

RESEARCH ARTICLE

Open Access



# Validation and reliability of a guideline appraisal mini-checklist for daily practice use

Andrea Siebenhofer<sup>1,2\*</sup>, Thomas Semlitsch<sup>1</sup>, Thomas Herborn<sup>2</sup>, Ulrich Siering<sup>3</sup>, Ina Kopp<sup>4</sup> and Johannes Hartig<sup>5</sup>

## Abstract

**Background:** The use of comprehensive instruments for guideline appraisal is time-consuming and requires highly qualified personnel. Since practicing physicians are generally busy, the rapid-assessment Mini-Checklist (MiChe) tool was developed to help them evaluate the quality and utility of guidelines quickly. The aim of this study was to validate the MiChe in comparison to the AGREE II instrument and to determine its reliability as a tool for guideline appraisal.

**Methods:** Ten guidelines that are relevant to general practice and had been evaluated by 2 independent reviewers using AGREE II were assessed by 12 GPs using the MiChe. The strength of the correlation between average MiChe ratings and AGREE II total scores was estimated using Pearson's correlation coefficient. Inter-rater reliability for MiChe overall quality ratings and endorsements was determined using intra-class correlations (ICC) and Kendall's W for ordinal recommendations. To determine the GPs' satisfaction with the MiChe, mean scores for the ratings on five questions were computed using a six-point Likert scale.

**Results:** The study showed a high level of agreement between MiChe and AGREE II in the quality rating of guidelines (Pearson's  $r = 0.872$ ;  $P < 0.001$ ). Inter-rater-reliability for overall MiChe ratings (ICC = 0.755;  $P < 0.001$ ) and endorsements (Kendall's  $W = 0.73$ ;  $P < 0.001$ ) were high. The mean time required for guideline assessment was less than 15 min and user satisfaction was generally high.

**Conclusions:** The MiChe performed well in comparison to AGREE II and is suitable for the rapid evaluation of guideline quality and utility in practice.

**Trial registration:** German Clinical Trials Register: DRKS00007480

**Keywords:** Guideline assessment, Validation, Reliability, AGREE II instrument, Mini-checklist, General practitioners

## Background

Clinical practice guidelines are defined by the Institute of Medicine as “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]. There is evidence to suggest that rigorously developed guidelines have the power to translate the complexity of scientific research

findings and other evidence into recommendations for healthcare action [2–8]. To increase guideline quality, several institutions [1, 9–23] have prepared manuals, that attempt to define standards for guideline developers. At the same time, tools have been developed to help potential guideline users to assess guideline quality. The most commonly used international guideline appraisal tool is the AGREE II Instrument [24], but its use is time consuming and demands considerable skill on the part of the guideline appraiser.

\* Correspondence: andrea.siebenhofer@medunigraz.at

<sup>1</sup>Institute of General Practice and Evidence-Based Health Services Research, Medical University of Graz, Auenbruggerplatz 2/9, 8036 Graz, Austria

<sup>2</sup>Institute of General Practice, Goethe University Frankfurt, Frankfurt, Germany  
Full list of author information is available at the end of the article

Graham 2000 identified and compared guideline appraisal tools in a systematic review [25], which was later updated by Vlayen in 2005 [26] and Siering in 2013 [27]. Siering identified 40 different appraisal tools that vary considerably in terms of the number of quality dimensions taken into account. In the opinion of the authors, appraisal tools containing many quality dimensions may not represent the best choice in all cases. Depending on the problem being addressed, a tool containing a few well thought out questions may well suffice.

To be effective, guidelines must be applied by clinicians. An appraisal tool that is quick and easy to use and assesses the most relevant quality dimensions of a guideline would generally encourage their wider use. We therefore developed and published a mini-checklist (MiChe) for the rapid appraisal of the usefulness and quality of a guideline for clinical practitioners. Detailed information on the development process is provided elsewhere [28]. However, the development was based on a systematic search in guideline directories and bibliographic databases for guideline appraisal instruments. The assessment criteria used in the retrieved instruments were identified, and their importance to the development of an effective rapid assessment tool was judged by German guideline experts. The key criteria for MiChe were then selected on the basis of the most commonly found criteria in the retrieved instruments and the ratings from the expert survey.

Our primary objective was to validate the MiChe vs. the AGREE II instrument and determine its reliability for daily users in terms of ability to rapidly assess the strengths and weaknesses of a guideline and dependability of content.

## Methods

Twelve general practitioners (GPs) were asked to use the MiChe to assess 10 eligible guidelines that had already been evaluated by 2 independent reviewers using the AGREE II instrument.

## Aims

### Primary outcomes

- Validate the overall quality rating of AGREE II as the gold standard vs. the overall quality rating of the MiChe.
- Estimate the inter-rater reliability of the overall quality rating assigned by different guideline assessors using the MiChe.

Secondary outcomes relating to the MiChe alone:

- Demonstrate the inter-rater reliability of endorsement: willing to recommend this guideline for use in practice (“yes”; “yes, with certain reservations” or “no”).
- Demonstrate user satisfaction to indicate whether the MiChe would help raters decide whether to use a specific guideline or not.
- Feedback to improve the MiChe.
- Time required for an assessment using MiChe.

### Tertiary outcomes:

Evaluate the correlation between overall quality rating and endorsement of the MiChe vs. quality ratings of individual items (domain 1 – 6) of AGREE II.

## Participants

During a quality circle that took place in November 2014, a convenience sample of GPs working as resident doctors was recruited from the more than 100 accredited general practices that make up GP Research Network Frankfurt (ForN) [29]. GPs with experience of guideline development or appraisal, i.e. members of guideline commissions or GPs in post-graduate training were excluded. All participants received 1.5-h of training on the basics of guideline development and appraisal at the Institute of General Practice in Frankfurt. In addition, a sample guideline was provided, along with instructions to read and appraise the guideline using the MiChe. Participants later received a folder with a printed version of 10 guidelines and were asked to use the MiChe to appraise them. Results were returned by mail.

In the Federal State of Hesse, Germany, the code of medical ethics allows formal ethical approval to be waived upon request if the biomedical research to be conducted on patients or healthy volunteers involves no risky procedures and is not invasive. We contacted the local ethics committee of Frankfurt University Hospital, who informed us that ethical approval could be waived. As we were not expecting ethical approval to be required, participating GPs were only required to provide their verbal consent before starting to review the guidelines.

## Guideline selection and guideline assessment tools

The selection process was initiated by choosing guidelines already known to the study team and by studying a list of 20 guidelines, sorted according to their characteristics. Of these, 10 guidelines were selected that covered subjects that are relevant to general practice, had varying AGREE II quality levels, varied in length and were written in either German or English. Two independent reviewers with professional

expertise in guideline appraisal from the Institute for Quality and Efficiency in Health Care (IQWiG) first assessed the guidelines using the AGREE II instrument. These assessments served as the gold standard for the validation [24].

The MiChe [28, 30] contains 8 key-criteria that focus on important methodological features (quality of guideline creation, quality of reporting, quality of presentation, quality of evidence synthesis), as well as a 3-level assessment scale (see Fig. 1).

#### Data management

The GPs had to complete 10 MiChes for the 10 different guidelines and a short questionnaire on their personal characteristics and previous experience of guidelines. To indicate whether the MiChe would help raters to decide whether or not to use a guideline, 5 questions addressed user satisfaction (satisfaction, frequency of future use, makes it easier to deal with guidelines, influence of guideline recommendations on future daily practice use, comprehensibility) using a six-point Likert scale from 1 – 6, with 1 indicating a strong positive response. The average time required for assessment was measured separately for each guideline and GP. Suggestions for improvement and notes were documented in a free text field.

Ethics approval was not required, since no patients were involved. The protocol for this validity and reliability study was registered in the German Clinical Trials Register: DRKS00007480.

#### Data analyses

##### Validity

The strength of the correlation between the average MiChe ratings of the guidelines and the AGREE II total score were estimated using Pearson's correlation coefficient. A correlation of more than 0.70 is considered desirable. Additionally, correlations between the average recommendation on the MiChe and the separate AGREE II domains were calculated using Spearman's rank order correlation.

##### Inter-rater reliability

Inter-rater agreement for the various MiChe ratings of the GPs was determined using intra-class correlations (ICC) and Kendall's W for ordinal recommendations for endorsement ("yes", "yes, with certain reservations", "no"). For both coefficients, we consider values over 0.75 as good, values between 0.40 and 0.75 as moderate, and values below 0.40 as poor [31].

##### Evaluation of the mini checklist

To determine the GPs' satisfaction with the MiChe, mean scores for the ratings on the five questions

quoted in the data management section were computed.

##### Determination of required sample size

The required sample size to estimate inter-rater agreement can be determined by defining a specific null hypothesis and a specific alternative hypothesis, and selecting a desired type I and type II error rate ( $\alpha$  and  $\beta$  level) [32]. We chose to set ICC = 0.50 as the lowest acceptable agreement for the null hypothesis and an expected value of ICC = 0.75 for the alternative hypothesis. For  $\alpha = 0.05$  and  $\beta = 0.20$ , 10 GPs would have to evaluate 14 guidelines. The 10 guidelines evaluated in our study would still yield a statistical power between 60 % and 70 %, and, as the guidelines are not sampled randomly but selected to elicit high variation in the AGREE II and MiChe ratings, it would probably be higher. This, in turn, makes it more likely to statistically confirm high inter-rater agreement.

## Results

### Characteristics of the GPs and the tested guidelines

Twelve GPs (6 female) participated in our study. Their mean age was 53 (SD 7) years and mean professional experience as a GP 19 (SD 7) years; 6 worked in a joint practice and 7 in a rural area with less than 60,000 inhabitants. None of the participants had used a guideline assessment tool before, but all 12 GPs had previously used guidelines as a source of information (Table 1).

The included guidelines were published between 2006 and 2013, and covered different areas of relevance to general practice. Six guidelines were in German and 4 in English. They differed in length from 4 to 278 pages. The overall quality of the guidelines as assessed by AGREE II varied between 2 and 6 points on the 7-point scale. Four of them received a recommendation of "yes", 4 of "yes, with certain reservations" and 2 were given a "no" recommendation. The average MiChe overall quality score across the 12 GPs ranged from 2.4 (SD 1.0) to 6.7 (SD 0.7) for the 10 guidelines. Based on the MiChe assessment, 6 guidelines received a majority recommendation of "yes", 1 of "yes, with certain reservations" and 3 were given a "no" recommendation by the majority of the GPs. The total AGREE II score was lower than the total MiChe score for 7 of the 10 guidelines [33–39] and higher for the remaining 3 [40–42]. The DEGAM guideline on heart failure [34] was ranked best overall by both instruments and 2 guidelines [39, 42] were poorly ranked by both assessment tools (Table 2).

**Methodological Guideline Quality – Mini-Checklist**

1. The guideline has been written in a generally comprehensible manner and its key recommendations are easy to identify.

YES                       TO SOME EXTENT                       NO

2. The guideline’s target audiences and scope of application were specified.

YES                       TO SOME EXTENT                       NO

3. The background, the objectives of the guideline, and the patients for whom the guideline is relevant were clearly described.

YES                       TO SOME EXTENT                       NO

4. The persons that developed the guideline are named, and their financial independence and any conflicts of interest are clearly documented.

YES                       TO SOME EXTENT                       NO

5. The search for evidence was systematic and the criteria used to select evidence were described.

YES                       TO SOME EXTENT                       NO

6. The guideline recommendations are unambiguous and the evidence they are based on is clearly presented.

YES                       TO SOME EXTENT                       NO

7. Different treatment options are presented that take account of potential benefits, side effects and risks.

YES                       TO SOME EXTENT                       NO

8. Clear information is provided on how up-to-date the guideline is and for how long this is expected to be the case.

YES                       TO SOME EXTENT                       NO

**Overall assessment of the quality of the guideline based on above results:**

	1	2	3	4	5	6	7	
<b>Very poor</b>								<b>Very good</b>

**Would you recommend others use the guideline?**

**Yes**                                      **Yes, with certain reservations**                                      **No**

**Fig. 1** Mini-Checklist (MiChe)

**Table 1** Characteristics of participating general practitioners

Characteristics (N = 12)	
Age [mean (SD)]	53 (7.0)
Gender	
Male [%]	50
Female [%]	50
Geographical delimitation	
Urban [%]	42
Rural [%]	58
Medical specialization	
General practitioner (GP) [%]	67
Specialist in Internal Medicine practising as GP [%]	33
Professional experience	
Years in practice [mean (SD)]	25 (7.0)
Years as GP [mean (SD)]	19 (7.3)
Guideline experience	
Used guidelines as source of information [%]	100
Former usage of e.g. AGREE or German equivalent instrument [%]	0 <sup>a</sup>

<sup>a</sup> one participant declined to respond

### Primary endpoints on validity and inter-rater reliability of the overall quality rating

The average MiChe quality rating of the guidelines was strongly related to the total AGREE II score (Pearson's  $r = 0.872$ ; one-tailed  $P < 0.001$ ), as were the recommendations to use the guidelines (Spearman's  $\rho = 0.909$ ; one-tailed  $P < 0.001$ ). Both results indicate a high level of validity in the MiChe ratings.

Inter-rater reliability for the overall MiChe quality ratings of the 12 GPs was ICC = 0.755 (one-tailed  $P < 0.001$ ; 95 % CI: 0.572 < ICC < 0.914), indicating good agreement between raters.

### Secondary endpoint for the assessment of the mini-checklist

For the inter-rater reliability of willingness to recommend the guidelines, or "endorsement" for use in practice, Kendall's W for ordinal ratings was 0.73 ( $P < 0.001$ ), also indicating good agreement between raters.

Concerning user satisfaction, the mean value for overall satisfaction with the MiChe was 1.7 (SD 0.65) on the six-point Likert scale. As an indicator of future use, the mean value for the MiChe was 2.8 (SD 0.75). The question whether the use of MiChe makes it easier to deal with guidelines resulted in a mean value of 2.0 (SD 0.85). For the question on possible influence on the future implementation of guideline recommendations in

daily practice work, the mean value was 2.2 (SD 0.83). For the question on the comprehensibility of the MiChe, the mean value was 1.3 (SD 0.65). For further details see Fig. 2a-e.

The 12 GPs required an average of 12.9 min (SD 9.2) for the MiChe assessment. The mean appraisal time for each guideline ranged from 6.8 to 20.1 min (Table 2).

Eight GPs provided feedback. They would have liked to have a more differentiated assessment scale and mentioned that questions 2 and 3 were rather similar in content. Another suggestion was to add a test question regarding the existence of a structured pocket-version of the guideline for use in practice. Some GPs reckoned that assessments may depend on the language in which the guideline was written and some criticized the questions for their focus on methodological and formal aspects, as they felt this may influence a result even when a recommendation was of proven efficacy. It was further mentioned that the MiChe does not assess the practical usefulness of a guideline on a day-to-day basis.

### Correlation between the domains of AGREE II and the MiChe

The average overall quality rating of the 10 guidelines using MiChe was highly correlated (Pearson's correlations between 0.74 and 0.87) with the expert ratings in the AGREE II domains II - IV and VI. Correlations for the domains I and V were not statistically significant (Table 3). The pattern of correlations on the level of recommendation with the individual AGREE II domains is very similar. Details of the AGREE II for assessment per domain are shown in Table 4 in the Appendix.

### Discussion

Guidelines have the potential to improve the quality and safety of health care, but are often not used in clinical practice. In order to be helpful, a guideline must be of high methodological quality. The use of comprehensive research-focused instruments such as AGREE II is time-consuming and requires highly qualified personnel. Since practicing physicians are generally very busy, a new rapid-assessment tool (MiChe) was developed to help them evaluate the quality and utility of a guideline quickly and on their own.

This paper presents the results of a validation-study for MiChe [28], as compared to the AGREE II instrument [24]. Ten guidelines that are relevant to general practice and reflect a spectrum of methodological quality ranging from low to high according to an appraisal using the AGREE II instrument were included and assessed using the MiChe by 12 GPs that

**Table 2** Mean overall rating scores for AGREE II and the MiChe

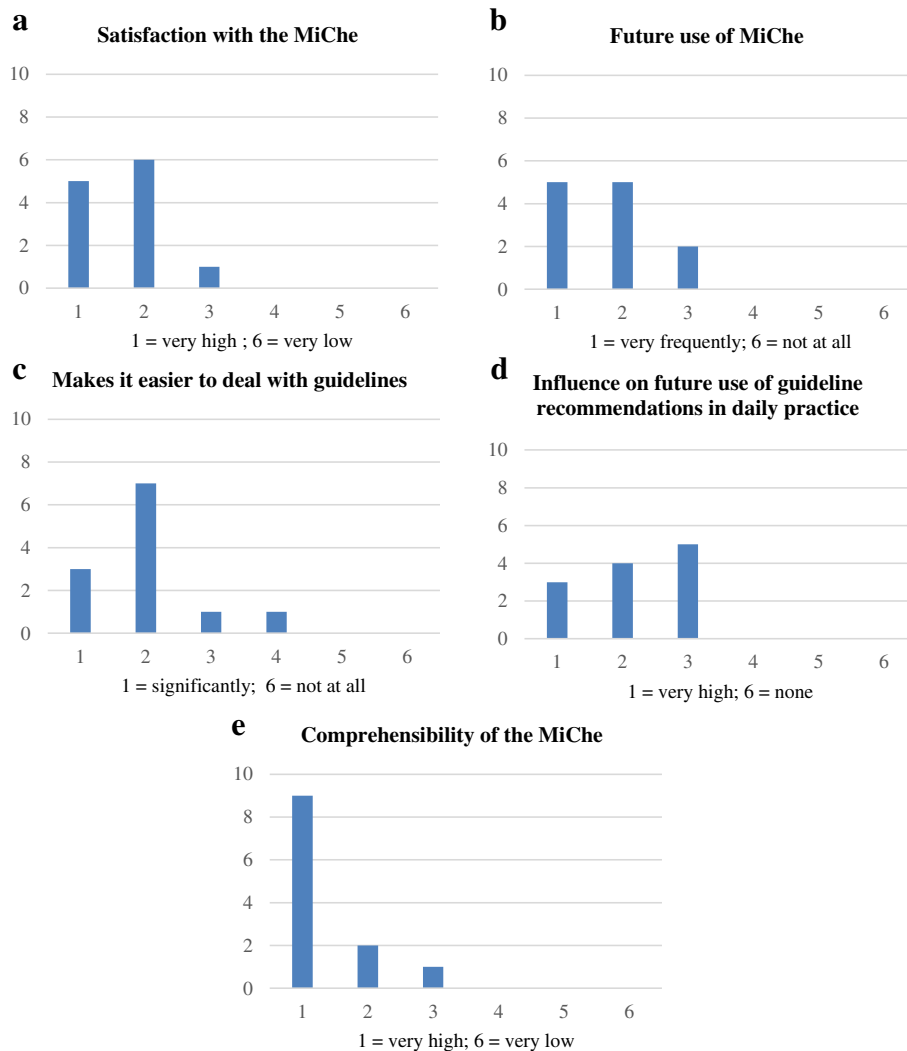
Guideline Title/Leading Organization (Abbreviation)	Year of publication/pages/ region/language	Topic	AGREE II		MiChe		
			Overall quality	Recommendation for use	Overall quality [mean (SD)]	Recommendation for use [%]	Mean duration of appraisal [minutes]
DEGAM Leitlinie Nr. 9, Herzinsuffizienz [Guideline no. 9 cardiac failure]/Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin (DEGAM 2006) [34]	2006/278/GER/German	Cardiac failure	6	Yes	6.7 (0.65)	Yes: 92 Yes, with certain reservations: 8 No: 0	12.1
Guideline for the Management of Acute Bronchitis/Alberta  Clinical Practice Guideline Working Group (ACPG 2008) [40]	2008/6/CDN/English	Bronchitis	3	Yes, with modifications	2.5 (1.20)	Yes: 17 Yes, with certain reservations: 33 No: 50	9.7
Diabetes und Schwangerschaft [Diabetes and pregnancy]/ Deutschen Diabetes-Gesellschaft (DDG 2008) [36]	2008/62/GER/German	Diabetes in pregnancy	3	Yes, with modifications	4.9 (1.31)	Yes: 67 Yes, with certain reservations: 33 No: 0	11.2
Suspected cancer in primary care: Guidelines for investigation, referral and reducing ethnic disparities/New Zealand Guidelines Group (NZGG 2009) [38]	2009/190/NZ/English	Cancer	4	Yes, with modifications	4.3 (1.15)	Yes: 8.3 Yes, with certain reservations: 83.3 No: 8.3	16.8
Guidelines for the practice of diabetes self-management education/American Association of Diabetes Educators (AADE 2010) [41]	2010/46/USA/English	Diabetes	4	Yes, with modifications	2.8 (1.10)	Yes: 75 Yes, with certain reservations: 17 No: 8	20.1
Nebennierenrinden-Insuffizienz [Suprarenal gland failure]/ Deutschen Gesellschaft für Endokrinologie; Gesellschaft für Kinderheilkunde und Jugendmedizin (DGKJ 2010) [42]	2010/4/GER/German	Suprarenal gland failure	2	No	1.8 (1.11)	Yes: 8 Yes, with certain reservations: 17 No: 75	6.8
Leitlinie Management der frühen rheumatoiden Arthritis [Guideline management of early rheumatoid arthritis]/Deutsche Gesellschaft für Rheumatologie (DGR 2011) [37]	2011/163/GER/German	Rheumatoid arthritis	5	Yes	6.7 (0.65)	Yes: 100 Yes, with certain reservations: 0 No: 0	13.3
Aszites, spontan bakterielle Peritonitis, hepatorenales Syndrom [Ascites, spontaneous bacterial peritonitis, hepatorenal syndrome]/Deutsche Gesellschaft für Verdauungs- und Stoffwechselkrankheiten (DGVS 2011) [33]	2011/39/GER/German	Renal failure	5	Yes	6.3 (0.90)	Yes: 92 Yes, with certain reservations: 8 No: 0	14.1

**Table 2** Mean overall rating scores for AGREE II and the MiChe (*Continued*)

Leitlinie Management der frühen rheumatoiden Arthritis [Guideline management of early rheumatoid arthritis]/Deutsche Gesellschaft für Rheumatologie (DGR 2011) [37]	2011/163/GER/German	Rheumatoid arthritis	5	Yes	6.7 (0.65)	Yes: 100 Yes, with certain reservations: 0 No: 0	13.3
Fieber unklarer Genese [Fever of unknown origin]/Gesellschaft für Kinder- und Jugendrheumatologie Deutschen Gesellschaft für Kinder- und Jugendmedizin (DGKJ 2013) [39]	2013/8/GER/German	Fever	2	No	2.4 (1.00)	Yes: 8 Yes, with certain reservations: 17 No: 75	6.8
Assessment and Management of Foot Ulcers for People with Diabetes/Registered Nurses' Association of Ontario (RNAO 2013) [35]	2013/160/USA/English	Diabetic foot lesions	5	Yes	5.5 (1.31)	Yes: 67 Yes, with certain reservations: 33 No: 0	18.2

CDN Canada, GER Germany, NZ New Zealand, SD Standard Deviation, USA United States of America





**Fig. 2** User satisfaction with the Mini-Checklist - 5 queries (**a** Satisfaction with the MiChe; **b** Future use of MiChe; **c** Makes it easier to deal with guidelines; **d** Influence on future use of guideline recommendations in daily practice; **e** Comprehensibility of the MiChe)

**Table 3** Correlation between mean AGREE II domain-scores and MiChe overall quality rate/MiChe recommendation for use

AGREE II Domain	Correlation with overall MiChe quality rating	Correlation with MiChe recommendation for use in practice
Domain 1: <i>Scope and Purpose</i>	0.55	0.48
Domain 2: <i>Stakeholder involvement</i>	0.80*	0.72*
Domain 3: <i>Rigor of development</i>	0.87*	0.80*
Domain 4: <i>Clarity of presentation</i>	0.74*	0.69*
Domain 5: <i>Applicability</i>	0.50	0.44
Domain 6: <i>Editorial independence</i>	0.79*	0.74*

\* P < 0.05

were inexperienced in guideline appraisal. The study showed a high level of agreement in the quality rating of guidelines between MiChe and AGREE II and recommendations to use the guideline. In addition, inter-rater-reliability for the overall MiChe quality ratings and MiChe recommendation for use in practice were high. With high user satisfaction and a mean time required for guideline assessment of less than 15 min, the MiChe was shown to be suitable for the rapid assessment of guideline quality and utility in practice.

Although the study shows high validity and inter-rater-reliability for the MiChe, it nevertheless has a number of limitations. The validation of the MiChe was performed using the AGREE II instrument as the gold standard for guideline appraisal. AGREE II is the most frequently used instrument for the assessment



of methodological guideline quality and has been validated in several studies [43–46]. Nevertheless it remains unclear whether all items and domains of AGREE II contribute equally to the quality of a guideline [25]. The results of our findings that the same individual AGREE II items (II-IV, VI) correlated with both average overall quality ratings and levels of recommendation should not be over-interpreted. The correlations were probably caused by chance, even though it was an interesting result that for domain 3 in particular (rigor of development), the correlation was very high. In addition, we clearly recognize that the questions on the MiChe cannot be seen as independent of the individual AGREE II items. Further empirical studies are needed to find out which items and quality dimensions are essential to the assessment of guideline quality.

Unfortunately we didn't measure the time it took to assess the guidelines using the AGREE II instrument. However, the AGREE II consortium recommends the use of at least 2 and preferably 4 appraisers and consists of 23 key items organized within 6 domains followed by 2 global rating items. Therefore, we assumed that it requires considerably more time and personnel resources to apply than are typically available to a GP.

Guideline appraisal instruments can be used to assess whether a guideline has been developed in a methodologically accurate and transparent way in accordance with international standards. Guidelines containing adequate information on these topics will therefore be judged to be of high (methodological) quality. But this appraisal is made regardless of whether all recommendations made in the guideline are correct or not. Thus some guidelines of high methodological quality may still contain individual recommendations that are not internally valid in terms of content. Equally, a guideline of low methodological quality may contain recommendations of high content validity [26, 47, 48].

Although the GPs involved in this study were inexperienced in guideline appraisal, they received short, basic training on guideline development and assessment before using the MiChe. A comparison between trained and untrained clinicians with regard to the usability and reliability of the MiChe was not part of this investigation. In addition, convenience sampling of the participants limits generalizability of the results. To achieve wider implementation, future research should assess whether clinicians with no prior training come to the same results as trained clinicians and apply random sampling techniques. To date only the German language version of the MiChe has been validated. It would be useful to know to what extent the use of an English

translation of the MiChe would lead to corresponding results.

A large number of manuals and instruments can be used for guideline development and quality assessment. A systematic review carried out by Siering et al in 2013 [27] identified a total of 40 different appraisal tools. Information on quality and validity was only available for 11 of these 40 tools, while detailed information concerning the validation process was reported for only 6. Among these, AGREE II was the most extensively validated instrument [43–46]. In recent years, a number of clinician-focused rapid assessment tools have been developed contemporaneously, and in addition to comprehensive research-focused instruments. Apart from MiChe, these include the iCAHE Guideline Quality Checklist [49], the Global Rating Scale (GRS) of the AGREE Collaboration [50], and the surgeons' checklist by Coroneos et al. [51]. Of these, MiChe is the only instrument of this type that is available in German and is thus more easily accessible for German speakers. It is also the only tool that has been validated for use in general practice. In 2014, Grimmer et al tested the validity, inter-rater reliability and clinical utility of the iCAHE Guideline Quality Checklist in comparison to the AGREE II instrument [49]. In their study they found a moderate to strong correlation between the iCAHE and the AGREE II scores. A comparison of these four tools was published by Semlitsch et al. in 2015 [30] and showed that, although developed independently, they all focus on a few, broad-based and very similar key questions. They can therefore only give a rudimentary impression of the value of a guideline. They are not intended to provide a comprehensive and detailed guideline appraisal, and include only a broad-based rating system.

## Conclusion

Physicians increasingly use guidelines to gain clinical knowledge. To be dependable, these guidelines need to be prepared using proper methods and to be of sufficiently high quality. The MiChe is a validated rapid-assessment instrument that allows busy physicians to assess the methodical quality of guidelines without the need for experts in guideline appraisal and judge whether a guideline is applicable in patient care or not. It thus increases the likelihood that guideline recommendations will be used in practice and contributes towards sustained improvement in patient health care.

## Declaration and availability statement

The dataset supporting the conclusions of this article are available from the corresponding author on request.

## Appendix

**Table 4** AGREE II domain-scores and overall quality for assessed guidelines

Guideline	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Domain 6	Overall Quality
DEGAM 2006 [34]	0.805	0.722	0.822	0.805	0.687	0.792	6
ACPG 2008 [40]	0.611	0.25	0.104	0.527	0	0	3
DDG 2008 [36]	0.305	0.305	0.416	0.888	0	0.583	3
NZGG 2009 [38]	0.916	0.75	0.562	0.861	0.083	0.625	4
AADE 2010 [41]	0.777	0.472	0.614	0.75	0.562	0.042	4
DGKJ 2010 [42]	0.138	0.027	0.052	0.1666	0	0	2
DGR 2011 [37]	0.694	0.666	0.75	0.777	0.333	0.833	5
DGVS 2011 [33]	0.777	0.638	0.708	0.805	0.145	0.791	5
DGKJ 2013 [39]	0.361	0.222	0.083	0.111	0	0.333	2
RNAO 2013 [35]	0.805	0.527	0.562	0.805	0.541	0.042	5

### Competing interests

Ina Kopp is Co-Author of the German Instrument for Methodological Guideline Appraisal (the German Equivalent to the AGREE-Instrument. The other authors declare they have no competing interests.

### Authors' contributions

TS and AS conceptualized and developed the MiChe instrument. AS, TH and JH conceptualized the validation and reliability study. US was involved in identifying clinical guidelines used for the validation processes and evaluated all guidelines with the AGREE II instrument. TH analyzed the GPs MiChe ratings. JH undertook statistical analysis and reporting. The paper was drafted by AS and TS, and all authors contributed to subsequent writing and review. All authors approved version of the paper submitted to review, and all have approved the revised version.

### Acknowledgements

We would like to thank Phillip Elliott for final editing of the manuscript and Alper Yurdakul from the Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany for evaluating the guidelines with the AGREE II instrument. Paul Glasziou extensively discussed the project idea during the research fellowship of Andrea Siebenhofer at the Centre for Research in Evidence-Based Practice, Bond University, Australia in 2013.

### Author details

<sup>1</sup>Institute of General Practice and Evidence-Based Health Services Research, Medical University of Graz, Auenbruggerplatz 2/9, 8036 Graz, Austria. <sup>2</sup>Institute of General Practice, Goethe University Frankfurt, Frankfurt, Germany. <sup>3</sup>Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany. <sup>4</sup>Association of Scientific Medical Societies' Institute of Medical Knowledge Management (AWMF-IMWi), Marburg, Germany. <sup>5</sup>Department of Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Frankfurt, Germany.

Received: 12 November 2015 Accepted: 22 March 2016

Published online: 02 April 2016

### References

- Graham RM, Mancher M, Miller-Wolman D, Greenfield S, Steinberg E. Clinical practice guidelines we can trust. Washington: National Academies Press; 2011.
- Grimshaw J, Eccles M, Tetroe J. Implementing clinical guidelines: current evidence and future implications. *J Contin Educ Health Prof.* 2004;24 Suppl 1:S31–7.
- Haines A, Jones R. Implementing findings of research. *BMJ.* 1994;308:1488–92.
- Hakkennes S, Dodd K. Guideline implementation in allied health professions: a systematic review of the literature. *Qual Saf Health Care.* 2008;17:296–300.
- 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;8:79–89.
- Shekelle PG, Kravitz RL, Beart J, Marger M, Wang M, Lee M. Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making. *Health Serv Res.* 2000;34:1429–48.
- Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande A. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. *Ann Intern Med.* 2003;139:493–8.
- Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ.* 1999;318:527–30.
- Methodology manual and policies from the ACCF/AHA Task Force on Practice Guidelines [[http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm\\_319826.pdf](http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf)].
- Baumann MH, Lewis SZ, Gutterman D. ACCP evidence-based guideline development: a successful and transparent approach addressing conflict of interest, funding, and patient-centered recommendations. *Chest.* 2007;132:1015–24.
- Coyne DW. Influence of industry on renal guideline development. *Clin J Am Soc Nephrol.* 2007;2:3–7.
- Eccles MP, Grimshaw JM, Shekelle P, Schünemann HJ, Woolf S. Developing clinical practice guidelines: target audiences, identifying topics for guidelines, guideline group composition and functioning and conflicts of interest. *Implement Sci.* 2012;7:60.
- Grol R, Cluzeau FA, Burgers JS. Clinical practice guidelines: towards better quality guidelines and increased international collaboration. *Br J Cancer.* 2003;89 Suppl 1:S4–8.
- Clinical practice guideline process manual: 2011 edition [<http://tools.aan.com/globals/axon/assets/9023.pdf>].
- The guidelines manual [<http://www.nice.org.uk/guidelinesmanual>].
- Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, Van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med.* 2012;156:525–31.
- Qaseem A, Snow V, Owens DK, Shekelle P. The development of clinical practice guidelines and guidance statements of the American College of Physicians: summary of methods. *Ann Intern Med.* 2010;153:194–9.
- Rosenfeld RM, Shiffman RN. Clinical practice guidelines: a manual for developing evidence-based guidelines to facilitate performance measurement and quality improvement. *Otolaryngol Head Neck Surg.* 2006;135(4 Suppl):S1–528.
- Scottish Intercollegiate Guidelines Network. SIGN 50: a guideline developer's handbook. Edinburgh: SIGN; 2011.
- Shekelle P, Woolf S, Grimshaw JM, Schünemann HJ, Eccles MP. Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development. *Implement Sci.* 2012;7:62.

21. Woolf S, Schünemann HJ, Eccles MP, Grimshaw JM, Shekelle P. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. *Implement Sci.* 2012;7:61.
22. Guidelines for WHO guidelines [http://whqlibdoc.who.int/hq/2003/EIP\_GPE\_EQC\_2003\_1.pdf].
23. AWMF Guidance Manual and Rules for Guideline Development, 1st Edition 2012. English version. [http://www.awmf.org/leitlinien/awmf-regelwerk.html].
24. Appraisal of guidelines for research and evaluation II: AGREE II instrument [http://www.agreerust.org/wp-content/uploads/2013/10/AGREE-II-Users-Manual-and-23-item-Instrument\_2009\_UPDATE\_2013.pdf].
25. Graham ID, Calder LA, Hebert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care.* 2000;16:1024–38.
26. 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;17:235–42.
27. Siering U, Eikermann M, Hausner E, Hoffmann-Esser W, Neugebauer EA. Appraisal tools for clinical practice guidelines: a systematic review. *PLoS One.* 2013;8:e82915.
28. Semlitsch T, Jeitler K, Kopp IB, Siebenhofer A. Development of a workable mini checklist to assess guideline quality. *Z Evid Fortbild Qual Gesundheitsw.* 2014;108:299–312.
29. Forschungsnetzwerk Allgemeinmedizin Frankfurt. [http://www.allgemeinmedizin.uni-frankfurt.de/forschung4/forn.html].
30. Semlitsch T, Blank WA, Kopp IB, Siering U, Siebenhofer A. Evaluating guidelines - a review of key quality criteria. *Dtsch Arztebl Int.* 2015;112:471–8.
31. Fleiss JL. Reliability of Measurement. In *The Design and Analysis of Clinical Experiments*. New York: John Wiley & Sons, Inc.; 1999: 1-32.
32. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17:101–10.
33. Aszites, spontan bakterielle Peritonitis und hepatorenales Syndrom [http://www.awmf.org/uploads/tx\_szleitlinien/021-017L\_S3\_Aszites\_2011.pdf].
34. Herzinsuffizienz, DEGAM Leitlinie Nr. 9 [http://www.degam.de/files/Inhalte/Leitlinien-Inhalte/Dokumente/DEGAM-S3-Leitlinien/LL-09\_Langfassung\_Herzinsuffizienz\_1\_E002BIndex.pdf].
35. Assessment and Management of Foot Ulcers for People with Diabetes [http://rmao.ca/sites/rmao-ca/files/Assessment\_and\_Management\_of\_Foot\_Ulcers\_for\_People\_with\_Diabetes\_Second\_Edition1.pdf].
36. Diabetes und Schwangerschaft [http://www.deutsche-diabetes-gesellschaft.de/fileadmin/Redakteur/Leitlinien/Evidenzbasierte\_Leitlinien/EBL\_Schwangerschaft\_2008.pdf].
37. Management der frühen rheumatoiden Arthritis [http://dgrh.de/fileadmin/media/Praxis\_\_\_Klinik/Leitlinien/2011/gesamt\_II\_ra\_2011.pdf].
38. Suspected Cancer in Primary Care - Guidelines for investigation, referral and reducing ethnic disparities [https://www.health.govt.nz/system/files/documents/publications/suspected-cancer-guideline-sep09.pdf].
39. Fieber unklarer Genese [http://www.awmf.org/uploads/tx\_szleitlinien/027-053L\_S1\_Fieber\_unklarer\_Ursache\_2013-01.pdf].
40. Guideline for the Management of Acute Bronchitis [http://www.topalbertadoctors.org/download/378/acute\_bronchitis\_guideline.pdf?\_20150722092250].
41. Guidelines for the practice of diabetes education [http://www.diabeteseducator.org/export/sites/aade/\_resources/pdf/research/Guidelines\_Final\_2\_1\_11.pdf].
42. Nebennierenrinden-Insuffizienz [http://www.awmf.org/uploads/tx\_szleitlinien/027-034\_S1\_Nebennierenrinden-Insuffizienz\_2010\_abgelaufen.pdf].
43. Klazinga N. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project [AGREE]. *Qual Saf Health Care.* 2003;12:18–23.
44. 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:E472–8.
45. 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;182:1045–52.
46. MacDermid JC, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham ID. Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. *BMC Health Serv Res.* 2005;5:18.
47. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ.* 2010;182: E839–842.
48. Watine J, Friedberg B, Nagy E, Onody R, Oosterhuis W, Bunting P, Charet JC, Horvath AR. Conflict between guideline methodologic quality and recommendation validity: a potential problem for practitioners. *Clin Chem.* 2006;52:65–72.
49. Grimmer K, Dizon JM, Milanese S, King E, Beaton K, Thorpe O, Lizarondo L, Luker J, Machotka Z, Kumar S. Efficient clinical evaluation of guideline quality: development and testing of a new tool. *BMC Med Res Methodol.* 2014;14:63.
50. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE, et al. The Global Rating Scale complements the AGREE II in advancing the quality of practice guidelines. *J Clin Epidemiol.* 2012;65:526–34.
51. Coroneos CJ, Voineskos SH, Cornacchi SD, Goldsmith CH, Ignacy TA, Thoma A. Users' guide to the surgical literature: how to evaluate clinical practice guidelines. *Can J Surg.* 2014;57:280–6.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit

