

Development of the AGREE II, part 2: assessment of validity of items and tools to support application

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

Previously published at www.cmaj.ca

∞ See related research article by Brouwers and colleagues in the July 13 issue (page 1045)

ABSTRACT

Background: We established a program of research to improve the development, reporting and evaluation of practice guidelines. We assessed the construct validity of the items and user's manual in the β version of the AGREE II.

Methods: We designed guideline excerpts reflecting high- and low-quality guideline content for 21 of the 23 items in the tool. We designed two study packages so that one low-quality and one high-quality version of each item were randomly assigned to each package. We randomly assigned 30 participants to one of the two packages. Participants reviewed and rated the guideline content according to the instructions of the user's manual and completed a survey assessing the manual.

Results: In all cases, content designed to be of high quality was rated higher than low-quality content; in 18 of 21 cases, the differences were significant ($p < 0.05$). The manual was rated by participants as appropriate, easy to use, and helpful in differentiating guidelines of varying quality, with all scores above the mid-point of the seven-point scale. Considerable feedback was offered on how the items and manual of the β -AGREE II could be improved.

Interpretation: The validity of the items was established and the user's manual was rated as highly useful by users. We used these results and those of our study presented in part 1 to modify the items and user's manual. We recommend AGREE II (available at www.agreetrust.org) as the revised standard for guideline development, reporting and evaluation.

As with any new assessment tool, ongoing development of the instrument was required to improve its measurement properties and advance the guideline enterprise. The AGREE Next Steps Consortium undertook a program of research to achieve these goals and create the next version of the tool, the AGREE II.⁸ The consortium completed two studies (parts 1 and 2). In part 1, also reported in this issue,⁹ we conducted an analysis of the performance of the new seven-point response scale, explored the usefulness of the AGREE items, and systematically identified ways in which the items and supporting document could be improved.

In part 2, reported here, we aimed to test the construct validity of the items and evaluate the new supporting documentation, which was intended to facilitate efficient and accurate application of the tool.

The validity of the original AGREE instrument was explored in three ways.⁵ Appraisers' attitudes about the instrument's usefulness and the helpfulness of the supporting documents (i.e., a user guide and training manual) were used as measures of face validity. The construct validity of the instrument was tested using three core hypotheses for each of the six domains; 3 of the possible 18 tests were supported. In retrospect, whether the hypotheses were generalizable across contexts was somewhat questionable. Finally, to establish criterion validity, correlations between users' overall global endorsement and quality ratings of individual items were calculated. Whether global endorsements were a reasonable proxy gold standard was somewhat questionable. Further, for both the construct validity and the criterion validity, guidelines chosen in these studies were nominated by members of the research team as representing a range of quality, creating significant risks of bias.

For clinical practice guidelines to achieve their full potential as tools to assist in clinical, policy-related and system-level decisions,¹⁻³ they need to be of high quality and developed using rigorous methods.⁴ Thus, strategies are required to facilitate the development and reporting of guidelines and tools able to distinguish guidelines of varying quality. The AGREE Collaboration (Appraisal of Guidelines, Research and Evaluation) was the first to create a generic tool to assess the process of guideline development and reporting,^{5,6} and it quickly became the standard for guideline evaluation.⁷

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 Environnement (Fervers), Université de Lyon – Centre Léon Bérard, Université Lyon 1, EA 4129, Lyon, France; and the Canadian Institutes of Health Research (Graham), Ottawa, Ont.

CMAJ 2010. DOI:10.1503/cmaj.091716

Together, these findings and methodological limitations illustrated the need for additional work to test and establish the instrument's validity. The most fundamental concept of construct validity, in particular, had not been yet addressed — are guidelines known to be of higher quality rated more favourably using the AGREE instrument than guidelines known to be of lower quality? In addition, no study to date has tested specifically whether the instructions for applying the tool are perceived to be appropriate, implementable, and helpful in differentiating among guidelines of varying quality. These perceptions are important components that contribute to the face validity of the tool.

We tested two specific research questions in this study. First, do the items in β -AGREE II differentiate between guideline content of known, varying quality? Second, is the new user's manual perceived by users as appropriate, easy to apply and helpful in differentiating good quality guidelines from poor quality guidelines?

Methods

Design and sample size

We used a two-level factorial design. Guideline quality (i.e., high and low) was the between-subjects factor. We sought to recruit 15 participants per group, to enable a two-sided test to have 80% power to detect an advantage of as little as one point on the seven-point scale between the high-quality and low-quality groups.

Participants

A convenience sample of guideline developers, researchers and clinicians was recruited to participate in this study. They were recruited from the Program in Evidence-based Care of Cancer Care Ontario, the Canadian Partnership Against Cancer and international coinvestigators of the research team. We oversampled by 33% to ensure we would receive data for our targeted sample size of 30.

Table 1: Examples of guideline excerpts purposefully designed to represent high- and low-quality content (part 1 of 2)

AGREE Item	High	Low
Domain 1: Scope and purpose		
3. The patients to whom the guideline is meant to apply are specifically described.	This recommendation applies to adult patients (> 18 years) with single or multiple radiographically confirmed bone metastases of any histology corresponding to painful areas in previously non-irradiated areas without pathologic fractures or spinal cord/cauda equine compression. It does not apply to the management of malignant primary bone tumour.	This recommendation applies to patients with bone metastases. It does not apply to the management of malignant primary bone tumour.
Domain 2: Stakeholder involvement		
6. The target users of the guideline are clearly defined.	This provincial guideline was initiated to summarize the evidence and to provide recommendations on the preferred standard radiotherapy fractionation schedule for the treatment of painful bone metastases. The intended audience for this radiotherapy guideline includes radiation oncologists and physicians who may refer patients for radiotherapy.	This provincial guideline was initiated to summarize the evidence and to provide recommendations on the preferred standard radiotherapy fractionation schedule for the treatment of painful bone metastases. This radiotherapy guideline was primarily intended for clinicians.
Domain 3: Rigour of development		
8. Systematic methods were used to search for evidence.	A search of MEDLINE, EMBASE, and the Cochrane Library (2002, Issue 4) was conducted to find randomized trials published between January 1998 and December 2002 using MeSH headings (radiotherapy, radiotherapy dosage, dose fractionation, bone neoplasms/sc [Secondary], explode Clinical Trials, clinical trial [publication type]), text words (bon; osseous, metasta; radiotherapy, irradiation, radiation, pain, analgesi; trial, and study) without language restrictions. Proceedings of the meetings of ASTRO (2001–2002) and the Canadian Association of Radiation Oncologists (2000), as well as reference lists of papers and review articles, were scanned for additional citations. Please see Appendix A for a detailed listing of the search strategy.	The Canadian Medical Association Infobase (http://www.cma.ca/cpgs/index.asp), the National Guidelines Clearinghouse (http://www.guideline.gov/index.asp) were searched for existing evidence-based practice guidelines prior to the development of this guideline report. A search of MEDLINE, EMBASE, and the Cochrane Library was conducted to find trials without language restrictions for this guideline. Terms specific to radiotherapy and metastatic disease were used. Proceedings of the meetings of ASTRO and the Canadian Association of Radiation Oncologists, as well as reference lists of papers and review articles, were scanned for additional citations. The search strategy is available upon request from the authors.

Table 1: Examples of guideline excerpts purposefully designed to represent high- and low-quality content (part 2 of 2)

AGREE Item	High	Low
Domain 4: Clarity of presentation		
15. The recommendations are specific and unambiguous.	For patients where the treatment objective is pain relief, a single 8 Gy treatment, prescribed to the appropriate target volume, is recommended as the standard dose-fractionation schedule for the treatment of symptomatic and uncomplicated bone metastases. There is insufficient evidence at this time to make a dose-fractionation recommendation for other treatment indications, such as long term disease control for patients with solitary bone metastasis, prevention/ treatment of cord compression, prevention/treatment of pathological fractures, and treatment of soft tissue masses associated with bony disease.	A single treatment, prescribed to the appropriate volume, is recommended as the standard dose-fractionation schedule for the treatment of symptomatic and uncomplicated bone metastases. There is insufficient evidence at this time to make a dose-fractionation recommendation for other treatment indications.
Domain 5: Applicability		
19. The potential organizational barriers in applying the recommendations have been discussed.	Based on written feedback from the external review, the radiation oncologists identified the need for a province-wide electronic medical record to identify areas where previous radiation occurred. Otherwise, no additional barriers impacting the implementation of the guideline were identified.	Appendix A outlines a list of policy considerations for implementation of this guideline.
Domain 6: Editorial independence		
23. Conflicts of interest of guideline development members have been recorded.	Members of the Supportive Care Guidelines Group disclosed potential conflict of interest information on standardized forms, addressing economic and academic conflicts (Appendix). The Supportive Care Guidelines Group Chair reviewed all reported conflicts in light of this guideline topic. One group member, who reported research funding from a pharmaceutical company producing anti-emetics, was excused from the development of recommendations. No other group members had conflicts of interest which precluded participation in the development of this guideline.	Members of the Supportive Care Guidelines Group disclosed potential conflicts of interest information. One group member, who disclosed pharmaceutical funding, was excused from the development of recommendations.

Creating guideline content of varying quality for assessment

An existing cancer-related guideline developed by an established guidelines program¹⁰ was used as the source guideline from which we purposefully designed excerpts of content of varying quality to reflect 21 of the 23 AGREE items. This guideline was chosen because it was of mid-range quality (as determined by two independent appraisers using the original AGREE instrument [MK, JM]), and thus enabled us to easily craft higher-quality content and lower-quality content. The AGREE instrument had not been explicitly used to facilitate its development. We did not test item 16 (i.e., the different options for management of the condition are clearly presented) because the source document only focused on one effective treatment option, and we did not want to introduce a recommendation that was fictitious or not based on evidence. Therefore, this item was excluded. Item 17 (i.e., key recommendations are easily identifiable) was not manipulated, because we

were presenting guideline excerpts related to each of the items one at a time rather than embedding all of the manipulated content to create a whole version of a guideline. Therefore, all participants received the same content as in the original source guideline for item 17.

In crafting guideline content, our objective was to reflect more nuanced differences that might typically be seen between guidelines rather than extreme examples of high and low content (Table 1). For each item, a high-quality version and a low-quality version of the content was pilot-tested, reviewed and refined by three members of the team (MB, MK, ER). From there, two versions of a study package were created. Excerpts of high- and low-quality content were randomly assigned to each version of the study package using a random number generator, such that in each package, only one version (high or low) was included for each item (except item 17 as per above). Version 1 included 14 high-quality items and 7 low-quality items. Version 2 included 7 high-

quality items and 14 low-quality items (i.e., the inverse of version 1 in quality).

Administration

After obtaining ethics approval, we distributed personalized letters of invitation and then reminders via email to participant-candidates. Participants were assigned a unique identifier code and were blinded to group and purpose of the study. They were randomly assigned to one of the two versions of the study package and sent a confidential username and password to access the web-based study platform. Once logged on, participants were asked to assess the guideline content, using the β -AGREE II items and user's manual to guide their assessment. Content relevant to each item was presented sequentially. Participants were then asked to fill out a survey to assess the usefulness of the user's manual.

Measures

β -AGREE II

The β -AGREE II comprised an item set and a user's manual. The set included the 23 items clustered into the six quality domains from the original AGREE instrument. However, the items were answered using the new seven-point response scale which was tested in part 1^{9,11} and replaces the original four-point scale.⁵ The most significant change to the β -AGREE II is the new user's manual that replaces the original supporting documentation. The user's manual is an extensively restructured revision of the original user guide and training manual. For each of the 23 items, the user's manual provides a definition of the concept, specific examples, suggestions for where to find the information in the guideline and clear direction (including criteria and considerations) on how to score the item.

Table 2: Demographic characteristics of participants

Characteristic	Group; no. of participants		
	Group 1 <i>n</i> = 16	Group 2 <i>n</i> = 14	Overall <i>n</i> = 30
Educational background or level			
Physician	2	4	6
Registered nurse	1	2	3
Bachelor's degree	3	2	5
Master's degree	7	7	14
PhD	6	2	8
Other	1	1	2
Primary role			
Clinician	0	3	3
Guideline developer or researcher	16	9	25
Policy- or decision-maker	0	2	2
Engaged in clinical practice	2	6	8
Engaged in clinical or health services research	12	11	23

Survey to assess the user's manual

A three-item scale was used to gather feedback on the user's manual based on previously published measures of clinical sensibility.¹² For each item represented in the manual, participants were asked to rate their agreement using a seven-point scale (i.e., 1 = strongly disagree, 7 = strongly agree) regarding item appropriateness, ease of application, and capacity to facilitate discrimination between good- and poor-quality guidelines. Participants were also asked to provide written feedback (i.e., qualitative, open-ended) on how the user's manual could be improved.

Analysis of data

To assess whether differences in item ratings existed between guideline content designed to be of high and low quality and to correct for multiple comparisons, we conducted a multivariable analysis of variance (MANOVA). This analysis included the 21 manipulated items as dependent measures. We report the results of both the MANOVA and the univariable analysis. A separate analysis of variance (ANOVA) was conducted to compare scores for item 17 between the two groups, where no difference was expected.

Descriptive statistics were calculated for each of the three assessment measures of the user's manual. For exploratory purposes, total scores were added across the AGREE items for each of the three assessment measures of the user's manual, and a one-way ANOVA was undertaken to determine if differences in overall assessments existed between version 1 and version 2 of the study packages. Our hypothesis was that no differences would exist between the two groups.

Results

Participants

Of 41 invited participants, we received data from 30 people (for a response rate of 73%), which met our requirement for sample size. One data point was missing for two of these participants. The demographic characteristics of participants are provided in Table 2. Almost three quarters of participants identified themselves as researchers, 28% engaged in clinical practice, and 83% were participants in some aspect of the guideline enterprise.

Assessment of guideline excerpts

Multivariable analysis of variance yielded a significant main effect for guideline quality ($p = 0.005$). Univariable analyses yielded significant differences in scores for 18 of the 21 manipulated items (Table 3). In all cases, content designed to be of high quality was rated higher than content designed to be of low quality. Whereas the mean scores were in the correct direction, the three items that did not yield significant univariable differences between the high- and low-quality item versions were item 10 (i.e., methods for formulating recommendations are clearly described), item 11 (i.e., health benefits, side effects and risks have been considered in formulating recommendations); and item 12 (i.e., there is an explicit link between the recommendations and the supporting evidence). As expected, item 17 (i.e., key recommenda-

tions are easily identifiable), for which participants in both groups received the same version, did not yield a significant difference between the high- and low-quality versions ($p > 0.05$) in the separate ANOVA.

Assessment of the user's manual

The results of the three usability assessments of the β version of the user's manual across each of the AGREE II items are presented in Table 4. Mean scores were high, with a range of 5.43–6.43 for the measure of appropriateness, 5.33–6.33 for that of ease of application, and 5.21–6.27 for that of ability to discriminate. No differences in total assessment scores were found between study package 1 and study package 2 ($p > 0.05$).

Final refinements

We received considerable written feedback from participants, including specific suggestions for improvements to the instrument (not presented). All feedback, in combination with feedback received in part 1, was formally discussed by the AGREE Next Steps Consortium, and final modifications were made to create the AGREE II.⁸

Interpretation

This study represents the first systematic analysis of the construct validity of the AGREE items. Our results show the capacity of the items to detect differences in guideline quality which the

Table 3: Differences in AGREE II β version scores as a function of high- and low-quality content

Domain	Item	High quality, mean score (SD)	Low quality, mean score (SD)	<i>p</i> value
Scope and purpose	1. The overall objective(s) of the guideline is (are) specifically described	5.67 (1.29)	3.92 (1.26)	0.001
	2. The clinical question(s) covered by the guideline is (are) specifically described	5.93 (1.71)	4.62 (1.19)	0.028
	3. The patients to whom the guideline is meant to apply are specifically described	6.47 (1.06)	4.15 (1.07)	< 0.001
Stakeholder involvement	4. The guideline development group includes individuals from all the relevant professional groups	6.77 (0.44)	4.20 (1.78)	< 0.001
	5. The patients' views and preferences have been sought	6.00 (1.25)	4.38 (1.19)	0.002
	6. The target users of the guideline are clearly defined	6.20 (0.86)	4.77 (1.79)	0.010
	7. The guideline has been piloted among end users	6.73 (0.59)	5.08 (1.26)	< 0.001
Rigour of development	8. Systematic methods were used to search for evidence	6.69 (0.63)	4.60 (1.64)	< 0.001
	9. The criteria for selecting the evidence are clearly described	6.00 (0.82)	4.07 (1.67)	0.001
	10. The methods for formulating the recommendations are clearly described	5.47 (1.51)	4.92 (1.26)	0.314
	11. The health benefits, side effects and risks have been considered in formulating the recommendations	4.27 (1.83)	3.85 (1.57)	0.524
	12. There is an explicit link between the recommendations and the supporting evidence	5.23 (1.74)	4.20 (1.94)	0.153
	13. The guideline has been externally reviewed by experts prior to its publication	6.20 (1.08)	4.54 (1.20)	0.001
	14. A procedure for updating the guideline is provided	6.07 (1.62)	4.00 (1.35)	0.001
Clarity of presentation	15. The recommendations are specific and unambiguous	5.73 (1.03)	3.92 (0.95)	< 0.001
	16. The different options for management of the condition are clearly presented	NA	NA	NA
	17. Key recommendations are easily identifiable*	5.53 (1.96)	5.57 (0.94)	0.948
	18. The guideline is supported with tools for application	6.67 (0.72)	3.62 (1.71)	< 0.001
Applicability	19. The potential organizational barriers in applying the recommendations have been discussed	5.00 (1.60)	3.46 (1.56)	0.017
	20. The potential cost implications of applying the recommendations have been considered	4.69 (1.65)	2.87 (2.07)	0.017
	21. The guideline presents key review criteria for monitoring and/or audit purposes	6.00 (1.16)	4.20 (1.66)	0.003
Editorial independence	22. The guideline is editorially independent from the funding body	6.92 (0.28)	6.00 (1.13)	0.008
	23. Conflicts of interest of guideline development members have been recorded	6.80 (0.56)	5.46 (1.27)	0.001

Note: NA = not applicable, SD = standard deviation.

*Participants in both groups received the same version of guideline text for assessment.

instrument purports to measure. Prior to this study, development work on the AGREE instrument had been done on real guidelines considered by researchers to reflect a range of quality. However, those results were confounded because the AGREE instrument served both as the measurement tool to evaluate the guidelines and as the object of the study intended to assess the instrument's capacity to evaluate guidelines. Differences in what looked like quality might have been confounded with other differences (e.g., guideline topic, intervention, organization, differ-

ences among researchers on criteria used to nominate good- and poor-quality exemplars). In this study, by removing these potential confounders, we were able to explicitly test the capacity and predictability of the AGREE items to distinguish among guideline information of known varying quality. By manipulating the quality of guideline excerpts, we have been able to determine how the scores relate to the operational definitions of the items.

Our results are encouraging, with all mean ratings falling in the intended direction, and 18 of the 21 means yielding sta-

Table 4: Assessment of usability of the β -AGREE II user manual

Domain	Item	Appropriate,* mean (SD)	Easy to apply,† mean (SD)	Facilitates successful discrimination,‡ mean (SD)
Scope and purpose	1. The overall objective(s) of the guideline is (are) specifically described	5.87 (1.53)	5.70 (1.34)	5.63 (1.45)
	2. The clinical question(s) covered by the guideline is (are) specifically described	5.90 (1.42)	5.50 (1.46)	5.60 (1.38)
	3. The patients to whom the guideline is meant to apply are specifically described	6.07 (1.41)	5.80 (1.45)	5.87 (1.50)
Stakeholder involvement	4. The guideline development group includes individuals from all the relevant professional groups	6.27 (1.23)	6.20 (1.27)	5.97 (1.38)
	5. The patients' views and preferences have been sought	6.23 (1.25)	5.87 (1.36)	6.03 (1.27)
	6. The target users of the guideline are clearly defined	6.20 (1.22)	5.87 (1.28)	5.83 (1.42)
	7. The guideline has been piloted among end users	6.03 (1.56)	5.80 (1.49)	5.67 (1.69)
Rigour of development	8. Systematic methods were used to search for evidence	6.40 (1.22)	6.07 (1.31)	6.27 (1.26)
	9. The criteria for selecting the evidence are clearly described	6.37 (1.22)	6.03 (1.27)	6.13 (1.20)
	10. The methods for formulating the recommendations are clearly described	5.70 (1.45)	5.50 (1.46)	5.67 (1.35)
	11. The health benefits, side effects and risks have been considered in formulating the recommendations	6.03 (1.30)	5.70 (1.42)	6.03 (1.25)
	12. There is an explicit link between the recommendations and the supporting evidence	5.63 (1.65)	5.33 (1.69)	5.67 (1.63)
	13. The guideline has been externally reviewed by experts prior to its publication	6.20 (1.42)	5.93 (1.44)	6.17 (1.23)
	14. A procedure for updating the guideline is provided	6.33 (1.18)	5.93 (1.44)	6.10 (1.30)
Clarity of presentation	15. The recommendations are specific and unambiguous	6.10 (1.24)	5.60 (1.40)	5.97 (1.27)
	16. The different options for management of the condition are clearly presented	5.59 (1.72)§	5.46 (1.57)§	5.61 (1.69)§
	17. Key recommendations are easily identifiable	6.13 (1.38)	6.03 (1.27)	5.77 (1.43)
Applicability	18. The guideline is supported with tools for application	5.80 (1.63)	5.77 (1.52)	5.21 (1.72)§
	19. The potential organizational barriers in applying the recommendations have been discussed	5.87 (1.33)	5.40 (1.54)	5.50 (1.41)
	20. The potential cost implications of applying the recommendations have been considered	5.60 (1.52)	5.70 (1.37)	5.60 (1.45)
	21. The guideline presents key review criteria for monitoring and/ or audit purposes	5.43 (1.74)	5.37 (1.54)	5.33 (1.69)
Editorial independence	22. The guideline is editorially independent from the funding body	6.13 (1.57)	6.33 (0.84)	6.20 (1.21)
	23. Conflicts of interest of guideline development members have been recorded	6.43 (1.22)	6.00 (1.39)	6.20 (1.21)

*The instructions for how to rate this item are appropriate.

†The instructions for how to rate this item are easy to apply.

‡The instructions for how to rate this item will facilitate successful discrimination between good and poor reporting in guidelines.

§The following observations are missing: Item 16: appropriate (3 observations), easy to apply (2 observations), successful discrimination (2 observations). Item 18: successful discrimination (1 observation).

tistically significant differences. In addition, this study established that the instructions of the β -AGREE II User Manual are appropriate, are easy to apply, and create confidence among users that good-quality guidelines will be differentiated from poor-quality guidelines.

Limitations

Our study has limitations. First, in testing of the β -AGREE II, participants were presented with excerpts of a guideline that reflected each item's concept and not an entire guideline. Whether the items would be sensitive in discriminating between differences in quality when users are presented with an entire guideline is a question for future research. Second, we chose a convenience sample of participants that was comprised primarily of guideline developers and researchers rather than a full range of potential users of AGREE II. As such, the generalizability of the findings may be limited. However, given that most of our participants (83%) were experienced in guideline development or research, they were uniquely situated as consumers who could be critical of the value of the user's manual. Third, although we met our sample-size goal of 30 participants for analytical purposes, this study is modestly sized. This fact may raise questions regarding the generalizability of our findings to a larger group of stakeholders. Fourth, by using a specialist (oncologic) guideline focused on one procedure as our source document, extrapolation of our findings to other clinical areas is contestable. Finally, for many of the items, the word count for high-quality items was larger than the word count for low-quality items. Thus, word count may be confounded with quality when interpreting the differences that emerged. This fact too may limit the generalizability of our findings.

Conclusion

Our study represents the first systematic assessment of the construct validity of the AGREE. Future research is warranted to reproduce these findings using a larger sample of stakeholders and including manipulated guideline content within the context of a whole report.

In combination with part 1,⁹ our results led to the final refinements and release of the AGREE II, the revised standard for guideline development, reporting and evaluation.⁸ The AGREE II is available at the website of the AGREE Research Trust (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 its assistance in the identification of eligible practice guidelines used in the research program of the consortium. The consortium also thanks Ms. Ellen Rawski for her support on the project as research assistant from September 2007 to May 2008.

Funding: This research was supported by the Canadian Institutes of Health Research (CIHR), which had no role in the design, analysis or interpretation of the data. 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. Whitworth JA. Best practices in use of research evidence to inform health decisions. *Health Res Policy Syst* 2006;4:11.
3. 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; 2009.
4. Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999;318:527-30.
5. The 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.
6. AGREE Research Trust. Hamilton (ON): The Trust; 2004. Available: www.agreetrust.org (accessed 2009 Sept. 17).
7. Vlayen J, Aertgeerts B, Hannes K, et al. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care* 2005;17:235-42.
8. Brouwers M, Kho ME, Browman GP, et al. for the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *CMAJ* 2010. In press.
9. Brouwers MC, Kho ME, Browman GP, et al. for the AGREE Next Steps Consortium. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. *CMAJ* 2010;182:1045-52.
10. Wu JS, Wong R, Johnston M, et al.; members of the Supportive Care Guidelines Group. *Radiotherapy fractionation for the palliation of uncomplicated painful bone metastases: practice guideline report #13-2*. Toronto (ON): Program in Evidence-Based Care, Cancer Care Ontario; 2003. Available: www.cancercare.on.ca/common/pages/UserFile.aspx?fileId=13922 (accessed 2009 Sept. 17).
11. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 3rd ed. Oxford (UK): Oxford University Press; 2003.
12. Atkins D, Briss PA, Eccles M, et al. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005;5:25.

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 Environnement, 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.