
Criteria for Evaluating Treatment Guidelines

American Psychological Association

This document presents a set of criteria to be used in evaluating treatment guidelines that have been promulgated by health care organizations, government agencies, professional associations, or other entities.¹ Although originally developed for mental health interventions, the criteria presented are equally applicable in other health service areas.

Two factors prompted this effort by the American Psychological Association (APA) to create a policy basis for evaluating guidelines. First, guidelines of varying quality, from both public and private sources, have been proliferating. Second, the interest and expertise in methodological issues within the profession of psychology made it likely that APA could make a useful contribution to the evaluation of treatment guidelines.

Generally, health care guidelines are pronouncements, statements, or declarations that suggest or recommend specific professional behavior, endeavor, or conduct in the delivery of health care services. Guidelines are promulgated to encourage high quality care. Ideally, they are not promulgated as a means of establishing the identity of a particular professional group or specialty, nor are they used to exclude certain persons from practicing in a particular area.

There are two different types of health care guidelines: practice guidelines and treatment guidelines. Practice guidelines, which are not addressed in this document, consist of recommendations to professionals concerning their conduct and the issues to be considered in particular areas of clinical practice rather than on patient outcomes or recommendations for specific treatments or specific clinical procedures at the patient level. Treatment guidelines, which are the focus of this document, provide specific recommendations about treatments to be offered to patients. That is, treatment guidelines are patient directed or patient focused as opposed to practitioner focused, and they tend to be condition or treatment specific (e.g., pediatric immunizations, mammography, depression).

The purpose of treatment guidelines is to educate health care professionals² and health care systems about the most effective treatments available. When there is sufficient information and the guidelines are done well, they can be a powerful way to help translate the current body of knowledge into actual clinical practice.

Many treatment guidelines are disorder based. The most common classification system is the *International Classification of Diseases (ICD-10; World Health Organization, 1992)* and, for mental disorders, the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994)*. The disorder-based approach has limitations: Patients³ commonly present issues

This document was approved as policy of the APA in August 2000. This document replaces as policy and is in part a revision of an earlier document, the *Template for Developing Guidelines: Interventions for Mental Disorders and Psychosocial Aspects of Physical Disorders* (APA, 1995), approved by the APA Council of Representatives in February 1995. The 1995 document was developed by the Task Force on Psychological Intervention Guidelines, which represented a collaborative effort of the APA Board of Professional Affairs (BPA), the APA Board of Scientific Affairs (BSA), and the APA Committee for the Advancement of Professional Practice (CAPP). The task force included David Barlow, chair; Susan Mineka, co-vice chair; Elizabeth Robinson, co-vice chair; Daniel J. Abrahamson; Sol Garfield; Mark S. Goldman; Steven D. Hollon; and George Stricker.

At its August 1997 meeting, the APA Council of Representatives approved a motion requesting that within a three-year time frame, BPA conduct a comprehensive evaluation of the *Template for Developing Guidelines* (APA, 1995) and the experience with its implementation. This motion also requested that BPA recommend revisions to the document based on this evaluation. The *Template* Implementation Work Group—a continuing collaboration among BPA, BSA, and CAPP—was charged with this task. The work group included Daniel J. Abrahamson, chair; Nancy C. Bologna; Steven D. Hollon; Ivan J. Miller; Elizabeth Robinson; and George Stricker. The work group's efforts were informed by extensive commentary from a wide range of reviewers, both within and outside APA governance.

We in the *Template* Implementation Work Group thank three chairs of BPA who provided encouragement, support, and detailed input over the course of the revisions: Robert A. Brown (1998), Ronald H. Rozensky (1999), and Suzanne Bennett Johnson (2000). Under the leadership of Russ Newman, APA Practice Directorate staff members Christopher J. McLaughlin, Georgia Sargeant, and Robert W. Walsh provided the horsepower needed to steer this endeavor through multiple revisions and logistical roadblocks. Finally, the work group expresses its deepest appreciation to Geoffrey M. Reed, without whose inspiration, intellectual challenge, sense of humor, and true leadership we could not have sustained this effort.

A checklist for use in applying the criteria contained in this document to the evaluation of guidelines is available online at <http://www.apa.org/practice/guidelines/treatcrit.html>.

Correspondence concerning this article should be addressed to Geoffrey M. Reed, Practice Directorate, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.

¹ It is important to recognize that the term *guidelines* generally refers to pronouncements that support or recommend but do not mandate specific approaches or actions. In this regard, guidelines differ from what are sometimes called *standards* in that standards are considered mandatory and may be accompanied by an enforcement mechanism. The *Criteria for Evaluating Treatment Guidelines* should be regarded as guidelines, which means that it is essentially aspirational in intent. It is intended to facilitate and assist the evaluation of treatment guidelines but is not intended to be mandatory, exhaustive, or definitive and may not be applicable to every situation.

² We have chosen to use the term *health care professional*, shortened at times to *professional*, to refer to the trained and legally authorized person who delivers health care services.

³ To be consistent with the context in which most guidelines are applied, we use the term *patient* to refer to the individual (child or adult), couple, family, or group receiving treatment. However, we also recognize that in many situations, there are important and valid arguments for using terms such as *client*, *consumer*, or *person* in place of *patient* to describe the recipient of services.

that cut across diagnostic lines, dual diagnoses (comorbidity) are common, and disorder-based diagnosis is often a weak basis for determining appropriate levels of care and other characteristics of treatment. Other classification systems, such as the World Health Organization's functionally based *International Classification of Functioning, Disability, and Health* (World Health Organization, 2001), might also provide a basis for the development of treatment guidelines. It is important for groups constructing or evaluating guidelines to consider the adequacy and limitations of the nosological systems on which they are based.

Health care professionals are in the best position to be aware of the unique characteristics of individual patients. The treatment strategy most likely to succeed usually combines the most effective specific interventions with a strong therapeutic relationship and a mutual expectation of and framework for improvement. Such factors, which are common to most treatment situations, can be powerful determinants of treatment success. Good guidelines allow for flexibility in treatment selection so as to maximize the range of choices among effective treatment alternatives. The judgment of health care professionals, although always needed, is particularly important in the treatment of conditions for which research data are limited. Guideline panels should take these factors into consideration and particularly should avoid encouraging an overly mechanistic approach that could undermine the treatment relationship.

It is often assumed that the use of treatment guidelines will significantly reduce the cost of services. This is not necessarily true. It is possible that guideline implementation may cause some services to be discontinued because of evidence documenting an intervention's lack of efficacy. However, it is also possible that the adoption of guidelines will lead to a shift toward more effective but not necessarily less costly services. And it is possible that more costly or additional treatments will be recommended.

Another common assumption is that standardizing treatment via guidelines will always be beneficial because it reduces practice variation. However, variation in clinical practice is often based on the needs of individual patients and their responses to specific treatments. When the application of guidelines results in a rigid system that eliminates the ability to respond to individual needs of the patient and the opportunity for self-correction in treatment, this can be detrimental to patient care.

In this document, it is not presumed that guidelines are inherently either beneficial or detrimental, and the document is not intended either to encourage or to discourage their development. However, the burden of proof remains on the makers of each guideline and those responsible for its implementation to establish that the application of the guideline is indeed beneficial and does not impair patient care.

The purpose of this document is to provide criteria to assist in the determination of the strengths and weaknesses of each guideline. These criteria are intended to provide structure and guidance for those individuals or groups that evaluate the quality and appropriateness of treatment guidelines. Each criterion describes an important issue that

guideline makers should aspire to address in the best possible manner.

The primary purpose of this document is to assist in the evaluation of treatment guidelines. Although it will be helpful to those wishing to construct treatment guidelines, it does not provide sufficient specificity to serve as the sole basis for such efforts. It is not intended to promote the application of a particular set of treatment techniques or approaches. Finally, this document is not intended to imply that the treatments provided by individual practitioners should be subject to the evaluative process described here for assessing treatment guidelines.

Treatment guidelines have the potential to influence the health care of many patients, and therefore the guidelines and the process used in their development should be open to public scrutiny. Moreover, failure to disclose the scientific justification for a guideline violates a basic principle of science, which requires open scrutiny and debate. Without the disclosure of adequate scientific information, guidelines are mere expressions of opinion.

This document is organized on the basis of two related dimensions for the evaluation of guidelines. The first dimension is *treatment efficacy*, the systematic and scientific evaluation of whether a treatment works. The second dimension is *clinical utility*, the applicability, feasibility, and usefulness of the intervention in the local or specific setting where it is to be offered. This dimension also includes determination of the generalizability of an intervention whose efficacy has been established. To encourage accountability, criteria for evaluating the process of guideline production are also provided.

Treatment Efficacy

This dimension asks the question, How well does the intervention work? and reflects information and data collected in the course of systematically evaluating the efficacy of a particular intervention. The term *treatment efficacy* refers to a valid ascertainment of the effects of a given intervention as compared with an alternative intervention or with no treatment, in a controlled clinical context. The fundamental question in evaluating efficacy is whether a beneficial effect of treatment can be demonstrated scientifically.

Methods for evaluating efficacy often begin with health care professionals' judgments and then progress through more highly systematized research strategies. For some treatments, the most accessible source of information on treatment efficacy may be the judgment of health care professionals and patients who have experience with the treatments. It is important to distinguish between the context of discovery of an intervention and the context of verification of its clinical efficacy. Historically, some interventions that were later proven by systematic evaluation to be very powerful have arisen from clinical innovations and case studies. The question of whether particular interventions have beneficial effects is best answered using research methodologies that have been refined over many years to reduce the uncertainties inherent in subjective judgment alone and to increase confidence in the strength

of the intervention. The systematic application of these research strategies also promotes the welfare of patients.

Without evidence of efficacy, health care professionals are forced to rely exclusively on their direct experience of the effects of different interventions—an approach that risks erroneous conclusions. For example, in some cases, disorders resolve themselves without formal treatment; an intervention that had coincidentally been applied in such a case might erroneously be judged effective. Some medical and psychological practices that initially appeared helpful and became widely accepted were subsequently found ineffective or even harmful. One purpose of this document is to provide a strategy for evaluating the level of confidence to be placed in judgments about the relative efficacy of different interventions.

Criterion 1.0 Guidelines should be based on broad and careful consideration of the relevant empirical literature.

Evaluation is necessary, regardless of the theoretical derivation of the intervention. Individual studies should be evaluated on the logic of their experimental design. Adequate studies may be compiled using qualitative approaches or quantitative methods such as meta-analysis. When guidelines are based in part on compilations of studies, both the analyses and the individual studies on which they are based should be examined carefully, and alternative hypotheses should be explored.

Criterion 2.0 Recommendations on specific interventions should take into consideration the level of methodological rigor and clinical sophistication of the research supporting the intervention.

Not all studies of a given intervention are equal: Some methodologies involve more stringent tests of internal validity and therefore are more persuasive arguments for efficacy. Guidelines should take into consideration information from the sources identified in Criteria 2.1, 2.2, and 2.3 (below), which are listed in ascending order as to their contribution to internal validity.

Criterion 2.1 Guidelines consider clinical opinion, observation, and consensus among recognized experts representing the range of views in the field. The efficacy of interventions included in treatment guidelines can be supported by multiple observations by trained, knowledgeable, and experienced individuals. Consensus, by which we mean agreement among recognized experts in a particular area, can always add information. As the sole basis for conclusions about efficacy, consensus is more compelling than individual observation but less compelling than carefully controlled empirical evaluation. For very infrequent behaviors and rare conditions, clinical consensus on appropriate treatment may be the only available data.

Criterion 2.2 Systematized clinical observation is weighted more heavily than unsystematized observation in evaluating treatment efficacy. Systematized observation has two advantages: (a) The intervention is generally applied in a naturalistic practice setting and (b) the evaluation typically includes examination of qualitative

data. These clinical observations may then form the basis for further systematic evaluation. Appropriate methodologies may include systematized clinical case studies and clinical replication series, in which the clinical efficacy of an intervention is examined with a series of diverse patients who have a given disorder.

Criterion 2.3 The evaluation of treatment efficacy places greatest emphasis on evidence derived from sophisticated empirical methodologies, including quasi experiments and randomized controlled experiments or their logical equivalents. Quasi experiments do not involve randomization but include other controls that are designed to rule out some threats to the internal validity of inferences regarding treatment efficacy. Some single-subject designs also include such controls. Randomized controlled experiments represent a more stringent way to evaluate treatment efficacy because they are the most effective way to rule out threats to internal validity in a single experiment. Random assignment of patients to conditions reduces the likelihood that the groups differ before treatment with respect to characteristics that could influence subsequent status. The advantage of randomized clinical trials is their ability to rule out rival plausible alternatives to the notion that the treatment produced an effect. However, they are potentially subject to several threats to their external and construct validity, some of which are described later in this document. Randomized controlled experiments are definitive only when all aspects of the experimental design, including the participant population, are fully representative of the phenomena of interest.

Criterion 3.0 Recommendations on specific interventions should take into consideration the treatment conditions to which the intervention has been compared.

Guidelines should take into consideration the nature of the comparisons identified in Criteria 3.1, 3.2, and 3.3 (below), which are listed in ascending order as to their contribution to the strength of a recommendation.

Criterion 3.1 Guidelines consider whether the treatment gets better results than doing nothing. It is often difficult to operationalize “doing nothing,” so assessment-only or wait-list controls are typically used, despite their inherent limitations. Comparing a treatment with non-treatment allows the determination not only of whether an intervention has any efficacy at all but also of whether it has adverse effects. This determination is often an important part of the treatment evaluation process.

Criterion 3.2 Guidelines consider whether the intervention offers the patient any benefit beyond simply being in treatment. Positive results of treatment may be due to such factors as the quality of the treatment relationship and the health care professional’s ability to create a mutual framework for change. The usual strategy in evaluating a psychological intervention involves creating a credible comparison treatment appropriate to the clinical trial, such as the provision of a caring relationship. Similarly, the provision of a placebo in a test of the efficacy of a pharmacological agent duplicates all the aspects of the medication regime except the medication itself. Both these

strategies have their strengths and weaknesses, and the results must be examined carefully.

Criterion 3.3 Guidelines consider whether an intervention's results are better than the results of other interventions. The strongest recommendations are based on demonstrations that the treatment under consideration is more effective than alternative interventions that are known or believed to be effective.

Criterion 4.0 Guidelines should consider available evidence regarding patient–treatment matching.

Some individuals with a given problem may respond better to certain treatments than to others, whereas a different patient with the same problem may show a different pattern of response. Patient–treatment matching may maximize efficacy.

Criterion 5.0 Guidelines should specify the outcomes the intervention is intended to produce, and evidence should be provided for each outcome.

In examining the outcomes assessed in efficacy studies, guideline makers are encouraged to attend to the following important issues:

1. *Participant selection.* It is important to consider the method and rationale for selecting participants and how closely the resulting sample represents the population and phenomena of interest.

2. *Treatment goals.* Different parties in an intervention may have different goals for treatment. For example, clinical practitioners, clinical scientists, patients, family members, purchasers, and third-party payors may each value different results.

3. *Quality of life, life functioning.* Outcomes evaluated in efficacy studies should ideally include valid measures of life functioning such as social and occupational functioning, family or couple functioning, subjective well-being, and freedom from symptoms.

4. *Attrition.* In evaluating treatment outcomes, panels should consider attrition due to dropout or refusal. Attrition can seriously undermine the internal validity of a study, compromising the equivalence of groups initially created by randomization and leading to experimental results that are confounded by individual differences. Additionally, a study's loss of a substantial number of patients, through either refusal or dropout, seriously compromises the ability to generalize from the study to other clinical settings.

5. *Long-term consequences of treatment.* Some interventions hold up better than others over time. All things being equal, treatments that have enduring effects following termination are to be preferred over those that do not.

6. *Indirect consequences of treatment.* In addition to direct consequences of treatment such as symptom reduction or disease prevention, treatments may have indirect consequences as well. For example, a corrective surgical procedure may enhance self-esteem and improve social functioning, or the choice of a behavioral rather than a pharmacological treatment may enhance feelings of per-

sonal control. Guidelines should take available data regarding such indirect consequences into account.

7. *Patient satisfaction with treatment.* Patients' subjective evaluation of treatment and its results is important in evaluating treatment outcome, even though it may not be strongly correlated with clinical improvement.

8. *Iatrogenic negative effects or side effects of treatment.* Thorough outcome evaluation not only considers potential benefits but also examines possible side effects or negative outcomes associated with treatment.

9. *Clinical significance.* Ideally, outcome descriptions should specify clinical significance (i.e., actual clinical benefit) in addition to reporting any statistical significance. The full range of responses to the intervention should be reported, including such outcomes as (a) functioning within normal limits, (b) much improved but not functioning within normal limits, (c) improved, (d) no change, and (e) deterioration. The mandate for a particular intervention is enhanced if it normalizes functioning.

10. *Methods.* Ideally, outcomes should be assessed using converging methods of measurement and sources of information.

11. *Treatment goals.* The outcomes selected should be consistent with the goals and orientation of the treatment.

Summary Comments and Cautionary Notes on Treatment Efficacy

Although randomized clinical experiments can make an important contribution to the evidentiary base for treatment guidelines, a single experiment from one setting does not provide sufficient evidence of efficacy. Replication across multiple studies and multiple settings is desirable.

Moreover, some questions are easier than others to address in controlled clinical experiments. For example, short-term, problem-focused treatments lend themselves more readily to controlled experimentation than do longer-term interventions aimed at more multifaceted concerns. Consequently, the extent of the available scientific literature may vary depending upon the ease with which the intervention can be tested using controlled clinical trials. Easily researched questions may have more literature supporting them than hard-to-research questions do. Paucity of literature does not necessarily imply that an intervention is ineffective.

Furthermore, the aggregate data produced by controlled trials do not necessarily predict individual responses. Even the most effective treatments do not work with every patient. In addition, some patients may respond best to a treatment that is not effective with the majority. Therefore, good treatment guidelines allow for some flexibility in treatment selection to accommodate individual responses.

Finally, any study is the product of many subjective judgments concerning whom to treat, how to treat them, and how to measure change. Each of these decisions can affect the study's construct validity—the extent to which the experiment truly addresses the underlying clinical question. As a consequence, even a treatment that is well supported in randomized controlled experiments may turn

out to be of little value clinically if those studies have poor external validity. Panels have a fundamental responsibility to evaluate all these considerations when developing treatment guidelines.

Clinical Utility

Clinical utility is the second dimension to be considered in evaluating treatment guidelines. Important components of this dimension include the generalizability of the intervention across settings and the feasibility of implementing the intervention with various types of patients and in various settings. The costs associated with the administration of the intervention may also be considered.

The clinical utility dimension addresses (a) the ability of health care professionals to use and of patients to accept the treatment under consideration and (b) the range of applicability of that treatment. This dimension reflects the extent to which the intervention will be effective in the practice setting where it is to be applied, regardless of the efficacy that may have been demonstrated in the clinical research setting.

The evaluation of clinical utility involves the assessment of interventions as they are delivered in real-world clinical settings. Many aspects of clinical utility are themselves increasingly the focus of systematic evaluation and even controlled experimentation.

Generalizability

The term *generalizability* refers to the extent to which an effect of a treatment is robust and therefore will be replicated even when details of the context are altered. Relevant factors include patients' characteristics, health care professionals' characteristics, variations across settings, and the interactions among these factors.

Criterion 6.0 *Guidelines should reflect the breadth of patient variables that may influence the clinical utility of the intervention.*

Factors such as age, gender, language, and ethnicity can all affect treatment outcomes. These factors may or may not have been assessed in the outcome literature for the treatment under consideration. To the extent possible, guidelines take into account the appropriateness of the treatment for patients characterized by each of the factors considered in Criteria 6.1 through 6.5 (below).

Criterion 6.1 **Guidelines take into account the complexity and idiosyncrasy of patients' clinical presentations, including severity, comorbidity, and external stressors.** Many patients evidence a variety of problems. For example, a depressed patient may also suffer from substance abuse and marital dysfunction, or a patient diagnosed with cancer may experience depression and social isolation. Successful treatment of the individual may well require attention to each problem. Good guidelines provide for the treatment of patients as they present themselves in real-world settings.

Criterion 6.2 **Guidelines take into consideration culturally relevant research and expertise.** Interventions that are of demonstrable efficacy with one ethnic, cultural,

or linguistic group may not be equally applicable to patients from other groups. In the absence of relevant research, panels should be cautious about generalizing to patients with varied cultural backgrounds. Good guidelines comment on evidence for the applicability of the treatment to different cultural groups.

Criterion 6.3 **Guidelines take into consideration research addressing the issue of the patient's gender (a social characteristic) and sex (a biological characteristic).** Interventions that are of demonstrable efficacy with male patients may not be applicable to female patients and vice versa. Good guidelines comment on whether there is evidence for the applicability of the treatment to both men and women.

Criterion 6.4 **Guidelines take into account research and relevant clinical consensus concerning the age and developmental level of the patient.** Interventions that are of demonstrable efficacy with middle-aged patients may not be equally applicable for children or geriatric patients. Good guidelines comment on evidence for the applicability of the treatment for different age groups.

Criterion 6.5 **It is recommended that guidelines take into account research and clinical consensus on other relevant patient characteristics.** Patient characteristics including but not limited to socioeconomic status, religion, language, sexual orientation, and physical condition may play important roles in determining the clinical utility of a particular intervention for that patient. Good guidelines comment on evidence for the applicability of the treatment to individuals with differing characteristics that are relevant to the success of the intervention.

Criterion 7.0 *It is recommended that guidelines take into account data on how differences between individual health care professionals may affect the efficacy of the treatment.*

Such factors as the professional's skill, experience, gender, language, and ethnic background can affect outcome in ways that are only partly understood.

Criterion 7.1 **It is recommended that guidelines take into account the effect of the health care professional's training, skill, and experience on treatment outcome.** The skill and experience levels of both the health care professionals who originally delivered the treatment and those now likely to deliver it are important factors. It is recommended that guidelines take into account whether the recommended treatment was originally implemented by health care professionals whose skill and experience were comparable to those for whom the guidelines are intended.

Criterion 7.2 **It is recommended that guidelines take into account the effects on treatment outcome of interactions between the patient's and the health care professional's characteristics, including but not limited to language, ethnicity, background, sex, and gender.** The effectiveness of an intervention may, but not necessarily, be affected by differences in backgrounds or ethnicities of the health care professional and the patient.

Criterion 8.0 *It is recommended that guidelines take into account information pertaining to the setting in which the treatment is offered.*

A treatment with proven effectiveness in one type of setting (e.g., the home, the school, day treatment, the clinic, the office, or the institution) may vary in effectiveness when it is offered in other settings. Good guidelines specify the settings in which the treatment has been documented to be effective.

Criterion 9.0 *Guidelines should take into account data on treatment robustness.*

A treatment's clinical utility may vary with alterations in administration. Data relevant to issues such as adherence to a protocol, differing time frames for delivering treatment, and differing modes of delivering treatment (e.g., individual treatment vs. group treatment) may influence components known to be critical to the treatment's effectiveness.

Feasibility

Feasibility refers to the extent to which a treatment can be delivered to patients in the actual setting. Feasibility evaluation addresses such factors as the acceptability of the intervention to potential patients, patients' ability and willingness to comply with the requirements of the intervention, the ease of dissemination of the intervention, and the ease of administration of the intervention. The examination of feasibility may also include consideration of the cost of the intervention.

Criterion 10.0 *Guidelines should take into account the intervention's level of acceptability to the patients who are to receive the service.*

There are many reasons why individual patients may prefer not to receive particular treatments, regardless of their demonstrated efficacy. These reasons may include such factors as pain, expense, duration, fear, side effects, adverse reactions, values, culture, and personal preferences.

Criterion 10.1 *Guidelines provide for informed patient choice among comparable interventions.* Good guidelines maximize patient choice among treatment alternatives. Patient choice may increase the clinical utility of a given intervention. Similarly, the unwillingness of a patient to accept a specific treatment may preclude its administration, regardless of its proven efficacy with other patients.

Criterion 10.2 *Guidelines consider patients' willingness and ability to participate in recommended interventions.* Some treatment interventions may require both in- and out-of-session activity on the part of the patient. If the patient is unwilling or unable to participate in treatment requirements, the intervention will not be effective. Sometimes patients do not adhere to treatment regimens because of negative side effects or concern about possible risks. Patients may also be unwilling or unable to self-monitor activities, engage in or sustain new behaviors, or take medications regularly.

Criterion 11.0 *Guidelines should explicitly note and evaluate possible adverse effects of interventions as well as their benefits.*

Treatments may have adverse effects. These should be explicitly documented and considered in the formulation of any guideline.

Criterion 12.0 *Guidelines should address the preparation of the health care professionals to deliver the intervention.*

Different interventions may require different levels of training and skill to achieve optimal effects. In a given setting, the clinical utility of a treatment may be reduced if too few sufficiently competent health care professionals are available. Guideline panels should consider what training is required of the health care professional and whether it is readily available. However, a current lack of sufficiently trained health care professionals or training opportunities should not lead a guideline panel to discount the utility of a promising treatment. It may also be helpful for guidelines to consider whether professionals might be reluctant to deliver an intervention because the cost of completing it exceeds the resources available, because the equipment is not available, or because it relies heavily on an incompatible treatment approach or theoretical orientation.

Consideration of Costs

Guidelines sometimes address the costs associated with treatment. When they do, costs need to be considered separately from effectiveness and determined broadly.

Criterion 13.0 *When guidelines include consideration of costs, it should be reported separately from consideration of effectiveness.*

Scientific and clinical evidence of the effectiveness of treatment and consideration of the costs of treatment are conceptually distinct. Conflating them compromises the scientific foundation of the guidelines. Any integration of cost and effectiveness must be open and explicit. Preserving this distinction is particularly important in discussions of medical necessity.

Criterion 14.0 *When guidelines consider costs, they should consider the direct, indirect, short-term, and long-term costs to the patient, to the professional, and to the health care system, as well as the costs associated with withholding treatment.*

Costs include such things as expense to the patient, expense to the health care professional, the cost of any technology or equipment involved in the intervention, and the cost of training the health care professional. Costs of withholding or delaying treatment may include the patient's loss of time from work and disability costs.

Cost savings associated with an intervention may include prevention of future disorders, as when an early intervention with a childhood disorder obviates the need for treatment later on. Savings may also accrue when an intervention makes other treatments unnecessary. For example, interventions such as smoking cessation programs and diabetic behavior management reduce the need for additional

medical treatment. Providing appropriate psychosocial services may also reduce medical visits in primary care.

Although costs can be partially reduced to monetary terms, some significant costs are not financial, including expenditure of patient time, suffering, and functional impairment. Good guidelines take nonmonetary costs into account.

The Guideline Development Process

The process by which guidelines are developed includes not only the deliberations of the guideline panel, discussed above, but also the steps by which the panel was formed and guideline development was undertaken. Review of guidelines is incomplete without review of the underlying development process.

Treatment guidelines directly impact the health and well-being of many consumers; therefore, panels must undertake guideline formation with careful deliberation. Care in constituting a balanced panel to develop guidelines can help ensure that panel members will be able to evaluate the relevant literature fully, assess standards of care, and weigh intervention costs and benefits fairly. Any potential conflicts of interest should be disclosed.

Criterion 15.0 It is recommended that guideline panels be composed of individuals with a broad range of documented expertise.

Criterion 15.1 It is recommended that guideline panels include one or more individuals with expertise in the delivery of services in the subject area or areas under consideration.

Criterion 15.2 It is recommended that guideline panels include one or more individuals with expertise in the scientific methodology of intervention evaluation in the diagnostic area or areas under consideration.

Criterion 15.3 It is recommended that guideline panels include representatives of the patient community (e.g., patients, advocates, and family members) who are familiar with the condition under consideration.

Criterion 15.4 It is recommended that guideline panels include experts from a broad range of relevant disciplines (e.g., the health care professions, health care economics, public health).

Criterion 15.5 It is recommended that guideline panels include members with expertise and sensitivity to relevant issues of diversity, including but not limited to sex, ethnicity, language, sexual orientation, age, and disability.

Criterion 16.0 Nominees for guideline panels and panel members should disclose potential, actual, and apparent conflicts of interest. Bodies with appropriate oversight authority should evaluate these conflicts of interest and take steps to eliminate or minimize them.

Criterion 17.0 Guideline panels should maintain the climate of openness and free exchange of views required by scientific objectivity.

Criterion 17.1 Selection criteria for guideline panelists, their qualifications for membership, and their

potential conflicts of interest should be described in the guidelines.

Criterion 17.2 It is recommended that guideline panel procedures and deliberations be made available for review by concerned parties.

Criterion 17.3 It is recommended that before being adopted, the guidelines be widely distributed to concerned parties, including consumer advocates and health care professionals. It is recommended that resulting comments be fully and fairly considered by the panel before it makes final recommendations and conclusions.

Criterion 17.4 It is recommended that a reference list of the information and documents reviewed in developing the guidelines be included with the guidelines or otherwise made available.

Criterion 17.5 It is recommended that when panel members disagree on the interpretation and significance of specific evidence, these disagreements be noted in the guidelines.

Criterion 18.0 It is recommended that guideline panels agree on specific goals for constructing the guidelines.

Criterion 18.1 It is recommended that guideline panels identify the audience for whom the guideline is intended. Some guidelines may be broadly targeted to health care professionals and consumers as well as to administrative agencies involved in the delivery of health care. Other guidelines may be more narrowly targeted. Determination of the target audience may affect the specific data-gathering strategies to be used in constructing the guideline, as well as the form and language of the final reports. Sets of guidelines may be published in multiple versions, each one suitable to the needs of the specific audience.

Criterion 18.2 Goals of guideline development other than improving patient care should be clearly identified in the guidelines. Institutional goals such as rationing services, minimizing legal liability exposure, and containing costs should be explicit.

Criterion 19.0 It is recommended that the guideline panel define the process and methods of guideline development as carefully as possible.

Criterion 19.1 It is recommended that the guideline panel specify the target condition or problems for the treatments under consideration.

Criterion 19.2 It is recommended that the guideline panel specify the patient population(s) for whom the treatments under consideration are intended.

Criterion 19.3 It is recommended that the guideline panel specify what clinical interventions will and will not be considered.

Criterion 19.4 It is recommended that the guideline panel specify the type of professional and the practice setting to which the guideline will be applicable.

Criterion 19.5 It is recommended that after selecting a general topic, the guideline panel decide on specific subsidiary goals around which literature reviews will be organized.

Criterion 20.0 *Guideline panels should specify the methods and strategies they have used for reviewing evidence. It is recommended that the information and documents reviewed be listed.*

The review process should be documented and organized so that both the process itself and the available evidence can be evaluated by others. The panel should consider the evidence appropriate to the treatments being examined.

Criterion 21.0 *It is recommended that guideline panels specify methods for evaluating the guidelines they produce.*

Criterion 21.1 **It is recommended that guideline panels make detailed recommendations to facilitate independent evaluation of the reliability of the guidelines they produce.** Ascertaining whether the guidelines are interpreted and applied consistently by health care professionals comprises one assessment of reliability. Typically, reliability would be approximated by independent review of the guidelines by alternative groups with equivalent expertise.

Criterion 21.2 **It is recommended that guideline panels make detailed recommendations to facilitate independent evaluation of the validity of the guidelines they produce.** The guidelines' validity may be evaluated retrospectively by independent consideration of the substance and quality of evidence cited, the methods chosen to evaluate the evidence, and the relationship between the evidence and the ultimate recommendations. Guidelines may be evaluated prospectively by determining whether they lead to better therapeutic outcomes in the target populations. Guideline panels and those responsible for convening them have the added responsibility of encouraging

the use of such criteria to evaluate the validity of the guidelines.

Criterion 21.3 **It is recommended that guideline panels make detailed recommendations to facilitate independent evaluation of the clinical utility of the guidelines they produce.** The clinical utility of a guideline may be evaluated through such mechanisms as examining the extent to which it leads to improved therapeutic outcomes in different populations and different settings. Clinical utility should also be evaluated through systematic feedback from health care professionals and perhaps from patients regarding their experiences related to the application of the guidelines.

Criterion 21.4 **It is recommended that guidelines be reviewed and revised periodically to ensure that they do not become obsolete. It is recommended that the panel specify a time frame for a revision of the guidelines.** Guidelines in treatment areas where knowledge is advancing quickly may need to be updated more frequently and thoroughly than those in other areas.

REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychological Association. (1995). *Template for developing guidelines: Interventions for mental disorders and psychosocial aspects of physical disorders*. Washington, DC: Author.
- World Health Organization. (1992). *International statistical classification of diseases and related health problems* (10th rev. ed.). Geneva, Switzerland: Author.
- World Health Organization. (2001). *International classification of functioning, disability, and health*. Geneva, Switzerland: Author.