AGREE II: advancing guideline development, reporting and evaluation in health care

Melissa C. Brouwers PhD, Michelle E. Kho BHSc(PT) MSc, George P. Browman MD MSc, Jako S. Burgers MD PhD, Francoise Cluzeau PhD, Gene Feder MD, Béatrice Fervers MD PhD, Ian D. Graham PhD, Jeremy Grimshaw MBChB PhD, Steven E. Hanna PhD, Peter Littlejohns MD, Julie Makarski BSc, Louise Zitzelsberger PhD, for the AGREE Next Steps Consortium

∞∞ See related research articles by Brouwers and colleagues, available at www.cmaj.ca

Clinical practice guidelines, which are systematically developed statements aimed at helping people make clinical, policy-related and system-related decisions,[1,2] frequently vary widely in quality.[3,4] A strategy was needed to differentiate among guidelines and ensure that those of the highest quality are implemented.

An international team of guideline developers and researchers, known as the AGREE Collaboration (Appraisal of Guidelines, Research and Evaluation), was established to create a generic instrument to assess the process of guideline development and reporting of this process in the guideline. Based on rigorous methodologies, the result of the collaboration's efforts was the original AGREE instrument, which is a 23-item tool comprising six quality-related domains that was released in 2003 (www.agreetrust.org).

As with any new assessment tool, ongoing development was required to improve its measurement properties, usefulness to a range of stakeholders and ease of implementation. Over the years, a number of issues were identified. For example, the original four-point response scale used to answer each item of the AGREE instrument is not in compliance with methodologic standards of health measurement design. This noncompliance threatens the performance and reliability of the instrument.[5] In addition, data on the usefulness of the AGREE items has never been gathered systematically from the perspectives of different groups of users. Further, we were interested in identifying strategies to make the evaluation process more efficient, such as reducing the number of items or the number of required raters, while ensuring the instrument was reliable and valid. Therefore, an exploration of the role of shorter versions of the AGREE instrument, comprising fewer items that are tailored to the unique priorities of different stakeholders, was warranted. Finally, there was a need to establish the fundamentals of construct validity — in other words, whether the AGREE items could measure what they purport to measure, and that is variability in quality of practice guidelines.

Key points

- AGREE II (Appraisal of Guidelines, Research and Evaluation), which comprises 23 items and a user's manual, offers refinements of a new way to develop, report and evaluate practice guidelines.
- Key changes from the original version include a new seven-point response scale, with modifications to half of the items, and a new user's manual.
- AGREE II is available online at the AGREE Research Trust (www.agreetrust.org).

Redesign of AGREE

In response to these issues, the AGREE Next Steps Consortium was established and undertook two studies.[6,7] As part of the first study, the consortium introduced a new seven-point response scale and evaluated its performance and measurement properties, analyzed the usefulness of the AGREE items for decisions made by different stakeholders, and systematically elicited stakeholders' recommendations for changes to the AGREE items and domains.[6] In the second study, the consortium evaluated the construct validity of the tool and designed and evaluated new supporting documentation aimed at facilitating efficient and accurate use of the tool.[7]

The following key findings emerged from the two studies:

- Ratings of the quality of the AGREE domains are good predictors of outcomes associated with implementation of guidelines.[6]
- Participants (i.e., guideline developers or researchers, policy-makers, and clinicians) evaluated AGREE items and domains as very useful, but no differences emerged in ratings of usefulness among groups,[6]

From McMaster University (Brouwers, Kho, Hanna, Makarski); the Program in Evidence-based Care, Cancer Care Ontario (Brouwers), Hamilton, Ont.; British Columbia Cancer Agency (Browman), Victoria, BC; the Dutch Institute for Healthcare Improvement CBO and IQ Healthcare (Burgers), Radboud University Nijmegen Medical Centre, the Netherlands; St. George's University of London (Cluzeau), London, UK; the University of Bristol (Feder), Bristol, UK; Unité Cancer et Environement (Fervers), Université de Lyon – Centre Léon Bérard, Université Lyon 1, EA 4129, Lyon, France; the Canadian Institutes of Health Research (Graham), Ottawa, Ont.; the Ottawa Hospital Research Institute (Grimshaw), Ottawa, Ont.; the National Institute for Health and Clinical Excellence (Littlejohns), London, UK; and the Canadian Partnership Against Cancer (Zitzelsberger), Ottawa, Ont.

CMAJ 2010. DOI:10.1503/cmaj.090449

DOI:10.1503/cmaj.090449

- No evidence exists to direct the development of shorter abridged versions of the instrument.[6]
- The psychometric properties of the seven-point response scale are promising.[6]
- The instrument successfully differentiates between high- and low-quality guideline content.[7]
- The new user's manual is well received by users.[7]
- Users provided considerable feedback on how to improve the instrument and the user's manual.[6,7]

Based on these results and three rounds of interpretation and consensus by the consortium, several refinements were made to the items and supporting documents, culminating in the release of AGREE II, which consists of 23 items, two overall assessment items and a user's manual (see Appendix 1, available at www.cmaj.ca/cgi/content/full/cmaj.090449/DC1).

Changes to AGREE II items

The 23 items in AGREE II are grouped into the same six domains as in the original AGREE instrument. These domains are scope and purpose, stakeholder involvement, rigour of development, clarity of presentation, applicability, and editorial independence. The key changes from the original document involved refinements to the purpose, response scale and items of the instrument.

The purpose of the AGREE II is more explicitly stated. The new version of the instrument is designed to assess the quality of practice guidelines across the spectrum of health, provide direction on guideline development, and guide what specific information ought to be reported in guidelines. The four-point response scale was replaced by a seven-point response scale, in compliance with key methodologic principles of test construction.[5] A score of 1 indicates an absence of information or that the concept is very poorly reported. A score of 7 indicates that the quality of reporting is exceptional and all of the criteria and considerations articulated in the user's manual were met. A score between 2 and 6 indicates that the reporting of the AGREE II item does not fully meet criteria or considerations. As more criteria are met and more considerations addressed, item scores increase (see user's manual below). Finally, modifications, deletions and additions were made to approximately half of the original 23 items (Table 1).

Changes to the AGREE II User's Manual

The user's manual (Appendix 1) was rewritten and extended with the following information linked to each item:

- Explicit descriptors for the different levels on the new seven-point scale
- A description that defines each concept underlying the item and inclusion of specific examples
- Direction on common places to look for desired information within the guideline document or accompanying documentation
- A list of common terms or labels to represent the concept
- Guidance on how to rate the item, including criteria and considerations. Criteria refer to explicit elements that reflect the operational definition of each item. Considerations aim to provide information on the nuances of the assessment.

The consortium recommends that the AGREE II replace the original AGREE instrument[8] as the preferred instrument for guideline development, reporting and evaluation. We used high-quality methods to direct the improvements made, with strong empirical evidence supporting the changes.[6,7]

Knowledge gaps

As with the first version of the AGREE, the items and domains in AGREE II focus on methodologic issues relevant to guideline development and reporting. However, they do not evaluate the clinical appropriateness or validity of the recommendations themselves. While rigorous development and explicit reporting are necessary, they do not guarantee optimal and acceptable recommendations or better health outcomes for patients and populations.[9,10] The new item assessing the description of strengths and limitations of the body of evidence (i.e., item 9) can be considered as a precursor for clinical validity or appropriateness of the recommendations. The consortium is targeting this area as its next priority for further study in the AGREE A3 initiative. This research initiative, funded by the Canadian Institutes of Health Research, is focused on the application, appropriateness and implementability of recommendations in clinical practice guidelines.

Similarly, some of the concepts in AGREE II could be improved. For example, the consortium considerably debated the representation of patient–public engagement in guideline development, as well as the items related to applicability and implementability in the instrument. These areas are also being targeted for future research.

Using AGREE II

Depending on the structure and length of the guideline document, quality-related assessment of a guideline using AGREE II will take 1.5 hours, on average, per appraiser. Although basic knowledge of the principles of evidence-based decision-making and health care methodology can facilitate its use, the new user's manual should allow novices to use the instrument with confidence. Furthermore, although content-specific expertise on the topic of a guideline is not necessary, it may improve the ease of interpretation of the findings. At this time, we recommend that at least two appraisers, and preferably four, rate each guideline to ensure sufficient reliability as the consortium continues its formal reliability testing.

The AGREE II has been used to evaluate several hundred guidelines related to the control of cancer (www.cancerview.ca; select "Services" in the menu bar and click on the "SAGE" link). It will be available on the AGREE Research Trust website (www.agreetrust.org).

AGREE II has myriad uses. Guideline developers can incorporate the concepts of the AGREE II framework into their development protocols, procedural documents and reporting templates. The instrument can also be used to evaluate the quality of guidelines that are candidates for use in clin-

Table 1: Comparison of original AGREE and AGREE II

Original AGREE Item	AGREE II Item
Domain 1: Scope and purpose	
The overall objective(s) of the guideline is (are) specifically described.	No change
The clinical question(s) covered by the guideline is (are) specifically described.	The health question(s) covered by the guideline is (are) specifically described.
The patients to whom the guideline is meant to apply are specifically described.	The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.
Domain 2: Stakeholder involvement	
The guideline development group includes individuals from all the relevant professional groups.	No change
The patients' views and preferences have been sought.	The views and preferences of the target population (patients, public, etc.) have been sought.
The target users of the guideline are clearly defined.	No change
The guideline has been piloted among end users.	Deleted and incorporated into description of item 19 in user guide
Domain 3: Rigour of development	
Systematic methods were used to search for evidence.	No change in item; renumbered to 7
The criteria for selecting the evidence are clearly described.	No change in item; renumbered to 8
	New item (9) added: The strengths and limitations of the body of evidence are clearly described.
The methods for formulating the recommendations are clearly described.	No change
The health benefits, side effects and risks have been considered in formulating the recommendations.	No change
There is an explicit link between the recommendations and the supporting evidence.	No change
The guideline has been externally reviewed by experts prior to its publication.	No change
A procedure for updating the guideline is provided.	No change
Domain 4: Clarity of presentation	
The recommendations are specific and unambiguous.	No change
The different options for management of the condition are clearly presented.	The different options for management of the condition or health issue are clearly presented.
Key recommendations are easily identifiable.	No change
Domain 5: Applicability	
The guideline is supported with tools for application.	• The guideline provides advice or tools on how the recommendations can be put into practice. • Change in domain (from Clarity of presentation) • Renumbered to 19
The potential organizational barriers in applying the recommendations have been discussed.	• The guideline describes facilitators of and barriers to its application. • Change in order (renumbered to 18)
The potential cost implications of applying the recommendations have been considered.	The potential resource implications of applying the recommendations have been considered.
The guideline presents key review criteria for monitoring or audit purposes.	The guideline presents monitoring or auditing criteria.
Domain 6: Editorial independence	
The guideline is editorially independent from the funding body.	The views of the funding body have not influenced the content of the guideline.
Conflicts of interest of members of the guideline development group have been recorded.	Competing interests of members of the guideline development group have been recorded and addressed.

ical practice, for formulating policy-related decisions or for adaptation of recommendations from one context to another. Journal editors and reviewers may use AGREE II as a framework to help define reporting requirements for guidelines submitted for publication, as has been done with the CONSORT (Consolidated Standards of Reporting Trials)[11] and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statements.[12] Finally, given the increasing number of guidelines developed worldwide, AGREE II provides a framework for reaching consensus on methodologic principles and reporting requirements for transnational cooperation.

Other tools to support the application of AGREE II are being developed, including a translation into French, an online version and an interactive online AGREE II training tool. The AGREE Research Trust, an independent body established in 2004, manages the interests of the AGREE project, supports an agenda of research regarding its development and formally endorses AGREE II.

The AGREE II, along with support tools and information about ongoing research-based initiatives associated with the instrument, is available at www.agreetrust.org.

This article has been peer reviewed.

Competing interests: Melissa Brouwers, Francoise Cluzeau and Jako Burgers are trustees of the AGREE Research Trust. No competing interests declared by the other authors.

Contributors: Melissa Brouwers conceived and designed the study, led the collection, analysis and interpretation of the data, and drafted the manuscript. All of the authors made substantial contributions to the study concept and the interpretation of the data, critically revised the article for important intellectual content and approved the final version of the manuscript to be published.

Acknowledgements: The AGREE Next Steps Consortium thanks the US National Guidelines Clearinghouse for helping to facilitate the identification of eligible practice guidelines for the research program. The consortium also thanks Ms. Ellen Rawski for her support on the project as research assistant from September 2007 to May 2008.

Funding: This work was supported by the Canadian Institutes of Health Research (CIHR). Michelle Kho is supported by a CIHR Fellowship Award (Clinical Research Initiative).

REFERENCES

1. Committee to Advise the Public Health Service on Clinical Practice Guidelines, Institute of Medicine. In: Field MJ, Lohr KN, editors. *Clinical practice guidelines: directions for a new program*. Washington (DC): National Academy Press; 1990.
2. Browman GP, Brouwers M, Fervers B, et al. Population-based cancer control and the role of guidelines-towards a "systems" approach. In: Elwood JM, Sutcliffe SB, editors. *Cancer control*. Oxford (UK): Oxford University Press; 2010.
3. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature [see comments]. *JAMA* 1999;281:1900-5.
4. Vigna-Taglianti F, Vineis P, Liberati A, et al. Quality of systematic reviews used in guidelines for oncology practice. *Ann Oncol* 2006;17:691-701.
5. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 3rd ed. Oxford (UK): Oxford University Press; 2003.
6. Brouwers MC, Kho ME, Browman GP, et al.; AGREE Next Steps Consortium. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. *CMAJ* 2010 May 31. [Epub ahead of print].
7. Brouwers MC, Kho ME, Browman GP, et al.; AGREE Next Steps Consortium. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ* 2010 May 31. [Epub ahead of print].
8. AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003;12:18-23.
9. Nuckols TK, Lim YW, Wynn BO, et al. Rigorous development does not ensure that guidelines are acceptable to a panel of knowledgeable providers. *J Gen Intern Med* 2008;23:37-44.
10. Watine J, Friedberg B, Nagy E, et al. Conflict between guideline methodologic quality and recommendation validity: a potential problem for practitioners. *Clin Chem* 2006;52:65-72.
11. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-91.
12. von Elm E, Altman DG, Egger M, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806-8.

Correspondence to: *Dr. Melissa C. Brouwers, Department of Oncology, McMaster University, Henderson site, G wing, Rm. 207, 711 Concession St., Hamilton ON L8V 1C3; mbrouwer@mcmaster.ca*

Members of the AGREE Next Steps Consortium: *Dr. Melissa C. Brouwers, McMaster University and Cancer Care Ontario, Hamilton, Ont.; Dr. George P. Browman, British Columbia Cancer Agency, Vancouver Island, BC; Dr. Jako S. Burgers, Dutch Institute for Healthcare Improvement CBO, and Radboud University Nijmegen Medical Centre, IQ Healthcare, Netherlands; Dr. Francoise Cluzeau, Chair of AGREE Research Trust, St. George's University of London, London, UK; Dr. Dave Davis, Association of American Medical Colleges, Washington, USA; Prof. Gene Feder, University of Bristol, Bristol, UK; Dr. Béatrice Fervers, Unité Cancer et Environement, Université de Lyon – Centre Léon Bérard, Université Lyon 1, EA 4129, Lyon, France; Dr. Ian D. Graham, Canadian Institutes of Health Research, Ottawa, Ont.; Dr. Jeremy Grimshaw, Ottawa Hospital Research Institute, Ottawa, Ont.; Dr. Steven E. Hanna, McMaster University, Hamilton, Ont.; Ms. Michelle E. Kho, McMaster University, Hamilton, Ont.; Prof. Peter Littlejohns, National Institute for Health and Clinical Excellence, London, UK; Ms. Julie Makarski, McMaster University, Hamilton, Ont.; Dr. Louise Zitzelsberger, Canadian Partnership Against Cancer, Ottawa, Ont.*