

PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Classifying Recommendations for Clinical Practice Guidelines
Steering Committee on Quality Improvement and Management
Pediatrics 2004;114:874-877
DOI: 10.1542/peds.2004-1260

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://www.pediatrics.org/cgi/content/full/114/3/874>

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2004 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



POLICY STATEMENT

Organizational Principles to Guide and Define the Child Health Care System and/or Improve the Health of All Children

Steering Committee on Quality Improvement and Management

Classifying Recommendations for Clinical Practice Guidelines

ABSTRACT. Clinical practice guidelines are intended to improve the quality of clinical care by reducing inappropriate variations, producing optimal outcomes for patients, minimizing harm, and promoting cost-effective practices. This statement proposes an explicit classification of recommendations for clinical practice guidelines of the American Academy of Pediatrics (AAP) to promote communication among guideline developers, implementers, and other users of guideline knowledge, to improve consistency, and to facilitate user understanding. The statement describes 3 sequential activities in developing evidence-based clinical practice guidelines and related policies: 1) determination of the aggregate evidence quality in support of a proposed recommendation; 2) evaluation of the anticipated balance between benefits and harms when the recommendation is carried out; and 3) designation of recommendation strength. An individual policy can be reported as a "strong recommendation," "recommendation," "option," or "no recommendation." Use of this classification is intended to improve consistency and increase the transparency of the guideline-development process, facilitate understanding of AAP clinical practice guidelines, and enhance both the utility and credibility of AAP clinical practice guidelines. *Pediatrics* 2004;114:874-877; *practice guidelines, evidence-based, recommendation, classification system.*

ABBREVIATION. AAP, American Academy of Pediatrics.

INTRODUCTION

Clinical practice guidelines are intended to reduce inappropriate variations in clinical care, minimize harm, promote cost-effective practice, and produce optimal health outcomes for patients. Evidence-based guidelines use a systematic process to select and review scientific evidence to develop policy. In a clinical practice guideline, policy is stated in terms of recommendations. Recommendations are the guideline components that are intended to influence practitioner and patient behavior.

The contemporary, evidence-based approach to guideline development differs from other methods of creating policy in several ways, including:

1. The high level of rigor with which the evidence in support of a policy is identified, appraised, and summarized, and

2. The explicit linkage between recommendations and the evidence that supports them.

Like all scientists, evidence-based guideline developers define their methods first and then allow their methods to lead to the results rather than deciding first on the outcome.

A variety of systems have been used to convey information to guideline readers regarding the quality of evidence that supports a given recommendation and the strength assigned to the recommendation by guideline developers. A large number of numeric and alphabetic codes contribute to general confusion about the meaning of these scales. A recent evidence report prepared by the Agency for Healthcare Research and Quality found 121 scales, checklists, or other types of instruments for rating evidence quality.¹

The American Academy of Pediatrics (AAP) develops clinical practice guidelines internally through various entities and in collaboration with other organizations and also considers for endorsement guidelines developed by external organizations. Clearly, the method used to classify recommendations in clinical practice guidelines should be consistent and explicit. A unified approach will facilitate communication between guideline developers and users and promote appropriate interpretation and application of guidelines by clinicians.

The objective of this statement is to describe a system for defining recommendation strength for AAP evidence-based practice guidelines that is clear, informative, and helpful to users. A common current approach is to append a term to describe the "level of consensus" or, in some instances, fervor achieved. Because this judgment may not reflect unanimity of the development team and can be influenced disproportionately by vociferous or persuasive team members, it represents a lapse in an otherwise explicit system to link guideline statements to the strength of evidence and recommendation.

The proposed system was derived from existing systems that support the principle of using explicit criteria for guideline development. This system is intended for use by the AAP in the process of developing evidence-based clinical practice guidelines. This statement describes 3 sequential processes in evidence-based policy setting: 1) determination of evidence quality in support of a proposed recom-

mentation; 2) evaluation of the balance between anticipated benefits and harms when the recommendation is carried out; and 3) designation of recommendation strength.

CLASSIFICATION PROCEDURE

1. Assess Evidence Quality: Individual and Aggregate

Quality appraisal of individual studies examines both the type of study and the rigor of the investigator's adherence to methodologic principles. Evidence quality refers to "the extent to which all aspects of a study's design and conduct can be shown to protect against systematic bias, nonsystematic bias, and inferential error."² Systematic errors include selection bias and confounding, in which values tend to be inaccurate in a particular direction. Nonsystematic errors are attributable to chance. Inferential errors result from problems in data analysis and interpretation, such as choice of the wrong statistical measure or wrongly rejecting the null hypothesis.

Highest-quality evidence for therapeutic interventions comes from well-designed and well-conducted randomized, controlled trials performed on a population similar to the guideline's target population. The lowest-quality evidence is derived from case reports, reasoning from first principles of pathophysiology, and expert opinion based on ill-defined "clinical experience." Intermediate-quality evidence is associated with a randomized, controlled trial with "nonfatal flaws" or methodologic limitations (for example, one performed on a group from a population different from the target population, therefore requiring that findings be extrapolated) or with an observational design such as a case-control or cohort study. For studies of diagnostic tests, the representativeness of the population studied, the adequacy of the description of the test, the appropriateness of the reference standard against which the test is compared, and the methods used to avoid bias in interpretation of results (such as blinded comparison with the reference standard) are criteria for judging quality.³

After systematically reviewing the literature for studies that bear on a policy decision, evidence-based guideline developers must carefully review each study, extract the findings, and appraise both the quality of the study design and its execution. The specific quality criteria applied depend on the design and type of study. For example, appraisal of controlled trials requires consideration of the adequacy of randomization and blinding and loss to follow-up; assessing case-control studies requires consideration of the appropriateness of matching cases and controls.

Next, guideline developers must consider the quality of the aggregate of studies that bear on the issue. Judging the strength of a body of evidence requires careful consideration of the consistency of the results of individual studies, the magnitude of the effect that the studies detect, and the individual and aggregate sample sizes of these studies.¹

2. Assess Anticipated Balance Between Benefits and Harms

The anticipated benefit, harm, risk, and cost of adherence to a guideline recommendation constitute the second factor that influences the strength of a recommendation. Guyatt et al⁴ suggest looking at the clarity of the balance between benefits and harms. When the evidence indicates a clear benefit not offset by important harms or costs or a clear harm not mitigated by important benefit, stronger recommendations are possible. On the other hand, when the magnitude of the benefit is small or benefits are present but offset by important adverse consequences, the equilibrium between benefits and harms prevents a strong recommendation. A clear preponderance of benefit or harm supports stronger statements for or against a course of action. When the benefit-harm assessment is balanced, no matter how good the studies, practitioners should be offered options rather than recommendations. Such cases mirror the situation described by Bass,⁵ in which "there is adequate evidence at hand to support those who wish to treat...yet the evidence is not so overwhelming as to suggest that those who do not choose to use this form of therapy are in error."

3. Assign Recommendation Strength

Recommendation strength communicates the guideline developers' (and the sponsoring organizations') assessment of the importance of adherence to a particular recommendation and is based on both the quality of the supporting evidence and the magnitude of the potential benefit or harm. The proposed classification defines 4 levels of policy (strong recommendation, recommendation, option, and no recommendation) based on:

- Four levels of aggregate evidence quality: A, B, C, and D (see Fig 1);
- Two benefit-harm assessments: clear (ie, substantial) preponderance of benefit or harm versus a relative balance of benefits and harms; and
- A category for recommendation under exceptional situations in which evidence cannot be obtained but clear benefits or harm are evident.

Because guideline recommendations are prescriptive or proscriptive (constraining variation in practice), guideline developers must follow an approach that has a high likelihood of doing more good than harm. The more restrictive the guidance (strong recommendation), the more certain the guideline developers and endorsers must be of its correctness. The AAP believes that its policy makers should be cautious about classifying a recommendation as strong, lest they jeopardize their credibility by making statements that do not stand up to scientific scrutiny. Most recommendations are likely to be just that: recommendations.

When the evidence is of low quality and the benefit-harm equilibrium is balanced, guideline developers generally should not constrain the clinician's discretion by making a recommendation but instead should designate acceptable alternatives as options.

Evidence Quality	Preponderance of Benefit or Harm	Balance of Benefit and Harm
A. Well-designed, randomized controlled trials or diagnostic studies on relevant populations	Strong Recommendation	Option
B. RCTs or diagnostic studies with minor limitations; overwhelmingly consistent evidence from observational studies	Recommendation	
C. Observational studies (case control and cohort design)	Option	No Recommendation
D. Expert opinion, case reports, reasoning from first principles		
X. Exceptional situations where validating studies cannot be performed and there is a clear preponderance of benefit or harm	Strong Recommendation	

Fig 1. Integrating evidence quality appraisal with an assessment of the anticipated balance between benefits and harms if a policy is carried out leads to designation of a policy as a strong recommendation, recommendation, option, or no recommendation.

Although options do not direct clinicians' actions toward one activity or another, they may place boundaries by delineating appropriate alternative practices.

When the evidence is scant and the balance of benefit and harm is unknown, as for example with some complementary and alternative medicine practices, no recommendation regarding therapy may be possible. Stating that no recommendation is possible provides information but little direction to the clinician. No-recommendation statements are therefore of limited utility and should be discouraged. In some cases, guideline developers still may be able to make policy or offer options based on evidence. For example, although there may be no evidence of effectiveness of a complementary and alternative medicine practice, developers may be able to recommend that clinicians should inquire about use of complementary and alternative medicine and counsel about potential interactions. In other circumstances, guideline developers might suggest individualization on the basis of risk and values.

In some cases, a recommendation or strong recommendation may be made when analysis of the balance of benefits and harms demonstrates an exceptional preponderance of benefit or harm and it would be unethical to perform clinical trials to "prove" the point. These are almost exclusively situations of a medical (not social or political) nature. For example, the anticipated benefit of a recommendation for prescribing anthrax prophylaxis to exposed patients clearly outweighs the expected harms and calls for a strong recommendation, although studies do not exist to support the practice. Such situations with poor evidence but a highly unbalanced benefit-harm equation must be unmistakably differentiated from other circumstances in which

high-quality evidence supports strong recommendations. Requiring the authors to explicitly state the benefit-harm proposition opens it up for constructive debate.

INTERPRETING GUIDELINE RECOMMENDATIONS

How should a clinician interpret recommendations from the AAP in light of the proposed criteria for guideline recommendations? Guidelines are never intended to overrule professional judgment; rather, they may be viewed as a relative constraint on individual clinician discretion in a particular clinical circumstance. Less frequent variation in practice is expected for a strong recommendation than might be expected with a recommendation. Options offer the most opportunity for practice variability.⁶ Clinicians should always act and make decisions on behalf of their patients' best interests and needs regardless of guideline recommendations. Guidelines represent the best judgment of a team of experienced clinicians and methodologists addressing the scientific evidence for a particular clinical topic.

A strong recommendation means that the committee believes that the benefits of the recommended approach clearly exceed the harms of that approach (or, in the case of a strong negative recommendation, that the harms clearly exceed the benefits) and that the quality of the evidence supporting this approach is either excellent or impossible to obtain. Clinicians should follow such guidance unless a clear and compelling rationale for acting in a contrary manner is present.

A recommendation means that the committee believes that the benefits exceed the harms (or, in the case of a negative recommendation, that the harms exceed the benefits), but the quality of the evidence on which this recommendation is based is not as

strong. Clinicians also generally should follow such guidance but also should be alert to new information and sensitive to patient preferences.

An option means either that the evidence quality that exists is suspect or that well-designed, well-conducted studies have demonstrated little clear advantage to one approach versus another. Options offer clinicians flexibility in their decision-making regarding appropriate practice, although they may set boundaries on alternatives. Patient preference should have a substantial role in influencing clinical decision-making, particularly when policies are expressed as options.

No recommendation is made when there is both a lack of pertinent evidence and an unclear balance between benefits and harms. Clinicians should feel little constraint in their decision-making when addressing areas with insufficient evidence. Patient preference should have a substantial role in influencing clinical decision-making.

The AAP believes that adoption of an explicit, consistent classification of recommendations will facilitate communication between the AAP entities that develop guidelines (committees, sections, and task forces and the board of directors) and pediatricians who apply them to clinical practice. We recognize that any classification system may be considered by payors in making reimbursement decisions. This classification is intended only to increase transparency and to enhance the utility and credibility of AAP clinical practice guidelines. Direct linkage of this classification to reimbursement decisions would be overly simplistic, because recommendation strength is one of many factors that should be considered in developing reimbursement policy. Experience and new knowledge will likely require periodic revision of the proposed classification system.

STEERING COMMITTEE ON QUALITY IMPROVEMENT AND
MANAGEMENT, 2003–2004

Charles J. Homer, MD, MPH, Chairperson
Carole M. Lannon, MD, MPH, Director

Norman Harbaugh, MD
Elizabeth Susan Hodgson, MD
*Edgar K. Marcuse, MD, MPH
*Richard N. Shiffman, MD
Lisa Simpson, MB, BCH

LIAISONS

Jay Berkelhamer, MD
Paul Darden, MD
Section on Epidemiology
Denise Dougherty, PhD
Agency for Healthcare Research and Quality
Ellen Schwalenstocker, MBA
National Association of Children's Hospitals and
Related Institutions

STAFF

Junelle P. Speller

*Lead authors

REFERENCES

1. West S, King V, Carey TS, et al. *Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment no. 47*. Research Triangle Park, NC: Research Triangle Institute-University of North Carolina Evidence-Based Practice Center; 2002. AHRQ Publication no. 02-E016
2. Lohr KN, Carey TS. Assessing "best evidence": issues in grading the quality of studies for systematic reviews. *Jt Comm J Qual Improv*. 1999; 25:470–479
3. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med*. 2003;138:40–44
4. Guyatt G, Sinclair J, Cook D, Jaeschke R, Schuenemann H, Pauker S. Moving from evidence to action: grading recommendations—a qualitative approach. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago, IL: American Medical Association Press; 2002:599–608
5. Bass JW. Corticosteroids in croup [letter]. *JAMA*. 1972;221:1164
6. Eddy DM. *A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach*. Philadelphia, PA: American College of Physicians; 1992

All policy statements from the American Academy of Pediatrics automatically expire 5 years after publication unless reaffirmed, revised, or retired at or before that time.

Classifying Recommendations for Clinical Practice Guidelines
Steering Committee on Quality Improvement and Management
Pediatrics 2004;114:874-877
DOI: 10.1542/peds.2004-1260

Updated Information & Services	including high-resolution figures, can be found at: http://www.pediatrics.org/cgi/content/full/114/3/874
References	This article cites 2 articles, 1 of which you can access for free at: http://www.pediatrics.org/cgi/content/full/114/3/874#BIBL
Citations	This article has been cited by 13 HighWire-hosted articles: http://www.pediatrics.org/cgi/content/full/114/3/874#otherarticles
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Office Practice http://www.pediatrics.org/cgi/collection/office_practice
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.pediatrics.org/misc/Permissions.shtml
Reprints	Information about ordering reprints can be found online: http://www.pediatrics.org/misc/reprints.shtml

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

