

Standardized Reporting of Clinical Practice Guidelines: A Proposal from the Conference on Guideline Standardization*

Richard N. Shiffman, MD, MCIS; Paul Shekelle, MD, PhD; J. Marc Overhage, MD, PhD; Jean Slutsky, PA, MSPH; Jeremy Grimshaw, MB, ChB, PhD; and Aniruddha M. Deshpande, MD

Despite enormous energies invested in authoring clinical practice guidelines, the quality of individual guidelines varies considerably. The Conference on Guideline Standardization (COGS) was convened in April 2002 to define a standard for guideline reporting that would promote guideline quality and facilitate implementation. Twenty-three people with expertise and experience in guideline development, dissemination, and implementation participated. A list of candidate guideline components was assembled from the Institute of Medicine Provisional Instrument for Assessing Clinical Guidelines, the National Guideline Clearinghouse, the Guideline Elements Model, and other published guideline models. In a 2-stage modified Delphi process, panelists first rated their agreement with the statement that “[item name] is a necessary component of practice guidelines” on a 9-point scale. An individualized report was prepared for each panelist; the report summarized the panelist’s rating for each item and the median and dispersion

of rankings of all the panelists. In a second round, panelists separately rated necessity for validity and necessity for practical application. Items achieving a median rank of 7 or higher on either scale, with low disagreement index, were retained as necessary guideline components. Representatives of 22 organizations active in guideline development reviewed the proposed items and commented favorably. Closely related items were consolidated into 18 topics to create the COGS checklist. This checklist provides a framework to support more comprehensive documentation of practice guidelines. Most organizations that are active in guideline development found the component items to be comprehensive and to fit within their existing development methods.

Ann Intern Med. 2003;139:493-498.

www.annals.org

For author affiliations, see end of text.

*For a list of participants in the COGS, see the Appendix.

Evidence-based clinical practice guidelines can reduce the delivery of inappropriate care and support the introduction of new knowledge into clinical practice (1–3). In many cases, guidelines encapsulate the most current knowledge about best practices. Rigorously developed guidelines can translate complicated research findings into actionable recommendations for clinical care (4). Over the past decade, a plethora of guidelines has been created and published by a multitude of organizations at substantial cost.

Despite the enormous energies invested in guideline authoring, the quality of individual guidelines varies considerably. In its landmark report, the Institute of Medicine (IOM) defined 8 “desirable attributes” of clinical practice guidelines: validity, reliability and reproducibility, clinical applicability, clinical flexibility, clarity, documentation, development by a multidisciplinary process, and plans for review (5).

However, critical information that would attest to validity or would document fulfillment of the other IOM criteria is regularly absent from published guidelines. In an evaluation of 279 guidelines developed by U.S. medical specialty societies, Shaneyfelt and colleagues (6) found that guidelines published in the peer-reviewed medical literature do not adhere to established methodologic standards. Likewise, Grilli and colleagues (7) found that of 431 guidelines produced by specialty societies, 82% did not apply explicit criteria to grade the scientific evidence that supported their recommendations, 87% did not report whether a systematic literature search was performed, and 67% did not describe the type of professionals involved in guideline development. Systematic reviews of guidelines for drug therapy (8), management of depression (9), and osteoporosis (10) have confirmed marked variation in qual-

ity. Both nonadherence to methodologic standards and failure to document development activities contribute to this variation.

We convened the Conference on Guideline Standardization (COGS) to define a standard for guideline reporting that would promote guideline quality and facilitate implementation. The proposed standard provides a checklist of components necessary for evaluation of validity and usability. The checklist is intended to minimize the quality defects that arise from failure to include essential information and to promote development of recommendation statements that are more easily implemented.

In contrast to other instruments that have been developed for post hoc evaluation of guideline quality, the COGS checklist is intended to be used prospectively by developers to improve their product by improving documentation. The COGS panel used a systematic and rigorous process to define content of the proposed standard and to achieve consensus. The COGS panel also included a wide variety of perspectives, deliberately bringing together representatives from medical specialty societies, government agencies, and private groups that develop guidelines; journal editors and the National Guideline Clearinghouse (NGC), which disseminate guidelines; guideline implementers, including managed care representatives and informaticians; and academicians.

METHODS

We actively sought people with diverse backgrounds from wide-ranging geographic areas to participate in the meeting. Selection criteria for participants were 1) activity in a wide variety of guidelines initiatives, 2) recognition as

leaders in their field, and 3) willingness to collaborate. To maximize interaction, the number of participants in the Conference was limited.

We set as a task for the panelists the specification and definition of a set of “necessary” guideline components that should be considered for reporting in all evidence-based practice guidelines. We defined necessary items as those that establish the validity of the guideline recommendations or facilitate practical application of the recommendations. We noted that many additional items might be considered appropriate components in guidelines, but sought to define a minimal set of essential elements.

We assembled a list of candidate guideline components from the IOM Provisional Instrument for Assessing Clinical Guidelines (5), the NGC (11), and the Guideline Elements Model (12), a hierarchical model of guideline content that was created from a systematic review of published guideline descriptions. These items were supplemented with items highlighted in the literature, for example, structured abstract (13) and conflict of interest (14).

We applied a modified Delphi approach to help focus group discussion. This approach has been widely applied for evaluation of expert opinion on appropriateness (15) and medical necessity (16), for policy development (17), and for prioritization (18). The technique has been described in detail (19). In brief, after we secured agreement to participate in the COGS panel, we gave all participants a bibliography of resources regarding guideline quality and its appraisal, guideline implementation, and the modified Delphi consensus development approach. Panelists were asked to rate their agreement with the statement that “[Item name] is a necessary component of practice guidelines” on a 9-point scale. Rating an item with a 9 indicated strong agreement with the statement that this item was necessary. Rating an item with a 1 indicated strong disagreement with the statement and suggested that the item was absolutely unnecessary in a guideline. A rating of 5 indicated neutrality or indifference.

We developed a password-protected Web site that panelists used to complete the first round of ratings online before the meeting. Online rating permitted accurate and efficient data capture and analysis. For each item, the median rating and the disagreement index (defined as the Interpercentile Range divided by the Interpercentile Range Adjusted for Symmetry) were calculated (19). The disagreement index, which can be calculated for panels of any size, describes the dispersion of ratings more effectively than the mean absolute deviation from the median. Index values greater than 1 indicate disagreement. We displayed summary statistics for each item on a form that was individualized for each panelist.

The COGS meeting was held on 26 and 27 April 2002 in New Haven, Connecticut. In a colloquy facilitated by 2 of the authors, each candidate item was discussed to ensure that all participants agreed on its definition and potential contribution to the COGS checklist and to high-

light empirical evidence of its value. When appropriate, additional items were added to the list. The group determined that in the second round of ratings, it would be valuable to rate each item’s necessity on 2 subscales: necessity to establish validity and necessity for practical application. The participants then rated each item on these 2 dimensions.

Analysis

We tallied the median score and the disagreement index for each item. We retained items with median scores of 7 or higher and disagreement indexes less than 1.0 on either scale as necessary guideline components on the checklist. To interpret necessity, the 9-point scale is divided into 3 ranges: items scoring 1 to 3 are considered unnecessary, items scoring 4 to 6 are neutral, and items scoring 7 to 9 are considered necessary (20). In this study, the threshold of 7 was chosen because it represented the lowest rating at which the participants indicated that an item was necessary for inclusion in guidelines.

Pilot Review

To field test the proposed checklist, we surveyed organizations that were active in guideline development. We identified all organizations that met the NGC Web site criteria for guideline display on 12 July 2002. From that list, we selected organizations that had developed 10 or more guidelines displayed on the NGC Web site. We excluded 1) organizations that participated in development of the COGS checklist (because of potential bias) and 2) government agencies and organizations based outside the United States (for logistic reasons). A draft COGS checklist and a brief survey were sent to the people identified as being responsible for guideline development at each eligible organization.

RESULTS

All 23 panelists submitted first- and second-round ballots. During the discussion, participants suggested consideration of 10 new items in the second round of balloting and refined definitions of several items. Thirty-six discrete items were considered necessary to establish guideline validity; they received ratings of 7 or greater and had disagreement indexes of less than 1. Twenty-four items were considered necessary for practical application of the guideline, each with a disagreement index less than 1. Several items were rated necessary on both dimensions. Overall, 44 discrete items were considered necessary.

Closely related items were then consolidated into 18 topics to create the COGS checklist for reporting clinical guidelines (Table). Appendix Tables 1 and 2 (available at www.annals.org) present a complete listing of all items rated and their scores.

Twenty-two organizations met eligibility criteria for evaluation of the draft checklist, and all completed the survey (100% response rate). Sixteen organizations (73%) responded that they believed the checklist would be helpful

Table. The COGS Checklist for Reporting Clinical Practice Guidelines*

Topic	Description
1. Overview material	Provide a structured abstract that includes the guideline's release date, status (original, revised, updated), and print and electronic sources.
2. Focus	Describe the primary disease/condition and intervention/service/technology that the guideline addresses. Indicate any alternative preventive, diagnostic or therapeutic interventions that were considered during development.
3. Goal	Describe the goal that following the guideline is expected to achieve, including the rationale for development of a guideline on this topic.
4. Users/setting	Describe the intended users of the guideline (e.g., provider types, patients) and the settings in which the guideline is intended to be used.
5. Target population	Describe the patient population eligible for guideline recommendations and list any exclusion criteria.
6. Developer	Identify the organization(s) responsible for guideline development and the names/credentials/potential conflicts of interest of individuals involved in the guideline's development.
7. Funding source/sponsor	Identify the funding source/sponsor and describe its role in developing and/or reporting the guideline. Disclose potential conflict of interest.
8. Evidence collection	Describe the methods used to search the scientific literature, including the range of dates and databases searched, and criteria applied to filter the retrieved evidence.
9. Recommendation grading criteria	Describe the criteria used to rate the quality of evidence that supports the recommendations and the system for describing the strength of the recommendations. Recommendation strength communicates the importance of adherence to a recommendation and is based on both the quality of the evidence and the magnitude of anticipated benefits or harms.
10. Method for synthesizing evidence	Describe how evidence was used to create recommendations, e.g., evidence tables, meta-analysis, decision analysis.
11. Prerelease review	Describe how the guideline developer reviewed and/or tested the guidelines prior to release.
12. Update plan	State whether or not there is a plan to update the guideline and, if applicable, an expiration date for this version of the guideline.
13. Definitions	Define unfamiliar terms and those critical to correct application of the guideline that might be subject to misinterpretation.
14. Recommendations and rationale	State the recommended action precisely and the specific circumstances under which to perform it. Justify each recommendation by describing the linkage between the recommendation and its supporting evidence. Indicate the quality of evidence and the recommendation strength, based on the criteria described in 9.
15. Potential benefits and harms	Describe anticipated benefits and potential risks associated with implementation of guideline recommendations.
16. Patient preferences	Describe the role of patient preferences when a recommendation involves a substantial element of personal choice or values.
17. Algorithm	Provide (when appropriate) a graphical description of the stages and decisions in clinical care described by the guideline.
18. Implementation considerations	Describe anticipated barriers to application of the recommendations. Provide reference to any auxiliary documents for providers or patients that are intended to facilitate implementation. Suggest review criteria for measuring changes in care when the guideline is implemented.

* COGS = Conference on Guideline Standardization.

for creating more comprehensive practice guidelines, and an additional 2 organizations (9%) responded that it might be helpful. Nineteen respondents (86%) indicated that documenting the proposed items would fit within their organizations' guideline development methods. Fifteen (68%) stated that they would use the proposed checklist in guideline development and an additional 4 (18%) indicated that they might. Reasons for not using the COGS checklist included "We already have a system that covers these points" and "we . . . are producing guidelines that must be useful for clinicians and for people doing systems improvement work. Thus they must be succinct and brief. These characteristics cannot be achieved simultaneously with comprehensiveness."

DISCUSSION

We created the COGS checklist to promote the systematic reporting of precise details that are critical for understanding a guideline's development, its recommendation statements, and potential issues in its application. This structured system for reporting is designed to reduce the considerable variation in guideline documentation conventions; it is not intended to dictate a particular guideline development methodology or to interfere with the creative process.

One factor that contributes to the variation of guidelines is the fact that guideline development panels are commonly composed of people who are inexperienced in guideline authoring. New committees of experts are often assembled to address each discrete clinical topic. The COGS checklist represents a common framework for guideline developers to help ensure that important information is included in the guideline documentation, for journal editors and other disseminators to determine that a guideline draft is ready for publication, and for implementers to inform the selection process and to facilitate the representation and translation of recommendations into tools that influence clinical behavior.

The COGS checklist specifies the nature of information that should be documented about guidelines, but the specification of appropriate content that might be used to fulfill a checklist item is intentionally left to developers and those interested in quality appraisal. Despite the label of "necessary," it is entirely possible that some guidelines rightfully might not include content for every item, but they would address explicitly whether the guideline development team considered that item.

Although many individual organizations have devised manuals and procedures for developing guidelines, we are

unaware of any consensus standard that has been proposed for use prospectively to promote the development of high-quality guidelines. Graham and colleagues (21) identified 13 instruments for evaluation of clinical practice guidelines, all of which were intended for retrospective application to guidelines that had already been created and released. Inclusion criteria of the NGC (22) have defined a de facto documentation standard for organizations desiring to post their guidelines on this Internet-accessible repository. The guidelines are represented using a structured summary that describes a limited set of specific guideline attributes. In the United Kingdom, a central guideline appraisal service has been implemented to assess all guidelines funded by the National Health Service to help ensure that guidelines are sound before they are deployed (23). An evolving appraisal instrument has been in use since 1997 to help determine which guidelines should be commended for use (24). Cluzeau and Littlejohns found that guidelines were better documented since the appraisal instrument was introduced and attributed the improvement in part to the use of the tool as an "aide-memoire" by guideline developers (24). Their work has formed the basis for an instrument for guideline appraisal that was further refined and tested by the AGREE (Appraisal of Guideline Research and Evaluation) Collaboration, an international partnership of researchers and policymakers (25).

Experience with the Consolidated Statement for Reporting Trials (CONSORT) statement exemplifies the benefits of standardization of documentation (26, 27). The CONSORT statement is mandated by many prominent journals, and investigators reporting clinical trials have increasingly used the statement to improve the comprehensiveness of their reports. CONSORT's effectiveness has been demonstrated by the fact that quality scores for reports of trials have improved in journals that require use of CONSORT, and specific threats to validity have diminished (28, 29). CONSORT was not intended to be used for quality appraisal of clinical trials. Likewise, we believe COGS can be used most effectively to identify necessary components that should be documented in guidelines but should not be used (alone) to judge guideline quality or adequacy.

A potential but real concern is the additional workload that might be imposed on already overburdened guideline development organizations by the institution of new requirements. This concern must be weighed against the substantial benefits of adopting a uniform framework for guideline documentation. In general, the organizations surveyed found the checklist to fit within their guideline development methods. It may not be possible to generalize the high approval levels for the COGS checklist indicated by the survey respondents because only organizations with a high level of guideline productivity development were surveyed. Organizations that have produced fewer guidelines may find greater conflicts between their approaches and the documentation requirements proposed by COGS,

and therefore, they might find more difficulty using the checklist.

Another difficulty with using a standardized documentation instrument is that a "one size fits all" approach may not work well for all guideline developers because of their development process or the topic area on which the guideline focuses. Some guideline developers are moving toward establishing iterative and ongoing guideline development, where the guideline is a "living document," constantly fine-tuned on the basis of evaluation of patient outcomes or performance measures. Other guideline developers may be concerned with very narrow, circumscribed areas (for example, oncology, radiology, or pathology) where several of the COGS items are not relevant. In addition, other countries may have differences that may limit the applicability of the COGS checklist to the guideline development and reporting process. Real value of the checklist in terms of better-quality guidelines and easier implementation has not been demonstrated.

Like the CONSORT statement, the COGS checklist will probably require maintenance revisions. Members of the group have expressed an interest in meeting at intervals to consider the experience in application of COGS. Comments from users of the COGS checklist will be collected on the COGS Web site (ycmi.med.yale.edu/COGS). We plan to measure the effect of the COGS statement on guideline documentation using a before-after evaluation.

An important goal of the COGS meeting was to begin a conversation among guideline developers, disseminators, and implementers about how guideline statements should be written to facilitate their implementation. Specifically, one of the purposes was to address problems with implementing guidelines electronically. Guideline authors strive to make recommendations that accurately reflect the scientific evidence—sometimes intentionally introducing ambiguity into the recommendations to reflect their uncertainties. Often, developers use terms that are not clearly defined, thereby presenting difficulties when recommendations are integrated into clinical decision support systems (30). Implementers must resolve these ambiguities in order to devise pragmatic strategies that will effectively influence clinicians' behavior. A sustained and productive discussion among guideline developers, disseminators, implementers, and knowledge managers about critical guideline items and clear statement of decidable and executable recommendations will help to overcome major impediments to guideline use. That work is ongoing and will be reported in the future.

APPENDIX: PARTICIPANTS IN THE CONFERENCE ON GUIDELINE STANDARDIZATION

Abha Agrawal, MD (Lown Cardiovascular Institute, Brookline, Massachusetts); David Bates, MD (Partners Healthcare System, Boston, Massachusetts); Stephen M. Downs, MD, MS (Indiana University School of Medicine, Indianapolis, Indiana);

Robert A. Greenes, MD, PhD (Harvard Medical School, Boston, Massachusetts); Jeremy Grimshaw, MB, ChB (Ottawa Health Research Institute, Ottawa, Ontario, Canada); Carla Herrerias, MPH (American Academy of Pediatrics, Elk Grove Village, Illinois); Charles J. Homer, MD, MPH (National Initiative for Children's Healthcare Quality, Boston, Massachusetts); Bryant T. Karras, MD (University of Washington, Seattle, Washington); Christine Laine, MD, MPH (*Annals of Internal Medicine*, Philadelphia, Pennsylvania); Mauricio Leon, MD (IDX Systems Corporation, Seattle, Washington); Blackford Middleton, MD, MPH, MSc (Partners Healthcare Corporation, Boston, Massachusetts); Christel Mottur-Pilson, PhD (American College of Physicians, Philadelphia, Pennsylvania); J. Marc Overhage, MD, PhD (Indiana University School of Medicine, Indianapolis, Indiana); Ian Purves, MD (Sowerby Center for Health Informatics, Newcastle, United Kingdom); Paul G. Shekelle, MD, PhD (Southern California Evidence-based Practice Center, Los Angeles, California); Richard N. Shiffman, MD, MCIS (Yale University School of Medicine, New Haven, Connecticut); Jean Slutsky, PA, MSPH (Agency for Healthcare Research and Quality, Rockville, Maryland); Sharon Smart, MD (Sowerby Center for Health Informatics, Newcastle, United Kingdom); Leif I. Solberg, MD (HealthPartners Medical Group, Minneapolis, Minnesota); Harold Sox, MD (*Annals of Internal Medicine*, Philadelphia, Pennsylvania); Samson Tu, MS (Stanford University School of Medicine, Stanford, California); Steven H. Woolf, MD, MPH (Virginia Commonwealth University, Richmond, Virginia); and Stephanie Zaza, MD, MPH (Centers for Disease Control and Prevention, Atlanta, Georgia). Not all panelists voted to include all items on the checklist.

From Yale University School of Medicine and Yale Center for Medical Informatics, New Haven, Connecticut; Southern California Evidence-based Practice Center and University of California, Los Angeles, Los Angeles, California; Regenstrief Institute and Indiana University School of Medicine, Indianapolis, Indiana; Agency for Healthcare Research and Quality, Rockville, Maryland; and Ottawa Health Research Institute and the University of Ottawa, Ottawa, Ontario, Canada.

Disclaimer: The opinions expressed are the authors and do not necessarily represent official policy of the Agency for Healthcare Research and Quality or the Department of Health and Human Services.

Acknowledgments: The authors thank the American Academy of Neurology, American Academy of Ophthalmology, American Academy of Periodontology, American Academy of Sleep Medicine, American College of Cardiology, American College of Chest Physicians, American College of Radiology, American Diabetes Association, American Medical Directors Association, American Optometric Association, American Society of Clinical Oncology, American Society of Plastic Surgeons, American Urological Association, Cincinnati Children's Hospital Medical Center, Eastern Association for the Surgery of Trauma, Infectious Diseases Society of America, Institute for Clinical Systems Improvement, Society of Nuclear Medicine, University of Michigan Health System, and the other organizations that participated in the survey.

Grant Support: The COGS meeting was supported by grant R13 HS10962 from the Agency for Healthcare Research and Quality. Dr.

Shiffman's work is also supported by grants R01 LM07199 and R29-LM05551 from the National Library of Medicine.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Richard N. Shiffman, MD, MCIS, Yale Center for Medical Informatics, 300 George Street, Suite 501, New Haven, CT 065120; e-mail, Richard.Shiffman@yale.edu.

Current author addresses are available at www.annals.org.

References

- Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet*. 1993;342:1317-22. [PMID: 7901634]
- Merritt TA, Palmer D, Bergman DA, Shiono PH. Clinical practice guidelines in pediatric and newborn medicine: implications for their use in practice. *Pediatrics*. 1997;99:100-14. [PMID: 8989346]
- Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ*. 1999;318:527-30. [PMID: 10024268]
- Haines A, Jones R. Implementing findings of research. *BMJ*. 1994;308:1488-92. [PMID: 8019284]
- Institute of Medicine. *Guidelines for Clinical Practice: From Development to Use*. Washington, DC: National Academy Press; 1992.
- Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA*. 1999;281:1900-5. [PMID: 10349893]
- Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet*. 2000;355:103-6. [PMID: 10675167]
- Graham ID, Beardall S, Carter AO, Glennie J, Hebert PC, Tetroe JM, et al. What is the quality of drug therapy clinical practice guidelines in Canada? *CMAJ*. 2001;165:157-63. [PMID: 11501454]
- Littlejohns P, Cluzeau F, Bale R, Grimshaw J, Feder G, Moran S. The quantity and quality of clinical practice guidelines for the management of depression in primary care in the UK. *Br J Gen Pract*. 1999;49:205-10. [PMID: 10343424]
- Cranney A, Waldeger L, Graham I, Man-Son-Hing M, Byszewski A, Ooi D. Systematic assessment of the quality of osteoporosis guidelines. *BMC Musculoskeletal Disord*. 2002;3:20. [PMID: 12174195]
- Guideline Summary Sheet. National Guideline Clearinghouse. Accessed at www.guideline.gov/about/GuidelineSummarySheet.aspx on 20 July 2003.
- Shiffman RN, Karras BT, Agrawal A, Chen R, Marengo L, Nath S. GEM: a proposal for a more comprehensive guideline document model using XML. *J Am Med Inform Assoc*. 2000;7:488-98. [PMID: 10984468]
- Hayward RS, Wilson MC, Tunis SR, Bass EB, Rubin HR, Haynes RB. More informative abstracts of articles describing clinical practice guidelines. *Ann Intern Med*. 1993;118:731-7. [PMID: 8460861]
- Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA*. 2002;287:612-7. [PMID: 11829700]
- Park RE, Fink A, Brook RH, Chassin MR, Kahn KL, Merrick NJ, et al. Physician ratings of appropriate indications for six medical and surgical procedures. *Am J Public Health*. 1986;76:766-72. [PMID: 3521341]
- Kahan JP, Bernstein SJ, Leape LL, Hilborne LH, Park RE, Parker L, et al. Measuring the necessity of medical procedures. *Med Care*. 1994;32:357-65. [PMID: 8139300]
- Lavis JN, Anderson GM. Appropriateness in health care delivery: definitions, measurement and policy implications. *CMAJ*. 1996;154:321-8. [PMID: 8564901]
- Brook RH, Kamberg CJ. Appropriateness of the use of cardiovascular procedures: a method and results of this application. *Schweiz Med Wochenschr*. 1993;123:249-53. [PMID: 8446857]
- Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, et al.

The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND; 2001.

20. **Leape LL, Hilborne LH, Kahan JP, Stason WB, Park RE, Kamberg CJ, et al.** Coronary artery bypass graft: a literature review and ratings of appropriateness and necessity. Report No. JRA-02. Santa Monica, CA: RAND; 1991.
21. **Graham ID, Calder LA, Hebert PC, Carter AO, Tetroe JM.** A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care.* 2000;16:1024-38. [PMID: 11155826]
22. National Guideline Clearinghouse Inclusion Criteria. Accessed at www.guideline.gov/contact/coninclusion.aspx on 20 July 2003.
23. **Cluzeau FA, Littlejohns P.** Appraising clinical practice guidelines in England and Wales: the development of a methodologic framework and its application to policy. *Jt Comm J Qual Improv.* 1999;25:514-21. [PMID: 10522232]
24. **Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE.** Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care.* 1999;11:21-8. [PMID: 10411286]
25. **AGREE (Appraisal of Guideline Research and Evaluation) Collaboration.** Accessed at www.agreecollaboration.org on 21 August 2002.
26. **Moher D, Schulz KF, Altman DG.** The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med.* 2001;134:657-62. [PMID: 11304106]
27. **Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al.** The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134:663-94. [PMID: 11304107]
28. **Egger M, Jüni P, Bartlett C.** Value of flow diagrams in reports of randomized controlled trials. *JAMA.* 2001;285:1996-9. [PMID: 11308437]
29. **Moher D, Jones A, Lepage L.** Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA.* 2001;285:1992-5. [PMID: 11308436]
30. **Tierney WM, Overhage JM, Takesue BY, Harris LE, Murray MD, Vargo DL, et al.** Computerizing guidelines to improve care and patient outcomes: the example of heart failure. *J Am Med Inform Assoc.* 1995;2:316-22. [PMID: 7496881]

Current Author Addresses: Drs. Shiffman and Deshpande: Yale Center for Medical Informatics, 300 George Street, Suite 501, New Haven, CT 06511.

Dr. Shekelle: Greater Los Angeles Veterans Affairs Healthcare System 111G, 11301 Wilshire Boulevard, Los Angeles, CA 90073.

Dr. Overhage: Indiana University School of Medicine, 1050 Wishard Boulevard, Indianapolis, IN 46202.

Ms. Slutsky: Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, John M. Eisenberg Building, 540 Gaither Road, Rockville, MD 20850.

Dr. Grimshaw: Ottawa Health Research Institute, 1053 Carling Avenue, Ottawa, Ontario K1Y 4E9, Canada.

Appendix Table 1. Items Ranked Necessary for Guideline Validity, Median Ratings, Distribution of Ratings by Tertiles, and Disagreement Index*

Item	Necessity for Validity (Median Score)	Ratings			Disagreement Index
		1–3	4–6	7–9	
Description of evidence collection	9	0	0	23	0.00
Criteria for selection of evidence to review	9	0	1	20	0.00
Eligibility	9	0	2	21	0.00
Recommendation	9	3	1	19	0.00
Reason	9	0	1	22	0.00
Evidence quality	9	0	0	23	0.00
Evidence time period	9	3	1	19	0.05
Evidence quality rating scheme	9	0	1	22	0.05
Description of evidence combination	9	0	1	21	0.09
Strength of recommendation rating scheme	9	1	1	20	0.13
Methods used to reach judgment	9	0	3	17	0.13
Update plan	9	4	0	18	0.13
Linkage between recommendations and evidence	9	2	1	20	0.13
Recommendation strength	9	1	1	21	0.29
Technical report	9	4	2	17	0.29
Main focus	9	4	4	15	0.59
Quantification of harm and benefit	8	4	7	12	0.65
Funding	8	0	5	18	0.13
Conflict of interest	8	1	1	21	0.13
Specification of harm and benefit	8	1	4	18	0.22
Member expertise	8	1	6	16	0.24
Role of sponsor	8	1	5	17	0.29
Peer review	8	2	3	18	0.29
Reference	8	1	5	17	0.29
Definitions	8	2	3	18	0.29
Outcomes	8	3	5	15	0.49
Developer name	8	5	3	15	0.59
Status	7	4	7	12	0.65
Objectives	7	4	5	14	0.79
Number of source documents	7	0	9	14	0.29
Role of patient preference	7	3	5	15	0.29
Committee member	7	5	6	12	0.39
Release date	7	1	8	14	0.48
Alternative strategies	7	4	6	13	0.48
Care setting	7	6	2	15	0.59

* This table displays the panelists' (second round) ratings for items considered necessary to establish guideline validity. For each item, the median rating, the frequency of scores by tertile (i.e., the number of panelists selecting ratings of 1 to 3, 4 to 6, and 7 to 9) and the calculated disagreement index are shown. For a small number of items, the number of ratings does not sum to 23 because of temporary absences of some panel members.

Appendix Table 2. Items Ranked Necessary for Guideline Usability, Median Ratings, Distribution of Ratings by Tertiles, and Disagreement Index

Item	Necessity for Usability (Median Score)	Ratings			Disagreement Index
		1-3	4-6	7-9	
Main focus	9	1	0	22	0.00
Eligibility	9	0	0	23	0.00
Recommendation	9	0	0	23	0.00
Reason	9	2	1	20	0.19
Recommendation strength	9	1	2	20	0.19
Users	8	1	3	19	0.29
Care setting	8	2	3	18	0.29
Definitions	8	0	2	21	0.29
Objectives	8	4	5	14	0.37
Availability of guideline	8	5	2	15	0.43
Updating plan	8	6	3	13	0.97
Structured abstract	7	2	4	17	0.16
Patient resource	7	2	4	17	0.16
Expected barriers	7	2	4	14	0.22
Quality measures	7	1	7	12	0.37
Pilot testing	7	3	6	14	0.37
Outcomes	7	3	6	14	0.49
Quick reference guide	7	2	6	15	0.49
Alternative strategies	7	4	7	12	0.52
Strength of recommendation rating scheme	7	4	7	11	0.56
Algorithm	7	5	5	13	0.65
Linkage between recommendation and evidence	7	6	3	14	0.83

* This table displays panelists' (second round) ratings for items considered necessary to promote guideline usability. For each item, the median rating, the frequency of scores by tertile (i.e., the number of panelists selecting ratings of 1 to 3, 4 to 6, and 7 to 9) and the calculated disagreement index are shown. For a small number of items, the number of ratings does not sum to 23 because of temporary absences of some panel members.