RESEARCH

Advancing guideline quality through country-wide and regional quality assessment of CPGs using AGREE: a scoping review

Marli Mc Allister[1*], Ivan D. Florez[2,3,4], Suzaan Stoker[1] and Michael McCaul[1]

Abstract

Background and objective Clinical practice guidelines (CPGs) are evaluated for quality with the Appraisal of Guidelines for Research and Evaluation (AGREE) tool, and this is increasingly done for different countries and regional groupings. This scoping review aimed to describe, map, and compare these geographical synthesis studies, that assessed CPG quality using the AGREE tool. This allowed a global interpretation of the current landscape of these country-wide or regional synthesis studies, and a closer look at its methodology and results.

Study design and methods A scoping review was conducted searching databases Medline, Embase, Epistemonikos, and grey literature on 5 October 2021 for synthesis studies using the later versions of AGREE (AGREE II, AGREE-REX and AGREE GRS) to evaluate country-wide or regional CPG quality. Country-wide or regional synthesis studies were the units of analysis, and simple descriptive statistics was used to conduct the analysis. AGREE scores were analysed across subgroups into one of the seven Sustainable Development Goal regions, to allow for meaningful interpretation.

Results Fifty-seven studies fulfilled our eligibility criteria, which had included a total of 2918 CPGs. Regions of the Global North, and Eastern and South-Eastern Asia were most represented. Studies were consistent in reporting and presenting their AGREE domain and overall results, but only 18% (*n* = 10) reported development methods, and 19% (*n* = 11) reported use of Grading of Recommendations Assessment, Development, and Evaluation (GRADE). Overall scores for domains *Rigor of development* and *Editorial independence* were low, notably in middle-income countries. *Editorial Independence* scores, especially, were low across all regions with a maximum domain score of 46%. There were no studies from low-income countries.

Conclusion There is an increasing tendency to appraise country-wide and regionally grouped CPGs, using quality appraisal tools. The AGREE tool, evaluated in this scoping review, was used well and consistently across studies. Findings of low report rates of development of CPGs and of use of GRADE is concerning, as is low domain scores globally for *Editorial Independence*. Transparent reporting of funding and competing interests, as well as highlighting evidence-to-decision processes, should assist in further improving CPG quality as clinicians are in dire need of high-quality guidelines.

*Correspondence:
Marli Mc Allister
marlibogaards@hotmail.com
Full list of author information is available at the end of the article

Introduction

Clinical practice guidelines (CPGs) play an integral part of medical practice, policy, and politics; and wherever possible should be informed by systematic reviews of the best available evidence, while considering benefits and harms. Clinical practice guideline development groups, researchers and clinicians enjoy considerable resources to support and inform guideline development [1], reporting [1–3] and critical appraisal, all developed by a variety of guideline organisations across the globe [4–7]. However, despite the advances of such international repositories, the variation in the quality of CPGs in different countries and regions around the world, speaks to its overall complexity and multiplicity [8–10]. Moreover, development approaches varies (and perhaps rightfully so), especially across countries, ranging from developing guidelines de novo (starting anew) to adopting or adapting CPGs to local contexts [11]. Approaches such as these that rely on using existing high-quality guidelines instead of de novo, provide opportunities to save time and resources, especially relevant in resource-poor settings.

For example, in a study evaluating African hypertension guidelines, recommendations were reported as a mixture of standard treatment guidelines (adapting), WHO guidelines (adopting) and de novo CPG development [12]. Notably, low- and middle-income countries (LMICs) continue to face increasing challenges and complexity, in terms of developing and implementing high quality CPGs; not only regarding capacity and funding, but also an increased burden of diseases (especially infectious diseases), healthcare worker shortages, and weaker health systems [13]. This was recently noted anew in relation to COVID-19 responses and its challenging effect on evidence synthesis groups in LMICs specifically [14], contributing to the call for guideline developers to stratify CPG recommendations more effectively for low-resource settings [15].

Over the last two decades, steady inroads have been made in terms of levelling CPG quality using quality appraisal tools [16]. These tools have become somewhat of a landscape architect, supporting journal editors when reviewing guidelines, underpinning and assessing guideline validity, and allowing the trust placed in guidelines, to be strengthened. The internationally accepted standard for the quality appraisal of CPGs, is the Appraisal of Guidelines for Research and Evaluation (AGREE) tool [17–19], notably the latest AGREE II tool and others (AGREE-REX and -GRS). This tool is comprised of six domains including 'Scope and purpose', 'Stakeholder involvement', 'Rigor of development', 'Clarity of presentation', 'Applicability', and 'Editorial independence'. Appraisal of guidelines using these six domains allows CPG developers, researchers, and decision makers to critically evaluate fundamental elements of guideline construction, quality, and implementation ability.

The AGREE tools have been used for a variety of reasons, from appraising individual CPGs for guideline adaptation, to appraising specific grouped CPGs, including: a sample from the National Guideline Clearinghouse when the AGREE-REX tool was developed [20]; several mixed medical topics [16], as well as targeted medical topics [21]. Multiple countries and regions worldwide have assessed their local, national, or regional CPGs with this tool and reported on this in methodological or review synthesis studies [10, 22–24]. Some countries periodically appraise CPGs using AGREE to gauge progression of guideline quality over time [8, 25, 26]. Added to this, studies have presented quality assessments of CPGs regarding specific disciplines, across various regions [12, 27, 28]. However, there is a paucity of evidence that has sought a worldwide overview; evaluating and comparing how studies focused on assessing the quality of guidelines in specific countries and/or regions, exploring whether the CPG quality differ among these jurisdictions.

This scoping review aimed to fill this knowledge gap by describing and mapping national and regional CPG synthesis studies that used the most recent versions of the AGREE tool (i.e., AGREE II, AGREE REX and AGREE GRS) towards unpacking how AGREE is used and reported in guideline assessment studies. This allowed a comprehensive global view of the characteristics of these national and regional synthesis studies; quantity and quality of the included CPGs; and a global and regional evaluation of the six AGREE domain scores. Additionally, it allowed a focus on specific domains *Rigor of development* and *Editorial independence*. These two important domains have been historically considered [20], and again recently indicated [29], as having the most direct effect on overall CPG content quality.

Methods

This scoping review described the methodology and characteristics of CPG synthesis studies and its included CPGs; and subsequently mapped and compared studies that used later versions of the AGREE tool, to assess CPG

quality country-wide and/or regionally. This includes describing the use of the AGREE tool, comparing domain scores, and ascertaining use of the overall assessment. A protocol was developed a priori (Additional file 1: Appendix A) and this study was conducted in accordance with the Joanna Briggs Institute methodology for scoping reviews [30], where results were reported according to the Preferred Reporting Items extension for Scoping Reviews (PRISMA-ScR) [31].

Search strategy
A predefined search strategy was used to conduct a comprehensive search in the following databases: Embase (Ovid), Medline (Pubmed), Epistemonikos, and grey literature (Web of Science grey literature, greylit.org, and contacting key experts). The search was conducted on 5 October 2021. Studies published in any language were included until full-text stage. The full Medline and Embase search strategies are listed in Additional file 1: Appendix B.

Eligibility criteria
Participants
Secondary research on CPG quality including scoping reviews; methods studies (including meta-epidemiological studies); reviews; systematic reviews of CPGs; and evaluation/analysis of quality of CPGs were considered. Synthesis studies on country-wide, regional, and topical CPGs were considered. Grey literature including thesis, dissertations and unpublished studies were considered. Exclusions based on study types, were international scoping reviews collating different countries into a topical review.

Concept
Any guideline synthesis study that used AGREE II, AGREE GRS (Global Rating Scale: a short item tool especially useful when time and resources are limited), and AGREE-REX (designed to evaluate the clinical credibility and implementation of CPGs) were considered. This tool uses six domains with 23 items, each scored 1–7 (strongly disagree to strongly agree) as well as two overall assessments. The overall assessment requires each assessor to firstly rate the overall quality of the guideline (on a scale of 1–7) and secondly to make a judgement as to whether this guideline is recommended for use (i.e., recommended; recommended with modifications; or not recommended). Exclusions included studies that used other tools or scores to appraise the quality of its CPGs.

Context
All countries or regions worldwide and their medical specialities and sub-specialities, including allied health and traditional medicine, were considered. Regions according to WHO, United Nation or Sustainable Development Goals (SDG) regional groupings were considered. Only CPGs answering human, health-related questions were considered. Exclusions included humanitarian, military combat, health-system related and non-human studies.

Study/source of evidence selection
We exported the retrieved records into a Mendeley Library and subsequently uploaded it to the Rayyan web platform [32] and removed duplicates. Two reviewers (MMA, SS) screened titles and abstracts independently for assessment against the inclusion criteria for the review. Potentially relevant records were retrieved in full, and citation details imported into a Microsoft Excel sheet. A single reviewer (MMA) assessed the full text of selected records in detail against the inclusion criteria. At full-text screening stage, only English (or studies translated into English), and Spanish studies were included. Reasons for exclusion of sources at full text screening stage are reported in the table of excluded studies (see Additional file 1: Appendix C). Any disagreements between the reviewers at each stage were resolved through discussion.

Data extraction and analysis
A single reviewer (MMA) extracted data from most included records using a data extraction tool (created in Microsoft Excel), assisted by a reviewer extracting the four Spanish records (IF) and checked by another reviewer (SS). This tool was piloted on a small sample of possible included studies, identified in a previous review. Only English and Spanish records were included, as the study was limited in its access to translation services. Data extracted included study types and methodology; characteristics of included CPGs; and AGREE tool use including domain and overall score results. Included synthesis studies were the units of analysis, and simple descriptive statistics was used to conduct the analysis using STATA 14.

Categorical variables were described as percentages; whereas continuous variables were described by means and standard deviations (sd) where data was normally distributed, otherwise reported as median and interquartile range. Data normality was determined graphically and using the skewness-kurtosis test. Studies that calculated median overall scores for AGREE domains were converted to a mean score, as recommended by Hozo et al.   [33]. This allowed one standard summary statistic across domains. Regions were measured according to United Nation Sustainable Development Goals (SDG) regional groupings [34], however other

regional groupings were also considered. SDG regions were chosen, due to the meaningful geographical presentation of different regions with comparable income status.

AGREE scores were analysed across subgroups into one of the seven SDG regions. A list of included studies is found in Additional file 1: Appendix D.

Results

The results of the search and study inclusion are presented in Fig. 1. A total of 2918 CPGs were included across 57 studies, accounting for all seven SDG regions. Best represented regions were Eastern and South-Eastern Asia, and Europe and Northern America; a well-represented region was Latin America and the Caribbean;
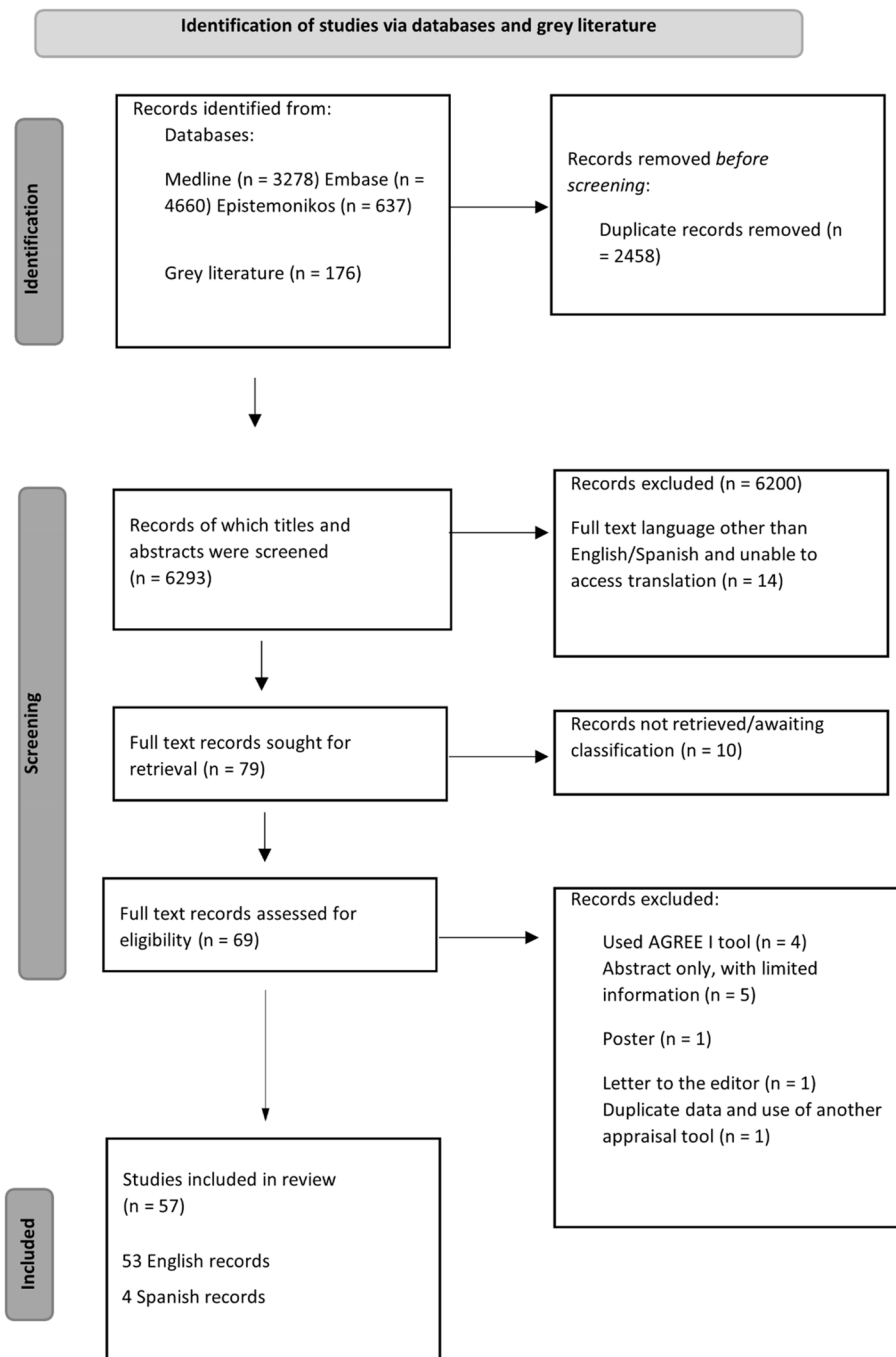

Fig. 1 Prisma flow chart of selecting included studies

whilst least represented areas were Northern Africa and Western Asia, and Oceania. Other regions outside the SDG grouping scope, included; Nordic countries, Asia, Middle East and North Africa, and Africa.

Characteristics of included studies

Fifty-seven studies were included in the analysis. The median number of included CPGs per study was 20 (IQR 44) and the median year-range of included CPGs was 8.5 years (IQR 6). The general aim of most studies was to describe and determine quality (or lack thereof) of CPGs, in either specific topical areas or country-wide; to enhance future CPG development and/or implementation of current CPGs. This included identifying areas of variability and vulnerability to speak to compliance and conformity. Most studies then used a cross-sectional design and topical purposeful sampling and were published by academic societies or researchers. However, a substantial number ($n = 14$) of studies evaluated all country-wide or regional CPGs. Most studies ($n = 34$, 60%) only searched local/regional databases and grey literature for including CPGs, whilst 20 studies (35%) searched both locally and internationally (including for example, Medline and guideline clearing houses). The number of included studies increased as the year of publication range increased, mostly published after 2019 (Table 1). One study only utilized the overall assessment and did not report domain scores and one study assessed one domain only.

Characteristics of included guidelines
Included CPGs

Table 2 illustrates the characteristics of included CPGs indicating that the included guidelines scoped a wide range of topics. Thirty-six studies (63%) did not use a formal definition of what it regarded as a CPG in its inclusion criteria. Subsequently, there was a poor reporting of development methods of CPGs and similarly, use of Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) to assess the certainty of the underlying evidence. Only 10 studies (18%) commented on the methods of development of included studies, where de novo (starting anew) development was the most prevalent ($n = 5$, 50% of the 10 studies). Additionally, overall, only 11 studies (19%) reported the use of GRADE in the development of the included CPGs.

AGREE use

AGREE use and completeness, was well reported and presented in all included studies. Most studies used two to four assessors to assess domain scores, and inter-observer agreement scores were used by 36 (63%). Almost three quarters of studies ($n = 43$, 74%) reported

Table 1 Broad summary of included studies

Characteristics	n (%)
Types of studies	
Cross-sectional analysis	34 (59.6)
Systematic review	17 (29.8)
Review[a]	6 (8.8)
Year of publication	
2010 – 2014	9 (15.8)
2015 – 2018	22 (38.6)
2019 – October 2021	26 (45.6)
Entity responsible for the appraisal	
Government related	4 (7)
Academic society/Researchers	49 (86)
Professional society	3 (5)
Non-governmental organization	1 (2)
Geographical locations by SDG region	
Sub-Saharan Africa	2 (3.7)
Northern Africa and Western Asia	1 (1.8)
Central and Southern Asia	2 (3.7)
Eastern and South-Eastern Asia	20 (38)
Latin America and the Caribbean	9 (17)
Oceania	1 (1.8)
Europe and Northern America	18 (34)
Sampling of CPGs	
No sample[b]	14 (24.5)
Random sample	1 (2)
Systematic sample	11 (19)
Topical purposeful sampling	29 (51)
Other purposeful sampling	2 (3.5)

[a] This category broadly includes literature surveys, methodological reviews, narrative reviews, and other non-systematic reviews

[b] All country/regional CPGs were included

the overall guideline assessment; most studies reporting only one overall assessment and modifying it.

Overall assessment

The method of formulating the overall quality assessment score of CPGs varied in most of our included studies. The majority of 43 studies (74%) made use of this overall assessment. Twenty-nine studies used one assessment only and thirteen studies used both assessments. However, 24 studies (57%) modified this overall assessment and did not apply it as per the AGREE manual, where 12 variations of modification were found. The three most common modifications included: i) calculating an overall average across the six AGREE domains ii) using a cut-off of *Rigor of development* domain score > 60% and iii) using a cut-off sliding scale of most domain scores > 60% being 'recommended', scores of 30–60% 'recommended with modifications', and if most scores were < 30% 'not recommended'. It was noted that even among these three

Mc Allister *et al. BMC Medical Research Methodology*   (2023) 23:283

Page 6 of 12

Table 2 Characteristics of *n* = 57 included studies and AGREE tool use

Characteristics	n (%)
Type of included CPGs	
All national/regional	9 (16)
Governmental	3 (5)
Topical	41 (72)
Professional society	4 (7)
Methods of CPG development	
None mentioned	47 (82)
De novo	5 (9)
Adopt	1 (2)
Miscellaneous	4 (7)
Healthcare topic/area	
Dermatology	1 (2)
Complementary and alternative medicine	4 (7)
Emergency medicine	1 (2)
Otolaryngology	3 (5)
"High priority diseases"	2 (3.5)
Internal medicine	10 (17)
Neurology	2 (3.5)
Nursing practice	2 (3.5)
Oncology	2 (3.5)
Orthopaedics	4 (7)
Paediatrics	2 (3.5)
Poverty-related	1 (2)
Primary care	1 (2)
Psychiatry	2 (3.5)
Surgery	1 (2)
Mixed	19 (33)
Strict inclusion criteria regarding the definition of a CPG	
Yes	21 (37)
No	36 (63)
AGREE use	
Used 2–4 assessors	56 (98)
Used overall assessment	43 (74)
Used one overall assessment only	29 (69)
Modified overall assessment	24 (57)

groupings, slight variation existed. Additionally, there was no regional pattern regarding this modification. Latin America and the Caribbean (*n* = 7, 78%), and Europe and Northern America (*n* = 15, 83%), used this overall assessment in most of their studies.

SDG regions
Overall domain scores were low for *Rigor of development* (ROD), *Applicability*, and *Editorial independence* (EI); moderate for *Stakeholder involvement*; and higher for *Scope and purpose* and *Clarity and presentation*

(Table 3). Figure 2 depicts the low global scores for ROD and EI per SDG region; simultaneously noting the low overall range of a maximum of 62.1% for ROD and a maximum of 46.3% for EI.

Figure 3 looks broadly at domain ROD, in relation to Gross Domestic Product (GDP) per capita [35], country-income profile [36], and population size. Here, there is a seemingly linear relationship between the ROD quality score, GDP per capita and country-income status. Lastly, we did not identify any studies representing low-income countries.

Discussion
Main findings
This scoping review provided a global view investigating the heterogeneity of CPG synthesis studies across different countries and regions, as noted in other studies [27, 37]. Despite the variation in quality, included CPG synthesis studies (and their included CPGs), were consistently described and adequately assessed and reported with the AGREE tool. This highlights the importance and value of quality assessment tools that are straightforward to use. Characteristics of these synthesis studies, mostly performed by researchers from the academia, were uniformly consisting of cross-sectional reviews where topical sampling was performed. Though high- and middle-income countries and regions were well represented, there were no synthesis studies from low-income countries. Generally, three AGREE domains of *Rigor of development*, *Applicability* and *Editorial independence* had the lowest domain scores, and this was most prevalent in middle-income countries and regions. *Editorial independence* repeatedly scored as the lowest domain, even in countries that had acceptable scores for other domains (including *Rigor of development*). Despite this, many AGREE domains across geography had either moderate or high-quality scores, and the tendency to appraise country-wide or regional CPGs appear to be increasing.

Characteristics of studies and CPGs
Regions of Eastern and South-Eastern Asia, and Europe and Northern America were the best represented in terms of quantity of studies included. This was recently seen in a living systematic review of Covid-19 rapid guidelines, where 45% of guidelines originated from high income countries (HICs) and no guidelines from low-income countries [38]. The included CPGs were either topical or all country-wide, regional, or governmental CPGs; mainly initiated by the medical society, corresponding to findings of previous reviews [16, 39]. Similar reviews found that high-quality guidelines were more often produced by government-supported organizations and/or within regulated, coordinated

Table 3 AGREE II domain scores for all SDG regions (*n* = 51)

Region	Domain (Mean% ± SD)					
	Scope and purpose (%)	Stakeholder involvement (%)	Rigor of development (%)	Clarity and presentation (%)	Applicability (%)	Editorial independence (%)
Central and Southern Asia	57.0 ± 31.5	34.5 ± 22.0	14.5 ± 17.7	53.2 ± 31.9	26.1 ± 25.6	46.1 ± 5.5
Eastern and South-Eastern Asia	53.5 ± 20.8	34.7 ± 15.6	28.6 ± 18.8	56.6 ± 17.9	21.3 ± 10.9	21.2 ± 20.3
Europe and Northern America	77.6 ± 17.8	55.1 ± 19.9	54.9 ± 26.3	77.2 ± 14.4	36.0 ± 13.4	45.4 ± 30.0
Latin America and the Caribbean	59.7 ± 21.0	28.3 ± 15.0	24.2 ± 13.0	63.5 ± 20.8	32.1 ± 25.4	33.0 ± 21.2
Northern Africa and West Asia*	64.0	37.9	35.3	37.9	49.3	46.3
Sub-Saharan Africa	56.6 ± 1.9	33.0 ± 8.5	11.8 ± 3.2	76.0 ± 9.9	18.6 ± 12.2	3.6 ± 5.1
Oceania*			62.1			
Total overall score for SDG regions	63.1 ± 21.7	40.4 ± 19.5	36.0 ± 24.5	64.9 ± 19.6	28.7 ± 16.7	32.1 ± 25.9

Rapid signal for quality of regional CPGs: Low quality: Red <40%; Moderate quality: Yellow 40%-59%; High quality: Green ≥60%

*Regions only contributing one study, thus no standard deviation derived

programmes [21, 40–42], in comparison to guidelines produced by professional societies.

Regarding the use of GRADE methods, either in evidence synthesis or as part of the evidence to decision process, only 11 synthesis studies (19%) reported using any GRADE methods, two of these primarily assessing GRADE adherence. Lack of reporting of GRADE and other methods of judging the certainty of evidence, along with how evidence is used to develop recommendations, is concerning. Reviewers of CPGs are encouraged to highlight the importance of how these evidence-to-decision processes evolve, to strengthen guideline development and assessment in the future. Regional quality assessments have shown that there is a slow but increasing incorporation of these requirements [43, 44]. Online tools, such as GRADEpro [45], can make the process transparent, faster, and more user-friendly in a fast-paced age.

Reciprocally, the low reporting rate of guideline development methods and absence of Covid-19 or rapid guidelines included in this scoping review, could signify areas of much needed growth. Through multiple online high-quality/tech savvy platforms created for the Covid pandemic, the guideline community can be inspired to make use of these types of resources to further advance guideline quality. This can assist in processes like guideline adaptation, that aim to reduce research waste and duplication of effort; by using living reviews and guideline repositories [46], recommendation mapping [47], and guideline development resources such as produced by e-COVID [48]. A future research focus on adaptation methodology such as GRADE Adolopment and others
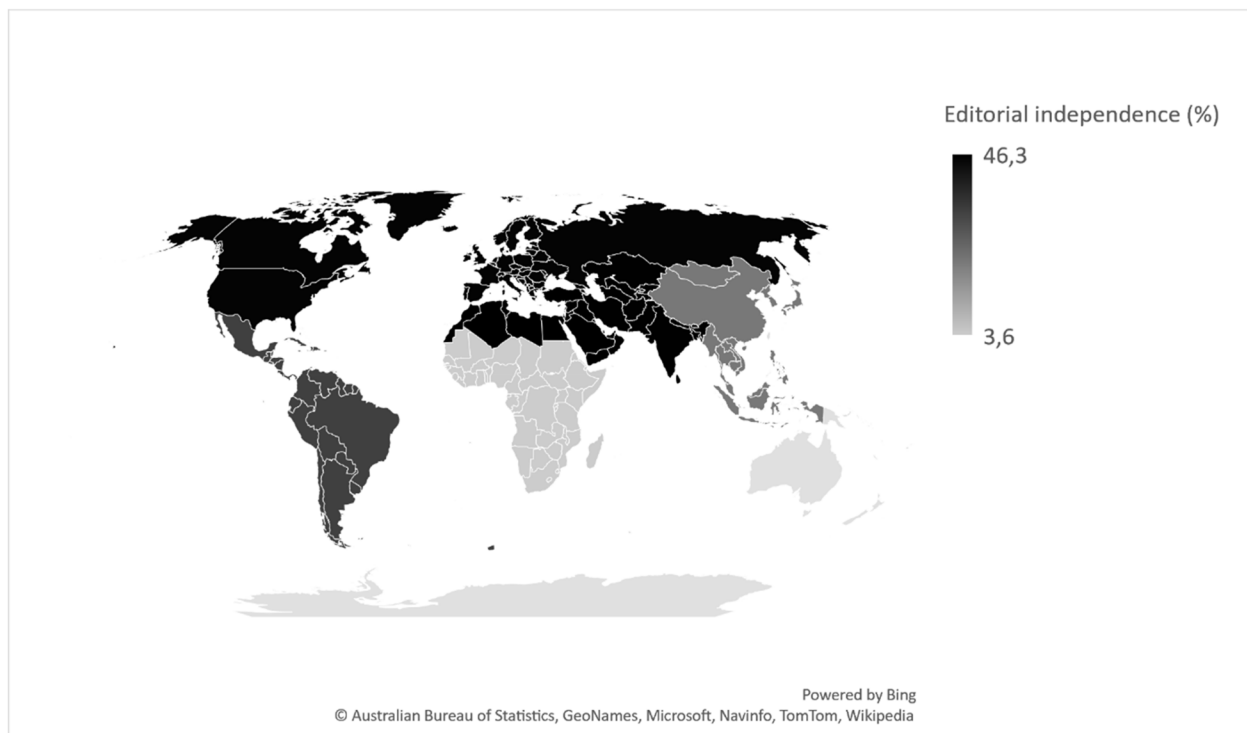
Mc Allister *et al. BMC Medical Research Methodology*   (2023) 23:283

Page 8 of 12

© Australian Bureau of Statistics, GeoNames, Microsoft, Navinfo, TomTom, Wikipedia

Fig. 2 Heatmaps of AGREE domains *Rigor of development* and *Editorial independence* in percentage (%) per SDG region
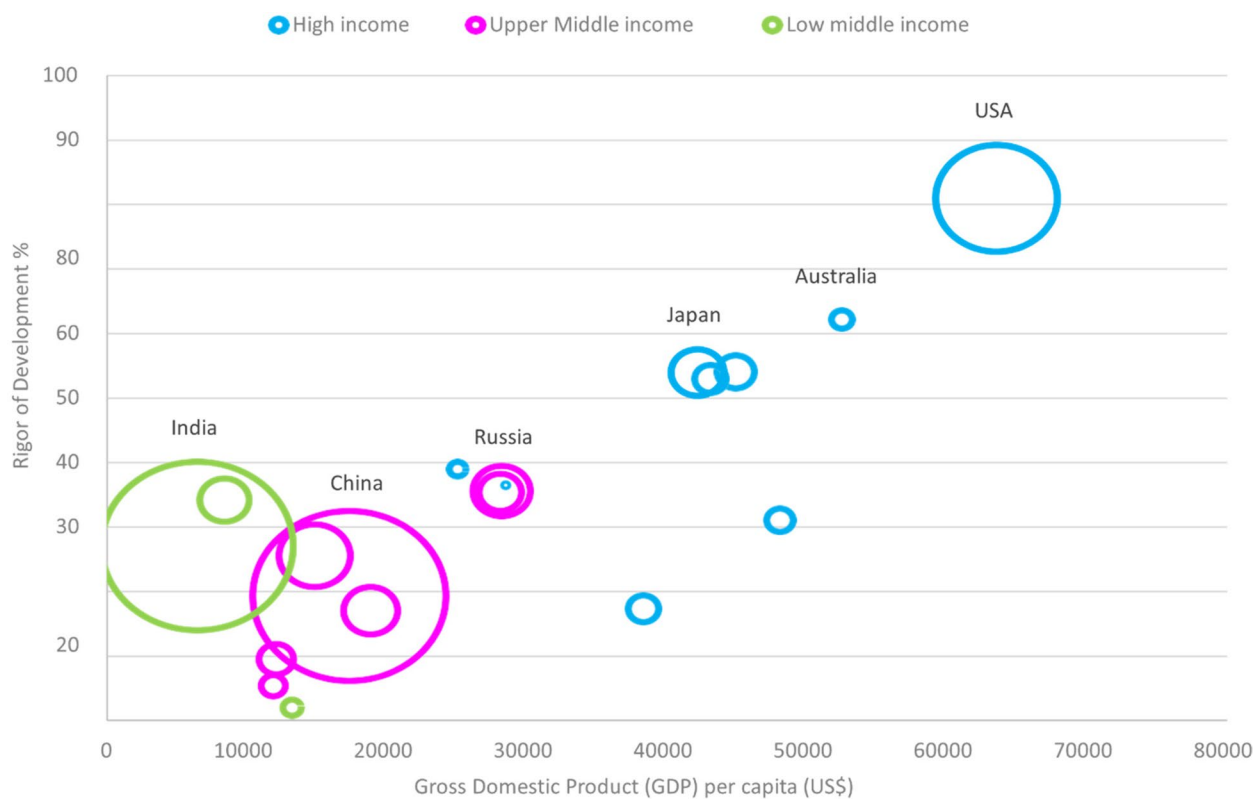
Fig. 3 Bubble plot comparing 4 variables, including ROD domain scores, Gross Domestic Product (GDP) per capita, different income-level countries, and population size (size of the bubble)

[49], can be further supported in LMICs to build capacity for guideline development in these regions.

The Inclusive Internet Index is an example of a starting point to plug LMICs into these resources. As the 2021 report [50] indicates, there is still a 'digital divide' preventing most LMICs from accessing adequate internet supplies, where sub-Saharan Africa was shown to have the most constrained connectivity globally. Future research can involve investigating the challenges of LMICs on internet connectivity, especially regarding uptake and implementation of CPGs. This can include investigating the uptake of evidence-to-decision online tools, and how this translates into CPG quality.

AGREE tool use and domains

AGREE tool use was well reported where most synthesis studies used a recommended two to four assessors per CPG, reported on interobserver agreement and assessed all 6 domains; similar to previous reviews [16, 39]. The idea of rating the overall quality of CPGs has been the subject of much debate. The AGREE tool recommends focusing on all six domains and not solely using the overall assessment to judge quality. Several authors

have suggested different thresholds for this overall score, mostly suggesting using the strength of the *Rigor of development* domain. Two recent studies attempted rating the importance of the different AGREE domains on the overall assessment, indicating that *Rigor of development*, *Clarity of presentation*, *Applicability*, and *Editorial independence* had a significant influence on the overall quality assessment [37, 51]. Almost three quarters (74%) of the synthesis studies included in this scoping review, used this overall assessment to define its CPGs' quality, and this is similar to the finding of two other publications [51, 52]. Those opting to use this overall assessment used multiple ways to modify it, similar to a review where only 23% of included studies provided clear criteria for generating the cut-offs applied [53]. This could indicate that several users would welcome an explicit distinction between high- and low-quality guidelines [5]. The important domain of *Rigor of development*, particularly used in the overall CPG quality consideration, is often scored as low [8, 10, 21], similarly found in this study. Notably, domain *Editorial independence* is becoming increasingly scrutinized as important for guideline quality [24]. In fact, Molino et al. showed that higher quality CPGs reported its funding

sources 10 times more often than those of lower quality [21]. Multiple recent studies reaffirmed the low scores across guidelines for this domain [54–56]; and this gap is crucial in LMICs especially [57], however, it can be argued that it is equally important for all countries world-wide.

Strengths and limitations

The scoping review design used in this study allowed us to attain a global overall view. This review focused on studies that used the AGREE tool for quality assessment and cannot comment on synthesis studies of guidelines using other appraisal tools. Additionally, there might be included studies that applied AGREE to other types of documents (such as end-user guidance documents or broad summary documents). There has been an indication that other types of 'guidance documents' have lower AGREE scores than true CPGs [58], and this could have overestimated the results of this study. This scoping review cannot conclude that CPGs from HICs are of higher quality than those of LMICs, as we are limited by the availability of studies summarizing the available guidelines per regions. In addition to this, the CPGs assessed for each region was often a sub-sample of all available CPGs in that region. Importantly, regional or national CPG synthesis studies, may be preferentially submitted to local journals, without indexing in international databases. Subsequently, these researchers may be inclined to not publish at all, simply posting results on association or organizational websites, resulting in potential missed studies. Cumulatively, this could have influenced the results of this scoping review, especially the associations with world regions and income groups, as LMICs could be more inclined to solely make use of local distributions of quality assessments of CPGs.

Conclusion

When looking at the landscape of guideline quality, there has been various attempts to level the playfield and inroads have been made. There is a current tendency to critically evaluate guideline quality in country-wide and regional approaches and AGREE is overarchingly used well in this practice. However, guideline *Rigor of development* varies between HICs and LMICs, necessitating building further guideline development capacity, including use of GRADE for guidelines. Improved reporting of funding and competing interests, as well as guideline development approaches and their underlying evidence sources, can further enhance regional quality of guidelines. Assessing country-wide or regional guidelines with quality appraisal tools, could advance overall guideline quality for all areas globally. This is an important step forward and toward global guideline uniformity, as clinicians are in dire need of high-quality guidelines to improve delivery and quality of care.

Supplementary Information

Additional file 1.

Declarations

Ethics approval and consent to participate
Not applicable as this is a scoping review of existing literature.

Consent for publication
Not applicable.

Competing interests
Ivan D. Florez (IDF) is the current lead of the AGREE collaboration and employed by the collaboration. The remainder of the authors (MMA, MMC and SS) have no competing interests to declare.

Author details
[1]Department of Global Health, Division of Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, Francie Van Zijl Drive TYGERBERG 7505, Cape Town, South Africa. [2]Department of Pediatrics, Faculty of Medicine, University of Antioquia, Medellin, Colombia. [3]School of Rehabilitation Science, McMaster University, Hamilton, Canada. [4]Pediatric Intensive Care Unit, Clinica Las Americas AUNA, Medellin, Colombia.

References
1. Schünemann HJ, Wiercioch W, Etxeandia I, Falavigna M, Santesso N, Mustafa R, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ. 2014;186(3):E123–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24344144. Cited 3 Sep 2021.
2. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P, et al. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med. 2012;156(7):525–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22473437. Cited 3 Sep 2021.
3. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice, Graham R, Mancher M, Miller Wolman D, et al. editors. Clinical Practice Guidelines We Can Trust. National Academies Press (US); 2011. Available from: https://www.ncbi.nlm.nih.gov/books/NBK209538/. Cited 27 Aug 2021.

4. National Institute for Health and Clinical Excellence. The guidelines manual | Guidance | NICE. NICE; 2012. Available from: https://www.nice.org.uk/process/pmg6/chapter/introduction. Cited 27 Aug 2021.
5. NHMRC NHaMRC. Guide to the development, evaluation and implementation of clinical practice guidelines | NHMRC. 2009. Available from: https://www.nhmrc.gov.au/about-us/publications/guide-development-evaluation-and-implementation-clinical-practice-guidelines. Cited 27 Aug 2021.
6. Square G. A guideline developer's handbook Scottish Intercollegiate Guidelines Network Scottish Intercollegiate Guidelines Network Citation text Scottish Intercollegiate Guidelines Network (SIGN). Complying with international standards. Revised ed. 2008. Available from: www.sign.ac.uk. Cited 27 Aug 2021.
7. World Health Organization. WHO handbook for guideline development, 2nd ed. 2014. Available from: https://www.who.int/publications/i/item/9789241548960.
8. Zhou Q, Wang Z, Shi Q, Zhao S, Xun Y, Liu H, et al. Clinical Epidemiology in China series. Paper 4: The reporting and methodological quality of Chinese clinical practice guidelines published between 2014 and 2018: A systematic review. J Clin Epidemiol. 2021; Available from: http://www.elsevier.com/locate/jclinepi.
9. Ariel Franco JV, Arancibia M, Meza N, Madrid E, Kopitowski K. Clinical practice guidelines: Concepts, limitations and challenges. Medwave. 2020;20(3):e7887–e7887. Available from: https://www.medwave.cl/link.cgi/Medwave/Revisiones/MetodInvestReport/7887.act.
10. Almazrou S, Alsubki L, Alsaigh N, Aldhubaib W, Ghazwani S. Assessing the quality of clinical practice guidelines in the middle east and North Africa (MENA) region: a systematic review. J Multidiscip Healthc. 2021;14:297–309 http://www.epistemonikos.org/documents/f0ee9ed4024cdea7b091caaa2f523ff3ccb0f6a0.
11. Dizon JM, Machingaidze S, Grimmer K. To adopt, to adapt, or to contextualise? The big question in clinical practice guideline development. BMC Res Notes. 2016;9(1):442. https://doi.org/10.1186/s13104-016-2244-7. Cited 1 Sep 2021.
12. Okwen PM, Maweu I, Grimmer K, Margarita Dizon J. Evaluation of all African clinical practice guidelines for hypertension: quality and opportunities for improvement. J Eval Clin Pract. 2019;25(4):565–74. https://doi.org/10.1111/jep.12954.
13. Iwelunmor J, Blackstone S, Veira D, Nwaozuru U, Airhihenbuwa C, Munodawafa D, et al. Toward the sustainability of health interventions implemented in sub-Saharan Africa: a systematic review and conceptual framework. Implement Sci. 2015;11(1):43. Available from: https://doi.org/10.1186/s13012-016-0392-8. Cited 8 Sep 2021.
14. Stewart R, El-Harakeh A, Cherian SA, LMIC members of COVID-END. Evidence synthesis communities in low-income and middle-income countries and the COVID-19 response. Lancet (London, England). 2020;396(10262):1539–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33096041. Cited 14 Sep2021.
15. Maaløe N, Ørtved AMR, Sørensen JB, Sequeira Dmello B, van den Akker T, Kujabi ML, et al. The injustice of unfit clinical practice guidelines in low-resource realities. Lancet Glob Heal. 2021;9(6):e875–9. Available from: http://www.thelancet.com/article/S2214109X21000590/fulltext. Cited 9 June 2023.
16. Armstrong JJ, Goldfarb AM, Instrum RS, MacDermid JC. Improvement evident, but still necessary in clinical practice guideline quality: a systematic review. J Clin Epidemiol. 2017;81:13–21. Available from: http://www.epistemonikos.org/documents/00574ff91ea8b3dbf56bec04feda6b2eaf82e250. Cited 8 Sep 2021.
17. The Appraisal of Guidelines and Research and Evaluation collaboration. AGREE Tools - AGREE Enterprise website. Available from: https://www.agreetrust.org/resource-centre/. Cited 14 Sep 2021.
18. 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. C Can Med Assoc J = J l'Association medicale Can. 2010;182(10):1045–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20513780. Cited 1 Sep 2021.
19. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. CMAJ. 2010;182(10):E472–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20513779. Cited 1 Sep 2021.
20. Brouwers MC, Spithoff K, Kerkvliet K, Alonso-Coello P, Burgers J, Cluzeau F, et al. Development and validation of a tool to assess the quality of clinical practice guideline recommendations. JAMA Netw Open. 2020;3(5):e205535. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2766238. Cited 28 Sep 2021.
21. Molino C de GRC, Leite-Santos NC, Gabriel FC, Wainberg SK, Vasconcelos LP de, Mantovani-Silva RA, et al. Factors associated with high-quality guidelines for the pharmacologic management of chronic diseases in primary care. JAMA Intern Med. 2019;179(4):553. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30776060. Cited 25 Nov 2021.
22. Kataoka Y, Anan K, Taito S, Tsujimoto Y, Kurata Y, Wada Y, et al. Quality of clinical practice guidelines in Japan remains low: a cross-sectional meta-epidemiological study. J Clin Epidemiol. 2021;138:22–31. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0895435621002031. Cited 9 Sep 2021.
23. Chang S-G, Kim D-I, Shin E-S, Jang J-E, Yeon J-Y, Lee Y-S. Methodological quality appraisal of 27 Korean guidelines using a scoring guide based on the AGREE II instrument and a web-based evaluation. J Korean Med Sci. 2016;31(5):682. https://doi.org/10.3346/jkms.2016.31.5.682.
24. Dans LF, Salaveria-Imperial MLA, Miguel RTD, Tan-Lim CSC, Eubanas GAS, Tolosa MTS, et al. Guidelines in low and middle income countries paper 3: appraisal of Philippine clinical practice guidelines using appraisal of guidelines for research and evaluation ii: improvement needed for rigor, applicability, and editorial independence. J Clin Epidemiol. 2020;127:184–90. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0895435619310297.
25. Loezar C, Pérez-Bracchiglione J, Arancibia M, Meza N, Vargas M, Papuzinski C, et al. Guidelines in low and middle income countries paper 2: quality assessment of Chilean guidelines: need for improvement in rigor, applicability, updating, and patients' inclusion. J Clin Epidemiol. 2020;127:177–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32853762. Cited 4 Sep 2021.
26. Gao Y, Wang J, Luo X, Song X, Liu L, Ke L, et al. Quality appraisal of clinical practice guidelines for diabetes mellitus published in China between 2007 and 2017 using the AGREE II instrument. BMJ Open. 2019;9(9):e022392. https://doi.org/10.1136/bmjopen-2018-022392.
27. Malherbe P, Smit P, Sharma K, McCaul M. Guidance we can trust? The status and quality of prehospital clinical guidance in sub-Saharan Africa: a scoping review. African J Emerg Med. 2021;11(1):79–86. Available from: https://www.sciencedirect.com/science/article/pii/S2211419X2030094X. Cited 5 Nov 2023.
28. Werner RN, Marinović B, Rosumeck S, Strohal R, Haering NS, Weberschock T, et al. The quality of <scp>E</scp>uropean dermatological guidelines: critical appraisal of the quality of <scp>EDF</scp> guidelines using the <scp>AGREE II</scp> instrument. J Eur Acad Dermatology Venereol. 2016;30(3):395–403. https://doi.org/10.1111/jdv.13358.
29. Florez ID, Brouwers MC, Kerkvliet K, Spithoff K, Alonso-Coello P, Burgers J, et al. Assessment of the quality of recommendations from 161 clinical practice guidelines using the Appraisal of Guidelines for Research and Evaluation-Recommendations Excellence (AGREE-REX) instrument shows there is room for improvement. Implement Sci. 2020;15(1):79. https://doi.org/10.1186/s13012-020-01036-5. Cited 28 Sep 2021.
30. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth. 2020;18(10):2119–26. https://doi.org/10.11124/JBIES-20-00167. Cited 5 Sep 2021.
31. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30178033. Cited 5 Sep 2021.
32. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210. https://doi.org/10.1186/s13643-016-0384-4. Cited 2 Dec 2021.
33. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol. 2005;5(1):13. https://doi.org/10.1186/1471-2288-5-13. Cited 12 Nov 2021.
34. United Nations. SDG Indicators — SDG Indicators. Available from: https://unstats.un.org/sdgs/indicators/regional-groups/. Cited 2 Dec 2021.
35. The World Bank. GDP per capita, PPP (current international $) | Data. data.worldbank.org. Available from: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?year_high_desc=true. Cited 30 Nov 2021.
36. The World Bank. World Bank Country and Lending Groups – World Bank Data Help Desk. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups. Cited 29 Nov 2021.

37. Hatakeyama Y, Seto K, Amin R, Kitazawa T, Fujita S, Matsumoto K, et al. The structure of the quality of clinical practice guidelines with the items and overall assessment in AGREE II: a regression analysis. BMC Health Serv Res. 2019;19(1):788. https://doi.org/10.1186/s12913-019-4532-0.

38. Amer YS, Titi MA, Godah MW, Wahabi HA, Hneiny L, Abouelkheir MM, et al. International alliance and AGREE-ment of 71 clinical practice guidelines on the management of critical care patients with COVID-19: a living systematic review. J Clin Epidemiol. 2021;0(0). Available from: http://www.ncbi.nlm.nih.gov/pubmed/34785346. Cited 6 Dec 2021.

39. Dijkers MP, Ward I, Annaswamy T, Dedrick D, Feldpausch J, Moul A, et al. Quality of rehabilitation clinical practice guidelines: an overview study of AGREE II appraisals. Arch Phys Med Rehabil. 2020;101(9):1643–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32437691. Cited 15 Nov 2021.

40. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. Int J Technol Assess Health Care. 2003;19(1):148–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12701947. Cited 3 Sep 2021.

41. Fervers B, Burgers JS, Haugh MC, Brouwers M, Browman G, Cluzeau F, et al. Predictors of high quality clinical practice guidelines: examples in oncology. Int J Qual Heal Care J Int Soc Qual Heal Care. 2005;17(2):123–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15665068. Cited 25 Nov 2021.

42. Burgers JS, Collaboration F the A, Grol R, Collaboration F the A, Klazinga NS, Collaboration F the A, et al. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. Int J Qual Heal Care. 2003;15(1):31–45. https://doi.org/10.1093/intqhc/15.1.31. Cited 3 Sep 2021.

43. Cabrera PA, Pardo R. Review of evidence based clinical practice guidelines developed in Latin America and Caribbean during the last decade: an analysis of the methods for grading quality of evidence and topic prioritization. Global Health. 2019;15(1):14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30782176. Cited 1 Dec 2021.

44. Barker TH, Dias M, Stern C, Porritt K, Wiechula R, Aromataris E, et al. Guidelines rarely used GRADE and applied methods inconsistently: a methodological study of Australian guidelines. J Clin Epidemiol. 2021;130:125–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33130237. Cited 17 Nov 2021.

45. GRADEpro. Available from: https://www.gradepro.org/. Cited 8 Dec 2021.

46. Australian National COVID-19 Clinical Evidence Taskforce. WHO living guidelines approach [Internet]. Available from: https://covid-nma.com/.

47. COVID19 Recommendations_recmap. Available from: https://covid19.recmap.org/about. Cited 9 Dec 2021.

48. McCaul M, Tovey D, Young T, Welch V, Dewidar O, Goetghebeur M, et al. Resources supporting trustworthy, rapid and equitable evidence synthesis and guideline development: results from the COVID-19 evidence network to support decision-making (COVID-END). J Clin Epidemiol. 2022;151:88–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35868494. Cited 2 Nov 2023.

49. Darzi A, Abou-Jaoude EA, Agarwal A, Lakis C, Wiercioch W, Santesso N, et al. A methodological survey identified eight proposed frameworks for the adaptation of health related guidelines. J Clin Epidemiol. 2017;86:3–10. Available from: https://pubmed.ncbi.nlm.nih.gov/28412463/. Cited 27 Jan 2022.

50. Executive Summary - The Inclusive Internet Index. Available from: https://theinclusiveinternet.eiu.com/summary. Cited 8 Dec 2021.

51. Hoffmann-Eßer W, Siering U, Neugebauer EAM, Brockhaus AC, Lampert U, Eikermann M, et al. Guideline appraisal with AGREE II: Systematic review of the current evidence on how users handle the 2 overall assessments. PLoS ONE. 2017;12(3):e0174831. Available from: http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0174831&type=printable. Cited 30 Aug 2021.

52. Bargeri S, Iannicelli V, Castellini G, Cinquini M, Gianola S. AGREE II appraisals of clinical practice guidelines in rehabilitation showed poor reporting and moderate variability in quality ratings when users apply different cuff-offs: a methodological study. J Clin Epidemiol. 2021;139:222–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34437947. Cited 1 Dec 2021.

53. Siering U, Lampert U, Hoffmann-Eser W, Neugebauer EAM, Eikermann M, Hoffmann-Eßer W, et al. Systematic review of current guideline appraisals performed with the appraisal of guidelines for research & evaluation II instrument-a third of AGREE II users apply a cut-off for guideline quality. J Clin Epidemiol. 2018;95:120–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29288133. Cited 4 Sep 2021.

54. Kent K, Jessup B, Marsh P, Barnett T, Ball M. A systematic review and quality appraisal of bereavement care practice guidelines. J Eval Clin Pract. 2020;26(3):852–62. https://doi.org/10.1111/jep.13225. Cited 1 Dec 2021.

55. Miguel RTD, Silvestre MAA, Salaveria-Imperial MLA, Tolosa MTS, Eubanas GAS, Dans LF. Disclosures of conflicts of interest in clinical practice guidelines. Clin Epidemiol Glob Heal. 2021;9:355–9. Available from: http://www.elsevier.com/journals/clinical-epidemiology-and-global-health/2213-3984.

56. Yao L, Chen Y, Wang X, Shi X, Wang Y, Guo T, et al. Appraising the quality of clinical practice guidelines in traditional Chinese medicine using AGREE II instrument: a systematic review. Int J Clin Pract. 2017;71(5):e12931. https://doi.org/10.1111/ijcp.12931.

57. Rohwer A, Young T, Wager E, Garner P. Authorship, plagiarism and conflict of interest: views and practices from low/middle-income country health researchers. BMJ Open. 2017;7(11):e018467. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29170291. Cited 1 Dec 2021.

58. Abell B, Glasziou P, Hoffmann T. Exploration of the methodological quality and clinical usefulness of a cross-sectional sample of published guidance about exercise training and physical activity for the secondary prevention of coronary heart disease. BMC Cardiovasc Disord. 2017;17(1):153. [cited 2023 Oct 28]. Available from: https://bmccardiovascdisord.biomedcentral.com/articles/10.1186/s12872-017-0589-z.

Publisher's Note