

Appraising the methodological quality of the clinical practice guideline for diabetes mellitus using the AGREE II instrument: a methodological evaluation

Mahmoud Radwan¹, Ali Akbari Sari², Arash Rashidian², Amirhossein Takian¹, Sanaa Abou-Dagga³ and Aymen Elsous¹

¹Department of Health Management and Economics, School of Public Health, International Campus, Tehran University of Medical Sciences, Tehran, Iran

²School of Public Health, International Campus, Tehran University of Medical Sciences, Tehran, Iran

³Department of Research Affairs and Graduates Studies, Islamic University of Gaza, Gaza Strip, Palestine

Corresponding author: Ali Akbari Sari. Email: akbarisari@tums.ac.ir

Summary

Objectives: To evaluate the methodological quality of the Palestinian Clinical Practice Guideline for Diabetes Mellitus using the Translated Arabic Version of the AGREE II.

Design: Methodological evaluation. A cross-cultural adaptation framework was followed to translate and develop a standardised Translated Arabic Version of the AGREE II.

Setting: Palestinian Primary Healthcare Centres.

Participants: Sixteen appraisers independently evaluated the Clinical Practice Guideline for Diabetes Mellitus using the Translated Arabic Version of the AGREE II.

Main outcome measures: Methodological quality of diabetic guideline.

Results: The Translated Arabic Version of the AGREE II showed an acceptable reliability and validity. Internal consistency ranged between 0.67 and 0.88 (Cronbach's α). Intra-class coefficient among appraisers ranged between 0.56 and 0.88. The quality of this guideline is low. Both domains 'Scope and Purpose' and 'Clarity of Presentation' had the highest quality scores (66.7% and 61.5%, respectively), whereas the scores for 'Applicability', 'Stakeholder Involvement', 'Rigour of Development' and 'Editorial Independence' were the lowest (27%, 35%, 36.5%, and 40%, respectively).

Conclusions: The findings suggest that the quality of this Clinical Practice Guideline is disappointingly low. To improve the quality of current and future guidelines, the AGREE II instrument is extremely recommended to be incorporated as a gold standard for developing, evaluating or updating the Palestinian Clinical Practice Guidelines. Future guidelines can be improved by setting specific strategies to overcome implementation barriers with respect to economic considerations, engaging of all relevant end-users and patients, ensuring a rigorous methodology for searching, selecting and synthesising the evidences and recommendations, and addressing potential conflict of interests within the development group.

Keywords

AGREE II, Clinical Practice Guideline, diabetes mellitus, psychometric properties, methodological quality

Introduction

Diabetes mellitus (DM) is a rapidly growing public health problem. It is a metabolic disease characterised by hyperglycaemia resulting from defects in insulin secretion, insulin action or both. 'Long-term complications of diabetes include retinopathy with potential loss of vision; nephropathy leading to renal failure; peripheral neuropathy with risk of foot ulcers, amputations and high incidence of cerebrovascular disease and hypertension'.¹ The World Health Organization estimates that about 347 million people globally have diabetes and projects that diabetes will be the seventh leading cause of death in 2030.² Most diabetes deaths (more than 80%) occur in low- and middle-income countries.² In Palestine, the prevalence rate was 10.5% in the West Bank and 11.8% in the Gaza Strip among the registered Palestinian refugees aged 40 years and older.³ Abu-Rmeileh et al.⁴ estimated the prevalence of DM in Palestine at 20.8% and 23.4% in 2020 and 2030, respectively.⁴ The Palestinian Clinical Practice Guideline (CPG) for DM was developed to promote evidence-based medicine in screening, diagnosis, and treatment and to standardise the care provided for diabetic patients type 1 and 2. The Palestinian CPG was adapted from international guidelines with extensive participation of local experts to ensure the acceptance in, applicability to and consistency with the Palestinian local context.⁵

CPGs are defined as 'systematically developed statements to assist practitioners and patient decisions about appropriate healthcare for specific circumstances'.⁶ It has been shown that CPGs are effective in promoting the quality of healthcare services and improving patients' outcomes.⁷ The large numbers of the published CPGs have raised many inquiries about their quality,⁸ and the method of developing CPGs has been exposed to extensive

criticism since high-quality CPGs should be evidence-based, valid and generalisable to the target population. In order to ensure standardised CPGs with high quality, the recommendations should be valid, transparent and applicable.⁹ Developing CPGs based on the best available evidence is a determinant for their trustworthiness.⁹ It is recognised that randomised clinical trials are the basic blocks for guideline development. Many guidelines developers, however, more likely developed the CPGs depending on expert opinion or based on lower levels of evidences.¹⁰ Such cases can result in biased recommendations.¹⁰

Previous studies showed that the quality of CPGs in different clinical areas was suboptimal.¹¹ Recent studies evaluating the trustworthiness of CPGs have found that guidelines contain incompatible recommendations, show an inconsistency in selecting the evidences, exhibit a lack of reliable and transparent methodology in their development processes and rarely have the declaration of any potential conflicts of interest among members of the guideline developers.¹² Such findings create CPGs with poor validity and restrict their utilisation as evidence-based resources for clinicians.¹²

The Appraisal of Guidelines for Research and Evaluation (AGREE II) assessment tool is a validated questionnaire used to assess the methodological quality of CPGs.¹³ It has been adopted by the World Health Organization for the assessment of CPGs and is recognised now as the new international tool for the assessment of CPGs.¹³ To the best of our knowledge, no study has investigated the quality of the Palestinian CPG for DM based on the AGREE II instrument. Therefore, and with the overall goal of improving the quality of the local and national future CPGs, this study aims to assess the methodological quality of the Palestinian CPG for DM using the AGREE II instrument.

Methods

Structure of AGREE II instrument

The AGREE II is a 23-item tool comprising six quality domains followed by two overall assessment items: Domain 1 'Scope and Purpose' is concerned with the overall aim of the guideline, the specific health questions and the target population (three items); Domain 2 'Stakeholder Involvement' focuses on the extent to which the guideline was developed by the appropriate stakeholders and represents the views of its intended users (three items); Domain 3 'Rigour of Development' relates to the process used to gather and synthesise the evidence, the methods to formulate

the recommendations and to update them (eight items); Domain 4 'Clarity and Presentation' deals with the language, structure and format of the guideline (three items); Domain 5 'Applicability' pertains to the likely barriers and facilitators to implementation, strategies to improve uptake and resource implications of applying the guideline (four items); and Domain 6 'Editorial Independence' is concerned with the formulation of recommendations not being unduly biased with competing interests (two items). A seven-point Likert scale (ranged between score 7 for strongly agree to score 1 for strongly disagree) is used to score each domain item. The overall assessment includes two items: the rating of the overall quality of the guideline (with a scale ranging from 7 for higher possible quality to 1 for lower possible quality) and whether the guideline would be recommended for use in practice, with a three-point scale (3 for not recommended; 2 for recommended with modifications; and 1 for recommended). In line with similar studies, scores of 50% or less were defined as poor quality.¹⁴

Translation process

The AGREE II instrument has been formally translated into several languages (available at: <http://www.agreetrust.org>). The Arabic translation process has been led by the first author (MR) and successfully published at the AGREE Enterprise website in 2015 (available at: <http://www.agreetrust.org/resource-centre/agree-ii-translations/>). The process of AGREE II translation from English to Arabic was mainly performed using the cross-cultural adaptation standardised guideline,¹⁵ which comprises the following steps: forward translation, back translation, review by a committee of experts and pre-testing (pilot study). Initially, the *forward translation* from English into Arabic language was independently done by two professional translators with Arabic mother tongue (one of them was familiar with study objectives as well guidelines development and evaluation). A consensual first draft translation was developed after examining by group of health experts. *Back translation* for the first draft was translated to English by two translators with English mother tongue (they did not participate in the first phase of translation and were neither familiar with AGREE II nor study objectives). Discrepancies between the original English and all translated versions of AGREE II were examined by a *reviewing committee* including the first author (MR), forward translators, one health expert, one researcher and one linguist professional to arrive at a consensual pre-final draft Arabic translation.

In order to verify the applicability of the pre-final Arabic translated AGREE II instrument and test the clarity and understandability of its items, *pilot testing* took place, with a sample of four health professionals (family physicians and staff nurses) being selected to use the pre-final instrument in evaluating the CPG for DM. Then they were asked to rate each item on a Likert-type scale ranging from 1 (not clear or understood) to 10 (completely clear and easily understood).¹⁶ Results showed a mean clarity index of 8.5 (scale 1–10), indicating that the questionnaire was easily understood by the target population.

Appraisal process

Sixteen appraisers were purposively selected (two endocrinologists, eight family physicians, four academic professors ('health experts') and two experienced nurses who work with chronic diseases) in order to assess the methodological quality of the existing CPG for DM. Members of the guideline development group were excluded to reduce any potential bias. All appraisers were familiar with at minimum the principles of guidelines development and research methodology. Sixteen copies of Translated Arabic Version of the AGREE II instrument accompanied with 16 copies of the CPG for DM were provided to the appraisers to independently assess its quality. Based on the manual guide of the Translated Arabic Version of the AGREE II (available at <http://www.agreetrust.org/resource-centre/agree-ii-translations/>), detailed descriptions and explanations on the meaning of the whole AGREE II items were attached with each copy (as advised by AGREE II enterprise) to ensure appraisers' understanding of the items and reduce the inter-rater disagreement.

Data analysis

Parameters of construct and criterion validity were used to assess the validity of the Translated Arabic Version of the AGREE II instrument, whereas the internal consistency and inter-rater agreement parameters were used to assess its reliability. The SPSS program was used for all statistical analyses; *p* values of 0.05 or less were considered significant. Descriptive analysis values were mean, standard deviation and minimum and maximum with 95% confidence intervals (CIs). Correlations analysis was performed using Spearman's rho and Kendall's Tau B coefficients. Domain scores were calculated as recommended by the AGREE Collaboration. Domain scores were calculated by summing up all the scores of the individual items in a domain and by scaling the total as a percentage of the maximum possible score for that

domain. The standardising was performed using the following formula recommended by AGREE II enterprise: (obtained score–minimum possible score)/(maximum possible score–minimum possible score).

Validity analyses

Face validity. The Translated Arabic Version of the AGREE II instrument with its translated user guide was sent to 16 experts (academics and researchers) to share their comments and feedback about clarity, simplicity and easiness to understand of AGREE II items and phrases of the user guide. The reviewers were asked to rephrase any ambiguous item or item which could not be understood.

Construct validity. Construct validity was assessed by evaluating two assumptions. The first one proposed that presenting the key recommendations in summarised tables, flow charts and algorithms might favourably affect the appraisers' perspective towards the quality of this guideline. The second assumption suggested that the explicit description of the guideline external reviewers (e.g. number, qualifications, experiences, affiliations) might affect the appraisers' perspective towards recommending the use of this guideline.

Criterion validity. Criterion validity was assessed in the same manner as reported by the AGREE Collaboration in their validation study.¹⁷ Assessment of criterion validity was assessed by calculating the Kendall's Tau B rank correlation coefficients between the appraisers' domains scores and the overall assessment scores.

Reliability analyses

Two measures of reliability were conducted: (1) Using mean item scores, the Cronbach α coefficient was calculated to measure internal consistency of each domain of the final instrument where scores over 0.70 were considered acceptable, reflecting the internal correlation between items of the same area. (2) ICC with a 95% CI was calculated to assess the reliability within each domain and to measure the inter-rater reliability for each domain. ICC values above 0.75 were considered to represent good, 0.40–0.75 moderate and <0.40 poor reliability.¹⁸

Results

Validity and reliability

Face validity. Items of the Translated Arabic Version of the AGREE II instrument with its Arabic

Table 1. Correlation of domain scores with overall assessment of AGREE II.

Domain	Kendall's Tau-B coefficient	Sig (p)
Domain 1 Scope and Purpose	0.21	0.32
Domain 2 Stakeholder Involvement	-0.18	0.38
Domain 3 Rigour of Development	0.62	0.003
Domain 4 Clarity of Presentation	0.31	0.13
Domain 5 Applicability	0.24	0.23
Domain 6 Editorial Independence	0.26	0.22

Kendall's Tau B correlations were conducted between the mean rating of the overall assessment of the CPG across all appraisers as compared to the mean of each domain score.

translated user guide were evaluated as clear and easy to understand by all health experts. Only very minor changes in the wording of a few items were performed.

Construct validity. We observed a significant association between the overall quality assessment scores and presenting the key recommendations in summarised tables, flow charts and algorithms (Spearman's $\rho = 0.51$, $p < 0.05$). The association between recommending the use of this guideline in practice and the clear description of the guideline external reviewers was significant (Spearman's $\rho = 0.73$, $p < 0.01$).

Criterion validity. The correlation between domain scores and overall quality assessment ranged between low (-0.18) to moderate (0.62). The highest correlation being observed was for the Rigour of Development which was the only domain whose correlation reached the significant level (Kendall's Tau B Coefficient = 0.62, $p = 0.003$; Table 1).

Reliability. The overall internal consistency for domain items was high (Cronbach's $\alpha = 0.87$) and for domains scales ranged between 0.67 and 0.88 (Table 2). The highest consistency was shown for the Rigour of Development (Cronbach's $\alpha = 0.88$). The overall agreement among reviewers was good (ICC = 0.82; 95% CI 0.67–0.92). The ICCs for domain scores ranged between moderate and good (0.56–0.88; Table 3). ICCs showed that appraisers reached higher agreement when assessing the Rigour of Development domain ($r = 0.88$).

Table 2. Internal consistency of AGREE II instrument domain scales.

Domain	Number of items	Cronbach's alpha
Domain 1 Scope and Purpose	3	0.74
Domain 2 Stakeholder Involvement	3	0.72
Domain 3 Rigour of Development	8	0.88
Domain 4 Clarity of Presentation	3	0.67
Domain 5 Applicability	4	0.83
Domain 6 Editorial Independence	2	0.70

Cronbach's alpha was calculated for if item deleted.

Table 3. Inter-rater reliability of AGREE II instrument domain scores.

Domain	Number of items	ICC coefficient	95% Confidence interval
Domain 1 Scope and Purpose	3	0.75	0.42–0.90
Domain 2 Stakeholder Involvement	3	0.56	0.05–0.83
Domain 3 Rigour of Development	8	0.88	0.87–0.95
Domain 4 Clarity of Presentation	3	0.65	0.24–0.86
Domain 5 Applicability	4	0.83	0.59–0.92
Domain 6 Editorial Independence	2	0.63	0.09–0.87

Intra-class correlation (ICC) coefficient calculated using absolute agreement.

Quality assessment

Table 4 shows the AGREE II domain scores of the CPG for DM as evaluated by the 16 appraisers. In general, it seems that the quality of the CPG for DM was questionable when using the AGREE II instrument. The overall average score was 45% and

Table 4. AGREE II domain scores of the CPG for DM.

Domain	Standardised scores (%)	Overall average AGREE II score (%)	Overall quality score (range 1–7) M ± SD	Overall assessment 'Recommendation of Use' (n = 16)
Domain 1 Scope and Purpose	66.7%	45%	4.19 ± 0.83	Recommended with modifications (n = 12) Not recommended (n = 4)
Domain 2 Stakeholder Involvement	35%			
Domain 3 Rigour of Development	36.5%			
Domain 4 Clarity of Presentation	61.5%			
Domain 5 Applicability	27%			
Domain 6 Editorial Independence	40%			

A standardised formula was used as recommended by AGREE Enterprise to assess the quality of guidelines.

domain standardised scores ranged between 27% and 66.7%. The highest domain scores were the Scope and Purpose (66.7%) and clarity of presentation (61.5%), whereas the other four domains (Applicability, Stakeholder Involvement, Rigour of Development and Editorial Independence) were less than 50%. This implies that less than 50% of rigorous criteria were met for developing this guideline. The overall mean score of guideline quality is 4.19. The use of this guideline in practice was recommended with modifications by 12 appraisers and not recommended by 4 appraisers.

Discussion

To the best of our knowledge, this is the first study critically evaluating the methodological quality of the current CPG for DM using the AGREE II instrument. An Arabic version of the AGREE II instrument was developed and tested in a standardised manner. The study findings suggest that the Translated Arabic Version of the AGREE II instrument was reliable and valid when used in assessing the CPG for DM. Obviously, it sounds that the reliability scores of the Translated Arabic Version of the AGREE II have been favourably influenced due to the Arabic translated user guide attached to each item of the instrument. A study aiming to assess the effects of Korean AGREE II score guide, which provides detailed evaluation criteria for each item, showed higher reliability at all guidelines.¹⁹ In this guideline, the internal consistency of each domain was acceptable and similar to findings of AGREE Collaboration.¹⁷ The inter-rater reliability was good with the domains of Rigour of Development ($r=0.88$), Applicability ($r=0.83$) and Scope and Purpose ($r=0.75$), and moderate with the rest of domains.

Improvements in reliability of these domains might be anticipated with further training to guidelines appraisers as training can minimise the opportunities of misinterpretation. The construct validity had two assumptions: in the first one, the findings showed a significant association between the overall quality assessment scores and presenting the key recommendations in summarised tables, flow charts and algorithms. Cautiously, it seems that the presence of summarised tables, flow charts or algorithms was more likely to favourably affect the appraisers' perspective towards the quality of this guideline. In the second assumption, we also observed that the association between recommending the use of this guideline in practice after modification and the clear description of the guideline external reviewers was significant. It can be said that the clear description of the guideline external reviewers (e.g. numbers, qualifications, experiences, affiliations) played a notable role in recommending the use of this guideline. For criterion validity, the correlation between the domain scores and the overall quality assessment was only significant for Rigour of Development (Kendal Tau B = 0.62, $p=0.003$). Interpreting the variation in the results of validity measures appears to be difficult. Even the AGREE Collaboration's studies were not able to demonstrate conclusively the validity of the AGREE instrument since they admitted that validation is a challenging task.¹⁷

This study shows that the methodological quality of the current CPG for DM was disappointingly low when using the AGREE II instrument. The low quality of this guideline could be explained by the fact that the AGREE II instrument was not used in developing or updating the current guideline. The use of the AGREE II instrument already showed evidence of improving the quality of other guidelines,

such as guidelines on the management of low back pain.¹⁴ Another possible explanation is poor reporting even if there were reasonable endeavours to develop it. Assessing the methodological quality of a guideline relies heavily on how well documented the guideline development process is.²⁰

The present study results are widely consistent with previous appraisals evaluating various topics of CPGs.²¹ In previous appraisals, both domains, the 'Scope and purpose' and 'Clarity of presentation' had the highest scores, whereas the scores for 'Applicability', 'Stakeholder Involvement', 'Rigour of Development', and 'Editorial Independence' were the lowest. Although both domains 'Scope and Purpose' and 'Clarity of Presentation' had acceptable scores, this guideline inadequately reported on the domains 'Stakeholder Involvement,' 'Rigour of Development,' 'Applicability' and 'Editorial Independence'. In particular, the low score on the largest domain Rigour of Development (36.5%) is causing a remarkable concern because explicit descriptions of how the available evidence was identified, selected and synthesised are important for the development of valid and reliable evidence-based recommendations. Systematic reviews should form the basis for all high-quality CPGs.²² Apparently, this was not the case with this guideline where searching and selecting the evidence as well as formulating transparent recommendations did not pass through a rigorous path. A more logical explanation for the low score of this crucial domain is the limited experience in synthesising the evidence of randomised clinical trials and systematic reviews. Poor guidelines might be a result of the lack of methodological expertise among guideline developers. The quality may be improved by involving search experts and methodologists in the guideline development process, as well as clarifying the methods of guideline development.²³ It is expected that the current guideline was largely based on experts' opinion rather than rigorous criteria for evidence synthesis. Guidelines developed by expert consensus or using nonsystematic methodology may create biased recommendations and malpractice.²⁴

Another extremely worrisome result is the lowest domain score of the Applicability (27%) since considering the tools and the potential resources for applying the recommendations as well as providing criteria for monitoring the performance can facilitate the implementation of the recommendations. In low-income countries, developing CPGs with high quality is more challenging because of the limited resources and capacity.²⁵ Inclusion of cost and resource information within CPGs can encourage the professionals to transparently select among treatment alternatives

and it is quite important in the light of progress in medical technology and increasing healthcare costs.²⁶ A meta-review of 42 studies in which 626 guidelines on different topics published in various countries from 1980 to 2007 were assessed with the AGREE instrument found that most guidelines achieved low scores for applicability (mean 22%, 95% CI 20.4%–23.9%) compared to all other domains.²¹ Poor applicability is not due to an inherent flaw of the well-validated AGREE II. This is most likely due to the poor defining of implementation strategies for addressing the organisational barriers and cost implications during the guidelines development process.²⁷ The lowest score of 'Applicability' in this appraisal suggests that the guideline developers did not pay sufficient attention to the possible factors affecting the practical implementation of guideline recommendations. Guidelines that fail to address these areas may lead to poor uptake by healthcare professionals and therefore have a limited contribution to improving healthcare quality.²⁸ Another possible explanation is the inadequate pilot testing of this guideline, and if it was piloted at a small scale, it seems that gathering and analysing the feedback from users was at the lowest levels. Feedback of pilot studies can help in formulating intervention strategies to support clinical guidelines.²⁹

A much more expected finding is the low score of Stakeholder Involvement (35%) which could reflect the inadequate consideration of patients' views and preferences during guideline development as well as the unclear description of the target users (e.g. doctors, nurses, pharmacists, etc.). It has been shown that inclusion of patients' views in a guideline development has crucial implications for the success of guidelines implementation.³⁰ It seems that in the Palestinian organisational culture, engagement of patients in the guidelines development process, healthcare planning or the process of strategic planning and considering their views or expectations all remain a challenge. Similarly, in other low- and middle-income countries, engagement of patients or end-users may not be considered in guidelines development that targets different places.³¹ Poor engagement of patient representatives during guideline development could make the patients uncommitted to guideline recommendations. In a disease such as DM, non-adherence to treatments can have significant impact on morbidity and mortality. Considering patient preferences on treatment decisions is associated with better clinical outcomes.³²

One last low score is the Editorial Independence (40%) which focuses on how far the views of a funding body affected the final recommendations and how guideline developers declared the conflict of interest.

Conflicts of interest are the most common source of bias in guideline development, which may result in biased recommendations.³³ The suboptimal score of this domain could be explained in two possible ways: either the guideline development group members were not aware of conflict of interest disclosures, or they preferred not to follow the explicit manner of declaration. It could be that guideline developers did not follow the systematic path of reporting the conflict of interest.³⁴ However, a previous study found that conflicts of interest are widely common among CPGs in a variety of clinical areas.³⁵ Although this guideline was funded by the World Bank, it is still unclear to what extent it has been influenced by the donor's agenda or what measures were taken to minimise the influence of competing interests on guideline development or formulation of the recommendations.

A limitation of the present study is that the appraisers did not receive a specific training course about the AGREE II instrument. However, the user guide was used to overcome such shortcoming. Another limitation was that the AGREE II instrument lacked any criteria necessary to assign a score for the last two overall quality assessments.

Conclusion

The Translated Arabic Version of the AGREE II instrument is a satisfactory, valid and reliable tool to assess the methodological quality of Palestinian CPG for DM. By using the Translated Arabic Version of the AGREE II instrument, this study shows that the quality of this guideline is disappointingly low. To improve the quality of our current and future guidelines, a more systematic and rigorous approach in synthesis and reporting of evidence and recommendations is extremely advised. The AGREE II instrument should be incorporated as a gold standard for developing, evaluating or updating the Palestinian CPGs. Instruments such as AGREE II, GRADE and PRISMA could be quite useful tools and contribute positively in the improvement of guidelines' quality and transparency. Considering the potential resources and designing monitoring measures can strengthen its implementation. A systematic involvement of different healthcare professionals and patients in the development and evaluation of guidelines is necessary to explore their experiences and expectations. Extra attention for updating procedures and managing conflict of interest are exceedingly recommended.

Declarations

Competing Interests: None declared.

Funding: None declared.

Ethics approval: Not required as no individual patient data were used or sought.

Guarantor: AAS.

Contributorship: MR (principal investigator) collected, analysed and interpreted the data and wrote the first draft of the manuscript. MR, AS, AR and AT significantly contributed in the study design and the critical review of the manuscript. SD and AS remarkably contributed in the analysis and interpretation of data and the critical review of the manuscript. Final approval was given by all the authors.

Acknowledgements: The authors would like to thank the 16 appraisers. Special thanks and appreciation are extended to all translators.

Provenance: Not commissioned; peer-reviewed by Alexis Lewis and Mathumalar Loganathan Fahrni.

References

1. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2014; 37: S67–S74.
2. WHO. *Diabetes factsheets*. Geneva: World Health Organization, 2014.
3. UNRWA. *Annual report of the Department of Health*. Amman: United Nations for Relief and Works Agency, 2007.
4. Abu-Rmeileh NME, Hussein A, O'Flaherty M, et al. Forecasting prevalence of type 2 diabetes mellitus in Palestinians to 2030: validation of a predictive model. *The Lancet* 2012; 380: S21.
5. Palestinian MOH. *Palestinian guidelines for diagnosis and management of diabetes mellitus*. Gaza, Palestine, 2004.
6. IOM. Committee to advise the Public Health Service on clinical practice, guidelines. In: Field MJ and Lohr KN (eds) *Clinical practice guidelines: directions for a new program*. Washington, DC: National Academies Press, 1990.
7. Vecchio AL, Giannattasio A, Duggan C, et al. Evaluation of the quality of guidelines for acute gastroenteritis in children with the AGREE instrument. *J Pediatr Gastroenterol Nutr* 2011; 52: 183–189.
8. Kung J, Miller RR and 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: 1628–1633.
9. Sacks DB, Arnold M, Bakris GL, et al. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Diabetes Care* 2011; 34: e61–e99.
10. Tricoci P, Allen JM, Kramer JM, et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009; 301: 831–841.
11. Zhang ZW, Liu XW, Xu BC, et al. Analysis of quality of clinical practice guidelines for otorhinolaryngology in China. *PloS ONE* 2013; 8: e53566.

12. Ransohoff DF and Sox HC. Guidelines for guidelines: measuring trustworthiness. *J Clin Oncol* 2013; 31: 2530–2531.
13. Brouwers MC, Kho ME, Browman GP, et al. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ* 2010; 182: E472–E478.
14. Bouwmeester W, van Enst A and van Tulder M. Quality of low back pain guidelines improved. *Spine* 2009; 34: 2562–2567.
15. Guillemin F, Bombardier C and Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46: 1417–1432.
16. Pasquali L. *Psychometrics: theory testing in psychology and education*. Rio de Janeiro: Vozes, 2003.
17. AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003; 12: 18–23.
18. Fleiss JL. *Reliability of measurement. The design and analysis of clinical experiments*. Toronto: John Wiley and Sons, 1986, pp.1–32.
19. Oh MK and Jo H. Improving the reliability of clinical practice guideline appraisals: effects of the Korean AGREE II scoring guide. *J Korean Med Sci* 2014; 29: 771–775.
20. Hayward RS, Wilson MC, Tunis SR, et al. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. *JAMA* 1995; 274: 570–574.
21. Alonso-Coello P, Irfan A, Sola 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: e58.
22. IOM. *Clinical practice guidelines we can trust*. Washington, DC: The National Academies Press, 2011.
23. Knai C, Brusamento S, Legido-Quigley H, et al. Systematic review of the methodological quality of clinical guideline development for the management of chronic disease in Europe. *Health Policy (Amsterdam, Netherlands)* 2012; 107: 157–167.
24. Antman EM, Lau J, Kupelnick B, et al. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992; 268: 240–248.
25. Rashidian A. Adapting valid clinical guidelines for use in primary care in low and middle income countries. *Primary Care Resp J* 2008; 17: 136–137.
26. Vale L, Thomas R, MacLennan G, et al. Systematic review of economic evaluations and cost analyses of guideline implementation strategies. *Eur J Health Econ* 2007; 8: 111–121.
27. Francke AL, Smit MC, de Veer AJE, et al. Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. *BMC Med Inform Decision Making* 2008; 8: 38.
28. Donnellan C, Sweetman S and Shelley E. Implementing clinical guidelines in stroke: a qualitative study of perceived facilitators and barriers. *Health Policy (Amsterdam, Netherlands)* 2013; 111: 234–244.
29. Gifford WA, Graham ID and Davies BL. Multi-level barriers analysis to promote guideline based nursing care: a leadership strategy from home health care. *J Nurs Manag* 2013; 21: 762–770.
30. Poitras S, Rossignol M, Avouac J, et al. Management recommendations for knee osteoarthritis: how usable are they? *Joint, Bone, Spine* 2010; 77: 458–465.
31. Polus S, Lerberg P, Vogel J, et al. Appraisal of WHO guidelines in maternal health using the AGREE II assessment tool. *PloS ONE* 2012; 7: e38891.
32. Umscheid CA. Should guidelines incorporate evidence on patient preferences? *J Gen Intern Med* 2009; 24: 988–990.
33. Norris SL, Burda BU, Holmer HK, et al. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. *J Clin Epidemiol* 2012; 65: 725–733.
34. Hu J, Chen R, Wu S, et al. The quality of clinical practice guidelines in China: a systematic assessment. *J Eval Clin Pract* 2013; 19: 961–967.
35. Norris SL, Holmer HK, Ogden LA, et al. Conflict of interest disclosures for clinical practice guidelines in the national guideline clearinghouse. *PloS ONE* 2012; 7: e47343.