

Attributes of ATS Documents That Guide Clinical Practice

Recommendations of the ATS Clinical Practice Committee

THIS REPORT WAS APPROVED BY THE ATS BOARD OF DIRECTORS, JUNE 1997.

The American Thoracic Society has an extensive history in the development and dissemination of reports that guide clinical practice. ATS official statements, position papers, and conference reports have provided a major contribution to public health and the care of patients with thoracic disorders and critical illnesses worldwide for more than 80 years (1). These documents have improved clinical practice by providing clinicians with scientifically valid indications for medical interventions in addition to diagnostic and therapeutic approaches to managing specific pulmonary and critical care problems. Formulation of these practice guidelines have rested on conclusions derived from scientifically sound clinical and basic science investigations whenever possible or on expert consensus in the absence of high grade investigative data. Central to the development of ATS documents has been an interest in constructing pragmatically sound statements that can be effectively applied by practicing physicians. This practical approach to guideline development intends to ensure rapid dissemination and prompt application of ATS recommendations with goals of promoting the health of patients with or at risk for respiratory and critical care conditions.

Recent national trends have accelerated the development of documents that guide clinical practice by professional medical societies, governmental bodies, private enterprises, insurers, hospitals, physician practice groups, academic centers, managed care organizations, and independent research centers (2). Expanding goals of stemming rising health care costs, lessening practice variation, and monitoring inappropriate care have supplemented the traditional intent of practice guidelines to communicate advances in care (3). Federal mandates to emphasize outcome studies in health services research further foster the development of clinical practice guidelines (4). Additionally, the individual physician faces increasing challenges to analyze and assimilate the mass of published clinical studies, which vary greatly in quality (5). And perhaps most importantly, physician groups are developing practice policies to maintain quality of care in the face of policies formulated by nonphysician organizations that are intended for use as cost-saving management tools (6).

Emerging from these trends is recognition that the development of valid and reliable documents for guiding clinical practice is a scientific process that is aided by application of an explicit methodology and documentation of its use within the text of a guideline. Furthermore, some practice guidelines

may require "testing" similar to new drugs to determine their effectiveness and safety (7, 8).

Recognition of the need for an expanded methodology has propelled documents that guide clinical practice toward a new formality termed "clinical practice guidelines," which implies new processes for document development (9). The emphasis on new development methodologies should not lose sight of the fact that clinical practice guidelines are not truly a departure from the past. Recommendations for guiding clinical practice have been published for centuries in medical textbooks. When defined as official statements or policies of major organizations on the proper indications for medical procedures or treatments or the proper management of specific clinical problems, practice guidelines have existed for more than 50 years (2).

Nevertheless, the methodology for developing practice guidelines has evolved in parallel with the changes in methodology that have occurred during the last 50 years for designing, analyzing, and reporting original investigative research. There has been a shift away from unexamined reliance on professional expert judgment toward more formal, structured, and accountable processes (10). Organizations sponsoring guideline development recognize that well-constructed and effectively disseminated statements incorporating the best available research and professional judgment have the greatest opportunity of achieving their goals. Furthermore, readers of guidelines expect the same rigor and quality that is applied to the other scientific literature they read (11).

The purpose of the present statement is to develop a series of recommendations-or guidelines-for developing ATS statements that guide clinical practice. These recommendations are intended to assist ATS committee members in the development of ATS statements and aid ATS members in their critical appraisal of clinical practice guidelines published by the ATS and other organizations. The recommendations of this statement should serve as guidelines and not absolute standards for developing ATS statements. The ATS recognizes that every clinical topic does not require complete application of a formal methodology in its entirety; different questions require different degrees of methodologic rigor. Moreover, full application of a formal, comprehensive methodology for every facet of every statement would be prohibitively expensive in terms of human and financial resources. Application of a formal methodology to a broad-based subject may require a year or more for document development (12). But whatever techniques are employed in guideline preparation, documentation within the statement text of an explicit methodology, as required for any scientific publication, assists ATS membership and other organizations in evaluating the validity of the recommended practice policies, thereby promoting their rapid adoption into clinical practice.

Correspondence and requests for reprints should be addressed to John E. Hefner, M.D., Professor of Clinical Medicine, University of Arizona Health Sciences Center, St. Joseph's Hospital and Medical Center, 350 West Thomas Road, Phoenix, AZ 85013.

DEFINITION OF TERMS

A variety of titles have been applied to statements that guide clinical practice, including clinical practice guidelines, practice parameters, and clinical policies, to name a few. Specific attributes have been variably attached to each of these different terms. Regardless of the term used, however, these documents are all ideally unambiguous guidelines for aiding physicians in selecting appropriate diagnostic and therapeutic plans for particular clinical problems in specific patient groups (13). These statements or supplemental documents paraphrased in lay terms may also assist patients in selecting appropriate care. To be effective, guidelines need to be practical, clinically important, convincing, and acceptable to those who will use them (14). Most importantly, guidelines are developed with an explicit methodology that ensures their effectiveness and acceptability (Table 1).

Recommendation: The ATS adopts the definition of the Institute of Medicine for its statements that guide clinical practice, which describes guidelines as *systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances* (9). For the purpose of the present statement, practice guidelines and statements that guide clinical practice will be used as general terms for any ATS document that provides patient care recommendations and practice policies.

GOALS OF GUIDELINE DEVELOPMENT

The overlying general goals of all ATS statements are to: (1) promote quality, effectiveness, and appropriateness of health care; (2) maintain access to quality care for patients with respiratory and critical care conditions; and (3) identify the limits

of existing investigative data that define guidelines in order to set research priorities (15). These goals are achieved by the development of clear and clinically applicable recommendations formulated from a synthesis of scientific evidence and expert opinions. Using the best synthesis of scientific evidence and expert judgment, ATS guidelines can counterbalance the market influences and economic incentives of patient care systems that primarily focus on decreasing healthcare costs. Well-constructed and valid guidelines can assist cost-control measures in decreasing the *overuse* of unnecessary healthcare interventions while safeguarding against the *underuse* of needed and valuable medical care (16).

Statements that guide clinical practice may have differing specific goals that depend on the problem or issue approached and the audience addressed. In one instance, a document may be directed toward bringing the results of new and important clinical or basic science investigation rapidly to clinical practice. In other instances, goals may center on affecting health care policy at the federal, state, or regional level. Specific goals may center on issues of disease prevention, screening techniques, diagnosis, treatment, or palliation. The audience for statements may include ATS membership, other physicians, payers, managed care organizations, hospital and clinic administrators, patients, patient families, and nonphysician health care providers.

Statements are made more effective by a clear understanding during the development and dissemination of documents of both their general and specific goals.

Guidelines are not substitutes for caregiver expertise. They should not be considered tools for allowing clinicians who are otherwise unprepared by training or experience to care for patients in the clinical settings addressed by a guideline (17).

TABLE 1
STEPS IN DEVELOPING CLINICAL PRACTICE GUIDELINES

Topic selection	Is it important? Does a sufficient body of applicable literature exist? Does it have economic impact? Does extensive practice variation exist?
Chairperson selection	Does the proposed chair provide general content expertise, risk for bias, demonstrated group facilitation and project organizational skills, and commitment to evidence-based methodologies?
Committee member selection	Is there a need for multidisciplinary and multiorganizational membership? Can focused expertise be provided by group membership or outside presentations to the committee? Is there a risk for conflict of interest? How broad should the definition of "expert" be? Were nonphysician experts considered for committee membership? Were patient preferences explicitly considered?
Searching for the evidence	How can a comprehensive search for evidence be performed? Can the search be reproduced as described in the final guideline?
Grading the evidence	What is the quality of the evidence? What is the magnitude of the effect? How heterogeneous is the evidence?
identifying relative health care and economic values	Was the relative importance of various health care values and outcomes for formulating recommendations stated? Are there economic assessments of recommendations? Were costs of alternative approaches considered? Are final recommendations based on a balance between the quality of the data, relative values of the outcomes, and subjective judgment of experts? Are methods for achieving balance between these factors explicitly described?
Seeking input from outside of the development committee	Were there efforts during guideline development to receive opinion from experts outside of the development committee? Was there a forum for presenting nearly completed recommendations?
Critical review of a document draft	Were outside reviewers used to critically analyze the document? Is the document valid, reliable and reproducible, clinically applicable, flexible, and clear in its use of language? Are all methodologies explicitly documented? Is there a structured abstract? Are there recommendations for scheduled review? Were all important stakeholders affected by the recommendations included in some stage of the process? Are there proposals for determining the clinical impact of the recommendations? Are there plans or recommendations for validating the recommendations after publication? Were the method and rationale for selecting the topic stated?
Implementation strategy	Are the plans for disseminating and implementing the guidelines sufficient to ensure their clinical application? Are there measures for monitoring performance measures? Are there provisions for periodic review? Will the information be disseminated in a broad-based manner? Will patient-oriented information be distributed to allow patients to exercise their preferences by selecting among alternative recommendations when none has been shown to be clearly superior? Have patient-oriented materials been validated to be effective in test populations?

Recommendation: All ATS statements that guide clinical practice have an overlying goal of promoting quality, effectiveness, and appropriateness of health care and maintaining access to quality care for patients with respiratory and critical care conditions. Effectiveness of statements is enhanced by an explicit description within the text of the specific goals of the document by clearly defining the major questions addressed (18). In the final analysis, ATS statements identify health care that is appropriate as defined by interventions wherein quantitative assessments confirm that clinical benefit obtained outweighs the harms and costs involved (19). Guidelines are intended to guide the clinical practice of qualified caregivers by providing recent advances in knowledge; they are not designed to substitute for experience and training or to facilitate care delivery by inadequately prepared practitioners.

ATTRIBUTES OF AN EFFECTIVE STATEMENT THAT GUIDES CLINICAL PRACTICE

Well-constructed guidelines have the greatest opportunity for being accepted by their target audiences and achieving their goals. Consensus is emerging regarding the attributes of guideline construction (Table 2). The following attributes are promoted by the Institute of Medicine and the Agency for Health Care Policy and Research (AHCPR) (18, 20).

Validity. Guideline validity is defined by the ability of a document to produce its intended health and cost outcome goals if the target audience follows the guideline's recommendations. A target audience can evaluate the potential validity of a guideline by critically analyzing relationships between the evidence provided and the recommendations made, the substance and quality of the scientific and clinical evidence cited, the means used to evaluate the evidence, and the methods (e.g., majority rule, ballots, Delphi techniques) used to arrive at consensus of the guideline committee members. In some instances, guidelines may describe proposed or planned steps for validation of the recommendations after publication of the document (21). Validity also may be considered in terms of face and content validity. Face validity is the document's appearance of being reasonable and appropriate to relevant experts. Content validity depends on how solidly the guidelines are based on the literature and current clinical knowledge (22).

Reliability and reproducibility. These attributes are defined by the probability that a different group of committee members of similar expertise would have developed similar guidelines if provided with the same evidence and methods. Also,

they measure whether the guidelines are interpreted and applied consistently by physicians and other health care providers in similar clinical circumstances. An explicit within-text description of the methodology for a guideline's development and methods for selecting committee members assists analysis of the likelihood of reliability and reproducibility.

Clinical applicability. Appropriate application of guidelines in clinical practice and avoidance of guideline misuse are aided by describing within the text the patient populations to which the guideline is directed. This definition should be as inclusive as permitted by available scientific and clinical evidence and expert judgment.

Clinical flexibility. Practice guidelines cannot define optimal care in all clinical settings with absolute certainty given the imperfect process of analyzing evidence and developing expert consensus, regional variations in expertise and resources, and the nonuniformity of patients and their diseases (15). Guideline documents should consider all sensible alternative practices and all reasonably likely potential outcomes to allow practitioners to adopt an appropriate approach to individual patients' clinical needs and personal values. Guideline recommendations should be identified as standards of clinical practice only if sound and extensive investigative data support no practice variation from the recommendation (e.g., influenza vaccination for patients at risk without contraindications to vaccine). Few practice policies have achieved sufficient evidence to qualify as standards. Recommendations listed as guidelines assume situations exist wherein all or some recommendations cannot or should not be followed. Recommendations that are options may depend on a consensus of expert opinion alone or weak evidence. Options are so flexible that they provide little guidance for clinical decision making but provide practitioners with one or more possible courses of action (6). Whenever possible, statements that guide clinical practice should state explicitly known or anticipated exceptions to recommendations. Guidelines should complement rather than substitute for practitioners' experience and judgment.

Clarity. To promote reliability and reproducibility and define the degree of clinical flexibility, guidelines should state clearly and precisely what they mean by their recommendations (1.5). Unambiguous and precise use of terms with concrete recommendations that use readily understandable modes of information communication are encouraged. Numeric thresholds and explicit decision points allow guidelines to be *decidable* (23). Established epidemiologic terms should be used where they exist rather than nonstandard descriptive phrases. Algo-

TABLE 2
ATTRIBUTES OF A WELL-CONSTRUCTED GUIDELINE

-
- Validity: Ability of a guideline to produce the health and cost outcome goals of the document.
 - Reliability and reproducibility: Probability another group of experts would develop the same guidelines given the same evidence and using the same methodology. Probability that physicians and other care givers would interpret and apply the guidelines similarly given the same clinical circumstances.
 - Clinical applicability: Explicitly state as much as possible the populations to which the guidelines apply.
 - Clinical flexibility: Avoid overstatement of recommendations, recognizing the complexity of clinical practice. Explicitly state known or anticipated exceptions.
 - Clarity: Unambiguous and precise terms, standard epidemiologic terms where they exist, concrete recommendations, easily understandable modes of information. Precision in describing strength of recommendations (i.e., standardized use of "musts," "shoulds").
 - Multidisciplinary process: Participation by key groups affected at various stages of development through various levels of participation.
 - Schedule review: Schedule or timeline for update and revision. Appointment of a subgroup of the guideline development committee to monitor advances in the field so as to advance the timeline for revision if necessary.
 - Documentation: Explicit description of methodology for guideline development.
 - Planned guideline assessment: Recommendations within the text for confirming the validity of the guideline.
-

rhythms, tables, and figures should be substituted for dense or ambiguous text whenever possible. The strength of a recommendation should be clearly indicated by precise usage of words, such as "must," "should," and "may" (Table 3). The inclination of committees to overstate recommendations based on opinion rather than high-grade evidence should be avoided (20).

Multidisciplinary process. The process of guideline development should include participation by representatives of the major groups affected by the recommendations. A broad definition of "expert" may include investigators, academicians, practicing physicians, nonphysician caregivers, payers, and patients. Participation in guideline development includes not only membership on a development committee but also activities in presenting evidence and opinions to the committee and reviewing drafts of planned methodologies and guideline drafts.

Scheduled review. Valid guidelines have a lifespan that varies depending on the topic addressed. Published guidelines should be periodically reviewed for update and revision. Anticipated schedules for review should be included into guideline texts to allow readers to recognize when recommendations may become outdated and obsolete. This "date stamping" warns readers when recommendations may be no longer valid in order to avoid unintentional promotion of outdated rather than optimal care (22). It is often difficult to anticipate when a guideline for a particular topic should undergo scheduled update and revision. A subgroup of the guideline development committee should be appointed to monitor the guideline after its publication so as to advance the timeline for revision should advances in the field make the guideline's recommendations prematurely obsolete.

Documentation. As emphasized throughout this statement, a well-constructed guideline requires a valid methodology for its development. Guidelines should explicitly state within their text the methodology used, the participants involved, the evidence used, the assumptions and rationales accepted, and the analytic techniques employed.

Recommendations: ATS statements that guide clinical practice should seek to attain the above attributes. Although a determination of validity is the most central aspect of guideline development to ensure the guideline's favorable impact on health care, the vast majority of guidelines developed in all fields of medicine have not undergone a validation process.

IDENTIFYING A TOPIC, FORMULATING QUESTIONS, AND DETERMINING THE END-USER

Topics appropriate for guideline development can be identified by their potential for limiting practice variation, improving patient outcome, and providing information on new, innovative, or not widely known health care interventions. Guidelines can focus on a condition (e.g., cystic fibrosis), a procedure (e.g., pulmonary function testing, fiberoptic bronchoscopy), or a process (e.g., utility of nurse practitioners in the ICU, ethical decision making). Most topics will arise out of some degree of valid clinical evidence or developed experience. The absence of extensive, high-quality evidence, however, does not obviate the selection of an important topic for guideline development. In such instances, recommendations would be based on expert judgments that will be subject to revision on the basis of future evidence. Recommendations based solely on expert judgments, however, are highly subject to bias. In some instances, guideline developers may be better off collecting new data prospectively rather than basing recommendations on inadequate literature (23).

Several professional societies and the Institute of Medicine have developed criteria for selection of topics for guideline development (24) (Table 4). Consensus methodologies exist for groups charged with selecting topics among many possibilities (25). Annual queries to ATS membership, assemblies, the Council of Chapter Representatives, and committees are commonly used to identify lists of potential topics. These topics are reviewed and prioritized by the Clinical Practice Committee, which will assist with the development of a work plan and budget for submission to the Program and Budget Committee. The Program and Budget Committee will prioritize funding, considering other ATS project requests and the potential for alternative sources of funding. The methodology employed and the rationale for selecting the topic should be explicitly stated in the guidelines to emphasize the importance and relevance of the recommendations.

After selecting the topic, precise definitions of the questions to be answered and establishment of clear objectives assist the guideline development process.

It is increasingly apparent that success or failure of a guideline is judged as much by its use as by its development process (26). A common problem is a failure to identify the needs and

TABLE 3

ATS DEFINITIONS FOR THE USE OF DIRECTIVE LANGUAGE IN GUIDELINE RECOMMENDATIONS

- Must be done: Practices that have been shown to be, or are felt to be, of such clear value that they should be offered to all applicable patients. A practitioner who chooses not to effect such a practice must provide clear, cogent reasons for the departure as part of a permanent medical record. Such practices make appropriate medical review criteria. This type of recommendation represents a "standard."
- Should be done: Practices that have been shown to be, or are felt to be, of value to the great majority of applicable patients. The operational definition for ATS guidelines is that at least 80% of an applicable population should be provided the practice. While considered necessary or essential for most patients, such a practice may be inappropriate for a small number of individuals. The practitioner must make that judgment. A decision not to effect such a practice should be documented in the permanent medical record. These recommendations are termed "guidelines."
- May be done: Practices that have been shown to be, or are felt to be, of value to a significant proportion of applicable patients. Their use or non-use in any particular patient should be at the discretion of the practitioner and would not require special documentation or justification. Such a practice is an "option."
- Should not be done: Practices that have been shown to be, or are felt to be, of value to only a small minority of applicable patients. The operational definition for ATS guidelines is that fewer than 20% of an applicable population would be provided such a practice. While considered unnecessary or inappropriate for most patients, such a practice may be essential for a small number of individuals. The practitioner must make that judgment. A decision to effect such a practice should be documented in the permanent medical record.
- Must not be done: Practices that have been shown to be, or are felt to be, of either no value or of such great harm that they should not be offered to any applicable patient. A practitioner who chooses to effect such a practice must provide clear, cogent reasons for the departure as part of a permanent medical record. Such practices make appropriate medical review criteria.

TABLE 4
EXAMPLES OF PRIORITY-SETTING CRITERIA FOR
SELECTING GUIDELINE TOPICS*

<i>Institute of Medicine</i>
Objective criteria:
Prevalence of the condition
Cost of technology used to manage the condition
Variation in use of this technology
Subjective criteria:
Burden of illness
Potential to change health outcomes
Potential to change costs
Potential to clarify ethical, legal, or social issues
<i>College of Family Physicians of Canada</i>
Frequency of the condition
Seriousness of health consequences
Effect of intervention
<i>American College of Physicians</i>
Potential significant health benefit
Potential risk
Potential wide application
Extent of interest to practitioners
<i>American Medical Association</i>
Potential impact on substantial patient population
Controversy within the medical community
Availability of scientific data to support evaluation

* From reference 24 with permission.

resources of the end-user. Often the end-user is not the physician but other members of the healthcare team, payers, administrators, or other critical components of health care systems. Use of guidelines may be delayed or ignored if the person who reads the document is not who will use it or is not in a position to change or assess the patient care addressed by the guideline. Structuring the document toward the needs of the end-user permits both physician acceptance and more rapid incorporation of the recommendations into clinical practice.

Recommendations: The explicit criteria and methodologies for selecting topics for guideline development should be described in the final guideline document. The specific questions to be addressed should be clearly stated. Precision in defining the guideline's questions, stating the objectives, and identifying the end-user before beginning guideline development provides clarification of purpose and facilitates the process.

APPOINTMENT OF AN EXPERT CONSENSUS DEVELOPMENT COMMITTEE

A strength of the ATS is its members, who can provide requisite content expertise, multidisciplinary outlooks, and analytic skills to participate in the development of valid statements that guide clinical practice. Many guidelines, however, affect stakeholders beyond the ATS membership and would benefit during their development from joint sponsorship with other organizations. A multidisciplinary and multiorganizational approach to selecting members for guideline development committees increases the likelihood that the final recommendations will be adopted into clinical practice.

The number of members selected for a guideline development committee varies depending on the complexity and breadth of the questions approached. Although large numbers decrease the likelihood of biases, overly large groups slow the consensus process and introduce too many compromises of opinion, resulting in bland statements of little practical clinical use. Groups composed of 10-15 members for large projects

promote adequate discussions and allow a multidisciplinary process.

Selecting members from different backgrounds (physicians, practitioners, academicians, nurses, respiratory therapists, payers) further promotes the validity of a guideline (27). Selection of participants with similar points of view predetermines the "consensus" of the group (28). In some situations, the definition of an "expert" should be broadened to include patients or patient representatives in order to obtain their unique perspective and preferences (29). Consideration of patient preferences is becoming increasingly important in guideline development (30, 31). Some topics may require special expertise in epidemiology, health services research, law, ethics, or health care economics. In other instances, this expertise can be provided by individuals appointed to advise the committee about specific issues.

Regardless of the constitution of the final group, the methods for selecting group members and the criteria for membership should be explicitly stated in the guideline text to allow readers to determine likelihood of group biases.

Protection against the appearance of conflicts of interest and maintenance of the integrity of the process requires potential members to submit curriculum vitae and sign the Statement on the Potential for Conflict of Interest of Members of the American Thoracic Society. Committee members should have no substantial financial interests or affiliations that might compromise the integrity of the development process or final document.

The success of a guideline committee often depends on the commitment, group facilitation skills, and organizational abilities of the committee chairperson. Chairs should have demonstrated leadership, experience, and training in the general field under discussion but need not have widely recognized content expertise in the specific topic of the guideline. In many instances, selecting a chair who is not an expert on the specific topic limits bias in the group process. The chairperson should be characterized by an interest in quality of care, commitment to the development of evidence-based documents through formal methodologies, experience in analytic processes that can be used in evidence-based methodologies, absence of conflict of interest, and a demonstrated ability to focus group interactions and develop recommendations by evidence and formal consensus-based processes.

Recommendations: A multidisciplinary and multiorganizational membership of the development committee should be sought when widened participation will improve the validity and effectiveness of the recommendations. The guideline should explicitly state the methods for appointing members to the committee to indicate how these methods provided balanced expertise and limited group bias. Qualifications as an expert may be broad and, in some instances, include stakeholders in the recommendations other than physicians. The need for focused expertise should be anticipated during the selection of members or presenters to the committee. Potential committee members should submit curricula vitae and conflict of interest statements. The committee chairperson should be selected on the basis of expertise and experience in the general field addressed, group facilitation skills, organizational capabilities, and commitment to evidence-based methodologies.

EXPLICIT EVIDENCE-BASED DEVELOPMENT

To varying degrees, all guidelines derive from the scientific evidence that following their recommendations will promote the desired outcomes. Guidelines are developed, therefore, with a

goal to consider all scientific evidence relevant to the major questions of the guideline. Explicit statements are required within guideline texts regarding the comprehensiveness of the literature search and the methods of the search so that it may be reproduced by others.

The strength of recommendations depends to a degree on the quality of the underlying scientific evidence. It is not entirely clear, however, what counts for evidence and how to assign levels of importance or weights to different types of evidence (32). Formal methodologies have been devised by the Canadian Task Force on the Periodic Health Examination (33), AHCPR guideline committees (32), and the United States Preventive Services Task Force (34) to grade the level of evidence. These statements generally consider that randomized prospective studies provide higher grade conclusions compared to observational investigations using cohort or case-control designs. The latter are believed to provide stronger evidence than expert opinion. Two of these grading systems are included in the APPENDIX and others have been recently employed with modification for specific projects (35).

Grading of the literature, however, requires additional critical analysis to determine the strength of the evidence. Considerations include sample size and design flaws such as recruitment bias, unmasked outcome assessment, verification bias, atypical subject groups, unusual or impractical patient settings, and unreproducible interventions. These design flaws weaken external validity and the strength of guideline recommendations that studies with such flaws can support (14). Moreover, independent, well-designed studies may each arrive at different conclusions so that a measure of heterogeneity of the evidence is required. Also, the precision and magnitude of the study results affect the strength of guideline recommendations. In evaluating study design and effect magnitude, consideration of overviews and meta-analyses assists guideline developers in determining the strength of guideline recommendations that can be supported by the evidence (36). The major elements of study design and execution that affect validity are listed in Table 5.

To incorporate the strength of the evidence into grading systems, the United States Preventive Services Task Force includes prose descriptions of design flaws in its grading of the evidence (37, 38). An extensive categorical assessment system developed by the AHCPR committee for developing recommendations for the management of patients with left ventricular dysfunction provides an explicit guide to grade the quality of randomized controlled and cohort studies for ATS guideline committees (32) (APPENDIX). These and other (39) proposed methodologies include an evaluation of the heterogeneity of study results (APPENDIX). Clear consensus regarding what constitutes a well-conducted or flawed study, however, remains incompletely defined (32).

TABLE 5
ELEMENTS IN STUDY DESIGN AND EXECUTION
THAT AFFECT VALIDITY*

Selection of patients
Allocation of patients to treatment groups
Therapeutic regimen
Study administration
Withdrawals from the study
Patient blinding (randomized clinical trials only)
Outcome measurement
Statistical analysis

* From reference 32 with permission.

The relative value placed on various health and economic outcomes may also determine the strength of a recommendation (40). Recommendations based on results from high-quality studies may be difficult to promote if the magnitude of a positive effect is marginal and associated with high economic costs, patient discomforts, or impracticalities. Conversely, weak evidence most often can only justify weak recommendations ("options") unless the evidence suggests a reasonable likelihood that major harm can be avoided with a low-risk, low-cost intervention (14).

Recommendations: A comprehensive search for and review of the evidence is an important process in guideline development. The literature search should be explicit and reported in guideline texts so it can be reproduced by others. The evidence should be graded by established methodologies as to the quality of the study design, the magnitude and consistency of positive compared to negative outcomes, the threshold effect magnitude over which negative outcomes outweigh the benefits of a positive effect, and the relative value placed on different health and economic outcomes (14). Because no specific grading system has undergone validation (32), the selection of a methodology should be left to the guideline development committee. The grading method should be explicitly described within the text so that another set of experts developing similar guidelines would derive similar gradings of the strength of the guideline recommendations. Those who graded the literature should be identified (28). To establish the currency of the guideline, the dates of the most recent literature review, the most recent group input, and the submission of the completed document for publication should be stated (28).

INCORPORATING EXPERT OPINION

An evidence-based approach to guideline development is occasionally misinterpreted to mean that expert opinion must be excluded from the formulation of recommendations. Grading of the evidence and the determination of relative values of different health care outcomes, however, are dependent on subjective expert opinion. Proponents of explicit criteria for determining quality of study design and grading of the evidence admit that little empirical support exists for assigning the importance of study flaws; judgments depend, to a degree, on subjective panel opinions (32). Moreover, there are simply not enough high-quality clinical trials or epidemiologic studies to guide all clinical practices (41). In the absence of evidence, professional judgment and group consensus fill the gaps of knowledge (42, 43). Recommendations derived partly or solely from expert opinion, however, should be explicitly identified in the guideline text as to their subjective nature. Additionally, methodologies used for developing expert consensus should be stated to determine the potential for expert bias in the group process.

These methodologies for synthesizing expert consensus should first define what is meant by "consensus" (43). Definitions include unanimity, majority rule, two-thirds agreement, scoring systems with an achieve threshold score, or other specific techniques (44). Once consensus is defined, processes are initiated to achieve consensus and develop the clinical practice guidelines. Historically, informal group discussions have been most widely used but are highly subject to bias and are the least valid method for developing clinical guidelines (42, 45).

Consensus conferences represent a more formal approach wherein a jury convenes to review the evidence and recommendations previously analyzed and developed by a smaller expert group (46, 47). Consensus conferences are most effective for specific topics, projects with a small number of precise

questions, and topics involving ethical and social considerations (5).

Quantitative assessment of expert opinion after an initial review of the literature can be done through questionnaire techniques with appropriateness scores (48) and variations of the Delphi technique (42, 43). In this method, questionnaires are sent to experts, results tallied, and questionnaires with tallied results are resubmitted to experts in an iterative process until stable consensus is achieved (42, 43, 49, 50).

Another approach uses an expert panel to review and grade the literature and another panel to subsequently formulate recommendations by blending the evidence with expert opinion (9, 38, 42, 51, 52). Opinions can be developed in informal group discussions, open voting, or anonymous balloting. More formal group processes with variations of Delphi methods and nominal group techniques prevent outspoken, charismatic, and opinionated members of the group from dominating the consensus process (42, 43, 53, 54). Iterative drafts of guideline statements can be developed and distributed for criticism with repeated revisions until no further criticism is elicited (43). This approach is expensive and laborious but produces the most valid recommendations and can be applied to broad subjects (5).

The definition of consensus and the processes used to achieve it should be explicitly stated in the final document. Often, important minority opinions persist after a consensus process and these opinions should be documented. Regardless of the consensus development process employed, adherence to a methodology creates a legitimacy for the final guidelines (55).

Recommendations: Expert opinion is an important component of guideline development. Evidence-based methodologies, however, are directed at reducing expert subjectivity to its minimum attainable value (32). The degree to which various recommendations are based on subjective expert opinions should be identified and explicitly stated. The structured group process for achieving expert consensus should be described. Establishing a formal methodology for developing expert consensus before embarking on guideline development is akin to the scientific method in any field of investigation that requires predetermined methods for collecting and analyzing data.

DEVELOPING THE DOCUMENT

Multiple resources provide guideline developers with templates for designing a consensus process (5, 9). These resources can assist small groups in project development when members of the group lack experience with consensus methodology. Larger groups with projects of broad scope benefit by selecting a chairperson and one or more group members who have experience in guideline development techniques. This information can also be provided during the initial meeting of the guideline development committee.

In general, after reviewing the evidence, considering the opinions of external experts, and completing a group consensus process, the committee writes a statement draft that synthesizes the information into conclusions and guidelines. The drafts include a complete list of references and an explicit description of the methodology for evaluating and grading the evidence, developing group consensus, incorporating opinions of experts, and conducting any surveys or studies. The nature of the evidence or expert opinion upon which each recommendation is based should be clearly stated. Recommendations based primarily on opinion or weak evidence should be highlighted as areas in need of further research and research funding.

Drafts of the guideline document are submitted to members of the committee, external experts, and related organizations for critical review to suggest revisions and gather support (42, 43). Suggested modifications are resubmitted to the members of the development committee. This process increases the validity of the final document and enhances diversity of opinion. Overly extensive revisions and repeated deliberations, however, may result in outdated statements with bland content (56). During this stage of validation, the document may be used in clinical practice for a brief period by a small group of sample practitioners to determine its clarity, applicability, and reliability.

To achieve official status, the document will be sent to the appropriate ATS Assembly chairperson. The document will then be sent to at least two independent and recognized experts for written critical review. These experts shall not be members of the ATS Board of Directors nor close associates of any member of the development committee. The Assembly chairperson will transmit the reviewers' comments and the manuscript to the development committee for revision.

Recommendations: Drafting the guideline document is an iterative process that first synthesizes the evidence and consensus of the group with opinions of outside experts into conclusions and recommendations that are then critically reviewed in draft form by committee members, external experts, and other organizations. Modifications to drafts are resubmitted to committee members for review and acceptance. Pretesting the document with practitioners evaluates clarity, applicability, and reliability.

STRUCTURED ABSTRACT

As published guidelines become more common, extracting and repository services have relied more on structured abstracts to evaluate and promulgate guideline recommendations. The structured abstract not only assists dissemination of the information through national and organizational databases but also promotes its adoption into clinical practice by assisting readers in quickly analyzing the purpose, applicability, and methodology of a guideline. The end-users for whom the document is intended should be clearly stated.

Recommendations: The ATS supports the incorporation of a structured abstract as developed in collaboration with the U.S. National Library of Medicine (NLM) and approved for online abstracts in NLM databases (Table 6) (57).

TABLE 6
COMPONENTS OF A STRUCTURED GUIDELINE ABSTRACT

- **Objective:** the primary objective of the guideline, including the health problem and the targeted patients, providers, end-users, and settings.
- **Options:** the clinical practice options considered in formulating the guideline.
- **Outcomes:** significant health and economic outcomes considered in comparing alternative practices.
- **Evidence:** how and when evidence was gathered, selected, and synthesized.
- **Values:** disclosure of how values were assigned to potential outcomes of practice options and who participated in the process.
- **Benefits, harms and costs:** the type and magnitude of benefits, harms and costs expected for patients from guideline implementation.
- **Recommendations:** summary of key recommendations.
- **Validation:** report of any external review, comparison with other guidelines or clinical testing of guideline use. Planned assessment of the guideline's validity.
- **Sponsors:** disclosure of the person(s) who developed, funded, or endorsed the guideline.

Modified and reproduced with permission (57).

IDENTIFYING RELATIVE HEALTH CARE AND ECONOMIC VALUES

The economic costs of recommended health care interventions and the costs of alternative strategies are important considerations in formulating guideline recommendations. Absolute costs and the importance of costs relative to health care outcomes constitute some of the underlying assumptions and values upon which recommendations are based. Economic analysis of a recommendation should consider the human, financial, and material resources required for its implementation and the potential savings if implementation occurs. Potential savings include reduction in morbidity and decreased needs for further treatments, diagnostic studies, or hospitalizations (5). Methods for performing economic analyses have been reviewed in several sources (58-61). These economic considerations should be explicitly stated within guideline documents.

Recommendations: Statements that guide clinical practice should review present and future health care costs attached to guideline recommendations and alternative choices. Standard terms for health care economics should be used for cost information to enhance clarity whenever possible.

MEDICAL-LEGAL CONSIDERATIONS

Although the promulgation of clinical practice guidelines was predicted to initiate a flood of litigation (62), so far it has not had a major impact on court decisions deciding whether physicians have violated the standard of care. A recent survey showed that guidelines play a central role in determinations of negligence in only 6.6% of actions (63). Guidelines may be used both for inculpatory and exculpatory purposes (63). United States court rulings have placed variable importance on guidelines, considering them either optional for consideration, of possible significance in determining the standard of care, or the foundation of care standards (64). The American Medical Association and the Institute of Medicine promote guideline use as evidence of the standard of care in preference to their use as mandatory and inflexible standards (65, 66). In Great Britain, published guidelines are considered hearsay evidence because they cannot be cross-examined as can a medical expert (67). The Canadian Medical Association in collaboration with the National Partnership for Quality in Health Care commissioned a study into the legal aspects of guidelines (68). This study emphasized the importance of flexibility in clinical practice guidelines that always keeps in mind the uniqueness of each patient's condition, environment, and preferences. They concluded that if guidelines are to be flexible and responsive to physician and patient judgment, they cannot be held up as absolute standards of care (68, 69).

To date, courts in the United States have critically analyzed guidelines and have applied them only so far as the critical analysis supports their authoritative nature and representation of customary clinical practice. Lawyers introduce guidelines as evidence for proof of what the standard of care should be in a given case, and opposing lawyers introduce evidence from multiple sources in rebuttal to indicate that the selected guideline is not applicable to the clinical situation. Published guidelines have not short-circuited the judicial process and do not decide in themselves what is reasonable and proper care. They serve as just another source of evidence supporting the standard of practice akin to textbooks, medical journals, and expert opinion (70). This judicial use of guidelines, which is balanced with a preserved importance of other sources of evidence for determining the standard of care, deemphasizes the need for physicians to be aware of all published guidelines in

their field of practice to protect themselves from malpractice suits.

Courts in the United States have ruled that guideline developers may be held liable for poor outcomes consequent to practitioners following faulty guidelines (71). Some professional organizations have incorporated disclaimers in their documents emphasizing that guidelines cannot apply to all patients at all times in all practice settings (72). Practitioners, however, cannot void their medicolegal liability by blaming faulty guidelines for clouding their clinical judgment (73). The joint liability of guideline developers and practitioners who follow published guidelines underscore the importance of ensuring the validity of statements that guide clinical practice. Valid guidelines may have a particular value in providing practice norms that serve as preliminary benchmarks for screening suits and decreasing frivolous claims (68).

Some state jurisdictions have created statutes that impose guidelines as a method for promoting malpractice reform (74). Physicians who follow guideline recommendations within these jurisdictions could not be held liable for any negative outcome. Some observers have criticized this use of guidelines in that guideline recommendations represent general statements regarding appropriate care and cannot address or anticipate the minute specifics of every clinical situation (75). Because of the needed flexibility of guidelines and the uniqueness of each patient's clinical circumstances, it is doubtful that clinical practice guidelines will ever become paramount in establishing the standard of care in broad jurisdictions.

Recommendations: Guideline committees should consider the medical-legal implications of their recommendations. Incorporation of sound methodologies in guideline development is the best safeguard for ensuring their validity and limiting the likelihood of disseminating flawed guidelines with medical-legal risk. Readers of guidelines should recognize that these documents are not developed to establish the legal standard of care. Guidelines can only represent the legal standard through legislation or through acceptance by the medical community as the standard of professional medical behavior (69).

GUIDELINE IMPLEMENTATION, DISSEMINATION, AND DIFFUSION

Practice guidelines are designed to expand clinical knowledge, change clinicians' attitudes, alter practice behavior, and improve health care-related outcomes (15). Studies give variable reports on the ability of published guidelines to achieve these goals (3, 56, 76-82). These studies emphasize that practice guidelines are not panaceas in themselves for health care problems and require supplementation with other interventions that alter clinical practice (15, 83, 84). Emerging data demonstrate that combining practice guidelines with computer systems, feedback, audits, reminders, continuing medical education, and financial incentives alter practice behavior and improve clinical outcome (15). Central to a guideline's success in altering physician practice patterns is the observation that the implementation strategy is as important as the strength of the guideline's recommendations; multiple approaches may be needed to effect changes in practice behavior (85, 86).

Fundamental to ensuring the effectiveness of a guideline is a comprehensive dissemination strategy to promote its wide diffusion and implementation into clinical practice. Depending on the topic, targets for dissemination include physicians and other health care workers, professional associations, scientific groups, health industries such as managed care organizations, patient associations, health administrations, funding

bodies, universities, and other training centers. Vehicles for dissemination vary depending on the document's purposes but may include publication in the *American Journal of Respiratory and Critical Care Medicine*, joint publication in other journals, publication as journal supplements, targeted mailings, and presentations at local, regional, national, and international meetings. Summary documents may be appropriate for dissemination in the form of lay information booklets, audio or video cassettes, organization web pages, CD-ROM, decisional aids, or continuing medical education materials. Surveys of practicing internists indicate that summaries of recommendations into brief, easily assimilated formats have the greatest likelihood of gaining physician acceptance (87).

Recommendations: Guideline developers should consider appropriate forums for dissemination, which may include the publication of quick reference guides, patient educational materials, media releases, online and electronic formats, booths at professional meetings, or presentations at the ATS annual scientific meeting. Recommendations for dissemination should be made to the ATS Board as a part of the final guideline report. The ATS recognizes that guideline development represents an arduous task and that "writing them is easier than making them work" (21). Suggestions within the document for performance measures related to the success of dissemination can assist in determining the document's impact. All ATS guidelines should be available in some form from the ATS worldwide web site. Every effort should be made, when appropriate, to create and distribute summary documents that communicate the most important components of the guidelines to practicing physicians.

CRITERIA FOR ASSESSMENT AND REFINEMENT OF ATS GUIDELINES

An ATS guideline is a beginning rather than an end to efforts for improving clinical practice in a specific area of patient care. The guideline document should include a description of structured performance measures that may be used to demonstrate the safety and effectiveness of the guideline's recommendations. These measures ultimately determine the validity of the document. Categories of performance measures include clinical aspects (symptom management, disease management, mortality, morbidity, length of hospital stay) and quality of life (general health perception, well-being, functional status). Assessment plans need not be elaborate or quantitative (26), but failure to identify a time for and manner of future assessment severely undercuts the efforts and expense expended on guideline development (9). Example measures may be before and after surveys of practice patterns or interviews of samples of professionals or patients affected by the guidelines. Some guidelines or components of guidelines may be tested in prospective clinical trials (79). Timelines for assessment will be different between guidelines. For instance, the evaluation of guidelines for cancer screening would require a longer timeline than guidelines for managing asthma exacerbations. Guides exist to help readers of published guidelines to determine the validity of their recommendations (88).

Local users of guidelines should consider testing them in their own practice settings to determine if the practice recommendations are transportable. Some decision aid guidelines shown to work in one setting have not worked in other settings (89, 90).

Recommendations: Performance measures by which the validity of a guideline will be determined should be stated in the document. All ATS guidelines should include a descrip-

tion of a structured validation process directed at the goals of the guideline.

CONCLUSIONS

Guideline developers increase the likelihood that their completed documents will be valid and successful in affecting clinical practice if they adhere to an explicit methodology. The marketplace for formal clinical practice guidelines and other statements that guide clinical practice has heightened this regard for methodological rigor. This rigor, however, preserves an emphasis on guideline flexibility; practice guidelines do not represent cookbook medicine but allow for and insist upon individualized practitioner treatment planning (91). Guidelines represent "reference points and not rigid criteria" (6). Published guidelines should be considered works in progress (92) rather than the final word on any subject except where supporting evidence for a recommendation is incontrovertible. Their development should emphasize as much what is not known on a subject as what recommendations can be supported by the existing evidence. In this way, ATS guidelines will not be overstated as dogma or undervalued as "cookbook" medicine. They serve to disseminate important advances in health care and establish a research agenda with goals of improving patient outcomes (92).

Adoption of all the comprehensive recommendations provided in this statement may not be achievable or appropriate for every ATS document. The scope and importance of a topic determines the human and financial resources that should be applied. The Clinical Practice Committee suggests, however, that these recommendations may serve as a contemporary review of issues pertaining to statements that guide clinical practice to assist ATS committees charged with creating these documents and ATS members in analyzing published guidelines from this and other organizations.

JOHN E. HEFFNER, M.D.

MOIRA AITKEN, M.D.

LOIS GEIST, M.D.

MOLLY OSBORNE, M.D.

YANCY PHILLIPS, M.D.

KINGMAN STROHL, M.D.

References

1. 1995. ATS statements, position papers, and conference reports. *ATS News* 21:1-6.
2. Woolf, S. H. 1990. Practice guidelines: a new reality in medicine, I: recent developments. *Arch. Intern. Med.* 150:1811-1818.
3. Lomas, J., G. M. Anderson, K. Donnick-Pierre, E. Vayda, M. W. Enkin, and W. J. Hannah. 1989. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N. Engl. J. Med.* 321:1306-1311.
4. Roper, W. L., W. Winkenwerder, G. M. Hackbarth, and H. Krakauer. 1988. Effectiveness in health care: an initiative to evaluate and improve medical practice. *N. Engl. J. Med.* 319:1197-1202.
5. Roche, N., and P. Durieux. 1994. Clinical practice guidelines: from methodological to practical issues. *Intensive Care Med.* 20:593-601.
6. Eddy, D. M. 1990. Practice policies-what are they? *J.A.M.A.* 263:877-880.
7. Pozen, M. W., R. B. D'Agostino, H. P. Selker, P. A. Sytkowski, and W. B. Hood. 1984. A predictive instrument to improve coronary-care unit admission practices in acute ischemic heart disease. *N. Engl. J. Med.* 310:1273-1278.
8. Chassin, M. R., J. Kosecoff, R. E. Park, C. M. Winlow, K. L. Kahn, N. J. Merrick, J. Keesey, A. Fink, D. H. Solomon, and R. H. Brook. 1987. Does appropriate use explain geographic variations in the use of health care services? A study of three procedures. *J.A.M.A.* 258:2533-2537.

9. Field, M. J., and K. N. Lohr. 1990. Clinical Practice Guidelines. National Academy Press, Washington, DC.
10. Garnick, D. W., A. M. Hendricks, and T. A. Brennan. 1991. Can practice guidelines reduce the number and costs of malpractice claims? *J.A.M.A.* 266:2856-2860.
11. Squires, B. P. 1991. Statements from professional associations, specialty groups, and consensus conferences: what editors expect. *Can. Med. Assoc. J.* 145:297-298.
12. Woolf, S. H. 1990. Manual for Clinical Practice Guideline Development: A Protocol for Expert Panels Convened by the Office of the Forum for Quality and Effectiveness in Health Care. U.S. Agency for Health Care Policy and Reform, Rockville, MD.
13. Eddy, D. 1982. Clinical policies and the quality of clinical practice. *N. Engl. J. Med.* 307:343-345.
14. Wilson, M. C., R. S. Hayward, S. R. Tunis, E. B. Bass, and G. Guyatt. 1995. User's 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 Evidence-Based Medicine Working Group. *J.A.M.A.* 274:1630-1632.
15. Woolf, S. H. 1993. Practice guidelines: a new reality in medicine: III. Impact on patient care. *Arch. Intern. Med.* 153:2646-2655.
16. Brook, R. H. 1996. Practice guidelines: to be or not to be. *Lancet* 348:1005-1006.
17. McDonald, R. C., and W. J. I. Martin. 1995. Health-care reform and pulmonary/critical care medicine. *Chest* 107:1190-1192.
18. U.S. Agency for Health Care Policy and Reform. 1993. Clinical Practice Guideline Development. U.S. Government Printing Office, Washington, DC. AHCPR Publication No. 93-0023.
19. Park, R. E., A. Fink, R. H. Brook, M. R. Chassin, K. L. Kahn, N. J. Merrick, J. Kosecoff, and D. H. Solomon. 1986. Physician ratings of appropriate indications for six medical and surgical procedures. *Am. J. Public Health* 76:766-772.
20. Audet, A. M., S. Greenfield, and M. Field. 1990. Medical practice guidelines: current activities and future directions. *Ann. Intern. Med.* 113:709-714.
21. Haines, A., and G. Feder. 1992. Guidance on guidelines: writing them is easier than making them work. *B.M.J.* 305:785-786.
22. Selker, H. P. 1993. Criteria for adoption in practice of medical practice guidelines (editorial). *Am. J. Cardiol.* 71:339-341.
23. McDonald, C. J., and J. M. Overhage. 1994. Guidelines you can follow and can trust: an ideal and an example. *J.A.M.A.* 271:872-873.
24. Evidence-Based Care Resource Group. 1994. Evidence-based care: I. Setting priorities: how important is this problem? *Can. Med. Assoc. J.* 150:1249-1254.
25. Kleindorfer, P. R., H. C. Kunreuther, and P. J. H. Schoemaker. 1993. Decision Sciences: An Integrative Perspective. Cambridge University Press, Cambridge, UK.
26. Berwick, D. M. 1996. Harvesting knowledge through improvement. *J.A.M.A.* 275:877-878.
27. Hillman, B. J. 1992. The consensus of committees. *Invest. Radiol.* 27:1.
28. Olson, C. M. 1995. Consensus statements: applying structure. *J.A.M.A.* 273:72-73.
29. Reiser, S. J. 1993. The era of the patient: using the experience of illness in shaping the missions of health care. *J.A.M.A.* 269:1012-1017.
30. Health Services Research Group. 1992. Studying patients' preferences in health care decision making. *Can. Med. Assoc. J.* 147:859-864.
31. Eddy, D. M. 1991. Rationing by patient choice. *J.A.M.A.* 265:105-108.
32. Hadorn, D. C., D. Baker, J. S. Hodges, and N. Hicks. 1996. Rating the quality of evidence for clinical practice guidelines. *J. Clin. Epidemiol.* 49:749-754.
33. Canadian Task Force on the Periodic Health Examination. 1979. The periodic health examination. *Can. Med. Assoc. J.* 121:1193-1254.
34. Woolf, S. H., R. N. Battista, G. M. Anderson, A. G. Logan, and E. Wang. 1990. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systemic methods in reviewing evidence and developing clinical practice recommendations. *J. Clin. Epidemiol.* 43:891-905.
35. Cook, D. J., G. H. Guyatt, A. Laupacis, D. L. Sackett, and R. J. Goldberg. 1995. Clinical recommendations using levels of evidence for antithrombotic agents. *Chest* 108(Suppl.):227S-230S.
36. Cook, D. J., G. H. Guyatt, A. Laupacis, and D. L. Sackett. 1992. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 102(Suppl. 4):305S-311S.
37. US. Preventive Services Task Force. 1989. Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions. Williams and Wilkins, Baltimore.
38. Battista, R. N., and S. W. Fletcher. 1988. Making recommendations on preventive practices: methodological issues. *Am. J. Prev. Med.* 4S:53-67.
39. Guyatt, G. H., D. L. Sackett, J. C. Sinclair, R. Hayward, D. J. Cook, and R. J. Cook. 1995. Users' guides to the medical literature: IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *J.A.M.A.* 274:1800-1804.
40. Kottke, T. E., and M. L. Brekke. 1990. Cholesterol policy. What should we do? How should we decide? *J. Clin. Epidemiol.* 43:1023-1027.
41. Sanazaro, P. J. 1976. Medical audit, continuing medical education, and high quality assurance. *West. J. Med.* 125:241-252.
42. Woolf, S. H. 1992. Practice guidelines, a new reality in medicine: II. Methods of developing guidelines. *Arch. Intern. Med.* 152:946-952.
43. Fink, A., J. Kosecoff, M. Chassin, and R. H. Brook. 1984. Consensus methods: characteristics and guidelines for use. *Am. J. Public Health* 74:979-983.
44. Perry, S. 1988. Consensus development: an historical note. *Int. J. Technol. Assess. Health Care* 4:481-484.
45. Eddy, D. M. 1990. Practice policies: where do they come from? *J.A.M.A.* 263:1265-1275.
46. Andreasen, P. B. 1988. Consensus conferences in different countries: aims and perspectives. *Int. Technol. Assess. Health Care* 4:305-308.
47. Office of Medical Applications of Research. 1988. Guidelines for the selection and management of consensus development conferences. *National Institute of Health* September. 1-10.
48. Brook, R. H., M. Chassin, A. Fink, and D. Solomon. 1991. A Method for the Detailed Assessment of the Appropriateness of Medical Technologies: A Rand Note. Rand, Santa Monica, CA.
49. Chassin, M. R. 1990. Practice guidelines: best hope for quality improvement in the 1990s. *J. Occup. Med.* 32:1199-1206.
50. Brook, R. H. 1989. Practice guidelines and practicing medicine: are they compatible? *JAMA* 262:3027-3030.
51. Eddy, D. M. 1990. Practice policies: guidelines for methods. *J.A.M.A.* 263:1839-1841.
52. Eddy, D. M. 1991. A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach. American College of Physicians, Philadelphia.
53. Scott, E. A., and N. Black. 1991. When does consensus exist in expert panels? *J. Public Health Med.* 13:35-39.
54. Tong, R. 1991. The epistemology and ethics of consensus: uses and misuses of 'ethical' expertise. *J. Med. Philos.* 16:409-426.
55. Moreno, J. D. 1991. Consensus, contracts, and committees. *J. Med. Philos.* 16:393-408.
56. Lomas, J. 1991. Words without action? The production, dissemination, and impact of consensus recommendations. *Annu. Rev. Public Health* 12:41-63.
57. Hayward, R. S. A., M. D. Wilson, S. R. Tunis, E. B. Bass, H. R. Rubin, and R. B. Haynes. 1993. More informative abstracts of articles on clinical practice guidelines. *Ann. Intern. Med.* 118:731-737.
58. Drummond, M., G. Stoddart, R. Labelle, and R. Cushman. 1987. Health economics: an introduction for clinicians. *Ann. Intern. Med.* 107:88-92.
59. Eddy, D. M. 1990. Connecting value and costs: whom do we ask, and what do we ask them? *J.A.M.A.* 264:1737-1739.
60. Eisenberg, J. M. 1989. A guide to the economic analysis of clinical practices. *J.A.M.A.* 262:459-471.
61. Guyatt, G., M. Drummond, D. Stoddart, B. Haynes, K. Bennett, R. Labelle, D. Feeny, and P. Tugwell. 1986. Guidelines for the clinical and economic evaluation of health care technologies. *Soc. Sci. Med.* 22:393-408.
62. Ellwood, P. M. 1988. Chattuck lecture: outcomes management, a technology of patient experience. *N. Engl. J. Med.* 318:1549-1556.
63. Hymans, A. L., J. A. Brandenburg, S. R. Lipsitz, D. W. Shapiro, and T. A. Brennan. 1995. Practice guidelines and malpractice litigation: a two-way street. *Ann. Intern. Med.* 122:450-455.
64. Hurwitz, B. 1995. Clinical guidelines and the law: advice, guidance, or regulation. *Journal of Evaluation in Clinical Practice* 1:49-60.
65. Hirshfield, E. 1993. Use of practice parameters as standards of care in health care reform: a view from the American Medical Association. *Journal of Quality Improvement* 19:322-329.
66. Institute of Medicine. 1992. Guidelines for Clinical Practice: From Development to Use. National Academy Press, Washington, DC.
67. Hurwitz, B. 1995. Clinical guidelines and the law: what is the legal status of guidelines? *B.M.J.* 311:1517-1518.
68. Jutras, D. 1993. Clinical practice guidelines as legal norms. *Can. Med. Assoc. J.* 148:905-908.
69. Wensel, R. H., A. O. Carter, D. Walters, and S. Beardall. 1996. Eying guidelines. *Can. Med. Assoc. J.* 155:1666-1668.

70. Hirshfeld, E. B. 1990. Practice parameters and the malpractice liability of physicians. *J.A.M.A.* 263:1556–1562.
71. Wilson v. Blue Cross of California. 1990. *California Reporter* 3rd Series 271:876–885.
72. Royal, H., R. J. Pierson, J. Fletcher, and G. Dillehay. 1996. Guidelines for guideline development. *J.Nucl. Med.* 37:878–881.
73. Wickline v. California. 1986. *California Reporter* 228:661–672.
74. Hirshfeld, E. B. 1991. Should practice parameters be the standard of care in malpractice litigation? *J.A.M.A.* 266:2886–2891.
75. Kluge, E. H. 1996. Clinical practice guidelines and the law. *Can. Med. Assoc. J.* 155:574–576.
76. Grimshaw, J. M., and I. T. Russell. 1993. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 342:1317–1322.
77. Grol, R. 1990. National standard setting for quality of care in general practice: attitudes of general practitioners and response to a set of standards. *Br. J. Gen. Pract.* 40:361–364.
78. Ellrodt, A. G., L. Conner, M. Riedinger, and S. Weingarten. 1995. Measuring and improving physician compliance with clinical practice guidelines: a controlled interventional trial. *Ann. Intern. Med.* 122:277–282.
79. Weingarten, S., L. Agocs, N. Tankel, A. Sheng, and A. G. Ellrodt. 1993. Reducing lengths of stay for patients hospitalized with chest pain using medical practice guidelines and opinion leaders. *Am. J. Cardiol.* 71:259–262.
80. Weingarten, S., B. Ermann, R. Bolus, M. S. Riedinger, H. Rubin, A. Green, K. Karns, and A. G. Ellrodt. 1990. Early ‘step-down’ transfer of low-risk patients with chest pain: a controlled interventional trial. *Ann. Intern. Med.* 113:283–289.
81. Eagle, K. A., A. G. Mulley, S. J. Skates, V. A. Reder, B. W. Nicholson, J. O. Sexton, G. O. Barnett, and G. E. Thibault. 1990. Length of stay in the intensive care unit: effects of practice guidelines and feedback. *J.A.M.A.* 264:992–997.
82. Feder, G., C. Griffiths, C. Highton, S. Eldridge, M. Spence, and L. Southgate. 1995. Do clinical guidelines introduced with practice based education improve care of asthmatics and diabetic patients? A randomised controlled trial in general practices in east London. *B.M.J.* 311:1473–1478.
83. Eisenberg, J. M., and S. V. Williams. 1981. Cost containment and changing physicians’ practice behavior: can the fox learn to guard the chicken coop? *J.A.M.A.* 246:2195–2201.
84. Greco, P. J., and J. M. Eisenberg. 1993. Changing physicians’ practices. *N.Engl. J. Med.* 329:1271–1274.
85. Rosner, F., J. A. Balint, and R. M. Stein. 1994. Remedial medical education. *Arch. Intern. Med.* 154:274–279.
86. Gallagher, R. E., P. M. Scalzi, C. Norris-Baker, and R. F. Bakemeier. 1986. Evaluating the cancer education program: examining a range of approaches. *J. Cancer Educ.* 1:141–151.
87. Hayward, R. S. A., M. C. Wilson, S. R. Tunis, G. H. Guyatt, K.-A. Moore, and E. B. Bass. 1996. Practice guidelines: what are internists looking for? *J. Gen. Intern. Med.* 11:176–178.
88. Hayward, R. S. A., M. C. Wilson, S. R. Tunis, E. B. Bass, and G. Guyatt. 1995. Users’ guides to the medical literature: VIII. How to use clinical practice guidelines, A: are the recommendations valid? *J.A.M.A.* 274:570–574.
89. Weingarten, S. R., B. Ermann, M. Riedinger, P. K. Shah, and A. G. Ellrodt. 1989. Selecting the best triage rule for patients hospitalized with chest pain. *Am. J. Med.* 87:494–500.
90. Young, M. J., L. F. J. McMahaon, and J. K. Stross. 1987. Prediction rules for patients with suspected myocardial infarction. *Arch. Intern. Med.* 147:1219–1222.
91. Kelly, J. T., and M. C. Toepp. 1992. Practice parameters: development, evaluation, dissemination, and implementation. *CRB Quality Review Bulletin* 1992:405–409.
92. Dans, P. E. 1994. Credibility, cookbook medicine, and common sense: guidelines and the college. *Ann. Intern. Med.* 120:966–968.

APPENDIX

Example Methodologies for Grading Evidence

- I. Grades of Recommendations for a Specified Level of Baseline Risk (39).
 - 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

Definition of abbreviations: RCT = randomized controlled trials; CI = confidence interval; NNT = number needed to treat.

- II. Level of evidence for guideline recommendations extracted from AHCPR-sponsored clinical practice guideline recommendations for patients with left ventricular dysfunction (32).
 - A. Supportive evidence from well-conducted randomized controlled trials that included 100 patients or more
 1. Evidence from a well-conducted multicenter trial
 2. Evidence from a meta-analysis that incorporated quality ratings in the analysis and included a total of 100 patients in its estimate of effect size and confidence intervals
 - B. Supportive evidence from well-conducted randomized controlled trials that included fewer than 100 patients
 1. Evidence from a well-conducted trial at one or more institutions
 2. Evidence from a meta-analysis that incorporated quality ratings in the analysis and included fewer than 100 patients in its estimate of effect size and confidence intervals
 - C. Supportive evidence from well-conducted cohort studies
 1. Evidence from a well-conducted prospective cohort study or registry
 2. Evidence from a well-conducted retrospective cohort study
 3. Evidence from a well-conducted meta-analysis of cohort studies
 - D. Supportive evidence from a well-conducted case-control study
 - E. Supportive evidence from poorly controlled or uncontrolled studies
 1. Evidence from randomized clinical trials with one or more major or three or more minor methodological flaws that could invalidate the results
 2. Evidence from observational studies with high potential for bias (such as a case series with comparison to historical controls)
 3. Evidence from case series or case reports
 - F. Conflicting evidence with the weight of evidence supporting the recommendation
 - G. Expert opinion

These 7 levels are collapsed into three levels: “A” evidence, “B” evidence, and “C” evidence. “A” evidence consists of levels 1-3 above; “B” evidence consists of levels 4-6; and “C” evidence consists of level 7.