Health Policy 123 (2019) 646–651

Contents lists available at ScienceDirect

Health Policy

journal homepage: www.elsevier.com/locate/healthpol

Assessment of health systems guidance using the Appraisal of Guidelines for Research and Evaluation – Health Systems (AGREE-HS) instrument

Melissa C. Brouwers[a,*], John N. Lavis[b], Karen Spithoff[b], Marija Vukmirovic[b], Ivan D. Florez[b,c], Marcela Velez[b,c], Mohammad Kibria[b], Nigar Sekercioglu[b], Elizabeth Kamler[b], Jillian Halladay[b], Jaspreet Sandhu[b], Ahmednur Ali[b], Anushka Jaffer[b], Ruhi Kiflen[b,d], Julia Pemberton[a]

[a] University of Ottawa, 75 Laurier Ave East, Ottawa, K1N 6N5, Ontario, Canada
[b] McMaster University, 1280 Main Street West, Hamilton, L8S 4L8, Ontario, Canada
[c] Universidad de Antioquia, Calle 67 # 53-108, Medellin, 050001, Colombia
[d] St. Michael's Hospital, 30 Bond Street, Toronto, M5B 1W8, Ontario, Canada

ARTICLE INFO

Article history:
Received 18 April 2018
Received in revised form 9 April 2019
Accepted 6 May 2019

Keywords:
Health systems
Health systems guidance
Health systems research
Health policy
Quality appraisal

ABSTRACT

Health systems guidance (HSG) documents contain systematically developed statements or recommendations intended to address a health system challenge. The concept of HSG is fairly new and considerable effort has been undertaken to build tools to support the contextualization of recommendations. One example is the Appraisal of Guidelines for REsearch and Evaluation - Health Systems (AGREE-HS), created by international stakeholders and researchers, to assist in the development, reporting and evaluation of HSG. Here, we present the quality appraisal of 85 HSG documents published from 2012 to 2017 using the AGREE-HS. The AGREE-HS consists of five items (Topic, Participants, Methods, Recommendations, and Implementability), which are scored on a 7-point response scale (1=lowest quality; 7=highest quality). Overall, AGREE-HS item scores were highest for the 'Topic' and 'Recommendations' items (means above the mid-point of 4), while the 'Participants', 'Methods', and 'Implementability' items received lower scores. Documents without a specific health focus and those authored by the National Institute for Health and Care Excellence group, achieved higher AGREE-HS overall scores than their comparators. No statistically significant changes in overall scores were observed over time. This is the first time that the AGREE-HS has been applied, providing a current quality status report of HSG and identifying where improvements in HSG development and reporting can be made.

1. Introduction

Health systems guidance (HSG) is becoming an increasingly important tool to support and strengthen health systems [1–3]. Informed by research evidence and jurisdictionally-relevant data, HSG provides systematically developed recommendations to manage challenges related to health system governance, financial and delivery arrangements, and the implementation strategies needed to get appropriate programs and services to those who need them. HSG aims to assist decision-making, particularly decisions made by health policy managers, health care managers, and professional leaders [1–3], whereas clinical practice guidelines (CPGs) provide recommendations on the most effective and safe clinical or public health actions. Examples of CPG topics include:

- Treatments for the management of breast cancer
- HPV vaccination to reduce incidence of cervical and oral cancers

In contrast, HSG provides recommendations on how systems, policies and finances must be organized to enable the implementation of these actions. Examples of HSG topics include:

* Corresponding author at: University of Ottawa, 75 Laurier Ave East, Room: 600 Peter Morand (Alta Vista), Room, Ottawa, K1N 6N5, Ontario, Canada.
E-mail addresses: melissa.brouwers@uottawa.ca (M.C. Brouwers), lavisj@mcmaster.ca (J.N. Lavis), spithokd@mcmaster.ca (K. Spithoff), vukmirov@mcmaster.ca (M. Vukmirovic), ivan.florez@udea.edu.co (I.D. Florez), velezcm@mcmaster.ca (M. Velez), kibrim1@mcmaster.ca (M. Kibria), sekercn@mcmaster.ca (N. Sekercioglu), kamlerek@mcmaster.ca (E. Kamler), halladje@mcmaster.ca (J. Halladay), sandhu_15@hotmail.com (J. Sandhu), alia76@mcmaster.ca (A. Ali), jaffea6@mcmaster.ca (A. Jaffer), kiflenr@mcmaster.ca (R. Kiflen), pemberj@mcmaster.ca (J. Pemberton).

https://doi.org/10.1016/j.healthpol.2019.05.004

- Guidance to protect health from climate change through health adaptation planning
- Guidance on country pharmaceutical pricing policies

HSG is a relatively new guidance tool compared to CPGs and other evidence-informed tools, such as health technology assessments (HTAs) [1–3]. For example, in the CPG field, there are tools to support the evaluation of CPG development processes (e.g., the original AGREE Instrument [4], AGREE II [5], IOM Standards [6,7]) and CPG recommendations (e.g. AGREE-REX [8]). There are tools to support reporting of CPG documents (e.g., AGREE Reporting Checklist [9], RIGHT [10], Check-UP [11]) and to provide specific methodological development direction on achieving quality goals (e.g., GRADE [12], Evidence to Decision Framework [13]).

Nonetheless, considerable effort has been undertaken in the last several years to build a consensus on the defining features of HSG, HSG development strategies, tools to support decisions by health systems and policy leaders, and methods to enable the contextualization of recommendations [1–3,14–17].

Similarly to the CPG and HTA fields, variation in the quality of HSG and the unintended and potentially harmful effects of implementing poor quality HSG (e.g. adoption of contextually inappropriate or ineffective system arrangements), are concerns. In response, an international team of researchers and stakeholders in the HSG field created the **A**ppraisal of **G**uidelines for **RE**search and **E**valuation - **H**ealth **S**ystems (AGREE-HS). The AGREE-HS is a newly released tool to support the development, reporting and evaluation of HSG; studies completed to date have indicated that it is usable, reliable, and valid [8,14,15,18].

Improving the quality of HSG documents and their adoption in decision-making can serve as a mechanism to optimize and strengthen health systems. We conducted a quality appraisal using the AGREE-HS to identify strengths and weaknesses of current HSG and to serve as a baseline upon which to measure future changes in HSG quality. In this article we present the results of that quality assessment.

2. Materials and methods

2.1. HSG search and selection

2.1.1. Search

The research team systematically searched for HSG documents from the following websites: World Health Organization (WHO; March 2016 and 2017) [19], National Institute for Health and Care Excellence (NICE; March 2017) [20], and Health Systems Evidence (McMaster Health Forum; February 2017) [21]. This was augmented by HSG collected from a member of the team for another project (i.e. personal library). All documents that met the definition of HSG were included regardless of their length or anticipated quality.

2.1.2. Eligibility criteria

HSG documents were included if the following criteria were met:

- provided specific advice or recommendations regarding a health systems challenge; governance, financial or delivery arrangements; and/or implementation strategies
- considered scientific research in the development of advice/recommendations
- considered local, national or regional data and/or contextual factors in the development of advice/recommendations
- targeted to stakeholders, health policy makers, health care managers, and professional leaders
- authored by an international agency, government body, or non-governmental organization (as opposed to independent investigators)
- published from 2012 to 2017
- English-language

2.1.3. Selection

Three members of the study team were involved in the selection and full review of the candidate documents, to identify those that met eligibility criteria. Disagreements were resolved by consensus.

2.2. Appraisers

A purposive sample of appraisers was sought to reflect the limited experience and exposure to HSG that we anticipated the expected users of the tool (e.g., junior staff and researchers) to have. To this end, 10 graduate students and three members of the study team from McMaster University (Canada) were recruited to participate in the study. To optimize variation in perspective, trainees were recruited from three graduate programs (Health Policy, Health Research Methodology, and Public Health), a range of countries of origin, and a range of past training (social sciences, health professional, other health sciences). An in-person orientation was provided to appraisers. Orientation included an overview of HSG and the AGREE-HS tool, as well as a practical group exercise (i.e. AGREE-HS appraisal of a sample HSG document). After the orientation, the HSG documents were randomly assigned to the 13 appraisers for assessment using the AGREE-HS tool. Appraisers received individualized emails with a downloadable copy of the AGREE-HS tool, a link to an online folder with their assigned HSG documents, and a link to an online survey to record their scores. The evaluations were completed in spring 2017. Two appraisers assessed each HSG document.

2.3. AGREE-HS tool

The final draft of the AGREE-HS tool was used to assess each HSG. The draft comprised five core quality items: Topic; Participants; Methods; Recommendations; and Implementability; and included an overall description and defining criteria for each item (Table 1). Each item was scored using a 7-point response scale (1 = lowest quality to 7 = highest quality), with higher scores indicating that more of the criteria were met. Appraisers also completed two general assessment statements:

- Rate the overall quality of this health systems guidance (1=lowest quality to 7=highest quality)
- I would recommend this health systems guidance for use in the appropriate context (yes, yes with modifications, no)

Each HSG document was assessed by two raters. The AGREE-HS overall score and mean AGREE-HS item scores were calculated for each document. The overall score was calculated by summing up all the scores of the individual items and by scaling the total as a percentage of the maximum possible score (with consideration given to the number of raters). The maximum AGREE-HS overall score was 100%. A mean AGREE-HS overall score (across all 85 HSG documents) was calculated.

The five AGREE-HS items are weighted equally given there are no empirical data to direct how items should be weighted; that is, there are no data to demonstrate that performance on certain items better predict, for example, the acceptance of HSG recommendations or their adoption.

Thus, the individual mean AGREE-HS item scores and the general assessment quality score were calculated by first averaging individual item ratings for each HSG between the two appraisers,

Table 1
AGREE-HS item information in the final draft[1].

Item name and description	List of criteria
Topic: This item addresses the description of the health system challenge, the causes of the challenge and the priority accorded to it, and relevance of the guidance.	The health system challenge is clearly described (i.e., the nature of the challenge; the magnitude, frequency or intensity of the challenge; the populations affected). The causes of the health system challenge are clearly described. The health system challenge is described in terms of its level of priority in the targeted health system and the affected population; arguments to support the priority classification should be provided. The guidance is relevant (i.e., appropriate in nature and timely in relation to when decisions will be made) to, and appropriate for, the health system challenge, system or sub-system needs, the target population(s), and the setting in which they will operate.
Participants: This item addresses the composition of the health systems guidance development team and management of competing interests and funder influence.	The health systems guidance development team includes members who have an interest or stake in the recommendations (e.g., decision makers/program managers, operational leaders, consumers and members of the public, other stakeholders). The health systems guidance development team is multidisciplinary (e.g., political scientists, economists, epidemiologists, methodologists). The health systems guidance development team is multi-sectoral (e.g., primary care, public health and, if appropriate to the challenge, finance and housing). Competing interests of the health systems guidance development team members (e.g., financial, professional) and the strategies used to identify and manage them are clearly described. Precautions have been taken to avoid or minimize the influence of a funding agency.
Methods: This item addresses the use of systematic methods and transparency in their reporting; the use of the best available and up-to-date evidence; consideration of effectiveness and cost-effectiveness of the potential options; and weighting of benefits and harms in the guidance document.	Systematic and transparent methods were used to identify and review the evidence (e.g., integrated review, scoping review, review of the grey literature, systematic review, etc.). The best available and most contextually relevant evidence was considered. The evidence base is current. Evidence of effectiveness of the potential options is clearly described, including descriptions of the contexts in which the options were tested. Evidence of cost and cost-effectiveness of the potential options is described. The weighting of the benefits and harms of the potential options is described. There is a link between the recommendations and evidence. The rationale behind the recommendations is clear. Systematic and transparent methods were used to agree upon the final recommendations (e.g., informal or formal consensus, Delphi method, nominal group methods).
Recommendations: This item addresses the outcomes orientation and comprehensiveness of the guidance; ethical and equity considerations drawn upon in its development; details for its operationalization; the sociocultural and political alignment of the guidance; and the updating plan.	The anticipated outcomes of implementing the recommendations are clearly described (including indicators, performance thresholds or targets, and standards to measure them). The recommendations are comprehensive and provide direction to all relevant health system levels (e.g., national, provincial/state), subsystems (e.g., cancer, mental health) and sectors (e.g., primary care, public health). The ethical principles used to develop the recommendations are described. The recommendations promote equity among the target population (e.g., in terms of age, sex, gender, culture, religion, race, sexual orientation). The recommendations' acceptability to and alignment with sociocultural and political interests were considered. The recommendations are easily identifiable, clear, and succinct. The recommendations are actionable and are sufficiently detailed to be operationalized. A plan for updating the recommendations is described.
Implementability: This item addresses the barriers and enablers to implementing the recommendations; the cost and resource considerations in implementing the recommendations; the affordability of implementation and anticipated sustainability of outcomes; the flexibility and transferability of the guidance; and the strategies for disseminating the guidance, monitoring its implementation and evaluating its impact.	Barriers and enablers to the implementation of the recommendations are described, including factors that are internal (e.g., resources, incentives, administrative structure) and external (e.g., legal system, social system, state of the economy, corruption, beliefs) to the health system. A plan to mitigate barriers and optimize enablers is included. Cost and resource considerations for the recommended actions are described (e.g., money, time, infrastructure, equipment, administrative capacity, supplies, staffing, and training). The stakeholders' acceptability of the recommendations is described. The affordability of the recommendations, in the context where implementation will take place, is described. The anticipated sustainability and requirements to maintain long-term outcomes is described. The recommendations are flexible and there is a description of how they can be adapted or tailored for specific contexts in which they will be implemented. A description of the degree to which the recommendations are transferable to other similar or different contexts is provided. Strategies for disseminating the health systems guidance are described. Strategies for assessing the implementation process and the impact of the recommendations are described.

[1] Very minor editorial modifications were made to these item descriptions and criteria before the AGREE-HS tool was published online [8].

Table 2
HSG search.

Database	Web Address	Date	Eligible HSG Full Text Review	Final Included HSG
WHO website	http://www.who.int/entity/en/	Mar 2016 (updated Mar 2017)	90	47
Health Systems Evidence database	https://www.nice.org.uk/guidance	Feb 2017	42	18
NICE website	https://www.healthsystemsevidence.org/.	Mar 2017	50	15
List From Previous Study	Not applicable	2016	18	5
Totals		–	200	85

Table 4
AGREE –HS overall score by HSG characteristics (n = 85 HSG).

HSG Characteristics		n	Mean	Standard Deviation	One-Way ANOVA
Authoring Group*	NICE	15	63.3	19.6	p = 0.002
	WHO	50	51.1	16.8	
	All Other	20	42.1	13.6	
Jurisdiction	Provincial / Subnational	12	47.5	13.3	p = 0.68
	National	23	53.1	21.9	
	Global	50	51.1	16.8	
Health Focus*	Disease Specific	54	47.9	16.5	p = 0.03
	Not Disease Specific	31	56.8	18.7	
Year of Publication	2012	23	46.3	17.6	p = 0.41
	2013	16	54.3	13.9	
	2014	17	49.4	20.1	
	2015	21	56.1	18.5	
	2016	8	49.8	18.1	

* Significant differences between groups were found.

and second, by averaging each AGREE-HS item across the 85 HSGs. The mean AGREE-HS item scores and the general assessment quality score ranged between one and seven. Frequencies for the general assessment question regarding recommendation of the HSG for use were calculated.

2.4. Analyses

A repeated measures analysis of variance (ANOVA) test was conducted, with AGREE-HS items serving as the repeated measure, to examine if there was a significant difference between the mean AGREE-HS item scores. Post-hoc tests were subsequently conducted to identify where the differences occurred between items; the Sidak correction was used to account for multiple comparisons [22].

For exploratory purposes only, a series of one-way ANOVA tests were used to examine mean differences in the AGREE-HS overall scores as a function of the following HSG characteristics: HSG authoring group, the type of jurisdiction in which the HSG was developed (global, national, or subnational/provincial), HSG health focus (specific disease/medical condition vs. not), and HSG publication year. Where significant overall differences between groups were found, post-hoc tests were conducted using the Tukey's correction to account for multiple comparisons [22].

Measurement properties of the tool have been fully reported in a separate publication [23]. In brief, internal consistency of the tool used was 0.79, with the Cronbach's alpha increasing to r = 0.85 with the deletion of the topic item. Intraclass correlations with two raters were 0.15, 0.82, 0.73, 0.63 and 0.48 for the Topic, Participants, Methods, Recommendations and Implementability items, respectively.

3. Results

3.1. HSG

The search yielded 85 eligible HSG documents for appraisal (see Table 2 and the supplementary material). WHO authored 50 documents (47 from the WHO website search and three from a previous study), NICE authored 15, and other guidance development groups authored the remaining 20. The majority were global in scope (n = 50) and disease-specific (n = 54).

Table 3
Mean AGREE-HS item scores.

AGREE-HS Item	Mean	Standard Deviation
Topic	5.2	1.0
Participants	3.5	1.9
Methods	3.3	1.7
Recommendations	4.5	1.2
Implementability	3.9	1.3

3.2. Mean AGREE-HS item scores

The repeated measures ANOVA, conducted to evaluate differences in mean item scores, yielded a significant difference ($p < 0.001$). Two of the five items had mean scores above 4.0, the mid-point of the scale: Topic (m = 5.2), and Recommendations (m = 4.5) (Table 3). Post-hoc analyses revealed significant differences between the mean item scores with the exception of Participants vs. Methods items and Participants vs. Implementation items, where no significant differences were found.

3.3. AGREE-HS overall scores

A series of one-way ANOVA tests were conducted to determine if there were differences between AGREE-HS overall scores as a function of HSG authoring group, jurisdiction, health focus, and year of publication. Descriptive statistics are reported in Table 4.

A statistically significant difference in overall scores between HSG authoring groups was found ($p = 0.002$). Post-hoc analyses revealed that HSG documents authored by NICE scored significantly higher than those authored by WHO or other organizations. In addition, a significant difference in overall scores based on HSG health focus was found ($p = 0.03$). HSG without a specific health focus (i.e. not targeted to a specific disease or medical condition) scored higher than HSG with a specific health focus. No significant differences were found for HSG jurisdiction ($p = 0.68$) or year of publication ($p = 0.41$).

3.4. General assessment scores

The mean general assessment quality score averaged across all the HSG documents was 4.0 (SD = 1.4). Based on the AGREE-HS assessments, only 13 HSG documents (15%) were recommended for use without modifications by both appraisers. The number (%) of HSG documents that were recommended, recommended with modifications, or not recommended, by both appraisers, was 13 (15%), 25 (29%), and 5 (6%), respectively; thus 50% had an exact match. Twenty-two (26%) HSG documents were recommended by one appraiser and recommended with modifications by the second; 18 (21%) were recommended with modifications by one appraiser and not recommended by the second; and 2 (2%) were recommended by one appraiser and not recommended by the second.

4. Discussion

In this study we evaluated 85 HSG documents with the final draft of the AGREE-HS tool. AGREE-HS item scores were highest for the Topic and Recommendations items (means above the mid-point of 4), with lower scores observed for the Participants, Methods, and Implementability items. HSG documents providing recommendations for system issues without a specific health focus and those authored by the NICE group achieved higher AGREE-HS overall scores than their comparators. No statistically significant changes in AGREE-HS overall scores were seen over time.

The averaged AGREE-HS item scores for all 85 HSG documents ranged from 3.3–5.2 on the 7-point response scale. The ranges in quality scores seen with the first application of the AGREE-HS tool appear to be similar to the ranges in quality scores seen with the first application of other AGREE-related tools used to evaluate CPGs. For example, the first application of the original AGREE instrument, and subsequently the AGREE II, resulted in mean domain scores ranging from 31% (applicability) to 66% (scope and purpose) and 45% (applicability) to 74% (scope and purpose), respectively [4,24]. Key areas for improvement in HSG development include transparency about the composition of the guidance development team, conflicts of interest, and influence of the funding body; use of systematic and transparent methods to review the evidence on which the guidance is based; and provision of sufficient information for users to be able to implement and evaluate the recommended actions in their own context.

Although the final draft of the AGREE-HS was applied by HSG appraisers, we are confident of the quality ratings. Differences in the item descriptions and criteria of the draft tool and the final version of the tool are minor and editorial in nature. The overall assessment question “Rate the overall quality of this health systems guidance” statement was removed in the final version, and another statement (i.e. I would recommend this health systems guidance for use in my context) was added, as optional to complete. The final version of the tool is available online [8].

As has been reported with the evaluation of AGREE II and CPGs, lower AGREE-HS scores could be a reflection of poor quality or inadequate reporting [24]. Specifically, HSG developers may have completed additional quality steps that were not described in the guidance document and thus were not considered in the evaluation. The absence of information leaves the HSG user guessing; comprehensive reporting, including the rationale for why things might not have been addressed, is warranted in those cases. Alternatively, it may be that certain quality criteria fall outside of the particular mandate of the HSG developers or are the responsibility of other units or groups with which the HSG team is affiliated (e.g. assessment of costs of implementing the recommendations may be the responsibility of another unit). For example, in Cancer Care Ontario guidelines, specific costing issues associated with the implementation of recommendations are considered, purposefully, by teams unique from the developers of recommendations. In addition, WHO guidance documents have resources and tools such as implementation guides and checklists that support application of recommendations, but that are distinct from the main HSG document. This disconnect between what was done during the HSG development process and what was reported in the HSG document may lead to tension in the evaluation process; however, transparency and explicitness in the methods used, including deviations from the preferred methods, and the rationale for these deviations, is required for HSG users to have confidence in the process. As a strategy to inform development and reporting, the AGREE-HS tool also provides direction about information to include in HSG documents, including any deviations from the norm and instructions on how to find additional information or complementary resources that are aligned with quality of recommendations. Based on these reasons, we believe the AGREE-HS has an important role for HSG development, reporting and evaluation, and not solely the latter.

The AGREE-HS overall score was highest for HSG authored by NICE in comparison to those authored by WHO or other groups. We hypothesize two different issues may be at play. As it relates to the NICE versus WHO difference, it is important to recognize that WHO HSG is global in scope with the expectation that recommendations will be contextualized and operationalized at the local level. Thus, some AGREE-HS criteria may not be relevant for inclusion until this second step. In contrast, NICE HSG is relevant to a single country. As it relates to NICE versus “all other” HSG, it is hypothesized that groups developing fewer HSG documents and less frequently will have less methodological expertise. It is not known why HSG that is not specific to a particular disease or medical condition (vs. HSG with a health focus) scored higher in the AGREE-HS assessment, although authoring group may be a confounding factor. Regardless, the study provides a snap shot of the current strengths and limitations of the HSG enterprise and can provide direction for where improvements can be made. HSG authoring groups, specifically, and the HSG field in general, can use these data and future data as a means to monitor change over time, to set quality score bench marks or goals, to inspire research projects designed to refine existing methods, and to generate new methods in HSG development that balance the goals of rigour, usability, and feasibility. These data can also be used by HSG developers to decide how the AGREE-HS tool can be used and tailored appropriately to their needs.

Key limitations of this study are linked to the methodological decisions made to ensure the study was feasible and completed in-budget. Specifically, only English-language HSG documents were included in the analysis. As a result, we have no data on the unique strengths or limitations of non-English HSG.

5. Conclusion

A five-item tool, the AGREE- HS, was created by international researchers and stakeholders to assist in the development, reporting and evaluation of HSG documents. The purpose of this study was to report on the appraisal of 85 publicly-available HSG documents using the AGREE-HS tool. Findings showed that documents without a particular health focus and those authored by NICE scored significantly higher than their comparators. Further, the mean general assessment score for all documents was 4.0 (the mid-point on a 7-point Likert scale), suggesting that there is room for improvement in overall quality of these documents. The results of this appraisal provide a status report of available HSG literature, presenting a baseline measurement for future large-scale appraisals. Ultimately, our research aims to positively impact the quality of HSG that is produced and implemented across the globe.

Conflict of interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgements

Funding for the AGREE-HS study was provided by the Canadian Institutes of Health Research (grant 201309MOP-311155-KTR-CEBA-40598). The funding body did not influence the design of the study, the collection, analysis, and interpretation of the data, or the writing of the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.healthpol.2019.05.004.

References

[1] Bosch-Capblanch X, Lavis JN, Lewin S, Atun R, Rottingen J, Droschel D, et al. Guidance for evidence-informed policies about health systems: rationale for and challenges of guidance development. PLoS Med 2012;9(3):e1001185.
[2] Lavis JN, Rottingen J, Bosch-Capblanch X, Atun R, El-Jardali F, Gilson L, et al. Guidance for evidence-informed policies about health systems: linking guidance development to policy development. PLoS Med 2012;9(3):e1001186.
[3] Lewin S, Bosch-Capblanch X, Oliver S, Akl EA, Vist GE, Lavis JN, et al. Guidance for evidence-informed policies about health systems: assessing how much confidence to place in the research evidence. PLoS Med 2012;9(3):e1001187.
[4] The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003;12:18–23.
[5] Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in healthcare. CMAJ 2010;182:E839–842.
[6] 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.
[7] Institute of Medicine. Clinical practice guidelines we can trust. Washington: The National Academies Press; 2011.
[8] AGREE Enterprise. AGREE: advancing the science of practice guidelines; 2009. Available from [Accessed 18 September 2017] http://www.agreetrust.org/.
[9] Brouwers MC, Kerkvliet K, Spithoff K. The AGREE reporting checklist: a tool to improve reporting of clinical practice guidelines. BMJ 2016;352:i1152.
[10] Chen Y, Yang K, Marusic A, Qaseem A, Meerpohl JJ, Flottorp S, et al. A reporting tool for practice guidelines in health care: the RIGHT statement. Ann Intern Med 2017;166(2):128–32.
[11] Vernooij RWM, Alonso-Coello P, Brouwers M, Garcia LM. Reporting items for updated clinical guidelines: checklist for the reporting of updated guidelines (CheckUp). PLoS Med 2017;14(1):e1002207.
[12] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924.
[13] Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. BMJ 2016;353:i2089.
[14] Akl EA, Fadlallah R, Ghandour L, Kdouh O, Langlois E, Lavis JN, et al. The SPARK Tool to prioritise questions for systematic reviews in health policy and systems research: development and initial validation. Health Res Policy Syst 2017;15:77.
[15] Ako-Arrey DE, Brouwers MC, Lavis JN. Giacomini MK on behalf of the AGREE-HS Team. Health system guidance appraisal – a critical interpretive synthesis. Implement Sci 2016;11:9.
[16] Ako-Arrey DE, Brouwers MC, Lavis JN. Giacomini MK on behalf of the AGREE-HS Team. Health system guidance appraisal – concept evaluation and usability testing. Implement Sci 2016;11:3.
[17] Moat KA, Lavis JN, Abelson J. How contexts and issues influence the use of policy-relevant research syntheses: a critical interpretive synthesis. Milbank Q 2013;91(3):604–48.
[18] Brouwers MC, Ako-Arrey D, Spithoff K, Vukmirovic M, Florez ID, Lavis JN, et al. Validity and usability testing of a health systems guidance appraisal tool, the AGREE-HS. Health Res Policy Syst 2018;16:51.
[19] Programmes and projects. World Health Organization; 2016. Available from [Accessed March 2016 and March 2017] http://www.who.int/entity/en/.
[20] NICE Guidance. National Institute for Health and Care Excellence; 2017. Available from [Accessed February 2017] https://www.nice.org.uk/guidance.
[21] McMaster Health Forum (McMaster University). Health Systems Evidence; 2017. Available from [Accessed February 2017] http://www.healthsystemsevidence.org.
[22] Kirk RE, In: Chapter 3 Multiple Comparison Tests. In: Experimental design: procedures for the behavioral sciences, 2nd ed. California, USA: Brooks/Cole Publishing Company; 1982.
[23] Brouwers MC, Ako-Arrey D, Spithoff K, Vukmirovic M, Florez ID, Lavis JN, et al. Validity and usability testing of a health systems guidance appraisal tool, the AGREE-HS. Health Res Policy Syst 2018;16:51.
[24] Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. CMAJ 2010;182:1045–52.