



Development and Validation of a Tool to Assess the Quality of Clinical Practice Guideline Recommendations

Melissa C. Brouwers, PhD; Karen Spithoff, MSc; Kate Kerkvliet, BScN; Pablo Alonso-Coello, MD, PhD; Jako Burgers, MD, PhD; Francoise Cluzeau, PhD; Beatrice Férvors, MD, PhD; Ian Graham, PhD; Jeremy Grimshaw, MBChB, PhD; Steven Hanna, PhD; Monika Kastner, PhD; Michelle Kho, PhD; Amir Qaseem, MD, PhD; Sharon Straus, MD, MSc; Ivan D. Florez, MD, MSc

Abstract

IMPORTANCE Clinical practice guidelines (CPGs) may lack rigor and suitability to the setting in which they are to be applied. Methods to yield clinical practice guideline recommendations that are credible and implementable remain to be determined.

OBJECTIVE To describe the development of AGREE-REX (Appraisal of Guidelines Research and Evaluation-Recommendations Excellence), a tool designed to evaluate the quality of clinical practice guideline recommendations.

DESIGN, SETTING, AND PARTICIPANTS A cross-sectional study of 322 international stakeholders representing CPG developers, users, and researchers was conducted between December 2015 and March 2019. Advertisements to participate were distributed through professional organizations as well as through the AGREE Enterprise social media accounts and their registered users.

EXPOSURES Between 2015 and 2017, participants appraised 1 of 161 CPGs using the Draft AGREE-REX tool and completed the AGREE-REX Usability Survey.

MAIN OUTCOMES AND MEASURES Usability and measurement properties of the tool were assessed with 7-point scales (1 indicating strong disagreement and 7 indicating strong agreement). Internal consistency of items was assessed with the Cronbach α , and the Spearman-Brown reliability adjustment was used to calculate reliability for 2 to 5 raters.

RESULTS A total of 322 participants (202 female participants [62.7%]; 83 aged 40-49 years [25.8%]) rated the survey items (on a 7-point scale). All 11 items were rated as easy to understand (with a mean [SD] ranging from 5.2 [1.38] for the alignment of values item to 6.3 [0.87] for the evidence item) and easy to apply (with a mean [SD] ranging from 4.8 [1.49] for the alignment of values item to 6.1 [1.07] for the evidence item). Participants provided favorable feedback on the tool's instructions, which were considered clear (mean [SD], 5.8 [1.06]), helpful (mean [SD], 5.9 [1.00]), and complete (mean [SD], 5.8 [1.11]). Participants considered the tool easy to use (mean [SD], 5.4 [1.32]) and thought that it added value to the guideline enterprise (mean [SD], 5.9 [1.13]). Internal consistency of the items was high (Cronbach $\alpha = 0.94$). Positive correlations were found between the overall AGREE-REX score and the implementability score ($r = 0.81$) and the clinical credibility score ($r = 0.76$).

CONCLUSIONS AND RELEVANCE This cross-sectional study found that the AGREE-REX tool can be useful in evaluating CPG recommendations, differentiating among them, and identifying those that are clinically credible and implementable for practicing health professionals and decision makers who use recommendations to inform clinical policy.

JAMA Network Open. 2020;3(5):e205535. doi:10.1001/jamanetworkopen.2020.5535

Open Access. This is an open access article distributed under the terms of the CC-BY License.

Key Points

Question Is it possible to create a tool to specifically evaluate the quality of clinical practice guideline recommendations?

Findings In this cross-sectional study of 322 international stakeholders, the Appraisal of Guidelines Research and Evaluation-Recommendations Excellence (AGREE-REX) tool was developed to appraise guidelines for clinical practice. All participants rated the tool as usable and agreed that it represents a valuable addition to the clinical practice guidelines enterprise.

Meaning A panel of stakeholders agrees that the AGREE-REX tool may provide information about the methodologic quality of guideline recommendations and may help in the implementation of clinical practice guidelines.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Introduction

Clinical practice guidelines (CPGs) are systematically developed statements informed by a systematic review of evidence and an assessment of the benefits and harms of care options designed to optimize patient care.¹⁻³ The potential benefits of CPGs, however, are only as good as their quality. Appropriate methods and rigorous development strategies are important factors in the successful implementation of CPG recommendations.⁴⁻¹⁰ Not all CPGs are alike; their quality is variable and often falls short of reported goals.¹¹⁻¹⁹

The Appraisal of Guidelines, Research and Evaluation revision (AGREE II) tool has become an accepted international resource to evaluate the quality of CPGs and to provide a methodologic framework to inform CPG development, reporting, and evaluation.^{5-7,20-22} The AGREE II tool targets the entire CPG development process and all components of the CPG report: the articulation of scope and practice, who is involved, methods used, applicability, editorial independence, and clarity.

Since the release of AGREE II, studies have reported that high AGREE II scores do not guarantee that the resulting CPG recommendations are optimal.²³⁻²⁷ For example, Nuckols et al²⁴ evaluated the technical quality and acceptability of 5 musculoskeletal CPGs. Use of the AGREE II tool resulted in high quality scores (eg, rigor domain scores >80%). However, participants reported that the CPGs omitted common clinical situations and contained recommendations of uncertain clinical validity. Similar results have been found with disability-related CPGs.²⁶

These studies suggest that a distinction exists between user perceptions of a CPG report and the report's recommendations. Hence, a barrier may exist if users rely solely on the AGREE II quality scores in making decisions about which CPG recommendations to implement or which CPGs to adapt to a specific context. For example, if a CPG provides insufficient information about the values of patients, health care professionals, and funders, or there is a lack of alignment across different viewpoints, that CPG may yield recommendations that are difficult to use and implement, even if the evidence base is solid or the methods used to create the CPG are of high quality. The CPGs that address controversial issues in which values clash (eg, medically assisted dying) may be especially susceptible to this concern. Inadequate consideration of different perspectives and varied implementation concerns are a common limitation in CPG appraisal tools.²⁸

The development of AGREE II focused primarily on methodologic quality and internal validity of the CPG report and to a lesser extent on the external validity of the recommendations. A more thorough investigation of the implementation science literature and the usability and relevance of recommendations was warranted. Our international team of CPG developers and researchers created the AGREE-REX (Appraisal of Guidelines Research and Evaluation-Recommendations Excellence) tool to evaluate the quality of CPG recommendations specifically, defined as credible and implementable recommendations.

Methods

Development of Draft AGREE-REX

The development process used international standards of measurement design.²⁹ Our first step required identification of candidate items. This step was completed and is described in previous studies.^{30,31} In brief, a realist review was conducted to identify attributes of CPGs associated with the implementation of their recommendations. The review resulted in the Guideline Implementability for Decision Excellence Model (GUIDE-M) that was vetted by the international CPG community.³⁰ This multilevel model comprises 3 core tactics, 7 domains, and approximately 100 embedded components. The model was evaluated by 248 stakeholders from 34 countries and refined.

A core domain of the model (deliberations and contextualization) provided content coverage of our concept of CPG recommendation quality. The domain is composed of 3 subdomains, 11 attributes, and many subattributes and elements: clinical applicability (clinical, patient, and

implementability relevance), values (perspectives of patient, health care professional, population, policy, developer), and feasibility (local, novelty, resources).

We derived candidate items from these data that 15 international CPG stakeholders evaluated. We used this feedback to refine the content and create the Draft AGREE-REX, used in this study (eAppendix 2 in the [Supplement](#)). The Draft AGREE-REX comprises 11 items (4 themes) and 2 overall items.

Three response scales were designed to rate each item of the Draft AGREE-REX. Two mandatory 7-point response scales (with 1 indicating strongly disagree and 7 indicating strongly agree) asked appraisers to rate the extent to which quality criteria are reported in the CPG (documentation scale) and then used to inform the CPG recommendations (consideration scale). An optional 7-point scale asked appraisers whether the documented and considered information aligned with, and was suitable for use in, their context (suitability scale). This scale was designed for use only when CPG recommendations from an authoring group are being considered for endorsement, adaptation, or implementation by another group. Two overall items asked appraisers for their overall ratings of the implementability of the CPG recommendations and their overall ratings of the clinical credibility of the CPG recommendations. Each item was answered according to a 7-point scale.

Participants

To test the Draft AGREE-REX tool, a cross-sectional study design was used. The CPG users, developers, researchers, or trainees were eligible to participate. Between December 2015 and March 2017, advertisements to participate were distributed through professional organizations (eg, the Guidelines International Network) as well as through the AGREE Enterprise social media accounts and their registered users. Given the nature of the recruitment strategy and the substantial number of cross-postings, an accurate number of individuals the advertisements reached is not available. Completion of the study implied consent and participants were offered a CAD\$50 gift card. The study received ethics approval from the Hamilton Integrated Research Ethics Board.

The CPGs were selected from the National Guideline Clearinghouse of the Agency for Healthcare Research and Quality. Selection criteria were as follows: English language, published between 2013 and 2015, and length of core CPG document less than 50 pages.

The target sample size was calculated based on the interrater reliability outcome, assuming 2 raters per CPG, an intraclass correlation coefficient of 0.6, and a CI from 0.5 to 0.7. On the basis of these assumptions, 316 participants were required to appraise 158 CPGs. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology ([STROBE](#)) reporting guideline for cross-sectional studies.

Procedures

Participants were required to read a single CPG, evaluate the entire set of recommendations with the Draft AGREE-REX, and complete the AGREE-REX Usability Survey. Individuals who responded to the advertisement were sent an email with an invitation letter, an electronic copy of the Draft AGREE-REX, the CPG to which they were randomly assigned, and access to LimeSurvey to submit AGREE-REX appraisal scores and to complete the AGREE-REX Usability Survey. Reminder emails were sent to nonrespondents at 2-week intervals up to 3 times.

Using the three 7-point scales, participants were asked to rate the items, the instructions, the response scale, their ability to apply the tool, and its usefulness. For each Draft AGREE-REX item, ratings from the documentation scale and the considerations scale were calculated as a mean between the 2 appraisers. Strong positive correlations between the 2 rating scales emerged (defined as an $r > 0.90$), and analyses produced identical patterns of results.

An overall AGREE-REX score was calculated by adding the mean item scores from the consideration scale and scaling the total as a percentage of the maximum possible score. These scores were used to assess the tool's measurement properties. The AGREE-REX ratings of the CPGs appraised in the study have been reported.³⁰

Two research staff members (K.S and K.K) with formal training and experience independently evaluated all the CPGs with the AGREE II tool. The AGREE II tool comprises 23 items within 6 domains. Each item is answered using a 7-point agreement scale with higher ratings indicating higher CPG quality.⁵ The AGREE II domain scores were used as part of the analytical framework to assess the performance of the Draft AGREE-REX.

Statistical Analysis

Quantitative data were analyzed using SPSS software, version 24 (IBM Corp). Means and SDs for each of the items in the AGREE-REX Usability Survey were calculated. Cronbach α and correlations-if-item-deleted were calculated to assess the internal consistency of the items. Intraclass correlations were calculated for 2 to 5 appraisers using the Spearman-Brown reliability adjustment to assess the reliability of the overall AGREE-REX score.^{29,32,33} A 2-tailed $P < .05$ was considered as statistically significant.

Differentiating itself from the AGREE II tool, the AGREE-REX tool evaluates the quality of CPG recommendations, defined as the extent to which they are credible and implementable. Thus, to explore construct validity, correlations between the overall AGREE-REX score and the implementability score and the clinical credibility score were calculated, with the expectation that positive correlations would emerge. As an exploratory measure of discriminant validation, the correlations between the overall AGREE-REX score and AGREE II domain scores, assuming the mean scores across 4 raters and correcting for the attenuation in the correlation due to measurement error, were also calculated. The correlations of the former were expected to be larger than those of the latter. No standard for CPG recommendation quality currently exists; thus measures of criterion validity were not appropriate.^{23,32,33}

Participants provided written feedback, and themes that emerged were noted. Formal thematic analysis was not undertaken.

Using the quantitative data and the written feedback from participants, the research team used an iterative process to refine the Draft AGREE-REX tool. This refinement was achieved through an in-person meeting, a feedback session with stakeholders at the 2017 Global Evidence Summit,³⁴ and multiple teleconference meetings with the AGREE-REX team (2017-2019). Decisions were reached by consensus.

Results

Of the 692 individuals who responded to the advertisement and were emailed a formal invitation, 322 (47.0%) completed the study. Of the 322 respondents, 202 (62.7%) were female, 252 (78.2%) had some experience with the AGREE II tool, 188 (58%) indicated that English was their first language, and 170 (53.8%) identified themselves as CPG developers (**Table 1**). Participants represented 6 geographic regions; 177 (55.0%) were from North America, 76 (24.0%) from Europe, 32 (10.0%) from South America, 24 (7.4%) from Asia, 7 (2.1%) from Africa, and 6 (2.0%) from Oceania.

As reported in **Table 2** and **Table 3**, participants rated the survey items as easy to understand (with a mean [SD] ranging from 5.2 [1.38] for the alignment of values item to 6.3 [0.87] for the evidence item on the 7-point scale) and easy to apply (with a mean [SD] ranging from 4.8 [1.49] for the alignment of values item to 6.1 [1.07] for the evidence item on the 7-point scale). Participants rated the tool's instructions on the 7-point scale as clear (mean [SD], 5.8 [1.06]), felt confident in applying the tool to a guideline (mean [SD], 5.1 [1.43]), regarded the tool as complete (mean [SD], 5.7 [1.18]), and agreed that the tool adds value to the CPG enterprise (mean [SD], 5.9 [1.13]). In addition, 229 (71%) of respondents intended to use the AGREE-REX tool for evaluation, 203 (63%) for endorsement, and 187 (58%) for development or reporting purposes.

Internal consistency of the items was high (Cronbach $\alpha = 0.94$); deleting an item did not alter this finding. Interrater reliability predicted for the mean of 2 was 0.47, of 3 was 0.57, of 4 was 0.64, and of 5 was 0.69.

Correlation between the overall AGREE-REX score and the implementability score was 0.81 and between the overall AGREE-REX score and the clinical credibility score was 0.76 and more robust

Table 1. Characteristics of 322 Participants

Demographic characteristic	Frequency, No. (%)
Sex	
Female	202 (62.7)
Male	115 (35.7)
Prefer not to disclose	5 (1.6)
Age, y	
19 or younger	2 (0.6)
20-29	49 (15.2)
30-39	100 (31.1)
40-49	83 (25.8)
50-59	63 (19.6)
60-69	23 (7.1)
≥70	2 (0.6)
Experience with AGREE II	
No experience	70 (21.7)
Some experience	122 (37.9)
Experienced	88 (27.3)
Very experienced	42 (13)
First language	
English	188 (58.4)
Spanish	51 (15.8)
Italian	14 (4.3)
Chinese	13 (4)
Dutch	10 (3.1)
Portuguese	7 (2.2)
French	4 (1.2)
Greek	3 (0.9)
Ukrainian	3 (0.9)
Other	29 (9)
Geographic location	
North America	177 (55)
Europe	76 (23.6)
Asia	24 (7.5)
South America	32 (9.9)
Africa	7 (2.2)
Oceania	6 (1.9)
Participants' role with clinical practice guidelines (as many as apply)	
Practice guideline developer	
Clinical expert	85 (26.4)
Patient/public representative	15 (4.7)
Methodologist	170 (52.8)
Practice guideline user	
Health care professional	102 (31.7)
Administrator/policy maker/manager	38 (11.8)
Patient/member of the public	20 (6.2)
Researcher	159 (49.4)
Other (eg, librarian, student)	25 (7.8)

Abbreviation: AGREE II, Appraisal of Guidelines, Research and Evaluation revision.

than the correlations between the overall AGREE-REX score and each of the AGREE II domain scores (for example, $r = 0.10$ for clarity of presentation and $r = 0.43$ for applicability) (Table 4).

Participants offered wording changes and editorial suggestions to help clarify concepts and ideas. Core themes emerged in the written feedback. For Draft AGREE-REX and AGREE II, some participants articulated concerns about how to use both tools, potential redundancy, and lack of

Table 2. AGREE-REX Section 1 Usability Survey Results From 322 Participants

Section 1 item ^a	Participant rating, mean (SD)	
	Easy to understand	Easy to apply
Evidence	6.3 (0.87)	6.1 (1.07)
Clinical relevance	6.2 (0.80)	5.9 (1.06)
Relevance to patients/populations	6.1 (0.89)	5.8 (1.07)
Implementation relevance	5.8 (0.99)	5.4 (1.31)
Guideline developer values	5.6 (1.20)	5.2 (1.37)
Target user values	5.7 (1.20)	5.3 (1.37)
Patient or population values	5.7 (1.15)	5.3 (1.35)
Policy values	5.4 (1.26)	5.1 (1.41)
Alignment of values	5.2 (1.38)	4.8 (1.49)
Local applicability	5.9 (1.05)	5.4 (1.33)
Resources, capacity and tools	6.0 (0.96)	5.6 (1.28)

Abbreviation: AGREE-REX, Appraisal of Guidelines for Research and Evaluation-Recommendations Excellence.

^a From Section 1 of the survey: asks agreement, with a response of 1 indicating strongly disagree and 7 indicating strongly agree.

Table 3. AGREE-REX Section 2 Usability Survey Results From 322 Participants

Section 2 item ^a	Participant rating, mean (SD)
The AGREE-REX instructions are clear	5.8 (1.06)
The AGREE-REX instructions are helpful	5.9 (1.00)
The AGREE-REX instructions are complete	5.8 (1.11)
The AGREE-REX was easy to use	5.4 (1.32)
I felt confident when applying the AGREE-REX to a guideline	5.1 (1.43)
The AGREE-REX is complete; there are no missing items	5.7 (1.18)
The use of multiple evaluation statements for each of the 11 items is appropriate	5.5 (1.52)
The use of a 7-point response scale is appropriate	5.9 (1.28)
The overall assessment questions are useful	5.9 (1.06)
The AGREE-REX would be useful for	
Evaluating a guideline	5.8 (1.29)
Guideline development and reporting	6.0 (1.19)
Deciding whether or not to adapt or endorse a guideline	5.7 (1.27)
Deciding whether or not to implement a guideline in clinical practice	5.7 (1.25)
The AGREE-REX adds value to the clinical practice guideline enterprise	5.9 (1.13)

Abbreviation: AGREE-REX, Appraisal of Guidelines for Research and Evaluation-Recommendations Excellence.

^a From Section 2 of the survey: asks agreement, with a response of 1 indicating strongly disagree and 7 indicating strongly agree.

Table 4. Correlations Between 161 Guidelines

Variable	Overall AGREE-REX score	
	Pearson r	P value
AGREE II domain score		
1. Scope and purpose	0.25	<.001
2. Stakeholder involvement	0.29	<.001
3. Rigor of development	0.27	.001
4. Clarity of presentation	0.10	.23
5. Applicability	0.43	<.001
6. Editorial independence	0.12	.12
AGREE-REX item score		
Overall implementability score	0.81	<.001
Overall clinical credibility score	0.76	<.001

Abbreviation: AGREE-REX, Appraisal of Guidelines for Research and Evaluation-Recommendations Excellence.

instruction. Some participants preferred having the tools separate and others suggested they be integrated. For Draft AGREE-REX content and usability, participants articulated challenges in applying some items in the values theme and offered suggestions for clarity. Most participants did not like the 2 response scales or could not differentiate the intent between them.

Final Refinements

Based on the study results and feedback from participants, changes were made to the tool. **Table 5** lists the final items and criteria. eAppendix 1 in the [Supplement](#) compares the draft with the final version 1 of the tool and eAppendix 2 provides the entire AGREE-REX User's Guide.

The original 11 items were edited to 9 items (2 items combined and 1 item deleted) and clustered into 3 conceptual categories: clinical applicability, values, and implementability.

The original 3 response scales were modified to 2. The mandatory quality assessment scale asked appraisers to rate on the 7-point scale the overall quality of the item by considering whether the item criteria were addressed in the CPG and influenced the recommendations—for example, the extent to which data on the values and preferences of the various stakeholders were obtained and reported and extent to which these data were explicitly considered in formation of the recommendation.

The optional 7-point suitability for use scale is appropriate when a CPG is being considered for endorsement, adaptation, or implementation. This response scale considers whether the content of the criteria and its consequences for recommendations align with what would be expected in the context in which the CPG recommendations would be applied—for example, whether the potential users of a CPG perceive that the values and preferences of patients and policy makers collected and used to inform the CPG recommendations align with those in their own context. Appraisers are asked to rate the suitability for use in their setting/context.

In response to feedback, the 2 overall assessment questions (implementability and clinical credibility) were replaced by 2 new overall assessment questions to align with the AGREE II overall assessment items. The first new question (required) asked raters whether they would recommend the CPG for use in an appropriate context and the optional second new question asked raters whether they would recommend the CPG for use in their own context. A categorical response scale of yes, yes with modifications, and no is used to answer these assessment questions.

There was debate whether to integrate the new items into the existing AGREE II or have a separate AGREE-REX tool. A decision was made to create a separate tool to provide optimal flexibility to potential users. A resource to provide directions for use of the AGREE suite of tools has been written (M. C. Brouwers, PhD, unpublished data, 2020).

Discussion

Key Results and Interpretation

Overall, results of the study indicated that AGREE-REX is a usable, reliable, and valid tool to evaluate CPG recommendations. The AGREE-REX tool is a complement rather than an alternative to the AGREE II tool. The AGREE II tool focuses on the quality of the entire CPG process. The AGREE-REX tool focuses specifically on the quality of the CPG recommendations.

We believe that AGREE-REX will be a useful tool to evaluate CPG recommendations (single, bundle), differentiate among them, and identify those that are clinically credible and implementable for practicing health professionals and decision makers who use recommendations to inform clinical policy. Appraising a CPG with the AGREE II tool and the AGREE-REX tool may help provide information about the methodologic quality and the quality of the guideline recommendations. The appraisal step using both tools may help mitigate challenges in moving directly to costly and complex implementation commitments with CPGs that may lack rigor and suitability to the setting in which they are to be applied.

Table 5. AGREE-REX (Version 1) Items and Criteria

Item	Criteria
<p>Item 1. Evidence</p> <p>Definition: To be of high quality, recommendation should be based on a thorough review of the quality and results of the available evidence^a</p>	<p>The guideline assesses any risk of bias related to the study designs of the supporting evidence</p> <p>The guideline describes the consistency of the results (ie, similarity of results across studies)</p> <p>The guideline addresses the directness of the evidence (ie, addresses the exact interventions, populations, and outcomes of interest) to the clinical/health problem</p> <p>The guideline indicates the precision of the results (eg, width of confidence intervals of individual studies or meta-analyses)</p> <p>The guideline describes the magnitude of the benefits and harms</p> <p>The guideline assesses the likelihood of publication bias</p> <p>The guideline addresses the possibility of confounding factors (if applicable)</p> <p>The guideline indicates the dose-response gradient (if applicable)</p>
<p>Item 2. Applicability to target users</p> <p>This item evaluates the degree to which the recommendations are applicable to the guideline's target users' practice context</p>	<p>The guideline addresses a clinical/health problem that is relevant to the intended target user(s)</p> <p>There is an alignment between the target user's scope of practice and targeted patients/populations</p> <p>Target user's scope of practice and recommended actions</p> <p>The direction of the recommendations (ie, in favor of or against a particular action) and the trade-offs between harms and benefits</p> <p>The definitiveness or strength of the recommendations and the trade-offs between harms and benefits</p>
<p>Item 3. Applicability to patients or populations</p> <p>This item assesses the extent to which the anticipated outcomes of the recommended action are relevant for, and valued by, the intended patients/populations</p>	<p>The guideline includes outcomes that are relevant to the targeted patients/populations. These outcomes are often referred to as patient-important outcomes, patient-centered outcomes, patient-reported outcomes, or patient experience</p> <p>Relevant outcomes were considered in the development of the evidence base</p> <p>Recommended actions have the potential to affect outcomes relevant to patients/populations (eg, improve desirable patient-relevant outcomes, mitigate undesirable patient-relevant outcomes)</p> <p>The guideline reports how the importance of outcomes to patients was determined</p> <p>The guideline describes how to tailor recommendations for application to individual (or subsets of) patients or populations (eg, based on age, sex, ethnicity, comorbidities)</p>
<p>Item 4. Values and preferences of target users</p> <p>Values and preferences of target users refers to the relative importance that the target users of the guidelines (eg, health care providers, policy makers, administrators) place on the outcomes of interest (eg, survival, adverse effects, quality of life, cost, convenience). Target user values and preferences are important to consider during the guideline development process because they influence whether the recommendations are acceptable and adopted into practice</p>	<p>Values and preferences of guideline target users, as they relate to the recommended actions, have been sought and considered</p> <p>Factors related to target user acceptability of the recommended actions have been considered (eg, the acceptability of learning new clinical skills or the need to adapt current routine)</p> <p>The guideline differentiates between recommended actions for which clinical flexibility and individual patient tailoring are more appropriate in the decision-making process and those for which they are less appropriate</p> <p>The guideline describes the range of recommended actions that are acceptable to the clinical community, including the preferred option (if relevant), and describing why it is the preferred choice</p>
<p>Item 5. Values and preferences of patients/populations</p> <p>Values and preferences of patients/populations refers to the relative importance that the recipients of the recommended actions place on the outcomes of interest (eg, survival, adverse effects, quality of life, cost, convenience). Patient or population values and preferences are important to consider during the guideline development process because they influence whether the recommendations are acceptable and adopted into practice</p>	<p>The guideline includes outcomes that are relevant to the targeted patients/populations. These outcomes are often referred to as patient-important outcomes, patient-centered outcomes, patient-reported outcomes, or patient experience</p> <p>Relevant outcomes were considered in the development of the evidence base</p> <p>Recommended actions have the potential to affect outcomes relevant to patients/populations (eg, improve desirable patient-relevant outcomes, mitigate undesirable patient-relevant outcomes)</p> <p>The guideline reports how the importance of outcomes to patients was determined</p> <p>The guideline describes how to tailor recommendations for application to individual (or subsets of) patients or populations (eg, based on age, sex, ethnicity, comorbidities)</p>

(continued)

Table 5. AGREE-REX (Version 1) Items and Criteria (continued)

Item	Criteria
Item 6. Values and preferences of policy/decision-makers	
Values and preferences of policy/decision-makers refers to the relative importance that policy stakeholders place on the outcomes of interest (eg, survival, adverse effects, quality of life, cost, convenience). The values and preferences of policy stakeholders can affect the implementation of guideline recommendations in the health care system (eg, provision of resources or funding to support the recommended actions)	Information about the needs of policy and decision-makers has been sought and considered in the formulation of the recommendations The effect of the recommendations on policy and system-level decision-making has been considered in the formulation of the recommendations The effect of the recommendations on health equities has been considered in the formulation of the recommendations The guideline describes where changes to policy should be made to align with the recommendations
Item 7. Values and preferences of guideline developers	
Values and preferences of guideline developers refers to the relative importance that developers place on the outcomes of interest (eg, survival, adverse effects, quality of life, cost, convenience). Guideline developer values can influence the selection of outcomes of interest, the choice of guideline development methods, the approach to integrating varying stakeholder perspectives, and the interpretation of the balance between benefits and harms.	There is a clear description of the values and preferences that guideline developers brought to the development process There is a clear description of how guideline developer values and preferences influenced their interpretation of the balance between benefits and harms The method used to integrate values and preferences, including when they differ between stakeholders (eg, target users, patients/population, policy makers), is described
Item 8. Purpose	
Practice guidelines can be developed to achieve several implementation goals, such as to influence health care decisions, to promote discussion in the clinical encounter, to provide rationale to create or refine clinical policy, or to identify actions that reflect clinical or population health goals.	The guideline recommendations align with the implementation goals of the guideline (eg, for advocacy or policy change) The anticipated effects of recommendation adoption on individuals (eg, patients, populations, target users), organizations, and/or systems are described
Item 9. Local application and adoption	
This item assesses the suitability of the guideline recommendations for the setting, patients/population, and/or the health care system in which they are being implemented. Guidelines that include advice or tools and resources to facilitate the implementation of the recommendations are easier to adopt in practice.	The guideline describes the types and degree of change required from current practice The guideline differentiates between recommendations for which local adaptation may be more or less relevant The guideline articulates relevant factors important to its successful dissemination The guideline developers considered the issues that can influence the adoption of the recommendations and provided tools and/or advice for guideline implementers related to: How to tailor recommendations for the local setting Resource considerations needed to implement the recommendations (eg, human resources, equipment) and their associated costs Economic analysis (eg, cost-effectiveness or cost-utility) of recommended actions (if appropriate) Competencies and/or training of personnel required to implement the recommended action Data required to implement and monitor the adoption of recommended actions Strategies to overcome barriers related to health care professional acceptability and/or patient/population and/or policy acceptability of the recommended actions Criteria that can be used to measure recommendation implementation and quality improvement

Abbreviation: AGREE-REX. Appraisal of Guidelines for Research and Evaluation-Recommendations Excellence. ^a Informed by GRADE Working Group criteria (www.gradeworkinggroup.org).

In addition to the evaluation version of the tool, we have created the AGREE-REX Reporting Checklist, which can be used to inform development and reporting standards. The criteria used for evaluation purposes are presented as quality concepts to be included and documented in the CPG as it is being developed and, moreover, to inform the development protocol. The checklist will help identify specific operational strategies to meet AGREE-REX quality criteria to incorporate from the outset. For example, the well-designed Evidence to Decision Framework reflects the utility of some of the AGREE-REX concepts.³⁵ In addition, the checklist can help researchers prioritize when there is an absence of rigorous and feasible operational methods so efforts can be directed to address those gaps.

The recently released Clinical Practice Guidelines Applicability Evaluation (CPGAE-V1.0) also addresses this area. Designed to evaluate CPG applicability,³⁶ the CPGAE-V1.0 has been used to assess traditional Chinese medicine guidelines but has not yet been tested by the international community, nor have its measurement properties been explored. Similarly, the recently released National Guideline Clearinghouse Extent of Adherence to Trustworthy Standards (NEATS instrument) is designed to measure CPG adherence to the Institute of Medicine standards for trustworthy guidelines.³⁷ The methods of development and scope of these tools are different; nonetheless, investigating how the AGREE-REX tool and these tools complement each other may be a valuable area of inquiry.

Strengths of the AGREE-REX tool include the use of methodologic standards of measurement design in its development^{29,32,33}; the use of multidisciplinary literature as a basis for the concepts underpinning AGREE-REX^{30,31}; and its development by a multidisciplinary international research team and engagement of 322 internationally representative participants involved in CPGs. The participants reaffirmed the need for this tool, and their participation was vital to ensure that the resource was tailored to the needs of the international CPG communities.

Limitations

This study has limitations. The measurement properties and usability surveys were performed with the penultimate draft version of the tool. Financial considerations prohibited the repetition of the studies to confirm that the changes made to the AGREE-REX tool were associated with improvements in measurement properties and usability. Nonetheless, we believe that decisions for modifications made were informed by evidence. Capturing information from in-the-field experiences on an ongoing basis will be essential in continuing to develop the evidence base to support use of the AGREE-REX tool. Additional supporting materials (eg, training tools) are being developed to improve interrater reliability of the tool. Another limitation is the criteria used to select the CPGs (<50 pages, English language only) and that the tool was applied to the whole set of recommendations in each report. Although the tool, and not the CPGs themselves, was the object of study, the criteria and unit of recommendation may affect the perceptions of the tool and its measurement properties. Continued application to a range of CPGs is required to better assess its generalizability.

Conclusions

The results of this study suggest that AGREE-REX is a reliable, valid, and usable tool designed to evaluate CPG recommendations specifically. It is a complement to the AGREE II tool.

ARTICLE INFORMATION

Accepted for Publication: March 19, 2020.

Published: May 27, 2020. doi:[10.1001/jamanetworkopen.2020.5535](https://doi.org/10.1001/jamanetworkopen.2020.5535)

Open Access: This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2020 Brouwers MC et al. *JAMA Network Open*.

Corresponding Author: Ivan D. Florez, MD, MSc, Department of Pediatrics, University of Antioquia, Calle 67, No. 53 - 108, Medellín 0500001, Colombia (ivan.florez@udea.edu.co).

Author Affiliations: University of Ottawa, Ottawa, Ontario, Canada (Brouwers); McMaster University, Hamilton, Ontario, Canada (Spithoff, Kerkvliet, Hanna); Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau-CIBERESP), Barcelona, Spain (Alonso-Coello); Dutch College of General Practitioners, Utrecht, the Netherlands (Burgers); Imperial College London, St Mary's Hospital, London, United Kingdom (Cluzeau); Département Cancer et Environnement, Centre Léon Bérard, Lyon Cedex 08, France (Férvors); Ottawa Hospital Research Institute, University of Ottawa, Ottawa, Ontario, Canada (Graham, Grimshaw); North York General Hospital, Toronto, Ontario, Canada (Kastner); Institute of Applied Health Sciences, McMaster University, Hamilton, Ontario, Canada (Kho); American College of Physicians, Philadelphia, Pennsylvania (Qaseem); Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, Ontario, Canada (Straus); Department of Pediatrics, University of Antioquia, Medellín, Colombia (Florez).

Author Contributions: Dr Brouwers had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Brouwers, Spithoff, Alonso-Coello, Burgers, Cluzeau, Férvors, Graham, Grimshaw, Kastner, Qaseem, Straus, Florez.

Acquisition, analysis, or interpretation of data: Brouwers, Spithoff, Kerkvliet, Burgers, Hanna, Kho, Qaseem, Straus, Florez.

Drafting of the manuscript: Brouwers, Burgers, Straus.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Brouwers, Kerkvliet, Alonso-Coello, Qaseem, Straus, Florez.

Obtained funding: Brouwers, Graham, Straus.

Administrative, technical, or material support: Kerkvliet, Straus, Florez.

Supervision: Brouwers, Spithoff, Burgers, Straus.

Other - International steering committee: Férvors.

Conflict of Interest Disclosures: Dr Brouwers reported receiving grants from the Canadian Institute for Health Research during the conduct of the study. Mss Spithoff and Kerkvliet reported receiving grants from the Canadian Institute for Health Research during the conduct of the study. Dr Burgers reported serving as Trustee of the AGREE Research Trust from 2004 to 2014. No other disclosures were reported.

Funding/Support: This project was funded by the Canadian Institutes of Health Research, grant 201209MOP-285689-KTR-CEBA-40598.

Role of the Funder/Sponsor: The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: The authors thank the following individuals for their contributions, advice, and input into this project: Onil Bhattacharyya, MD, PhD, University of Toronto, Canada; George Browman, MD, MSc, FRCPC, Retired, Canada; Anna Gagliardi, PhD, University of Toronto, Canada; Peter Littlejohns, MD, FRCPC, King's College London, United Kingdom; Holger Schunemann, MD, PhD, McMaster University, Canada; Louise Zitzelsberger, PhD, Health Canada, Canada. Contributors advised on the concept and proposed protocol and the early stages of the development of the beta version of the tool. No contributor was financially compensated, and all contributors provided permission to be acknowledged.

Additional Information: The AGREE suite of tools is available on the AGREE Enterprise website (<http://www.agreetrust.org>).

REFERENCES

1. Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. *Ann Intern Med*. 2003;139(6):493-498. doi:10.7326/0003-4819-139-6-200309160-00013
2. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P; Board of Trustees of the Guidelines International Network. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med*. 2012;156(7):525-531. doi:10.7326/0003-4819-156-7-201204030-00009
3. Institute of Medicine. *Clinical Practice Guidelines We Can Trust*. National Academies Press; 2011.
4. 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(1):18-23. doi:10.1136/qhc.12.1.18

5. Brouwers MC, Kho ME, Browman GP, et al; AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ*. 2010;182(18):E839-E842. doi:[10.1503/cmaj.090449](https://doi.org/10.1503/cmaj.090449)
6. 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;182(10):1045-1052. doi:[10.1503/cmaj.091714](https://doi.org/10.1503/cmaj.091714)
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;182(10):E472-E478. doi:[10.1503/cmaj.091716](https://doi.org/10.1503/cmaj.091716)
8. 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(9198):103-106. doi:[10.1016/S0140-6736\(99\)02171-6](https://doi.org/10.1016/S0140-6736(99)02171-6)
9. 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(1):21-28. doi:[10.1093/intqhc/11.1.21](https://doi.org/10.1093/intqhc/11.1.21)
10. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 16. Evaluation. *Health Res Policy Syst*. 2006;4:28. doi:[10.1186/1478-4505-4-28](https://doi.org/10.1186/1478-4505-4-28)
11. Graham ID, Beardall S, Carter AO, et al. What is the quality of drug therapy clinical practice guidelines in Canada? *CMAJ*. 2001;165(2):157-163.
12. 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(440):205-210.
13. Brouwers M, Browman G. *Assessment of the American Society of Clinical Oncology (ASCO) practice guidelines*. *J Clin Oncol*, Classic Reports and Current Comments; 2000:1081-1088.
14. Burgers JS, Fervers B, Haugh M, et al. International assessment of the quality of clinical practice guidelines in oncology using the Appraisal of Guidelines and Research and Evaluation Instrument. *J Clin Oncol*. 2004;22(10):2000-2007. doi:[10.1200/JCO.2004.06.157](https://doi.org/10.1200/JCO.2004.06.157)
15. Brouwers MC, Rawski E, Spithoff K, Oliver TK. Inventory of Cancer Guidelines: a tool to advance the guideline enterprise and improve the uptake of evidence. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):151-161. doi:[10.1586/erp.11.11](https://doi.org/10.1586/erp.11.11)
16. Kung J, Miller RR, Mackowiak PA. Failure of clinical practice guidelines to meet Institute of Medicine standards: two more decades of little, if any, progress. *Arch Intern Med*. 2012;172(21):1628-1633. doi:[10.1001/2013.jamainternmed.56](https://doi.org/10.1001/2013.jamainternmed.56)
17. Reames BN, Krell RW, Ponto SN, Wong SL. Critical evaluation of oncology clinical practice guidelines. *J Clin Oncol*. 2013;31(20):2563-2568. doi:[10.1200/JCO.2012.46.8371](https://doi.org/10.1200/JCO.2012.46.8371)
18. Armstrong JJ, Goldfarb AM, Instrum RS, MacDermid JC. Improvement evident but still necessary in clinical practice guideline quality: a systematic review. *J Clin Epidemiol*. 2017;81:13-21. doi:[10.1016/j.jclinepi.2016.08.005](https://doi.org/10.1016/j.jclinepi.2016.08.005)
19. Alonso-Coello P, Irfan A, Solà I, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Qual Saf Health Care*. 2010;19(6):e58. doi:[10.1136/qshc.2010.042077](https://doi.org/10.1136/qshc.2010.042077)
20. Qaseem A, Lin JS, Mustafa RA, Horwath CA, Wilt TJ; Clinical Guidelines Committee of the American College of Physicians. Screening for breast cancer in average-risk women: a guidance statement from the American College of Physicians. *Ann Intern Med*. 2019;170(8):547-560. doi:[10.7326/M18-2147](https://doi.org/10.7326/M18-2147)
21. Qaseem A, Denberg TD, Hopkins RH Jr, et al; Clinical Guidelines Committee of the American College of Physicians. Screening for colorectal cancer: a guidance statement from the American College of Physicians. *Ann Intern Med*. 2012;156(5):378-386. doi:[10.7326/OO03-4819-156-5-201203060-00010](https://doi.org/10.7326/OO03-4819-156-5-201203060-00010)
22. Qaseem A, Barry MJ, Denberg TD, Owens DK, Shekelle P; Clinical Guidelines Committee of the American College of Physicians. Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. *Ann Intern Med*. 2013;158(10):761-769. doi:[10.7326/OO03-4819-158-10-201305210-00633](https://doi.org/10.7326/OO03-4819-158-10-201305210-00633)
23. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care*. 2005;17(3):235-242. doi:[10.1093/intqhc/mzi027](https://doi.org/10.1093/intqhc/mzi027)
24. 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(1):37-44. doi:[10.1007/s11606-007-0440-9](https://doi.org/10.1007/s11606-007-0440-9)
25. 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(1):65-72. doi:[10.1373/clinchem.2005.056952](https://doi.org/10.1373/clinchem.2005.056952)

26. Nuckols TK, Shetty K, Raaen L, et al. Technical quality and clinical acceptability of a utilization review guideline for occupational conditions: ODG Treatment Guidelines by the Work Loss Data Institute. RAND Corporation; 2017. Accessed August 7, 2018. https://www.rand.org/pubs/research_reports/RR1819.html
27. Brouwers MC, Kerkvliet K, Spithoff K; AGREE Next Steps Consortium. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *BMJ*. 2016;352:i1152. doi:10.1136/bmj.i1152
28. Siering U, Eikermann M, Hausner E, Hoffmann-Esser W, Neugebauer EAM. Appraisal tools for clinical practice guidelines: a systematic review. *PLoS One*. 2013;8(12):e82915. doi:10.1371/journal.pone.0082915
29. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press; 2015. doi:10.1093/med/9780199685219.001.0001
30. Kastner M, Bhattacharyya O, Hayden L, et al. Guideline uptake is influenced by six implementability domains for creating and communicating guidelines: a realist review. *J Clin Epidemiol*. 2015;68(5):498-509. doi:10.1016/j.jclinepi.2014.12.013
31. Brouwers MC, Makarski J, Kastner M, Hayden L, Bhattacharyya O; GUIDE-M Research Team. The Guideline Implementability Decision Excellence Model (GUIDE-M): a mixed methods approach to create an international resource to advance the practice guideline field. *Implement Sci*. 2015;10:36. doi:10.1186/s13012-015-0225-1
32. Fleiss JL. The measurement of interrater agreement. In: *Statistical Methods for Rates and Proportions*. John Wiley & Sons; 1981.
33. John OP, Benet-Martinez V. Measurement: reliability, construct validation, and scale construction. In: Reis HT, Judd CM, eds. *Handbook of Research Methods in Social and Personality Psychology*. Cambridge University Press; 2000:339-370.
34. Brouwers M, Florez ID, Spithoff K, Kerkvliet K. Evaluating the clinical credibility and implementability of clinical practice guideline recommendations using the AGREE-REX tool [workshop]. Abstracts of the Global Evidence Summit, Cape Town, South Africa. *Cochrane Database Syst Rev*. 2017;9(suppl 2). doi:10.1002/14651858.CD201702
35. Alonso-Coello P, Schünemann HJ, Moher J, et al; GRADE Working Group. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 2016;353:i2016. doi:10.1136/bmj.i2016
36. Li H, Xie R, Wang Y, Xie X, Deng J, Lu C. A new scale for the evaluation of clinical practice guidelines applicability: development and appraisal. *Implement Sci*. 2018;13(1):61. doi:10.1186/s13012-018-0746-5
37. Jue JJ, Cunningham S, Lohr K, et al. Developing and testing the Agency for Healthcare Research and Quality's National Guideline Clearinghouse Extent of Adherence to Trustworthy Standards (NEATS) instrument. *Ann Intern Med*. 2019;170(7):480-487. doi:10.7326/M18-2950

SUPPLEMENT.

eAppendix 1. Draft AGREE-REX vs AGREE-REX Version 1 (V1)

eAppendix 2. AGREE-REX: Recommendation Excellence User's Guide