eddy DM, ed. Common Screening Tests. Philadelphia, Pa: American College of Physicians; 1991. [Context Link]

7. Canadian Task Force on the Periodic Health Examination. Periodic health examination, 1991 update: lowering blood cholesterol to prevent coronary heart disease. Can Med Assoc J. 1993;148:521-538. [Context Link]

8. Kottke TE, Brekke ML. Cholesterol policy: what should we do? how should we decide? J Clin Epidemiol. 1990;43:1023-1027. **Bibliographic Links** [Context Link]

9. Woolf SH, Battista RN, Anderson GM, Logan AG, Wang E. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. J Clin Epidemiol. 1990;43:891-905. **Bibliographic Links** [Context Link]

10. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest. 1986;89(suppl 2):2S-3S. **Bibliographic Links** [Context Link]

11. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest. 1992;102 (suppl 4):305S-311S. **Bibliographic Links** [Context Link]

12. Acute Pain Management Panel. Acute Pain Management: Operative or Medical Procedures and Trauma: Clinical Practice Guideline. Rockville, Md: Agency for Health Care Policy and Research, Public Health Service, US Dept of Health and Human Services; 1992. [Context Link]

13. Canadian Task Force on the Periodic Health Examination. The periodic health examination. Can Med Assoc J. 1979;121:1193-1254. **Bibliographic Links** [Context Link]

14. Jaeschke R, Guyatt G, Sackett DL, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, III: how to use an article about a diagnostic test, A: are the results of the study valid? JAMA. 1994;271:389-391. **Ovid Full Text** **Bibliographic Links** [Context Link]

15. Battista RN, Fletcher SW. Making recommendations on preventive practices: methodological issues. Am J Prev Med. 1988;4S:53-67. **Bibliographic Links** [Context Link]

16. Guyatt GH, Sackett DL, Sinclair J, for the Evidence-Based Medicine Working Group. User's guides to the medical literature, IX: a method for grading health care recommendations. JAMA. In press. [Context Link]

17. Weingarten SR, Reidinger MS, Conner L, et al. Practice guidelines and reminders to reduce duration of hospital stay for patients with chest pain. Ann Intern Med. 1994;120:257-263. [Context Link]

18. Richardson WS, Detsky AS, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, VII: how to use a clinical decision analysis, B: what are the results and will they help me in caring for my patients? JAMA. 1995;273:1610-1613. **Ovid Full Text** **Bibliographic Links** [Context Link]

19. Rush AJ. Depression in primary care: detection, diagnosis and treatment. Am Fam Physician. 1993;47:1776-1788. **Bibliographic Links** [Context Link]

20. Sox HC Jr, Littenberg B, Garber AM. The role of exercise testing in screening for coronary artery disease. Ann Intern Med. 1989;110:456-469. **Bibliographic Links** [Context Link]

21. Sox HC Jr, Garber AM, Littenberg B. The resting electrocardiogram as a screening test: a clinical analysis. Ann Intern Med. 1989;111:489-502. **Bibliographic Links** [Context Link]

22. American College of Cardiology/American Heart Association Task Force on Assessment of Cardiovascular Procedures. Guidelines for exercise testing. J Am Coll Cardiol. 1986;8:725-738. [Context Link]

23. American College of Physicians. Screening for asymptomatic coronary artery disease: the resting electrocardiogram In: Eddy DM, ed. Common Screening Tests. Philadelphia, Pa: American College of Physicians; 1991. [Context Link]

24. American College of Physicians. Screening for asymptomatic coronary artery disease: exercise stress testing. In: Eddy DM, ed. Common Screening Tests. Philadelphia, Pa: American College of Physicians; 1991. [Context Link]

25. Hayward RSA, Tunis SR, Wilson MC, Bass EB, Rubin HR, Haynes RB. More informative abstracts of articles describing clinical practice guidelines. Ann Intern Med. 1993;118:731-737. **Ovid Full Text** | **Bibliographic Links** | [Context Link]

Practice Guidelines; USERS' GUIDES TO THE MEDICAL LITERATURE (Rennie D, ed)

Accession Number: 00005407-199511220-00042
