Methodologies for the Development of CHEST Guidelines and Expert Panel Reports

Sandra Zelman Lewis, PhD; Rebecca Diekemper, MPH; Joseph Ornelas, PhD; and Kenneth R. Casey, MD, MPH, FCCP

BACKGROUND: American College of Chest Physicians’ (CHEST) new Living Guidelines Model will not only provide clinicians with guidance based on the most clinically relevant and current science but will also allow expert-informed guidance to fill in any gaps in the existing evidence. These guidance documents will be updated, as necessary, using one or more of three processes: (1) evidence-based guidelines, (2) trustworthy consensus statements, and (3) a hybrid of the other two. The new Living Guidelines Model will be more sustainable and will encourage maintenance of current and targeted recommendations and suggestions.

METHODS: Over recent years, the Guidelines Oversight Committee (GOC), which consists of CHEST members with methodologic experience and other stakeholders, developed a rigorous process for evidence-based clinical practice guidelines. This guideline methodology will be used to the greatest extent permitted by the peer-reviewed literature. However, for some important problems clinicians seek guidance but insufficient research prevents establishing guidelines. For such cases, the GOC has created a carefully structured approach permitting a convened expert panel to develop such guidance. The foundation of this approach includes a systematic review of current literature and rigorously vetted, entrusted experts.

RESULTS: Existing evidence, even if insufficient for a guideline, can be combined with a Delphi process for consensus achievement resulting in trustworthy consensus statements. This article provides a review of the CHEST methodologies for these guidance documents as well as the evidence-based guidelines.

CONCLUSIONS: These reliable statements of guidance for health-care providers and patients are based on a rigorous methodology and transparency of process.

CHEST 2014; 146(1):182-192

ABBREVIATIONS: CER = comparative effectiveness research; CHEST = American College of Chest Physicians; COI = conflict of interest; CPG = clinical practice guideline; CS = consensus statement; EBM = evidence-based medicine; GOC = Guidelines Oversight Committee; GRADE = Grading of Recommendations Assessment, Development, and Evaluation; IOM = Institute of Medicine; LGM = Living Guidelines Model; PICO = patient population, intervention, comparator, and outcome; RCT = randomized controlled trial
Evidence-based medicine (EBM) has advanced a science of guideline development addressing perceptions that guidelines foster “cookbook medicine” and disavow the art of medicine. Early “guidelines” were criticized for variation in quality and lack of transparency regarding the crafting of recommendations and did not provide convincing logic to permit selection of the best of multiple competing documents. Resulting from these criticisms and the parallel development of online access and the capability to search large collections of scientific studies, guideline methodology developed into a rigorous scientific approach, culminating in 2011 with the Institute of Medicine (IOM) standards defining “trustworthy” clinical practice guidelines (CPGs).1

As the medical community was increasingly exposed to guidelines based on systematic literature reviews and evidence syntheses, the respectability of evidence-based CPGs was enhanced. Consequently, published consensus-based documents were perceived to be inherently less reliable than CPGs. These perceptions were based on the absence of evidence-based foundations, lack of confidence that panelists were free of industry associations, and concern that there was no serious review to ensure the content was bias-free.

The need to address important clinical challenges in areas without strong evidence remained unfulfilled, because CPGs require supportive evidence. The scientific research hierarchy, with randomized double-blinded controlled trials (RCTs) at the apex,2 was criticized for disqualifying complex patients with multiple morbidities, yielding results not generalizable to broader populations.3,4 Observational studies were recognized to be especially important in identifying harms due to treatments.5

Because pharmaceutical companies infrequently conduct head-to-head comparisons of drugs in the same class or for the same conditions, there was insufficient evidence upon which to base informed choices between alternative interventions. In 2007, the Congressional Budget Office called for federal investment in comparative effectiveness research (CER) on medical treatments to reduce costs (projected to account for 12% of the gross domestic product by 2050)6–7 and improve care. Congress approved the American Recovery and Reinvestment Act of 2009 (Title 8, Public Law 111-5) granting the Agency for Healthcare Research and Quality authority and funding to implement a CER program. Congress then commissioned the IOM to include stakeholders, such as the general public, researchers, physicians, and professional organizations, to define CER and recommend 100 national priorities.8 The resulting infrastructure would sustain CER and support guidelines with more robust intervention comparisons.

These realizations suggest a need for guidance even when evidence is insufficient to inform evidence-based CPGs. Well-conducted consensus statements (CSs) can be at least as important as guidelines, since they address important questions where no guidance exists because of insufficient or imperfect evidence. American College of Chest Physicians (CHEST) an established guideline developer, sought to identify key characteristics that would be critical for CSs to be trustworthy and respected by users.

This article profiles the CHEST methodologies used in rigorous evidence-based guidelines, trustworthy consensus-based statements, and hybrids of these two approaches to provide clinically relevant and dependable guidance. The process for continual updates will also be described.

The Living Guidelines Model
Parallel to guideline development, implementation and maintenance are paramount. A CHEST Guidelines Oversight Committee (GOC) taskforce devised a plan for nimbly updating guidelines while maintaining methodologic rigor. The narrowed focus and other process efficiencies offered economies of time permitting targeted recommendations to be updated when necessary, with available resources. The taskforce’s proposal, based on contributions from multiple stakeholders, approved by the Board of Regents in March 2011, resulted in the Living Guidelines Model (LGM) that is applicable to all current and future guidelines, CSs, and hybrid projects (using both approaches).

Under this new LGM, large comprehensive guidelines spanning topics from prevention to end-of-life care, as in the past model,9,10 are replaced with more targeted guidelines in which individual key clinical questions or specific recommendations define the scope. Selecting the key questions or recommendations for updating is a...
complex challenge. CHEST members, through clinical NetWorks or other entities, are encouraged to annually review existing guidelines and CSs and propose recommendations or suggestions that warrant a review. Key questions or recommendations from previous iterations of the guidelines proposed for selection and prioritization for updating these guidelines are assessed based on eight criteria (Table 1) that primarily focus on whether there are new data, drugs, or devices. New key clinical questions may also be proposed for inclusion in the guidelines, assessed for prioritization using the same criteria, and later refined to include PICO (patient population, intervention, comparator, and outcome) elements.

Living guidelines, distinguished by their targeted scope and continual updates, provide the most current recommendations to health-care providers and patients. As publications are continually updated in the LGM, no single article will encompass a comprehensive set of recommendations on any single topic, so guideline readers need an online source to locate the most up-to-date recommendations.

CHEST Guidelines

The LGM is dependent on the online repository (CHEST Guidelines) currently in development. It will facilitate quick and easy access to all current CHEST recommendations and suggestions, supporting evidence, and related resources, available at the point of care. CHEST Guidelines will be a critical tool to provide the right information to the right people at the right time. It will have advanced search and browse capabilities backed by a large database of all current CHEST recommendations. Once users locate the information they seek, they can access supporting evidence, related tools and resources, and other relevant information. Mobile phone or tablet users will appreciate the adaptive responsiveness to properly size content pages for their device. CHEST Guidelines is currently under development, but readers should watch for announcements on the CHEST website (http://www.chestnet.org).

Evidence-Based Clinical Practice Guidelines

CHEST’s long history of guideline development using the comprehensive model includes nine editions of antithrombotic therapy guidelines beginning in 198611 and culminates with the third edition of the lung cancer guidelines in 2013.10 The new LGM is now inaugurated for all current and future projects. CHEST is committed to maintaining both the currency of these guidelines and the methodologic rigor permitting users to have confidence in the recommendations.

Funding

Guidelines development is expensive, especially the supporting systematic reviews. Previously, CHEST, like most other guideline developers, actively pursued unrestricted funding from industry while actively enforcing strong firewalls between the development work and acquisition of funding.12 However, as the conflict of interest (COI) review process advanced, the GOC requested guideline support from CHEST general funds rather than industry. Although a major expense, multiple surveys confirm that CHEST guidelines continue to be the favorite member benefit and second most important reason for joining the organization. Consequently, the Board of Regents committed all necessary funds to support credible evidence-based CPGs.

Scope

In the LGM, approval of topics for de novo development and prioritization of guideline updates requires meeting most of the criteria in Table 1. The narrow scope with specific key clinical questions broken down into PICO elements permits readers to quickly understand the clinical content areas, patient populations, and interventions addressed.
Panel Composition and Terms
Physicians, nurses, pharmacists, respiratory care practitioners, and other allied health providers who are experts in respective clinical content areas provide their important perspectives on panels. Frontline clinicians suggest challenges they encounter without existing guidance on diagnosis and/or treatment of the topic condition. Additional expertise in resource considerations, medical ethics, and related specialties may be incorporated.

Patient values and preferences are addressed in part by consumer representatives. There is a fine distinction between consumer representatives and patient advocates. Patient advocates (those promoting a particular cause) are precluded from participation because they harbor biases by definition. CHEST works diligently to prevent or remove real or perceived bias from the scientific and objective guideline processes. Consumers educated in EBM are ideal contributors.

CHEST methodologists conduct or oversee all evidence reviews. Approved volunteer methodologists sometimes support evidence reviews.

Other related societies and organizations are invited to participate by appointing representatives to provide input during discussions and peer review of drafts. Panelists who are also members of the invited associations may be appointed to represent these stakeholder interests. As panelists, they have voting rights (see Internal and External Peer Review and Voting section for more information about the voting process). Nonpanelist association representatives are not vetted by the GOC for their COIs and, therefore, do not have voting rights. At the end of the review process, these same organizations are offered the opportunity to endorse the guidelines and be listed in the publication.

LGM panelists serve renewable 3-year terms, provided the Panel Chair renominates them and they are reapproved by GOC based on qualifications and COI reviews. Calls for applications are widely disseminated when new topics are initiated.

COI Reviews and Approvals
COI review processes have evolved substantially in recent years. In many countries, there is recognition that collaborations with pharmaceutical and device manufacturers can affect an individual’s judgments. However, financial support from industry has been so pervasive that it has taken years for this recognition to be widely accepted. In the United States, intensive scrutiny by the media and Congress investigating guideline panels and their recommendations influenced guideline developers to eliminate, insofar as possible, or manage COIs.

Since the early 2000s, the GOC recognized that COI disclosures alone were not sufficient and must be minimized, especially for individual guideline Executive Committee members. The Policies and Procedures Subcommittee instituted practices to provide serious review of disclosed COIs relative to the topic and proposed role of the nominee, with subsequent follow-up during development and 1 year post-publication. Prior to the IOM report on COIs in 2008, there were no rules to guide these decisions. Consequently, the GOC developed its own practices and labored to make proper decisions. These deliberations were complicated by resistance from some nominees, especially those from countries where the scrutiny had not yet surfaced. Adding further complexity, medical researchers appreciated pharmaceutical companies’ support and were unaccustomed to having their objectivity questioned.

Specifying Key Research Questions and Search Strategies
CHEST guidelines are based on comprehensive and systematic reviews of the published medical literature in accordance with the highest standards. Key clinical questions are initially posed by expert panelists and frontline clinicians, then refined by methodologists and Executive Committee members into PICO elements, defining the appropriate patient populations (P), interventions (I), comparators (C), and patient-important outcomes (O) (not research end points). These PICO elements may be rephrased into questions or remain in tabular formats, but they define the terms of the guideline research based on the questions faced by practicing clinicians. Methodologists create search strategies from these terms and synonyms to comb MEDLINE, the Cochrane Library, and other relevant databases (eg, CINAHL, EMBASE, Google Scholar, or Web of Science), as well as the National Guidelines Clearinghouse and Guidelines International Network library. Each search strategy is run in at least two databases, without date limitations unless a topic is being updated. Exceptions are noted in individual guideline methods sections.

Manuscripts report the number of studies identified, screened, included, and excluded. Data from all included studies are extracted into evidence tables.
Study Quality Assessment

Content and methodology experts choose inclusion and exclusion criteria for selecting studies based on the PICO elements. Papers identified during literature searches undergo title and abstract screening, and those selected undergo full text screening. Final studies are assessed using appropriate quality assessment tools based on study design. Guidelines are compared against the Appraisal of Guidelines, Research and Evaluation (AGREE) II instrument\(^1\) and IOM standards.\(^1\) CHEST assesses systematic reviews with the Document and Appraisal Review Tool (DART).\(^15\) Another unnamed quality assessment tool for interventional RCTs and observational studies was adapted from an instrument created by R. Diekemper, MPH; B. Ireland, MD; and L. Merz, PhD, MPH, and two other published tools.\(^16,17\) QUADAS is used to assess the quality of diagnostic studies.\(^18\) RCTs generally address benefits of interventions, but observational studies meeting inclusion criteria are helpful in identifying harms or risks.\(^5\) The full body of evidence for any given recommendation is also assessed for overall quality, displayed in the evidence profiles (see Grading the Recommendations section for more information about grading the body of evidence).

Meta-analyses and Summary Profiles

Meta-analyses inform the recommendations when available. Methodologists performing a systematic review or updating an existing review use the Cochrane’s Review Manager (RevMan) for synthesizing data for a meta-analysis and creating forest plots. Data must be homogeneous to be included in a pooled analysis. GRADE profiler (Cochrane) software generates evidence profiles that examine the data by outcome. Evidence tables, profiles, and other data displays (eg, forest plots) are available in the supplementary materials accompanying guideline publications.

Finalizing the Recommendations

Recommendations are formulated by content experts informed about the evidence for that specific topic. Methodologists provide evidence tables, meta-analyses, forest plots (when appropriate), and evidence profiles to support those formulations. Authors consider the body of evidence, balance of benefits and harms associated with the interventions, and confidence in the relative effect for the specific patient population addressed.

Cost considerations may be included, especially if financial constraints or availability might impact the direction or strength of a recommendation, but only if published formal cost-benefit analyses are available.\(^19\) Resource consultants or health economists conduct cost-benefit analyses, as necessary. CHEST guidelines are implemented worldwide, so when cost constraints exist, guideline implementers are encouraged to use ADAPTE\(^20\) strategies.

Patient values and preferences\(^21\) are reflected in either the recommendations or the associated remarks, whenever applicable. These remarks and other caveats appear below the recommendation and grading but are considered integral to the recommendation. Patient preferences are especially pertinent in weaker recommendations when wide variability in patient choice is anticipated.

Recommendations must be specific and actionable, including as much detail as the evidence allows. Measure developers are cautioned that recommendations based on lower levels of evidence should not be converted into performance measures. Members of the panels’ Executive Committees, association and consumer representatives, and other panelists provide feedback on early drafts, but recommendations and grading are voted upon in a more formal process (see Internal and External Peer Review and Voting section).

Grading the Recommendations

The CHEST grading system\(^22\) encompasses two dimensions: (1) balance of benefits to harms, risks, or burdens, including confidence in the estimate of effect, and (2) level of evidence for the body of research supporting the recommendation (Table 2). Recommendations are strong when benefits clearly outweigh harms or vice versa. In the latter case, there could be a strong negative recommendation (eg, a strong recommendation not to use a specific drug or therapy). Strong recommendations (Grade 1) include the persuasive language “we recommend.” However, when benefits and harms are closely balanced and it is possible additional research might change either the direction or strength of a recommendation, it is considered weak, and statements read, “we suggest.” When benefits clearly outweigh harms, most if not all patients would choose the intervention; the recommendation is clearly Grade 1. But when there is considerable variability in patients’ preferences, tradeoffs between desirable and undesirable consequences are less clear, and recommendations are weaker (ie, Grade 2).

The other grading dimension is a calculation of the quality of the body of evidence underlying the recommendation, reflected in A, B, or C scores. Grading evidence is usually the purview of the methodologists, but...
TABLE 2 | Strength of the Recommendations Grading System

<table>
<thead>
<tr>
<th>Grade of Recommendation</th>
<th>Strength of the Recommendation: Benefits vs Risks, Harms, and Burdens, Grade 1 or 2</th>
<th>State of the Scientific Evidence: Methodologic Strength of Supporting Evidence, Grade A, B, or C</th>
<th>Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong recommendation, high-quality evidence (1A)</td>
<td>Benefits clearly outweigh risk and burden, or vice versa</td>
<td>Consistent evidence from randomized controlled trials without important limitations or exceptionally strong evidence from observational studies.</td>
<td>Recommendation can apply to most patients in most circumstances. Further research is very unlikely to change our confidence in the estimate of effect.</td>
</tr>
<tr>
<td>Strong recommendation, moderate-quality evidence (1B)</td>
<td>Benefits clearly outweigh risk and burden, or vice versa</td>
<td>Evidence from randomized controlled trials with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or very strong evidence from observational studies.</td>
<td>Recommendation can apply to most patients in most circumstances. Higher quality research may well have an important impact on our confidence in the estimate of effect and may change the estimate.</td>
</tr>
<tr>
<td>Strong recommendation, low or very low-quality evidence (1C)</td>
<td>Benefits clearly outweigh risk and burden, or vice versa</td>
<td>Evidence for at least one critical outcome from observational studies, case series, or from randomized, controlled trials with serious flaws or indirect evidence.</td>
<td>Recommendation can apply to most patients in many circumstances. Higher quality research is likely to have an important impact on our confidence in the estimate of effect and may well change the estimate.</td>
</tr>
<tr>
<td>Weak recommendation, high-quality evidence (2A)</td>
<td>Benefits closely balanced with risks and burden</td>
<td>Consistent evidence from randomized controlled trials without important limitations or exceptionally strong evidence from observational studies.</td>
<td>The best action may differ depending on circumstances or patients’ or societal values. Further research is very unlikely to change our confidence in the estimate of effect.</td>
</tr>
<tr>
<td>Weak recommendation, moderate-quality evidence (2B)</td>
<td>Benefits closely balanced with risks and burden</td>
<td>Evidence from randomized controlled trials with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or very strong evidence from observational studies.</td>
<td>Best action may differ depending on circumstances or patients’ or societal values. Higher quality research may well have an important impact on our confidence in the estimate of effect and may change the estimate.</td>
</tr>
<tr>
<td>Weak recommendation, low or very low-quality evidence (2C)</td>
<td>Uncertainty in the estimates of benefits, risks, and burden; benefits, risk, and burden may be closely balanced</td>
<td>Evidence for at least one critical outcome from observational studies, case series, or from randomized controlled trials with serious flaws or indirect evidence.</td>
<td>Other alternatives may be equally reasonable. Higher quality research is likely to have an important impact on our confidence in the estimate of effect and may well change the estimate.</td>
</tr>
</tbody>
</table>

The CHEST grading system defines the grades based on two key dimensions: (1) the balance of benefits of the proposed action as compared with the possible harms or risks, and (2) the methodologic quality of the supporting evidence. Both factors are represented in the overall grade, with the former represented by 1 or 2 and the latter by A, B, or C. When the benefits clearly outweigh the harms or risks (or vice versa), additional research is unlikely to change our confidence in the estimate of effect. When the balance is less clear, patient preferences will play a larger role in determining the best treatment of an individual patient. The methodologic quality is initially based on the hierarchy of study design, but studies can be upgraded or downgraded based on specific criteria (e.g., existence of methodologic flaws, directness, precision, consistency of results). (Adapted with permission from Guyatt et al.25)
Evidence-Based Medicine

meet design standards and (1) there is a large magnitude of effect, (2) there is a statistically significant effect even with the presence of bias, or (3) there is a dose-response gradient.

The CHEST grading system was previously a modification of Grading of Recommendations Assessment, Development, and Evaluation (GRADE), in that the “very low” category in GRADE was not permitted in the CHEST system. In February 2014, the GOC made the decision to start using unmodified GRADE and discontinue use of the CHEST-modified version of GRADE. CHEST has set a minimal threshold for evidence such that it must be published in peer-reviewed publications. When key questions have no supporting evidence or very weak evidence, CHEST would label it as insufficient evidence and follow a consensus-based process or hybrid approach (see later), whereas the GRADE approach would label it as very low evidence but permit the formulation of guideline recommendations. Some guideline developing organizations (eg, the United States Preventive Services Task Force) would refrain from providing any guidance.

Since CHEST guideline methodology has evolved to appraise the quality of the underlying evidence more critically, there are fewer 1A-graded recommendations. This primarily results from the judicious evaluations of the level of evidence. For many topics, the research gaps require extrapolating indirect evidence to the patient populations addressed in the recommendations, resulting in downgrading of the evidence. Some guideline readers will incorrectly believe lower graded recommendations are not strong. Recommendation strength is reflected in the numerical grades (1 or 2). Strong recommendations with low evidence (eg, 1C) are still highly recommended for the designated patient populations.

Internal and External Peer Review and Voting Executive Committees (Panel Chair, CHEST staff, GOC Liaison, and others appointed for their content expertise) perform the initial reviews. The full panel, including association representatives but excluding managed panelists with voting restrictions, then provide feedback on the manuscripts and recommendations.

Controversial recommendations and grades, with supporting evidence, are presented for anonymous voting, and aggressive lobbying is not permitted. Panelists with voting restrictions are tracked using a COI grid displaying each panelist’s COI status relative to each recommendation (Fig 1). These are completed by the project manager, topic editors, and panelists and published with the guidelines, representing the panelists’ conflicts at the time of voting. Additional COI disclosures, published as narrative summaries, are disclosed at the time of publication but are not specific to each recommendation.

Virtual meetings are organized with online polling options so panelists can participate in discussions and vote on recommendations remotely. Panels rarely have face-to-face meetings, and, when they occur, they usually coincide with the annual CHEST meeting. Face-to-face meetings include anonymous online polling synchronized with audience response voting for attendees.

Recommendations are presented with supporting evidence and voted upon in their entirety, including grades and associated remarks. Under extraordinary circumstances, if requested, specific grades may be discussed and voted upon separately from the recommendations. The GRADE grid (Fig 2) defines the level of agreement or disagreement with the proposed recommendations. All votes require 75% participation of those eligible to vote. Approval requires at least 80% of the votes combined for “strongly agree” or “agree.” Negative recommendations are similarly approved, with negative connotation in the recommendation itself and not by votes in disagreement. Disagreement is defined as voting against the proposed recommendation as written.

![Figure 1 – COI grid template. This spreadsheet is used to keep track of which panelists have conflicts of interests relevant to which recommendations. It allows staff and panel leadership to prevent voting by conflicted panelists whose conflicts are being managed. COI = conflict of interest.](http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/930460/ on 06/28/2017)
Recommendations achieving only 67% to 79% agreement may be published with a section labeled “additional remarks,” permitting those with minority opinions to address their concerns. Recommendations achieving < 67% agreement are not published; however, the authors may state that these recommendations were controversial and further research might solidify consensus. Final voting tallies are available upon request at science@chestnet.org.

Manuscripts are revised after panel feedback and voting. Once approved, recommendations may not be significantly modified. The Executive Committee, primarily the Chair, appraises revisions for consistency and completeness before the manuscripts and supporting materials are reviewed internally by CHEST.

As a diverse organization with multiple clinical specialties, clinically focused member NetWorks provide content reviews. The GOC simultaneously provides methodology oversight along with members of the Board of Regents in the first round of reviews. The CHEST Journal peer review process overlaps with the second round of reviews but uses ScholarOne technology and has its own COI processes for reviewers. Guidelines are considered final when approved by the four CHEST Presidents in the line of succession.

All internal reviewers are approved by GOC to ensure they have no COIs in the relevant guideline area. Anonymity is always maintained. Reviewer feedback forms address the AGREE II instrument\textsuperscript{14} domains for assessing guideline quality and clinical relevance, usability, and feasibility. All reviewer comments are denoted as mandatory or suggested, and the authors are required to address the former either with revisions or by providing a written justification explaining why they disagree with the reviewer. In the second round, reviewers consider the written justifications and modifications to determine acceptability. Additional cycles are sometimes necessary. The complete review process has included as many as 30 reviewers, but in the LGM this number is expected to be less since each document will be smaller and more focused.

After approval by the GOC and CHEST Presidents, only editorial changes to recommendations are allowed. The final version is forwarded to invited organizations that sign nondisclosure agreements to consider endorsement and listing in the final publication. As with all CHEST publications, CHEST guidelines undergo additional independent journal peer review.

**Limitations**

A major limitation of EBM today is that complex patients with multiple comorbidities have not been included in most RCTs. Despite this, there are some instances in which common comorbidities can be included in recommendations by providing careful definitions of these complex patient populations and incorporating specific advice regarding administration, dosing, duration of the intervention, and options. CHEST encourages increased funding of health-care research prioritizing patients with multiple comorbidities. As this scientific base evolves, guidelines will increasingly become more relevant for these complex patients.

**Dissemination and Implementation**

CHEST guidelines are published in CHEST online primarily, occasionally in print, and as mobile apps and submitted to the National Guidelines Clearinghouse and Guidelines International Network library. Many health-care systems and electronic medical record vendors worldwide use CHEST guidelines.

**Maintenance**

In the new LGM, all guideline recommendations will be reviewed for currency annually after publication. The criteria for these considerations are in Table 1.

**Trustworthy Consensus Statements and Hybrid Documents**

Many topic areas have no or weak supporting evidence; however, there is often need for credible guidance regarding clinical questions when there is insufficient evidence to support a formal guideline. In 2012, GOC approved a trustworthy CS development process aimed to address this very need. GOC also recognized the need for a hybrid approach combining evidence-based CPG methodologies and consensus-based
methods when evidence is variable across subtopics within a larger project scope. If a developer is transparent about which recommendations or suggestions were produced using which process and why, both can be used in the same guidance document. The varying approach to providing clinical guidance can be thought of as three overlapping processes on a continuum (Fig 3) spanning the lowest level of supporting evidence (consensus-based) to the highest level of supporting evidence (guidelines), with the hybrid approach in the middle using both processes. For some questions, initially the hybrid process will be most common, but as more research is published, the guideline methodology becomes more appropriate. This movement toward greater precision works well in the LGM because key clinical questions are continually reassessed for new levels of published evidence.

The key to credibility in consensus-based documents is the level of rigor in four key areas:

- Systematic review to identify any existing evidence or build the foundation for a consensus-based approach when evidence is lacking;
- Structured panel of experts representing all relevant stakeholders, carefully reviewed for conflicts of interest;
- Sound process for achieving consensus using the modified Delphi technique; and
- Comprehensive and thorough review process.

With these elements, rigorous guideline development methodologies are used to the fullest extent possible, given the current state of the science.

**Systematic Review**

To ensure all relevant evidence is identified, PICO-based literature searches of multiple databases are performed just as for systematic reviews for guidelines. If comprehensive and systematic PICO-based searches of the literature disclose insufficient evidence, the case is made for a consensus-based approach, although gray literature (eg, conference abstracts) may help inform the deliberations. Research studies surviving screening are individually assessed for quality, and data are extracted into evidence tables. In hybrid projects, sufficient evidence may exist to inform some areas and allow development of an evidence-based CPG process followed, including meta-analyses, quality assessment of the body of evidence, and evidence profiles. Publication must be transparent about the process followed for deriving each recommendation and suggestion.

Where to draw the line between evidence-based and the consensus-based approaches is a challenge. GOC has set the threshold for a guideline at two RCTs, two observational studies, or one of each, that quantitatively address comparable interventions and outcomes, with a quality level assessed as fair or good. This minimum permits pooling of the data for meta-analysis. GOC also recognized one exceptional study of very high quality might accommodate the evidence-based guideline approach, but only with approval of the GOC Guidelines Subcommittee.

**Panel Selection**

For readers to trust CS suggestions, they must have confidence in the experts providing guidance, perhaps to an even greater degree than for evidence-based recommendations. The same meticulous process used to select and approve guideline nominees is used for CSs.

**Consensus Achievement**

An explicit consensus achievement process based on the Delphi technique develops suggestions (since they are not strong, they are not called recommendations) for voting by all eligible panelists. Up to three rounds of voting may occur using the GRADE grid (Fig 2) until consensus is achieved. This process is similar to the voting in the evidence-based guideline process. The minimal response rate for each suggestion is 75% of the panel, with 80% of respondents voting to agree or strongly agree with the suggestion to obtain consensus. Open fields permit anonymous feedback on improving the suggestions. Between rounds, writing committees may revise suggestions not meeting the 80% threshold and resubmit them for the next round. Suggestions not achieving consensus in three rounds are dropped. All voting is anonymous, and final tallies are available upon request.

**Comprehensive Review**

Readers can be assured that CSs undergo the same meticulous evaluations as guidelines. CHEST CSs are reviewed by members of the relevant NetWork(s) for

---

**Figure 3 – Continuum of supporting evidence.** This illustrates the continuum of research data that can support an evidence-based guideline when the evidence is of higher-level quality (toward the right side) or will require the consensus of an expert panel when the evidence is too low (toward the left side) to support the more rigorous methodology of a guideline. The hybrid approach is used when there are topics that span this continuum in the same project. As the scientific evidence evolves, it might increase the likelihood that an evidence-based guideline methodology could be followed. Thus, evolving evidence moves toward the right on this continuum as more scientific papers are published on a topic.
content appraisal and by GOC members for appropriate methods. Additional evaluations are provided by members of the Board of Regents, the Presidential succession line, and CHEST peer reviewers.

As with guidelines, association representatives are invited to provide early feedback on drafts and endorsements of the final documents. This escalates the breadth of stakeholder feedback and improves the applicability for diverse audiences. Performance measures should not be based on consensus-based suggestions.

Looking to the Future
Innovations must be tested, evaluated, and improved by users and developers. As providers realize consensus-based guidance can be meticulously crafted, dependable, and credible, negative preconceptions will be lifted. Evidence-based guideline developers are compelled to adopt the living guidelines model to sustain production and maintenance of quality guidelines with realistic resources. Guideline readers must be able to discern guideline quality and trustworthiness. The LGM will help guideline developers meet standards of excellence and focus on finely targeted key clinical questions, while prudently maintaining quality, precision, and currency.

As the LGM is more universally used and inclusion criteria of the National Guidelines Clearinghouse continue to evolve, certain questions arise. Will allowances be made for recommendations reviewed annually but not updated if there are insufficient developments to justify a revised systematic review? Will consensus-based suggestions be acceptable if created through a credible process? How will quality improvement and measure developers react to reliable consensus-based suggestions? As the science of guideline development has evolved, so too have methodologic innovations. The need for trusted guidance will force organizations in the EBM community to adapt to these trends.

Acknowledgments

Financial/nonfinancial disclosures: The authors have reported to CHEST the following conflicts of interest: Dr Lewis makes public statements and give presentations about the CHEST Guideline Methodology at conferences and other meetings on this topic. Her expenses are sometimes reimbursed. She received one small honorarium from the Institute of Medicine in 2011. Ms Diekemper is an author of the DART tool but receives no compensation for it. Drs Ornelas and Casey have reported that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Role of sponsors: CHEST was the sole supporter of this article and the innovations addressed within.

Other contributions: We thank Mark Metersky, MD, FCCP; Dan Ouellette, MD, FCCP; and Ian Nathanson, MD, FCCP, as GOC executive leadership, for their direction and guidance as these innovations were created and implemented, as well as for their review of this article. All members of the GOC current and past (too numerous to name individually) have helped to define the challenges and debate the solutions leading to the current methodologies discussed in this article. We thank them for their significant contributions. We also thank Richard S. Irwin, MD, Master FCCP, whose careful review and suggestions improved this article. The positions expressed in this paper do not necessarily reflect those of the Department of Veterans Affairs or the United States Government.

References


