

Practice Guideline Development, Grading, and Assessment

Joseph E. Cruz, PharmD, BCPS; Germin Fahim, PharmD, BCPS; and Kelly Moore, PharmD

ABSTRACT

Published clinical practice guidelines (CPGs) are abundant and more easily accessible than ever because of advances in technology and communication and the widespread use of electronic medical literature. To make educated therapy recommendations based on published guidelines, it is important for clinicians to understand the process of guideline development and to be familiar with the resources used to assess guidelines. To do this, clinicians must know the commonly used systems for grading guideline recommendations and for evaluating guideline quality. By increasing their understanding of guideline development, grading, and evaluation processes, clinicians can appropriately use guidelines and provide the highest level of patient care.

BACKGROUND

CPGs are recommendations based on a summary of current best evidence that are systematically developed to assist practitioners and to improve patient care.1 CPGs are used in evidence-based medicine to help synthesize clinical experience and the best current scientific data when creating individualized patient-care plans. Developers of CPGs include government agencies (e.g., the Centers for Disease Control and Prevention), professional societies (e.g., the Infectious Diseases Society of America and the American College of Clinical Pharmacy), managed care organizations, third-party payers, expert panels, and quality assurance organizations (e.g., the National Quality Forum).2 To ensure quality, guidelines must be developed in a systematic manner. As a result of the 2008 Medicare Improvements for Patients and Providers Act, the Institute of Medicine (IOM) published standards for guaranteeing CPG dependability. These standards include establishing transparency and evidence foundations for rating the strength of recommendations (Table 1).3 Other health care organizations and governing bodies have created “handbooks” to manage CPG development, but they all agree on the same core concepts laid out by the IOM, i.e., a multidisciplinary guideline-development group (GDG) should be created; consumers and patients should be involved and consulted; important clinical topics must be identified (often using the Patient–Intervention–Comparison–Outcome [PICO] model); systematic literature searches and syntheses must be performed; recommendations should be drafted using a structured evidence evaluation; and continued updates and revisions should be performed post-publication.24

Table 1 Summary of the Institute of Medicine CPG Development Standards

<table>
<thead>
<tr>
<th>Standard 1—Establishing Transparency</th>
<th>• Funding sources should be openly disclosed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard 2—Management of COI</td>
<td>• Individuals selected to the GDG should disclose active or planned COIs before joining.</td>
</tr>
<tr>
<td></td>
<td>• When possible, GDG members and their families should divest financial COIs.</td>
</tr>
<tr>
<td></td>
<td>• The number of GDG members with COIs should be kept to a minimum.</td>
</tr>
<tr>
<td>Standard 3—GDG Composition</td>
<td>• The GDG should comprise experts and clinical stakeholders with a multidisciplinary makeup.</td>
</tr>
<tr>
<td></td>
<td>• Patients, the public, and consumer organizations should participate in and have representation on GDGs.</td>
</tr>
<tr>
<td></td>
<td>• Patient and consumer participation in GDGs should be increased.</td>
</tr>
<tr>
<td>Standard 4—CPG: Systematic Review Intersection</td>
<td>• Systematic reviews used in CPG development should meet IOM standards.</td>
</tr>
<tr>
<td></td>
<td>• Systematic reviews conducted for CPG-specific use should have GDG involvement.</td>
</tr>
<tr>
<td>Standard 5—Establishing Evidence Foundations for and Rating Strength of Recommendations</td>
<td>• Benefits and harms should be offered for each recommendation.</td>
</tr>
<tr>
<td></td>
<td>• Ratings should be assigned for confidence and strength of recommendations.</td>
</tr>
<tr>
<td></td>
<td>• Contrary opinions within recommendations should be mentioned.</td>
</tr>
<tr>
<td>Standard 6—Articulation of Recommendations</td>
<td>• Standardized forms should be used to express recommendations.</td>
</tr>
<tr>
<td></td>
<td>• Recommendations should be evaluable for compliance.</td>
</tr>
<tr>
<td>Standard 7—External Review</td>
<td>• External reviewers should come from diverse backgrounds.</td>
</tr>
<tr>
<td></td>
<td>• External reviews should be confidential.</td>
</tr>
<tr>
<td></td>
<td>• All reviewer comments should be addressed.</td>
</tr>
<tr>
<td></td>
<td>• A draft should be made available to the public for comment before CPG publication.</td>
</tr>
<tr>
<td>Standard 8—Updating</td>
<td>• Publication, systematic review, and future revision dates should be stated.</td>
</tr>
<tr>
<td></td>
<td>• There should be an ongoing assessment of new evidence.</td>
</tr>
<tr>
<td></td>
<td>• Updates should occur when recommendations become clinically outdated.</td>
</tr>
</tbody>
</table>

Dr. Cruz is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy at Rutgers University in Piscataway, New Jersey. He is also a Clinical Coordinator–Internal Medicine at Englewood Hospital and Medical Center in Englewood, New Jersey. Dr. Fahim is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy and a Clinical Pharmacist–Internal Medicine at the Monmouth Medical Center in Long Branch, New Jersey. Dr. Moore is a Clinical Pharmacy Specialist at the Hospital for Special Surgery in New York, New York. COI = conflicts of interest; CPG = clinical practice guideline; GDG = guideline development group; IOM = Institute of Medicine.

Disclosure: The authors report no commercial or financial relationships in regard to this article.

Dr. Cruz is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy at Rutgers University in Piscataway, New Jersey. He is also a Clinical Coordinator–Internal Medicine at Englewood Hospital and Medical Center in Englewood, New Jersey. Dr. Fahim is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy and a Clinical Pharmacist–Internal Medicine at the Monmouth Medical Center in Long Branch, New Jersey. Dr. Moore is a Clinical Pharmacy Specialist at the Hospital for Special Surgery in New York, New York. COI = conflicts of interest; CPG = clinical practice guideline; GDG = guideline development group; IOM = Institute of Medicine.

Disclosure: The authors report no commercial or financial relationships in regard to this article.

Dr. Cruz is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy at Rutgers University in Piscataway, New Jersey. He is also a Clinical Coordinator–Internal Medicine at Englewood Hospital and Medical Center in Englewood, New Jersey. Dr. Fahim is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy and a Clinical Pharmacist–Internal Medicine at the Monmouth Medical Center in Long Branch, New Jersey. Dr. Moore is a Clinical Pharmacy Specialist at the Hospital for Special Surgery in New York, New York. COI = conflicts of interest; CPG = clinical practice guideline; GDG = guideline development group; IOM = Institute of Medicine.

Disclosure: The authors report no commercial or financial relationships in regard to this article.

Dr. Cruz is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy at Rutgers University in Piscataway, New Jersey. He is also a Clinical Coordinator–Internal Medicine at Englewood Hospital and Medical Center in Englewood, New Jersey. Dr. Fahim is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy and a Clinical Pharmacist–Internal Medicine at the Monmouth Medical Center in Long Branch, New Jersey. Dr. Moore is a Clinical Pharmacy Specialist at the Hospital for Special Surgery in New York, New York. COI = conflicts of interest; CPG = clinical practice guideline; GDG = guideline development group; IOM = Institute of Medicine.

Disclosure: The authors report no commercial or financial relationships in regard to this article.

Dr. Cruz is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy at Rutgers University in Piscataway, New Jersey. He is also a Clinical Coordinator–Internal Medicine at Englewood Hospital and Medical Center in Englewood, New Jersey. Dr. Fahim is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy and a Clinical Pharmacist–Internal Medicine at the Monmouth Medical Center in Long Branch, New Jersey. Dr. Moore is a Clinical Pharmacy Specialist at the Hospital for Special Surgery in New York, New York. COI = conflicts of interest; CPG = clinical practice guideline; GDG = guideline development group; IOM = Institute of Medicine.

Disclosure: The authors report no commercial or financial relationships in regard to this article.

Dr. Cruz is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy at Rutgers University in Piscataway, New Jersey. He is also a Clinical Coordinator–Internal Medicine at Englewood Hospital and Medical Center in Englewood, New Jersey. Dr. Fahim is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy and a Clinical Pharmacist–Internal Medicine at the Monmouth Medical Center in Long Branch, New Jersey. Dr. Moore is a Clinical Pharmacy Specialist at the Hospital for Special Surgery in New York, New York. COI = conflicts of interest; CPG = clinical practice guideline; GDG = guideline development group; IOM = Institute of Medicine.

Disclosure: The authors report no commercial or financial relationships in regard to this article.

Dr. Cruz is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy at Rutgers University in Piscataway, New Jersey. He is also a Clinical Coordinator–Internal Medicine at Englewood Hospital and Medical Center in Englewood, New Jersey. Dr. Fahim is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy and a Clinical Pharmacist–Internal Medicine at the Monmouth Medical Center in Long Branch, New Jersey. Dr. Moore is a Clinical Pharmacy Specialist at the Hospital for Special Surgery in New York, New York. COI = conflicts of interest; CPG = clinical practice guideline; GDG = guideline development group; IOM = Institute of Medicine.

Disclosure: The authors report no commercial or financial relationships in regard to this article.

Dr. Cruz is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy at Rutgers University in Piscataway, New Jersey. He is also a Clinical Coordinator–Internal Medicine at Englewood Hospital and Medical Center in Englewood, New Jersey. Dr. Fahim is a Clinical Assistant Professor at the Ernest Mario School of Pharmacy and a Clinical Pharmacist–Internal Medicine at the Monmouth Medical Center in Long Branch, New Jersey. Dr. Moore is a Clinical Pharmacy Specialist at the Hospital for Special Surgery in New York, New York. COI = conflicts of interest; CPG = clinical practice guideline; GDG = guideline development group; IOM = Institute of Medicine.

Disclosure: The authors report no commercial or financial relationships in regard to this article.
The widespread use of CPGs by health care practitioners may promote evidence-based clinical practice, improve the consistency of care, and minimize harm to patients. The CPG recommendations are graded using a standardized method of evaluation. The National Guideline Clearinghouse (NGC) lists the compatibility of CPGs with the IOM system and offers a concise summary of several CPGs. The mission of the NGC is to provide an accessible mechanism for obtaining objective information on CPGs and to disseminate this information for clinical use.

The American College of Cardiology and the American Heart Association (ACC/AHA) Task Force created an approach that uses the letters A, B, and C to indicate the quality of evidence for a given treatment. The letter A indicates that the data were derived from multiple randomized clinical trials or meta-analyses; the letter B indicates that the data were derived from one randomized trial or from nonrandomized studies; and the letter C indicates that the data were derived from expert opinions, case studies, or standards of care. These recommended classifications are then divided into levels I, II, and III. Level I indicates that a consensus based on clinical evidence and expert opinions has found that the treatment is useful and effective. Level II is applied when there is conflicting evidence or differences of opinion, and it is further divided into levels IIa (in favor of the treatment) and IIb (the evidence and/or opinions are less well established). Level C indicates that the treatment is not useful or effective, and that, in some cases, it may even be harmful.

The present review focuses on methods proposed by the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) working group, which set out to create more-direct language regarding the strength of recommendations (“strong” or “weak” instead of letters and numbers) and the quality of recommendations (high, moderate, or low). This system allows clinicians to evaluate more effectively the quality of clinical evidence and the applicability of current recommendations to the care of their patients.

Overreliance on CPGs can have drawbacks, however. Despite the convenience of the current technological age, with its rapid communication via the Internet and the popularity of social media, CPGs can still become outdated and obsolete virtually overnight as new data become available. In addition, rising drug-acquisition costs, continued drug shortages, and the ability of technology to easily integrate decision support can limit the use of medications recommended in CPGs. Occasionally, multiple sources may issue differing CPGs on the same therapeutic area. Moreover, clinicians might fear potential litigation or the loss of reimbursement from third-party payers if they fail to follow current CPGs.

The objectives of the present paper are to help clinicians interpret CPGs using the various evaluation methods that are available, and to offer guidance for P&T committee purposes.

GUIDELINE INTERPRETATION

Guideline developers use several systems to determine the quality of evidence and the strength of recommendations in the literature. However, because numerous systems are available, their recommendations can be inconsistent. This variability makes it difficult for clinicians to assess each individual guideline. The GRADE working group initially began as an informal collaboration in an effort to create a standardized method of rating evidence in the literature. In 2004, the group started formally developing “guidelines for the guidelines” used in health care, and these recommendations have become increasingly popular for their consistency. The advantages of the GRADE system over other available methods include guideline development by international experts; differentiation between the quality and strength of the evidence; and the provision of clear interpretations of recommendations for stakeholders (i.e., patients, clinicians, and policy-makers).

“Quality of evidence” refers to the level of confidence in results across outcome studies. The “strength of a recommendation” refers to the degree of confidence that the benefits of an intervention will outweigh its potential risks, based on the quality of evidence. To evaluate evidence from the literature, the GRADE reviewers considered the quality of the information (observational studies versus randomized controlled trials [RCTs]); the risk of bias (methods and execution); inconsistency (the size of the effect and statistical significance); indirectness (the similarity of the population, interventions, and outcomes to those of interest); imprecision (sample size and confidence intervals), and publication bias.

Assigning a rank to quality depends on the level of the available evidence. High, moderate, and low (or very low) levels of evidence reflect the likelihood that conclusions about the direction or magnitude of an effect could change as a result of further research. As noted previously, these levels of evidence are reported with the letter grades of A, B, and C, respectively, which can change based on the factors given above.

The strength of a recommendation usually takes into account values, preferences, and cost. A strong recommendation is stated as “we recommend;” a weak recommendation is stated as “we suggest;” and “no recommendation” is used when the evidence is insufficient. For example, a strong recommendation with high-quality evidence may state that a treatment’s desirable effects clearly outweigh its adverse effects, whereas a weak recommendation with moderate-quality evidence may state that a treatment’s desirable effects are offset by its undesirable effects.

GUIDELINE ASSESSMENT

The Appraisal of Guidelines for Research and Evaluation (AGREE) project sought to develop a standardized method for grading clinical practice guidelines. Since its initial publication in 2003, the AGREE tool has been used internationally and has been cited more than 100 times in the scientific literature. In 2010, the AGREE tool was replaced by AGREE II, and the latter was updated in 2013 to improve its reliability, validity, and usability.

AGREE II contains six quality domains with a total of 23 items, which are used to quantify the rating of a guideline. Each of the 23 items is assessed on a seven-point Likert scale ranging from 1 (strongly disagree; guideline does not state the necessary information) to 7 (strongly agree; guideline meets all conditions). The scores are calculated using the number of appraisers, the minimum potential score, the maximum potential score, and the observed score for all of the items.
Practice Guideline Development, Grading, and Assessment

Table 2 Summary of the AGREE II Tool

<table>
<thead>
<tr>
<th>Domain</th>
<th>Description</th>
</tr>
</thead>
</table>
| Domain 1—Scope and Purpose | - The overall objectives of the guideline are specifically described.  
- The health questions covered by the guideline are specifically described.  
- The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described. |
| Domain 2—Stakeholder Involvement | - The GDG includes individuals from all relevant professional groups.  
- The views and preferences of the target population (patients, public, etc.) have been sought.  
- The target users of the guideline are clearly defined. |
| Domain 3—Rigor of Development | - Systematic methods were used to search for evidence.  
- The criteria for selecting the evidence are clearly described.  
- The strengths and limitations of the body of evidence are clearly described.  
- The methods for formulating the recommendations are clearly described.  
- The health benefits, side effects, and risks have been considered in formulating the recommendations.  
- There is an explicit link between the recommendations and the supporting evidence.  
- The guideline has been externally reviewed by experts prior to its publication.  
- A procedure for updating the guideline is provided. |
| Domain 4—Clarity of Presentation | - The recommendations are specific and unambiguous.  
- The different options for management of the condition or health issue are clearly presented.  
- Key recommendations are easily identifiable. |
| Domain 5—Applicability | - The guideline describes facilitators and barriers to its application.  
- The guideline provides advice and/or tools on how the recommendations can be put into practice.  
- The potential resource implications of applying the recommendations have been considered.  
- The guideline presents monitoring and/or auditing criteria. |
| Domain 6—Editorial Independence | - The views of the funding body have not influenced the content of the guideline.  
- Competing interests of the GDG members have been recorded and addressed. |
| Global Assessment Questions | - Rate the overall quality of this guideline.  
  • Scale of 1 (lowest possible quality) to 7 (highest possible quality)  
  • I would recommend this guideline for use.  
    • Yes  
    • Yes, with edits  
    • No |

AGREE = Appraisal of Guidelines for Research and Evaluation; GDG = guideline development group.

in a given domain (Table 2). In addition, the overall quality rating is assessed, and the appraisers are asked whether they would recommend use of the guideline. Clinicians who wish to evaluate guidelines in a systematic manner are encouraged to consider using the AGREE II tool and to visit the AGREE Enterprise website at http://www.agreetrust.org.

IMPLICATIONS FOR P&T COMMITTEE MEMBERS

Formulary decision-making is guided by many factors, including operational considerations, practicality, relative costs, and clinical applicability of the current best evidence (i.e., a synthesis of CPGs, RCTs, meta-analyses, observational studies, and case reports). The AGREE II tool is most useful in situations where multiple guideline recommendations conflict or when recommendations differ significantly from accepted clinical practice. In such situations, it would be prudent for P&T committees to use the AGREE II tool to assess the quality of CPGs.

Since AGREE II works best when domain scores are compiled from multiple graders, a subcommittee could be appointed to appraise CPGs using AGREE II with the goal of validating recommendations that are discordant with individual committee members’ opinions or comparing CPGs when there are multiple CPGs on a similar topic. While this process may be difficult to implement initially because of the lack of established AGREE II score cut-offs that differentiate high- and low-quality CPGs, it could help guide decision-making when protocols are developed at the institutional level. Further, since domain scores must be compiled individually, P&T committees could set institution-specific cut-offs for each domain, or they could provide benchmarks based on previous experiences using the tool.

Likewise, using the GRADE method may help P&T committees classify evidence for or against a drug in a standardized way. GRADE scores the literature based on its relevance to the clinical question at hand. For example, a high-quality RCT may not receive a high score if the study did not address the guideline question adequately. Even if guidelines did not use the GRADE approach, P&T committees may employ that method to make their own inferences regarding items to be added or deleted from the formulary. This may be a burdensome chore for the entire P&T committee to undertake; therefore, the appointed subcommittee could be put in charge of spearheading this initiative as well.

CONCLUSION

Health care practitioners must critically evaluate CPGs in order to make well-informed decisions regarding treatment recommendations and formulary management. While CPGs have certain limitations, their use is expected to increase as technology makes them more accessible. Thus, it is crucial that clinicians understand GRADE rankings and use the AGREE II tool to aid in the critical evaluation of CPGs and to quantify quality metrics. The interpretation and assessment of guidelines using a fully informed, systematic approach can help reduce biases and misunderstandings concerning the available clinical evidence.
REFERENCES


