

RESEARCH ARTICLE

Open Access



Guideline appraisal with AGREE II: online survey of the potential influence of AGREE II items on overall assessment of guideline quality and recommendation for use

Wiebke Hoffmann-Eßer^{1,2*}, Ulrich Siering¹, Edmund A. M. Neugebauer³, Anne Catharina Brockhaus¹, Natalie McGauran¹ and Michaela Eikermann⁴

Abstract

Background: The AGREE II instrument is the most commonly used guideline appraisal tool. It includes 23 appraisal criteria (items) organized within six domains. AGREE II also includes two overall assessments (overall guideline quality, recommendation for use). Our aim was to investigate how strongly the 23 AGREE II items influence the two overall assessments.

Methods: An online survey of authors of publications on guideline appraisals with AGREE II and guideline users from a German scientific network was conducted between 10th February 2015 and 30th March 2015. Participants were asked to rate the influence of the AGREE II items on a Likert scale (0 = no influence to 5 = very strong influence). The frequencies of responses and their dispersion were presented descriptively.

Results: Fifty-eight of the 376 persons contacted (15.4%) participated in the survey and the data of the 51 respondents with prior knowledge of AGREE II were analysed. Items 7–12 of Domain 3 (rigour of development) and both items of Domain 6 (editorial independence) had the strongest influence on the two overall assessments. In addition, Items 15–17 (clarity of presentation) had a strong influence on the recommendation for use. Great variations were shown for the other items. The main limitation of the survey is the low response rate.

Conclusions: In guideline appraisals using AGREE II, items representing rigour of guideline development and editorial independence seem to have the strongest influence on the two overall assessments. In order to ensure a transparent approach to reaching the overall assessments, we suggest the inclusion of a recommendation in the AGREE II user manual on how to consider item and domain scores. For instance, the manual could include an a-priori weighting of those items and domains that should have the strongest influence on the two overall assessments. The relevance of these assessments within AGREE II could thereby be further specified.

Keywords: Clinical practice guidelines, Methodological guideline appraisal, Methodological quality, Systematic reviews

* Correspondence: wiebke.hoffmann-esser@iqwig.de

¹Institute for Quality and Efficiency in Health Care (IQWiG), Im Mediapark 8, 50670 Cologne, Germany

²Institute for Research in Operative Medicine (IFOM), University of Witten/Herdecke, Campus Cologne, Cologne, Germany

Full list of author information is available at the end of the article



Background

According to the definition of the US Institute of Medicine (IOM), “clinical practice guidelines are statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options” [1, 2]. Various studies have shown that guidelines can improve health care [3–9]; however, their quality is variable and often unsatisfactory [10–14]. In order to be able to use guidelines as a reliable basis for decision-making, their quality, i.e. their methodological rigour and transparency, needs to be ensured. Guideline appraisal tools are applied for this purpose.

In 2003, an international group of guideline developers and researchers developed the Appraisal of Guidelines for Research & Evaluation (AGREE) instrument [15]. The revised version, AGREE II [16], was published in 2009 and is currently the most commonly applied and comprehensively validated guideline appraisal tool worldwide [17–19]. It consists of 23 appraisal criteria (items) organized into six domains (Table 1), each of which “captures a unique dimension of guideline quality” [16].

The items within each domain are rated on a seven-point scale (“strongly disagree” to “strongly agree”).

In addition, AGREE II includes two global rating items (overall assessments). In the first assessment, the overall guideline quality is rated on a seven-point scale (“lowest possible quality” to “highest possible quality”). In the second assessment, a recommendation is provided on whether to use the guideline or not (“yes”, “yes with modifications”, “no”). Both assessments should consider the items evaluated beforehand and the resulting domain scores, but should not be calculated from them: it is explicitly noted that the “six domain scores are independent and should not be aggregated into a single quality score” [16]). Beyond this information, AGREE II does not provide a specific approach to reaching the two overall assessments. The lack of operationalization for the conduct of the two overall assessments results in inconsistent approaches by guideline users, leading to subjective assessments [20–24].

In a recently published systematic review based on publications reporting guideline appraisals with AGREE II, we investigated how often AGREE II users conducted the two overall assessments and to what extent the six

Table 1 Items and domains of the AGREE II instrument^a

Item	Content	Domain
1	The overall objective(s) of the guideline is (are) specifically described.	Scope and Purpose
2	The health question(s) covered by the guideline is (are) specifically described.	
3	The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.	
4	The guideline development group includes individuals from all relevant professional groups.	Stakeholder Involvement
5	The views and preferences of the target population (patients, public, etc.) have been sought.	
6	The target users of the guideline are clearly defined.	
7	Systematic methods were used to search for evidence.	Rigour of Development
8	The criteria for selecting the evidence are clearly described.	
9	The strengths and limitations of the body of evidence are clearly described.	
10	The methods for formulating the recommendations are clearly described.	
11	The health benefits, side effects, and risks have been considered in formulating the recommendations.	
12	There is an explicit link between the recommendations and the supporting evidence.	Clarity of Presentation
13	The guideline has been externally reviewed by experts prior to its publication.	
14	A procedure for updating the guideline is provided.	
15	The recommendations are specific and unambiguous.	Applicability
16	The different options for management of the condition or health issue are clearly presented.	
17	Key recommendations are easily identifiable.	
18	The guideline describes facilitators and barriers to its application.	
19	The guideline provides advice and/or tools on how the recommendations can be put into practice.	
20	The potential resource implications of applying the recommendations have been considered.	Editorial Independence
21	The guideline presents monitoring and/or auditing criteria.	
22	The views of the funding body have not influenced the content of the guideline.	
23	Competing interests of guideline development group members have been recorded and addressed.	

^aExtracted from [16]

domain scores influenced these assessments [25]. We found that the two overall assessments were underreported by guideline assessors. Domains 3 (rigour of development) and 5 (applicability) had the strongest influence on the results of the two overall assessments, while the other domains had a varying influence.

Despite the deficits described above, the two overall assessments of AGREE II provide important information on whether a user can regard a guideline to be reliable, for example, as a basis for guideline development [26] or for application in clinical practice.

The above systematic review only investigated how strongly the six domains (and not the individual items) influenced the two overall assessments and was based on the published literature. The present analysis is an extension of the systematic review and aimed to provide a more detailed examination with a more practical orientation: on the basis of a survey of guideline users we investigated how strongly the 23 individual AGREE II items influenced the two overall assessments.

Methods

Conduct of the survey

We performed a systematic search to identify publications reporting results of guideline appraisals with AGREE II. We then asked the corresponding authors of these publications, as well as a group of further guideline users (all members of the Guidelines Section of the German Network for Evidence-based Medicine, DNEbM), to participate in an online survey conducted via Survey Monkey between 10 February and 30 March 2015. The link to the survey was included in the e-mail. The DNEbM members received a version including an introductory text and explanations in German plus the original AGREE II items in English; the corresponding authors of publications received a completely English version (see Additional file 1). A reminder e-mail was sent two weeks before the end of the deadline.

The focus of the survey was on the assessment of the strength of the potential influence of the AGREE II items on the two overall assessments (overall guideline quality and recommendation for use). For each of the 23 AGREE II items, respondents rated the strength of the influence on a Likert scale (0 = no influence to 5 = very strong influence). In addition, respondents were asked to provide information on characteristics such as their profession, knowledge of AGREE II, practical experience with the original AGREE instrument (AGREE I) or AGREE II, the purpose of guideline appraisals with AGREE I or II, and any prior involvement in guideline development. Furthermore, the survey contained an open question on which items respondents used in the overall assessment of guideline quality.

Data analysis

We analysed the combined results of the German and English versions of the survey using SPSS (PASW Statistics 18 [frequencies]) and SAS.

We presented the results descriptively; the respondents' characteristics were presented in a table; the respondents' evaluation of the influence of the AGREE II items on the two overall assessments was presented in box plots.

To determine the impact of potential confounding factors on the overall results, we also performed separate descriptive analyses according to profession, practical experience with AGREE I or II (number of guidelines appraised, experience in years), and any prior involvement in guideline development.

Before conducting the survey, we had formed the following three categories to assess the strength of the influence of the items on the two overall assessments and to enable clearer interpretation of the results: weak, medium, and strong influence (0–1, 2–3, and 4–5 points; median values).

Results

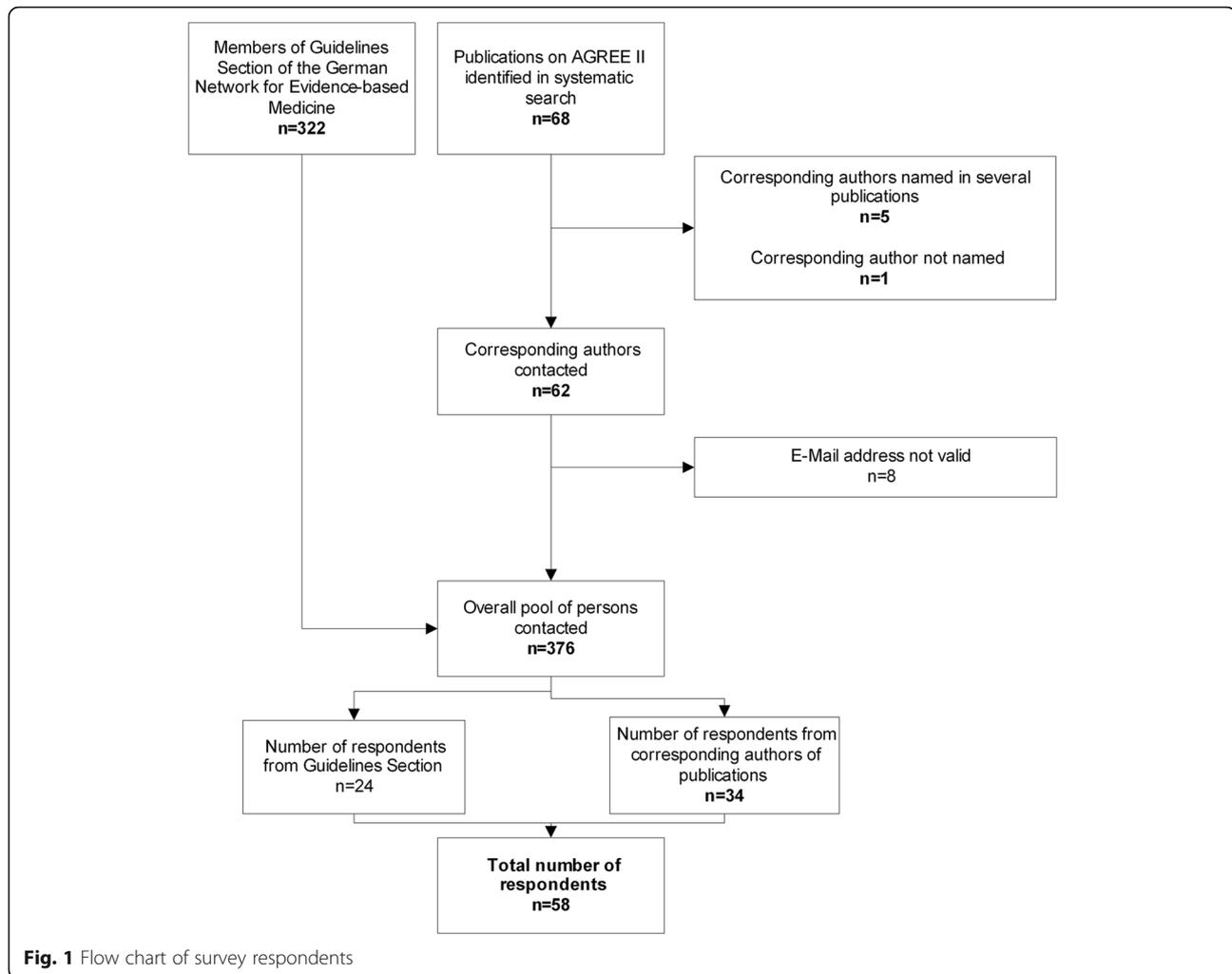
Response to online survey

A total of 376 guideline users with valid e-mail addresses were contacted: the German version of the survey was sent to 322 members of DNEbM and the English version was sent to 54 corresponding authors of publications on guideline appraisals (Fig. 1). Fifty-eight of the 376 persons contacted (15.4%) participated in the survey (see the raw data in Additional file 2): 34 of the 54 corresponding authors of publications (63.0%) and 24 of the 322 DNEbM members (7.5%).

Characteristics of respondents

Thirty-two (55.2%) of the 58 respondents were physicians of whom 10 (17.2%) were also methodological experts (Table 2). A further 10 respondents (17.2%) were solely methodological experts and 16 (27.6%) were from other professions (e.g. health scientists, pharmacologists, psychologists). 49 (84.5%) had previously performed guideline appraisals with AGREE I or II: 27 (46.6%) had performed less than 10 appraisals, nine (15.5%) had performed 10 to 20 appraisals and 13 (22.4%) had performed more than 20 appraisals.

Six (10.3%) of the respondents had less than one year experience with AGREE I or II appraisals, 35 (60.3%) had one to five years' experience, and eight (13.8%) had more than five years' experience. 35 (60.3%) had already been involved in guideline development. The most commonly reported reason for application of AGREE I or II was appraisal of guideline quality (24 respondents, 41.4%) followed by development of guidelines (seven



respondents; 12.1%) and writing of guideline synopses (seven; 12.1%).

Open question on use of items and domains

Twenty-one of the 58 respondents (36.2%) answered the open question on which items they use for the overall assessment of guideline quality: 10 (17.2%) stated that all items were used in equal measure and one (1.7%) stated that no item was used. Nine respondents (15.5%) named domains, not items. All nine named Domain 3 (rigour of development); four named this domain as the only domain and five named Domain 3 in combination with other domains. The second most named domain was Domain 6 (editorial independence). Only one respondent (1.7%) specified items (Items 9 and 12 of Domain 3).

It should be noted that seven respondents reported that they had no knowledge of AGREE II. However, two of them still answered the further questions; it is unclear whether their first answer was incorrect or whether they provided answers without having knowledge of AGREE

II. For this reason, both of these respondents were excluded from further analysis; the following results were thus provided by 51 respondents.

Evaluation of the influence of the AGREE II items

Not all of the 51 respondents included in the analysis evaluated all items with regard to their influence on the two overall assessments of AGREE II: four respondents provided no such evaluation and two respondents discontinued their evaluation at Item 7 and Item 18.

The boxplot shows great variations in the results for Items 1 to 3, 6, 14, 18, and 21 regarding both overall assessments (Fig. 2). For Items 19 and 20, the values vary greatly regarding guideline quality, but not regarding the recommendation for guideline use. The items with the strongest influence on the two overall assessments were reported to be Items 7 to 12 of Domain 3 (rigour of development) as well as both items (22 and 23) of Domain 6 (editorial independence). For Items 1, 15, 16 and 17–20, greater variations were notable for the

Table 2 Characteristics of respondents

Characteristics	Respondents N = 58 (%)
Profession	
Physician	22 (37.9)
Physician/methodological expert	10 (17.2)
Methodological expert	10 (17.2)
Other	16 (27.6)
Knowledge of the AGREE II instrument	
Yes	51 (87.9)
No	7 (12.1)
Performance of appraisals using the AGREE I or II instrument	
Yes	49 (84.5)
No	9 (15.5)
Number of appraised guidelines using the AGREE I or II instrument ^a	
< 10 guidelines	27 (46.6)
10–20 guidelines	9 (15.5)
> 20 guidelines	13 (22.4)
Experience in years ^a	
< 1 year	6 (10.3)
1–5 years	35 (60.3)
> 5 years	8 (13.8)
Involvement in guideline development	
Yes	35 (60.3)
No	23 (39.7)
Purpose of conducting appraisals using the AGREE I or II instrument ^b	
Assessment of guideline quality	24 (41.4)
Development of guidelines	7 (12.1)
Writing of guideline synopses	7 (12.1)
Research	3 (5.2)
Adaptation of guidelines	2 (3.4)
Application in clinical practice	2 (3.4)
Further training	2 (3.4)
Development of knowledge tools	1 (1.7)
Publication of scientific articles	1 (1.7)
Project work	1 (1.7)
Updating of guidelines	1 (1.7)
No response	5 (8.6)

^aNine survey respondents (15.5%) did not answer this question

^bThe question was formulated as an open question; we summarized the response options presented here from the individual responses given. Some of the respondents provided more than one response

influence on overall guideline quality than for the recommendation for use. A strong influence of these items can only be inferred for Items 15 to 17 of Domain 4 (clarity of presentation) with regard to the recommendation for use. The lowest scores were shown for the items of Domain 5 (applicability) and Item 14 of Domain 3, albeit with great variations.

The separate analyses of subgroups showed that the number of responses per subgroup (in most cases clearly fewer than 20 respondents) was too small to be able to draw valid conclusions on subgroup effects (data not shown). All in all, however, no marked deviations from the overall results were shown.

Discussion

On the basis of a survey of guideline users, the aim of our analysis was to investigate how strongly the individual AGREE II items influenced the two overall assessments (overall guideline quality and recommendation for use). Our findings indicate that Items 7 to 12 (Domain 3; rigour of development) and both items of Domain 6 (editorial independence) had the strongest influence on the two overall assessments. In addition, Items 15 to 17 (clarity of presentation) had a strong influence on the recommendation for use. Great variations in respondents' judgements were shown for the other items.

The importance of rigour of development (Domain 3) to guideline appraisers is not surprising, as this domain is regarded to be the strongest indicator of quality [10, 27], a high score for this domain indicating minimum bias and evidence-based guideline development [27]. The importance of editorial independence (Domain 6) highlights the relevance of conflicts of interest (COI) of guideline authors as a potential source of bias. Although the IOM clearly states that “To be *trustworthy*, guidelines should ...[b]e based on an explicit and transparent process that minimizes distortions, biases, and conflicts of interest” [2], most guidelines fail to disclose authors' COI, or if they do, numerous COI are reported [28–30].

In contrast to our systematic review [25], a strong influence of Domain 6, not Domain 5, was determined in the present analysis. This difference may have been caused by the different methods of data collection and data analysis: the data in our systematic review were based on actual applications of the AGREE II instrument whereas the data in the present analysis were based on more subjective assessments related to AGREE II collected by means of a survey. Therefore, some deviations in results are to be expected. We suggest considering Domain 6 in the weighting of results in order to achieve a more objective AGREE II assessment (see “Limitations”).

The finding that clarity of presentation (Domain 4) in a guideline had a strong influence on the recommendation for use is also not surprising, as “the main advantage of a well-reported guideline is that flaws in the methodology are more easily detected, so that inherent biases can be considered more explicitly and scrutinized by the potential users” [31].

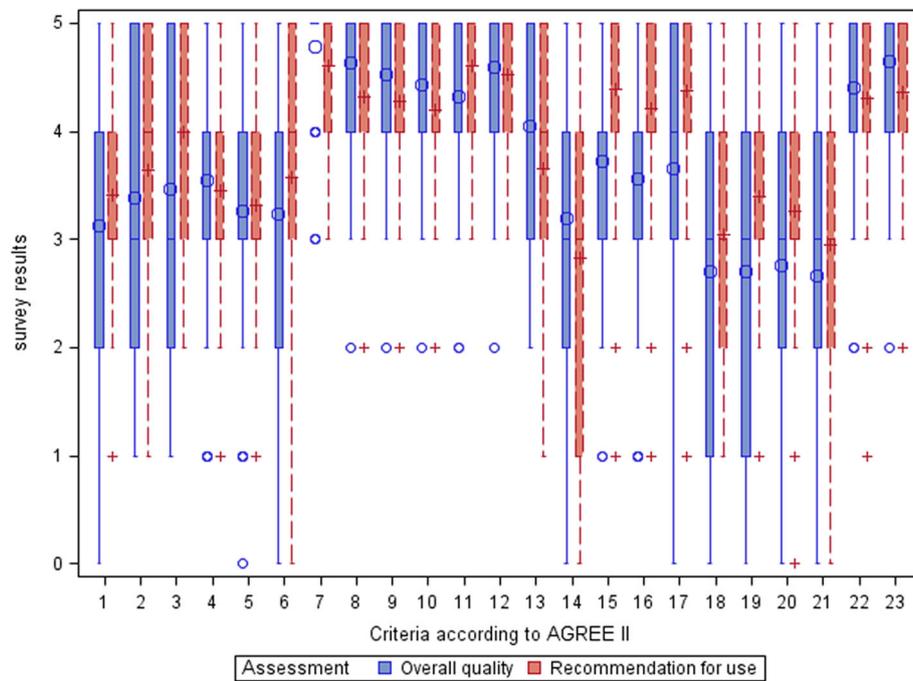


Fig. 2 Influence of the AGREE II items on guideline quality and recommendation for use (overall data)

Previous and potential future approaches to overall assessments in AGREE II

The results of our survey show that the overall assessments of AGREE II are highly subjective and a standardized approach to reaching these assessments is lacking. This is in line with previous research: the publications identified in our systematic literature search showed considerable variations in how the results from appraisals with AGREE II are used to reach the two overall assessments. For instance, in contrast to the recommendation in AGREE II, some users apply cut-offs to distinguish between high and low-quality guidelines [20, 21, 27, 32–55]. Others calculate a score for overall quality from the six domain scores; however, this no longer represents a separate assessment as foreseen by AGREE II [24, 44, 49, 56–59]. Further users weight items or domains without clearly presenting how this weighting affects the overall assessments [33, 34, 37, 44, 45, 60–62]. This issue was also addressed by Alonso-Coello et al. in 2010 in their review on guideline quality, who noted that “... the validity of the overall assessment may be limited, as there were no clear rules on how to weigh the different domain scores in making a decision about whether or not to recommend the guidelines” [10]. As stated, it has not yet been investigated in detail to what extent the individual AGREE II items influence the two overall assessments; our recently published systematic review [25] and the present analysis thus represent the first research to investigate this question.

The AGREE II user manual does not require transparent reporting with regard to how users reach their overall assessments and the approach applied is thus at the discretion of the users. This means that it is unclear how and to what extent these assessments are influenced by the individual assessments of items and domains. To ensure a transparent approach, the AGREE II user manual could include an a-priori weighting of those items and domains that should have the strongest influence on the two overall assessments. This would mean specifying which items are more (or less) useful regarding the operationalization of the conduct of the two overall assessments. This weighting approach could be included in an update of AGREE II to achieve more transparent operationalization, thus increasing objectiveness and leading to more comparable results of different appraisals of the same guideline. Ultimately, this would help to distinguish more clearly between high and low-quality guidelines. Additionally, the weighting approach could be used in the development of a rapid appraisal instrument including only the most useful items for the two overall assessments, and thus help to save resources.

In this context one could consider the findings by Fervers et al. [31], who examined characteristics of guidelines and guideline developing organizations to identify predictors of high-quality guidelines. They identified the availability of background information, that is, “explicit and detailed information about the objectives and context of the guideline development, including the

methods used, and the people and organizations involved in the development process” [31] as the strongest predictor of guideline quality, in particular for Domain 3 (rigour of development). The components cited could be used to help weight items in AGREE II.

Limitations

Our analysis is the first to investigate the influence of individual AGREE II items on overall guideline quality and recommendation for use. However, due to the low response rate of the survey (15.5%), only indications but no robust conclusions can be drawn from our findings. We had contacted members of the guideline section of a German scientific network, as we had expected a high response rate from this large pool of guideline users. However, the opposite was the case; the response rate in this group was actually far lower than in the group of authors of guideline appraisal articles (7.5% vs. 63.0%). One potential explanation could be that not all members of the guideline section of the German scientific network are actually involved in guideline development, but belong to this section due to their basic interest in clinical practice guidelines. Furthermore, some members of this section also belong to other working groups, so it is possible that some responses represent feedback from a whole working group rather than from a single respondent. In addition, non-responses are not necessarily limited to individual respondents, but can be associated with whole organizations choosing not to participate in a study [63].

In addition, German guideline appraisers primarily use the German adaptation of AGREE I (DELBI, [64]) and not the English-language instrument AGREE II – we did not consider DELBI in our survey, as it is not validated and is based on AGREE I. In contrast, the guideline appraisal articles identified in our systematic search referred primarily to AGREE II and one can thus assume a greater interest of these respondents in the survey. A further reason for the overall low response rate could be the type of survey conducted; web-based surveys often have lower response rates than those conducted by letter or phone [65].

Although nearly two-thirds of the respondents were not methodological experts, the results show a strong influence of Domain 3 (rigour of development); in our opinion a higher response rate including a higher proportion of methodological experts would therefore not necessarily have changed the results of the survey. However, we did not systematically assess the non-responses and our comments above are thus based on assumptions: ultimately, the extent to which the responses of the non-respondents would have changed the initial results is unclear and we cannot exclude potential bias.

Conclusions

The results of our survey indicate that in guideline appraisals using AGREE II, items representing the rigour of guideline development and the editorial independence of authors seem to have the strongest influence on the overall assessment of guideline quality and recommendation for use. In addition, items representing the clarity of presentation have a strong influence on the recommendation for use. Great variations in respondents' judgements exist regarding the other AGREE II items.

In order to ensure a transparent and consistent approach to reaching the two overall assessments, besides encouraging transparent reporting, we suggest the inclusion of a recommendation in the AGREE II user manual on how to consider item and domain scores. For instance, the user manual could include an a-priori weighting of those items and domains that should have the strongest influence on the 2 overall assessments so as to help distinguish more clearly between high and low-quality guidelines.

In addition, the weighting approach could be used in the development of a short (and economical) form of guideline appraisal including only the most important items and domains. In the next update of AGREE II, our study could thus help to contribute to determining which items and domains are most important for the operationalization of the two overall assessments. The relevance of the two overall assessments within AGREE II could thereby be further specified.

Additional files

Additional file 1: Questionnaire. (PDF 42 kb)

Additional file 2: Results (raw data of 58 respondents) of the assessment of the strength of the potential influence of the AGREE II items on the two overall assessments. (PDF 308 kb)

Abbreviations

AGREE: Appraisal of Guidelines for Research and Evaluation; COI: Conflict of interest; DELBI: German Guideline Appraisal Instrument; DNEbM: German Network for Evidence-based Medicine; IFOM: Institute for Research in Operative Medicine; IOM: Institute of Medicine; IQWiG: Institute for Quality and Efficiency in Healthcare; MDS: Medical Advisory Service of the German Social Health Insurance; PASW: Predictive Analysis SoftWare; SPSS: Superior Performing Software Systems

Acknowledgements

We thank Verena Wekemann for checking the format of the citations.

Consent for publications

No personal details of individual participants are presented in the manuscript since the survey was conducted anonymously. Therefore it was not necessary to obtain written informed consent from the participants of the survey.

Funding

This research received non-financial support from the Institute for Quality and Efficiency in Health Care (IQWiG). Four of the six authors are IQWiG employees and were thus involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Availability of data and materials

All data generated or analysed during this study are included in this published article and in the Additional file 2.

Authors' contributions

WHE, MEi, US and EN conceived and designed the experiment. WHE performed the experiment. WHE, ACB analysed the data. ACB and MEi contributed materials/analysis. WHE and NMG drafted the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The introductory text of the survey informed the participants that the survey was part of a scientific project and that the results were to be presented anonymously; participation was optional. Approval from the institutional review board or equivalent committee(s) was not obtained, as the online survey represented non-interventional research with anonymized data that did not include patients, but medical professionals and methodologists. There is no legal obligation to obtain ethics approval for this type of research in Germany; guidance provided by regional ethics committees (e.g. [66]) refers to interventional research. No informed consent was obtained from participants, as the survey was completely voluntary and anonymous, i.e. no personal data were collected. The survey assessors could not therefore draw any conclusions on the respondents. Although a list of the respondents' IP addresses was available in the Survey Monkey analysis, it was not possible for the assessors to allocate these addresses to individual respondents.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Institute for Quality and Efficiency in Health Care (IQWiG), Im Mediapark 8, 50670 Cologne, Germany. ²Institute for Research in Operative Medicine (IFOM), University of Witten/Herdecke, Campus Cologne, Cologne, Germany. ³Senior Professor for Health Services Research, University of Witten/Herdecke, Alfred-Herrhausen-Straße 50, 58448 Witten, Germany. ⁴Medical Advisory Service of the German Social Health Insurance (MDS), Theodor-Althoff-Straße 47, 45133 Essen, Germany.

Received: 16 July 2017 Accepted: 21 February 2018

Published online: 27 February 2018

References

- Field MJ, Lohr KN, editors. *Clinical practice guidelines: directions for a new program*. Washington: National Academy Press; 1990.
- Graham RM, Mancher M, Miller-Wolman D, Greenfield S, Steinberg E, editors. *Clinical practice guidelines we can trust*. Washington: National Academies Press; 2011.
- Hakkennes S, Dodd K. Guideline implementation in allied health professions: a systematic review of the literature. *Qual Saf Health Care*. 2008;17(4):296–300.
- Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess*. 2004;8(6):iii–v. 1–72.
- Medves J, Godfrey C, Turner C, Paterson M, Harrison M, MacKenzie L, Durando P. Systematic review of practice guideline dissemination and implementation strategies for healthcare teams and team-based practice. *Int J Evid Based Healthc*. 2010;8(2):79–89.
- Ray-Coquard I, Philip T, Lehmann M, Fervers B, Farsi F, Chauvin F. Impact of a clinical guidelines program for breast and colon cancer in a French cancer center. *JAMA*. 1997;278(19):1591–5.
- Smith TJ, Hillner BE. Ensuring quality cancer care by the use of clinical practice guidelines and critical pathways. *J Clin Oncol*. 2001;19(11):2886–97.
- Ray-Coquard I, Philip T, De Laroche G, Froger X, Suchaud JP, Voloch A, Mathieu-Daude H, Fervers B, Farsi F, Browman GP, et al. A controlled "before-after" study: impact of a clinical guidelines programme and regional cancer network organization on medical practice. *Br J Cancer*. 2002;86(3):313–21.
- Grimshaw J, Eccles M, Tetroe J. Implementing clinical guidelines: current evidence and future implications. *J Contin Educ Heal Prof*. 2004;24(Suppl 1):S31–7.
- Alonso-Coello P, Irfan A, Sola I, Gich I, Delgado-Noguera M, Rigau D, Tort S, Bonfill X, Burgers J. 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.
- 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):103–6.
- Kryworuchko J, Stacey D, Bai N, Graham ID. Twelve years of clinical practice guideline development, dissemination and evaluation in Canada (1994 to 2005). *Implement Sci*. 2009;4:49.
- 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):1628–33.
- Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA*. 1999;281(20):1900–5.
- AGREE Collaboration. *The appraisal of guidelines for research & evaluation (AGREE) instrument*. London: AGREE Research Trust; 2006.
- Appraisal of guidelines for research and evaluation II: AGREE II instrument [http://www.agreertrust.org/wp-content/uploads/2013/10/AGREE-II-Users-Manual-and-23-item-Instrument_2009_UPDATE_2013.pdf].
- Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ*. 2010;182(18):E839–42.
- Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Hanna SE, Makarski J. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. *CMAJ*. 2010;182(10):1045–52.
- Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Hanna SE, Makarski J. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ*. 2010;182(10):E472–8.
- Brousseau L, Rahman P, Toupin-April K, Poitras S, King J, De Angelis G, Loew L, Casimiro L, Paterson G, McEwan J. A systematic critical appraisal for non-pharmacological management of osteoarthritis using the appraisal of guidelines research and evaluation II instrument. *PLoS One*. 2014;9(1):e82986.
- Lee GY, Yamada J, Kyololo OB, Shorkey A, Stevens B. Pediatric clinical practice guidelines for acute procedural pain: a systematic review. *Pediatrics*. 2014;133(3):500–15.
- Polus S, Lerberg P, Vogel J, Watananirun K, Souza JP, Mathai M, Gulmezoglu AM. Appraisal of WHO guidelines in maternal health using the AGREE II assessment tool. *PLoS One*. 2012;7(8):e38891.
- Sabharwal S, Gauher S, Kyriacou S, Patel V, Holloway I, Athanasiou T. Quality assessment of guidelines on thromboprophylaxis in orthopaedic surgery. *Bone Joint J*. 2014;96-B(1):19–23.
- Sabharwal S, Patel NK, Gauher S, Holloway I, Athanasiou T. High methodologic quality but poor applicability: assessment of the AAOS guidelines using the AGREE II instrument. *Clin Orthop*. 2014;472(6):1982–8.
- Hoffmann-Eßer W, Siering U, Neugebauer EA, Brockhaus AC, Lampert U, Eikermann M. Guideline appraisal with AGREE II: systematic review of the current evidence on how users handle the 2 overall assessments. *PLoS One*. 2017;12(3):e0174831.
- Resource Toolkit for Guideline Adaptation: Version 1.0 [n.a.]
- Brousseau L, Rahman P, Poitras S, Toupin-April K, Paterson G, Smith C, King J, Casimiro L, De Angelis G, Loew L, et al. A systematic critical appraisal of non-pharmacological management of rheumatoid arthritis with appraisal of guidelines for research and evaluation II. *PLoS One*. 2014;9(5):e95369.
- Bindselev JB, Schroll J, Gotzsche PC, Lundh A. Underreporting of conflicts of interest in clinical practice guidelines: cross sectional study. *BMC Med Ethics*. 2013;14:19.
- Feuerstein J, Gifford A, Akbari M, Goldmann J, Leffler D, Sheth S, Cheifetz A. Systematic analysis underlying the quality of the scientific evidence and conflicts of interest in gastroenterology practice guidelines. *Am J Gastroenterol*. 2013;108(11):1686–93.

30. Norris SL, Holmer HK, Ogdan LA, Burda BU. Conflict of interest in clinical practice guideline development: a systematic review. *PLoS One*. 2011;6(10):e25153.
31. Fervers B, Burgers JS, Haugh MC, Brouwers M, Browman G, Cluzeau F, Philip T. Predictors of high quality clinical practice guidelines: examples in oncology. *Int J Qual Health Care*. 2005;17(2):123–32.
32. Acuna-Izcaray A, Sanchez-Angarita E, Plaza V, Rodrigo G, Montes de Oca M, Gich I, Bonfill X, Alonso-Coello P. Quality assessment of asthma clinical practice guidelines: a systematic appraisal. *Chest*. 2013;144(2):390–7.
33. Arevalo-Rodriguez I, Pedraza OL, Rodriguez A, Sanchez E, Gich I, Sola I, Bonfill X, Alonso-Coello P. Alzheimer's disease dementia guidelines for diagnostic testing: a systematic review. *Am J Alzheimers Dis Other Demen*. 2013;28(2):111–9.
34. Bekkering GE, Aertgeerts B, Asueta-Lorente JF, Autrique M, Goossens M, Smets K, Van Bussel JC, Vanderplasschen W, Van Royen P, Hannes K. Practitioner review: evidence-based practice guidelines on alcohol and drug misuse among adolescents; a systematic review. *J Child Psychol Psychiatry*. 2014;55(1):3–21.
35. Bragge P, Pattuwage L, Marshall S, Pitt V, Piccenna L, Stergiou-Kita M, Tate RL, Teasell R, Wiseman-Hakes C. Quality of guidelines for cognitive rehabilitation following traumatic brain injury. *J Head Trauma Rehabil*. 2014;29(4):277–89.
36. Huang TW, Lai JH, Wu MY, Chen SL, Wu CH, Tam KW. Systematic review of clinical practice guidelines in the diagnosis and management of thyroid nodules and cancer. *BMC Med*. 2013;11:191.
37. Kim SG, Jung HK, Lee HL, Jang JY, Lee H, Kim CG, Shin WG, Shin ES, Lee YC. Guidelines for the diagnosis and treatment of helicobacter pylori infection in Korea, 2013 revised edition. *J Gastroenterol Hepatol*. 2014;29(7):1371–86.
38. Langton JM, Pearson SA. eViQ cancer treatments online: how does the web-based protocol system fare in a comprehensive quality assessment? *Asia Pac J Clin Oncol*. 2011;7(4):357–63.
39. Larmer PJ, Reay ND, Aubert ER, Kersten P. Systematic review of guidelines for the physical management of osteoarthritis. *Arch Phys Med Rehabil*. 2014;95(2):375–89.
40. Lopez-Vargas PA, Tong A, Sureshkumar P, Johnson DW, Craig JC. Prevention, detection and management of early chronic kidney disease: a systematic review of clinical practice guidelines. *Nephrology*. 2013;18(9):592–604.
41. Norberg MM, Turner MW, Rooke SE, Langton JM, Gates PJ. An evaluation of web-based clinical practice guidelines for managing problems associated with cannabis use. *J Med Internet Res*. 2012;14(6):e169.
42. Parisi P, Vanacore N, Belcastro V, Carotenuto M, Del Giudice E, Mariani R, Papetti L, Pavone P, Savasta S, Striano P. Clinical guidelines in pediatric headache: evaluation of quality using the AGREE II instrument. *J Headache Pain*. 2014;15(1):57.
43. Piano V, Schalkwijk A, Burgers J, Verhagen S, Kress H, Hekster Y, Lanteri-Minet M, Engels Y, Vissers K. Guidelines for neuropathic pain management in patients with cancer: a European survey and comparison. *Pain Practice*. 2013;13(5):349–57.
44. Rios E, Seron P, Lanas F, Bonfill X, Quigley EMM, Alonso-Coello P. Evaluation of the quality of clinical practice guidelines for the management of esophageal or gastric variceal bleeding. *Eur J Gastroenterol Hepatol*. 2014;26(4):422–31.
45. Rohde A, Worrall L, Le Dorze G. Systematic review of the quality of clinical guidelines for aphasia in stroke management. *J Eval Clin Pract*. 2013;19(6):994–1003.
46. Sanclemente G, Acosta JL, Tamayo ME, Bonfill X, Alonso-Coello P. Clinical practice guidelines for treatment of acne vulgaris: a critical appraisal using the AGREE II instrument. *Arch Dermatol Res*. 2014;306(3):269–77.
47. Santos F, Sola I, Rigau D, Arevalo-Rodriguez I, Seron P, Alonso-Coello P, Berard A, Bonfill X. Quality assessment of clinical practice guidelines for the prescription of antidepressant drugs during pregnancy. *Curr Clin Pharmacol*. 2012;7(1):7–14.
48. Schildmann EK, Schildmann J, Kiesewetter I. Medication and monitoring in palliative sedation therapy: a systematic review and quality assessment of published guidelines. *J Pain Symptom Manag*. 2015;49(4):734–46.
49. Schoenmaker NJ, Tromp WF, Van der Lee JH, Offringa M, Craig JC, Groothoff JW. Quality and consistency of clinical practice guidelines for the management of children on chronic dialysis. *Nephrology Dialysis Transplantation*. 2013;28(12):3052–61.
50. Seron P, Lanas F, Rios E, Bonfill X, Alonso-Coello P. Evaluation of the quality of clinical guidelines for cardiac rehabilitation: a critical review. *J Cardiopulm Rehabil Prev*. 2014;35(1):1–12.
51. Shen J, Sun M, Zhou B, Yan J. Nonconformity in the clinical practice guidelines for subclinical Cushing's syndrome: which guidelines are trustworthy? *Eur J Endocrinol*. 2014;171(4):421–31.
52. Siering U, Hoffmann-Eßer W, Neugebauer EA, Eikermann M. Is there a cut-off for high-quality guidelines? A systematic analysis of current guideline appraisals using the AGREE-II instrument [poster]. G-I-N Conference; 7th - 10th October 2015; Amsterdam.
53. Wang Y, Luo Q, Li Y, Wang H, Deng S, Wei S, Li X. Quality assessment of clinical practice guidelines on the treatment of hepatocellular carcinoma or metastatic liver cancer. *PLoS One*. 2014;9(8):e103939.
54. Yan J, Min J, Zhou B. Diagnosis of pheochromocytoma: a clinical practice guideline appraisal using AGREE II instrument. *J Eval Clin Pract*. 2013;19(4):626–32.
55. Ye ZK, Li C, Zhai SD. Guidelines for therapeutic drug monitoring of vancomycin: a systematic review. *PLoS One*. 2014;9(6):e99044.
56. Burnett HF, Tanoshima R, Chandranipapongse W, Madadi P, Ito S, Ungar WJ. Testing for thiopurine methyltransferase status for safe and effective thiopurine administration: a systematic review of clinical guidance documents. *Pharmacogenomics J*. 2014;14(6):493–502.
57. Haran C, Van Driel M, Mitchell BL, Brodribb WE. Clinical guidelines for postpartum women and infants in primary care: a systematic review. *BMC Pregnancy Childbirth*. 2014;14:51.
58. Gesamtbewertung der Leitlinienqualität mit AGREE II: wird das AGREE II-Instrument vollständig umgesetzt? [<http://www.egms.de/static/de/meetings/ebm2015/15ebm095.shtml>].
59. White PE, Shee AW, Finch CF. Independent appraiser assessment of the quality, methodological rigour and transparency of the development of the 2008 international consensus statement on concussion in sport. *Br J Sports Med*. 2014;48(2):130–4.
60. Al-Ansary LA, Tricco AC, Adi Y, Bawazeer G, Perrier L, Al-Ghonaim M, AlYousefi N, Tashkandi M, Straus SE. A systematic review of recent clinical practice guidelines on the diagnosis, assessment and management of hypertension. *PLoS One*. 2013;8(1):e53744.
61. Stacey D, Macartney G, Carley M, Harrison MB, Costars TP. Development and evaluation of evidence-informed clinical nursing protocols for remote assessment, triage and support of cancer treatment-induced symptoms. *Nurs Res Pract*. 2013;2013:171872.
62. Tian H, Gou Y, Pan Y, Li Q, Wei D, Wang Z, Niu X, Liang W, Zhang Y. Quality appraisal of clinical practice guidelines on glioma. *Neurosurg Rev*. 2015;38(1):39–47.
63. Halbesleben JR, Whitman MV. Evaluating survey quality in health services research: a decision framework for assessing nonresponse bias. *Health Serv Res*. 2013;48(3):913–30.
64. Deutsches Instrument zur methodischen Leitlinien-Bewertung (DELBI): Fassung 2005/2006 + Domäne 8 (2008) [<http://www.leitlinien.de/mbd/edocs/pdf/literatur/delbi-fassung-2005-2006-domaene-8-2008.pdf>].
65. Sinclair M, O'Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. *BMC Med Res Methodol*. 2012;12:132.
66. Geschäftsordnung: Ethik-Kommission der Universität Witten/Herdecke [<http://www.ethik-kommission-uw.de/Geschaeftsordnung/geschaeftsordnung.html>].