



Methodological Quality of Pulmonary Arterial Hypertension Treatment Evidence-Based Guidelines: A Systematic Review Using the AGREE II and AGREE REX Tools

Ana Paula Oliveira Vilela¹ · Flávia Deffert² · Rosa Camila Lucchetta³ · Yara Maria da Silva Pires⁴ · Felipe Fernando Mainka⁴ · Fernanda S. Tonin^{4,5} · Roberto Pontarolo⁶

Accepted: 26 June 2024 / Published online: 9 July 2024
© The Author(s) 2024

Abstract

Purpose Pulmonary arterial hypertension (PAH) is a progressive disease with a poor prognosis, and its management should be grounded in well-developed clinical practice guidelines (CPG). Thus, we critically assess the methodological quality of the available CPG for pharmacological treatments for PAH.

Methods A systematic review (CRD42023387168) was performed in PubMed, Cochrane, Embase, and Tripdatabase (Jan-2023). Eligible records were appraised by four reviewers using the Appraisal of Guidelines, Research, and Evaluation Collaboration tool (AGREE II) and the complementary tool for assessing recommendations' quality and certainty, AGREE REX. Descriptive statistics were used to summarize the data.

Results Overall, 31 guidelines, mainly authored by professional societies (90%), targeting only physicians as primary users (84%), were identified. Guidelines presented a moderate overall quality (scores of 63% and 51% in AGREE II and AGREE REX, respectively), with a few domains showing slight improvements over the years. AGREE II "Scope and Purpose" (94%) and "Presentation Clarity" (99%) domains obtained the highest scores. The items related to "Stakeholder involvement," "Editorial independence," and "Clinical applicability" (AGREE REX) were fairly reported. Conversely, CPG lacks rigor in development (32% score, AGREE II), scarcely discusses the role of stakeholders, and provides deficient data on the implementation of recommendations (scores of 35% and 46% in AGREE II and AGREE REX, respectively). No differences in the quality of guidelines published by different developers or countries were observed ($p > 0.05$).

Conclusion Methodological weaknesses are common among guidelines addressing PAH treatment, especially regarding scientific rigor, stakeholders' values and preferences, and facilitators and barriers to implementability. Particular attention should be given to developing future guidelines.

Keywords Practice guideline · Quality assessment · Pulmonary arterial hypertension · AGREE II · AGREE REX

✉ Fernanda S. Tonin
fernanda.tonin@estesl.ipl.pt

Ana Paula Oliveira Vilela
paulavili@hotmail.com

Flávia Deffert
flaviadeffert@gmail.com

Rosa Camila Lucchetta
rc.lucch@yahoo.com.br

Yara Maria da Silva Pires
yarapires@ufpr.br

Felipe Fernando Mainka
felipemainka@hotmail.com

Roberto Pontarolo
pontarolo@ufpr.br

¹ Pharmaceutical Assistance Postgraduate Program, Federal University of Paraná, Curitiba, Brazil

² Pharmaceutical Sciences Postgraduate Program, Universidade Federal Do Paraná, Curitiba, Brazil

³ Health Technology Assessment Unit, Oswaldo Cruz German Hospital, São Paulo, Brazil

⁴ Pharmaceutical Sciences Postgraduate Program, Federal University of Paraná, Curitiba, Brazil

⁵ H&TRC-Health & Technology Research Center, ESTeSL-Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, Av. Dom João II Lote 4.69 01, 1990-096 Lisbon, Portugal

⁶ Department of Pharmacy, Federal University of Paraná, Curitiba, Brazil

Introduction

Pulmonary arterial hypertension (PAH), a subtype of pulmonary hypertension (i.e., group 1), is hemodynamically defined as a mean pulmonary arterial pressure > 20 mmHg, a pulmonary vascular resistance > 2 Wood units, and a pulmonary artery wedge pressure ≤ 15 mmHg at right heart catheterization [1]. It is a rare, highly complex, and progressive disorder that can lead to right ventricular failure and ultimately premature death [2–4]. The prevalence of PAH is around 15 to 50 cases per million within the USA and Europe, with an incidence of approximately five cases per population of 1 million per year, mainly affecting individuals between 30 and 60 years old [5, 6]. However, the triggering etiology of PAH pathogenesis is likely multifactorial (i.e., including inappropriate angiogenesis, metabolic derangements, DNA damage, and impaired vasoreactivity), idiopathic, heritable, and anorexigenic-induced PAH, which make up around 53% of the diagnosis [7–9].

Despite the remaining challenge to diagnose and treat, findings from recent clinical studies have shifted how PAH is managed—by adding further approved target drugs (yet at a remarkably high cost) to the symptomatic treatment to reduce right ventricle afterload. These practices are now reflected in recently revised clinical practice guidelines (CPG) jointly developed by international societies (e.g., the European Society of Cardiology (ESC)/European Respiratory Society (ERS)) [3, 8, 10]. CPGs are systematically developed statements built upon evidence-based principles and intend to provide trustworthy information with reduced risk of bias to aid practitioners and decision-makers about appropriate care and health policies within a specific clinical scenario [11]. Nevertheless, as their conduct and report may vary widely, appraisal tools have been developed to ensure that a minimum set of methodological quality features, including guidelines' development, context, content, and application, are followed [12–14].

The refined and currently used version of the Appraisal of Guidelines, Research, and Evaluation Collaboration tool (AGREE II), published in 2010, includes 23 quality items within six domains (scope and purpose, stakeholder involvement, rigor of development, clarity of presentation, applicability, and editorial independence) [15–19], while the complementary tool for assessing recommendations' quality and certainty (AGREE REX, Recommendations Excellence, published in 2020) comprises nine items within three domains (clinical applicability, values and preferences, local application and adoption—implementability) [17, 20]. Although these validated approaches are widely used [18], the available literature on systematic evaluations of CPGs using both tools (AGREE II and AGREE REX) is still limited to some clinical conditions [20–24].

Considering the increased number and recently updated guidelines addressing PAH treatment, a more comprehensive evaluation of their quality and generalizability is needed to identify potential weaknesses and strengths that may guide the development and improvement of future documents and practical frameworks for managing this rare disease—including new treatments' prioritization, access, and use in practice. Thus, our goal was to critically appraise the methodological quality of CPG on pharmacological treatments for PAH using a broad systematic review and AGREE II and AGREE REX assessments.

Methods

A systematic review was conducted following the recommendations of Joanna Briggs and the Cochrane Collaboration [25, 26] and reported following the updated PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [27]. The protocol is available at PROSPERO (CRD42023387168) and OSF platform (<https://doi.org/10.17605/OSF.IO/E9DQ3>).

Search Strategy and Selection of Guidelines

A systematic search was performed in PubMed, Cochrane, Embase, and Tripdatabase electronic databases (updated 01/2023) using terms related to “practice guideline” and “pulmonary arterial hypertension” combined with Boolean operators AND and OR. Searches were limited (i.e., by filters) to registers published after 2012 (given the last update of the PAH treatment algorithm, including all available therapeutic approaches) (see complete search strategy in Supplementary Information). A manual search was also performed in the reference lists of the included studies and websites of governmental and non-governmental organizations (Australian Clinical Practice Guidelines, Brazilian Ministry of Health, Canadian Agency for Drugs and Technologies in Health, Chilean Ministry of Health, Colombian Ministry of Health and Social Protection, Guidelines International Network (GIN), Institute for Clinical Systems Improvement, Portal Guía Salud, Scottish Intercollegiate Guidelines Network, and the National Institute for Health and Care Excellence, WHO, and other governmental or non-governmental websites). All other documents (e.g., appendix, development guidance, declaration of conflict of interests) related to the eligible guidelines were manually obtained (when available) from the sources described in the main report.

Studies were included if they met the following eligibility criteria: designed as a CPG including explicit recommendations (i.e., statements from an expert panel based on the best evidence) on the pharmacological treatment (i.e., drugs at any dose, regimen, or combination) of patients diagnosed

with PAH. Only documents available online and authored by professional societies or governmental organizations were considered for eligibility. During the screening (title and abstract reading) and full-text eligibility phases, articles were excluded if they were considered irrelevant to the study goals. Exclusion criteria are as follows: records written in non-Roman characters.

Data Extraction and Guideline Assessment

Retrieved records were imported into Rayyan [28] for duplicate removal, screening, and full-text appraisal. Two authors did all study selection and data extraction steps independently, with a third author to resolve discrepancies. Excel spreadsheets were used to extract guidelines' data on overall information (i.e., organization or governmental parties involved, publication year, geographic location, and evidence grading tools).

The methodological appraisal of the included guidelines (methods, evidence synthesis, and recommendations) was performed independently by four reviewers (APV, FD, FM, YMP), using both AGREE II, 23 quality items, and AGREE REX, a nine-item instrument. Disagreements were resolved by discussion among referees. Both tools were evaluated using a 7-point Likert scale: 1 = strongly disagree (poor or no fulfillment of the criteria) and 7 = strongly agree (all criteria were disclosed entirely). The domains' grades were calculated as described in the users' manual ((obtained score – minimum score possible)/(maximum score – minimum score)), and the overall assessment was calculated through the mean of the respective instrument domains. The final rate was achieved through a consensus of instruments' domain grades between the evaluators. Results were expressed in percentages (standardized scores for each domain range from 0 to 100%).

Data Analysis and Synthesis

Descriptive statistics were used to summarize the data. The Kolmogorov–Smirnov test assessed data normality with additional visual inspection of the QQ plots, confirming that the data was non-normal. For dichotomous variables, absolute and relative frequencies were reported, while continuous variables were expressed in the median, interquartile range (IQR), and minimum–maximum values.

Subgroup differences were explored using the Whitney test for overall and domain scores regarding the type of CPG developer (i.e., professional or governmental institutions) and publication date (categorized as before or after the median of the publication date). The Kruskal–Wallis test was used to compare general scores and domains between continents, following the geographical origin of publications. The differences' sizes were evaluated using the Monte

Carlo method with 95% confidence intervals (95% CI) as effect size measures. Spearman's correlation was used to assess the relationship between the distinct domains from the AGREE II and AGREE REX tools. The intraclass correlation coefficient (ICC) with a two-way mixed effects model was used to verify interrater agreement based on the 95% CI of the ICC estimate (values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively) [29]. Analyses were performed in IBM SPSS Statistics v.28, with a significance level set at 0.05.

In addition to the statistical analysis, an informal analysis was carried out by visual inspection of the graphs, considering recommendations from the PAH guidelines. In this analysis, only recommendations focused on target therapy for PAH were considered, according to the risk classification (low, intermediate, or high risk).

Results

Overall Characteristics of the Guidelines

A total of 5243 registries were obtained after duplicate removal through database searches. During the screening process (title/abstract reading), 5163 records were considered irrelevant to the study. The remaining 80 had their full-text appraised, of which 50 were excluded as they did not meet the eligibility criteria (full exclusion details are shown in Supplementary Information). One document was found by manual search (website), totaling 31 guidelines for synthesis (see Fig. 1).

These 31 guidelines were published between 2012 and 2022, with 2018 being the median year of publication. Most documents are from North America ($n = 14$; 45.2%), followed by Asia ($n = 7$; 22.6%) and Europe ($n = 6$; 19.4%). South America and Africa published three (9.7%) and one (3.2%) guideline. Most guidelines ($n = 28$; 90.3%) were published by professional societies; only three (9.7%) PAH-MS 2014, CHINA 2020, PROSTACYCLIN were developed by governmental institutions (references to the guidelines can be found in Supplementary Information). Of the 31 guidelines, only five—ATS 2014, CHEST 2014, AHA-ATS 2015, CHEST 2019, ESC ERS 2022—(16%) followed the recommendations of GRADE—Grading of Recommendations Assessment, Development and Evaluation—to evaluate the evidence and strength of recommendation, and three guidelines—CCS-CTS 2020, KSC-KATRD 2020, HAT 2021—(10%) followed AGREE for more careful elaboration of the guidelines. Regarding the development of recommendations, ten guidelines (32%) used expert consensus, another ten (32%) carried out systematic reviews, and eleven (36%) were based on other previously published documents. The

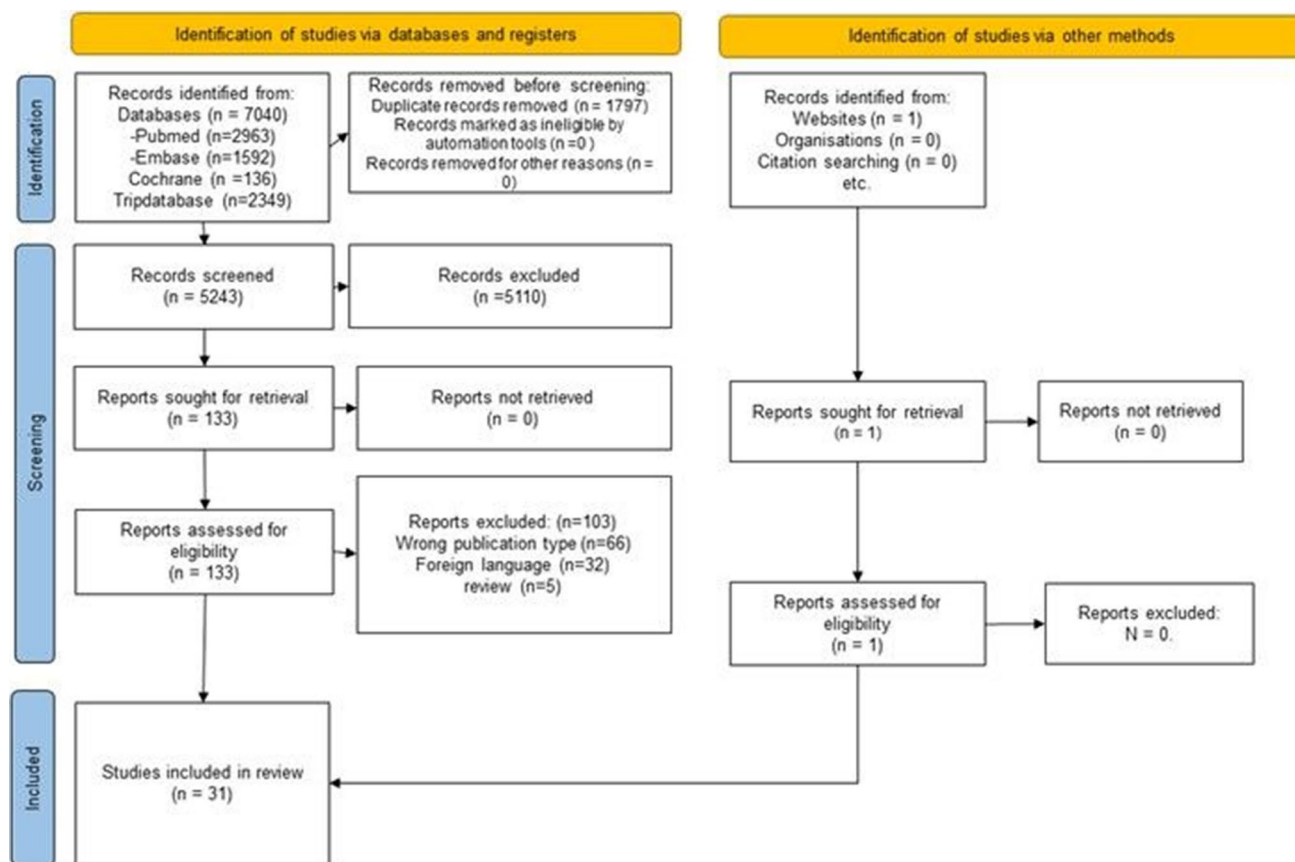


Fig. 1 Systematic review process

disease is associated with high treatment costs, and using resources for the healthcare system is a significant challenge; 29 guidelines (94%) did not report on this issue. The main characteristics of the included CPG are available in Supplementary Information. ICC among the four reviewers for the assessment of AGREE II and AGREE REX was 0.769 (95% CI, 0.722–0.811) and 0.578 (95% CI, 0.480–0.672), showing good and moderate overall interrater agreements, respectively.

AGREE II

The median overall AGREE II score for the 31 CPG was 63 (IQR 58–76), with domains 1, 2, 4, and 6 presenting higher scores. Most guidelines ($n = 20$; 64.5%) obtained a moderate-to-high quality overall score (50–75%), while 3 (9.7%) were judged to have low-moderate quality (25–50%) and 8 (25.8%) high quality (> 75%). No guidelines obtained an overall low-quality score (< 25%) (see Table 1).

In domain 1 (scope and purpose), the median score was 94 (IQR 83–97), with all guidelines scoring above 68%, except PCICS 2014 (a North American pediatric guideline developed by a professional society) with 46%. The Thai

Guideline (HAT) 2021 received the highest score in this domain by meeting 100% of the required criteria. Similarly, high overall scores were obtained in domain 4 (clarity of presentation) (median 99 (IQR 92–100) with almost all documents scoring above 85%) and in domain 2 (stakeholder involvement; median score 63 (IQR 54–90)). Although most guidelines were judged as having moderate-to-high quality in domain 6 (editorial independence) (overall 81 (IQR 54–90)), a critical dispersion exists—with scores ranging from 0% (six guidelines scoring < 25%) to 98%.

Although three guidelines (Chest 2014 and Chest 2019 from North America and the European ESC ERS 2022) were judged as having high quality in domain 3—rigor of development (> 85% scores), around one-third of CPG presented low quality—an overall median of only 32 (IQR 25–51), the SA HEART 2015 guideline from Africa fulfilled only 4% of the criteria of this domain. Equally poor results were found for domain 5 (applicability) with a median score of 35 (IQR 18–52); nine guidelines (29.0%) presented low quality. The European ESC ERS 2022 was the only guideline to obtain scores > 75% in this domain.

Despite these slight differences, no significant changes in CPG scores were observed over the years (Fig. 2).

Table 1 AGREE II and AGREE REX scores of the 31 PAH¹ treatment guidelines

Guideline	AGREE II							AGREE REX			
	AII	AII D1	AII D2	AII D3	AII D4	AII D5	AII D6	AR	AR D1	AR D2	AR D3
ESC ERS 2022	91	96	96	86	100	76	92	71	92	51	71
C.RIOCIGUAT 2022	63	97	65	31	100	30	54	51	58	49	46
C.CARE 2022	65	97	96	33	97	17	50	56	65	50	54
HAT 2021	71	100	57	36	99	49	83	68	79	49	75
C.MANAGEMENT 2021	58	97	51	32	86	10	73	44	53	38	40
SBPT 2020	53	86	65	16	99	03	50	41	83	10	29
CCS-CTS 2020	59	72	63	46	100	54	21	55	79	15	71
CHINA 2020	62	81	68	25	99	06	92	48	68	28	48
KSC-KATRD 2020	76	96	75	49	96	50	90	65	76	47	71
PROSTACYCLIN 2020	70	97	71	70	90	21	71	64	76	44	71
C.TREPROSTINIL 2020	57	83	63	32	89	18	56	60	64	53	63
EPPVDN 2019	81	94	99	61	99	52	83	68	86	44	75
JCS 2019	55	78	79	30	92	34	19	47	78	16	46
6th SYMPOSIUM 2019	61	68	56	21	97	35	90	45	60	43	33
CHEST 2019	89	96	94	96	100	52	94	81	99	71	73
CSC 2019	63	89	43	26	99	33	85	52	71	25	60
CCC 2018	70	90	51	33	92	61	92	42	78	14	35
TSOC 2018	64	97	54	26	99	53	56	46	72	19	46
SEPAR 2018	71	92	68	30	99	50	88	52	76	19	60
EPPVDN 2016	63	80	53	51	90	33	73	39	72	13	31
USA 2016	60	94	60	25	94	32	56	38	56	11	46
AHA-ATS 2015	84	92	99	67	100	61	85	43	76	12	41
ESC ERS 2015	81	99	90	51	100	69	75	60	75	49	56
SA HEART 2015	44	93	56	04	74	16	21	23	39	15	15
PCICS 2014	48	46	47	16	85	08	88	32	56	06	33
SAUDI 2014	67	82	54	27	100	48	88	44	60	35	38
TSOC 2014	50	97	56	24	99	25	0	34	50	19	33
ATS 2014	87	97	99	71	99	65	90	63	92	53	44
CHEST 2014	87	96	92	94	99	42	98	69	92	58	58
PAH-MS 2014	58	97	56	45	100	44	06	60	83	31	67
WHO 2012	53	90	51	19	72	06	81	35	57	10	38

AGREE II domains: (AII) global assessment; (AII D1) domain 1—scope and purpose; (AII D2) domain 2—stakeholder Involvement; (AII D3) domain 3—rigor of development; (AII D4) domain 4—clarity of presentation; (AII D5) domain 5—applicability; (AII D6) domain 6—editorial independence

AGREE REX domains: (AR) global assessment; (AR D1) domain 1—clinical applicability; (AR D2) domain 2—values and preferences; (AR D3) domain 3—implementability

Note: Green boxes refer to high-quality scores (above 75%), blue boxes for moderate-high quality scores (between 50 and 75%), yellow boxes at low-moderate quality indices (between 25 and 50%), and red boxes for low-quality scores (below 25%)

Abbreviations: *AHA-ATS* American Heart Association and American Thoracic Society, *CCC* Cologne Consensus Conference, *EPPVDN* European Pediatric Pulmonary Vascular Disease Network, *ESCERS* European Society of Cardiology and European Respiratory Society, *JCS* Japanese Circulation Society, *PCICS* Pediatric Cardiac Intensive Care Society, *SAUDI* Saudi Arabia, *TSOC* Taiwan Society of Cardiology, *6th Symposium* 6th World Symposium on Pulmonary Hypertension, *ATS* American Thoracic Society, *SBPT* Brazilian Society of Pulmonology and Phthisiology, *CCS-CTS* Canadian Cardiovascular Society-Canadian Thoracic Society, *CHEST* Chest Guideline and Expert Panel Report, *CHINA* Chinese expert-based consensus, *CSC* Sociedad Colombiana de Cardiología, *USA* United States of America, *PAH-MS* Pulmonary Arterial Hypertension—Ministry of Health of Brazil, *HAT* Heart Association of Thailand, *KSC-KATRD* Korean Society of Cardiology and the Korean Academy of Tuberculosis and Respiratory Diseases, *SA HEART* South African Heart Association, *SEPAR* Spanish Society of Pulmonology and Thoracic Surgery, *C.Treprostinil* Consensus Treprostinil, *C. Riociguat* Consensus Riociguat, *C. Management* Consensus Management, *C.Care* Consensus Care, *WHO* World Health Organization

Similar scores were also obtained for guidelines published either by governmental bodies or professional societies (Fig. 3) and by documents from different countries (geographical regions) (see Supplementary Information for additional figures and complete analyses).

AGREE REX

The median overall AGREE REX score for all included CPG was 51 (IQR 42–63), with most guidelines ($n = 15$; 48.4%) presenting moderate-to-high quality, followed by those with

Fig. 2 AGREE II and AGREE REX results according to guidelines' date of publication. AGREE II domains: domain 1—scope and purpose; domain 2—stakeholder involvement; domain 3—rigor of development; domain 4—clarity of presentation; domain 5—applicability; domain 6—editorial independence; AGREE REX domains: domain 1—clinical applicability; domain 2—values and preferences; domain 3—implementability

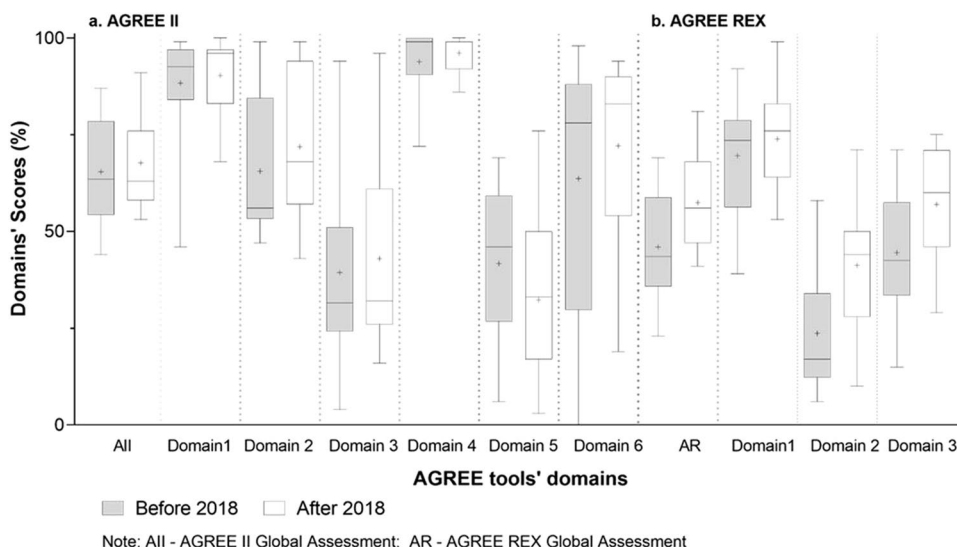
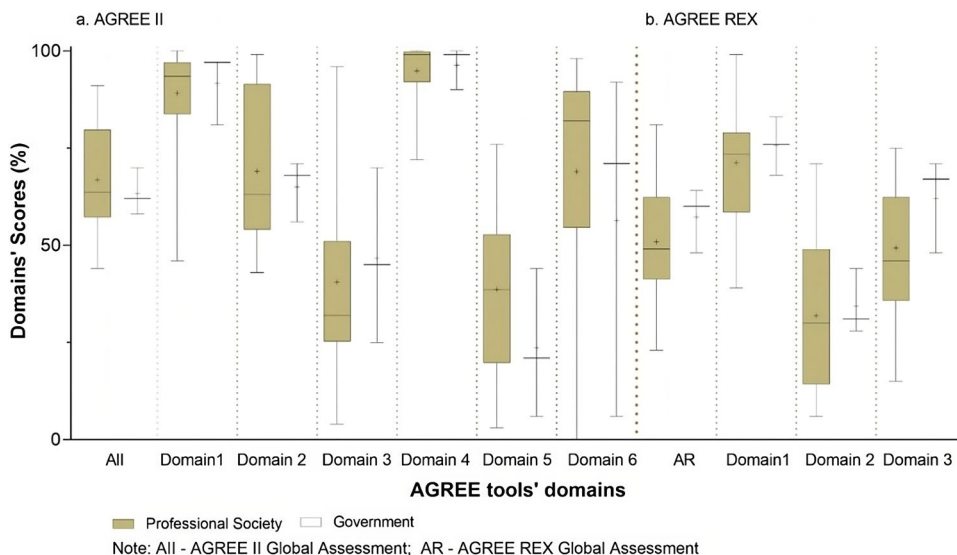


Fig. 3 AGREE II and AGREE REX results according to guidelines' developers. AGREE II domains: domain 1—scope and purpose; domain 2—stakeholder involvement; domain 3—rigor of development; domain 4—clarity of presentation; domain 5—applicability; domain 6—editorial independence; AGREE REX domains: domain 1—clinical applicability; domain 2—values and preferences; domain 3—implementability



low-moderate quality ($n = 14$; 42.2%). Low-quality and high-quality scores were found for one guideline each. Nonetheless, guidelines scored better only in domain 1 (clinical applicability) with a median result of 75.0 (IQR 60–79) and no CPG with low quality in this domain. Median scores in domains 2 (values and preferences) and 3 were 31.0 (IQR 15–49) and 46.0 (IQR 38–67), respectively, with 6 and 14 guidelines (mainly from North America and Europe) achieving scores between 50 and 75% in these domains, respectively (see Table 1).

Yet, significant improvements over time on the overall AGREE REX results were observed in the association analyses (see Fig. 2): publications before 2018 scoring 43.5 (IQR 35.8–58.8) and those after 2018 scoring 56.0 (IQR 47.0–68.0) (Mann–Whitney $U = 59$; $p = 0.016$). Similar trends were obtained for both individual domains 2 (values

and preferences) and 3 (implementability) ($U = 58$; $p = 0.014$ and $U = 65$; $p = 0.029$, respectively) of this tool. No other significant differences according to guidelines' developers (governmental or society; Fig. 3) nor geographical origin were observed (see Supplementary Information for additional figures and complete analyses).

Finally, the correlation matrix (see Supplementary Information) confirmed that similar patterns between AGREE II and AGREE REX exist, with a high correlation (Spearman $\rho = 0.722$) between the overall scores of the two instruments. All individual domains have a moderate-to-high correlation with their respective tools.

No relationship between the quality of guidelines (AGREE II or AGREE REX) and the recommendations, considering the options between monotherapy and combination therapy (Fig. 4) or pharmacological classes

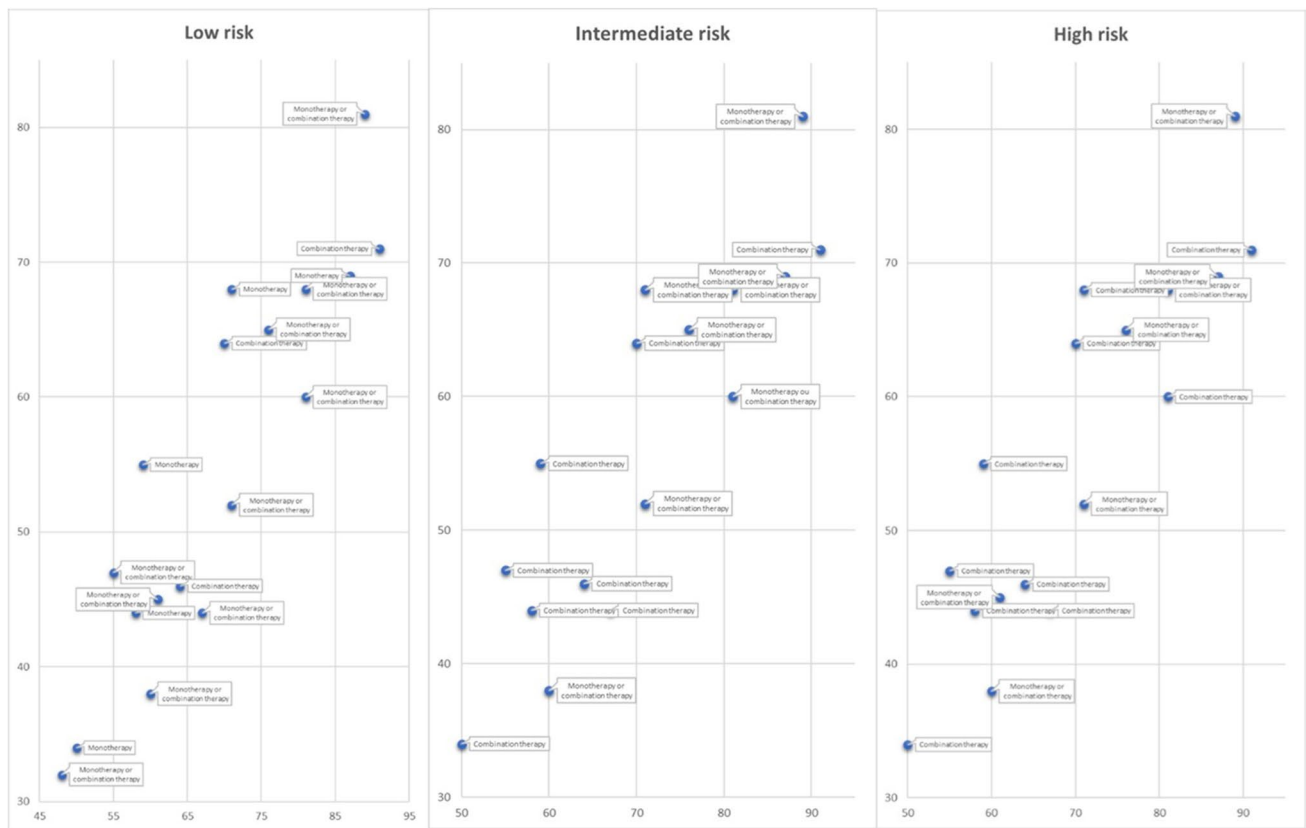


Fig. 4 Summary of recommendations from PAH treatment guidelines according to AGREE II and AGREE REX scores. Only recommendations that considered the WHO functional class for risk categorization were considered. Guidelines that did not formulate

(Supplementary Information), was found in the PAH target therapy recommendations analysis.

Discussion

This study systematically and critically appraised the methodological quality of 31 CPG on PAH treatments published by four continents using AGREE II and AGREE REX tools. The results of these instruments correlate to the guidelines' credibility and the validity of their recommendations. While adherence to clinical guidelines is paramount for healthcare providers and patients to improve the quality of care and achieve positive outcomes, periodic assessment of guidelines' quality is also important, including to address situations where treatment acquisition is restricted and associated with high resource utilization and financial strain—as is often the case in PAH management (e.g., high cost of new pharmacological treatments, challenges with prior authorization, tiered use) [4, 5, 10]. Yet, as moderate-to-good agreement rates among reviewers were obtained when using the

recommendations according to risk categorization (not applicable) or that did not make any pharmacological recommendations explicit (not reported) were excluded from this analysis (see Supplementary Information)

AGREE tools in our study, findings can be considered reliable and consistent in similar scenarios [17, 30, 31].

Overall, we found higher scores for AGREE II's domains 1 (scope and purpose) and 4 (clarity of presentation) and for AGREE REX's domain 1 (clinical applicability), which is similar to the findings reported by previous studies for other clinical conditions, including nutrition care, mental disorders, and medical oncology [21, 32–36]. These results mean a CPG document written in comprehensive language with a clear structure and presentation of goals regarding specific health issues and target populations. Guidelines written in concise, unambiguous language appear to be the most accessed by clinicians and are more likely to be implemented into clinical practice [33]. Conversely, several methodological issues were identified in domains 3 (rigor of development) and 5 (applicability) of AGREE II, and domains 2 (values and preferences) and 3 (implementability) of AGREE REX as also previously highlighted in the biomedical literature, including in areas of intensive care units, skin disorders, and lipid management for patients with coronary heart disease [21, 24, 32, 34, 35]. Insufficient scientific rigor for developing guidelines, lack of standard procedures

for evidence synthesis and recommendations' formulation, scarce information on the implementation process, and failure to consider patients' and other target users' opinions may negatively affect the findings of these documents. This is concerning, especially for rare diseases (such as PAH) with deficient evidence. Skewed guidelines towards researchers' or health professionals' values and preferences may lead to less meaningful patient guidelines and a low CPG implementation rate, contributing to knowledge waste in various clinical settings (results are inaccessible or unused) [11, 37].

Although the information on "stakeholder involvement" (domain 2 AGREE II) and "editorial independence" (domain 6 AGREE II) was fairly provided by most PAH guidelines, important pitfalls were detected on AGREE REX domains 2 and 3 (values and preferences and implementability). These findings are consistent with previous research in actinic keratosis, pressure injury, and uveal melanoma, highlighting the need to improve the proximity of clinical practice to patients' values and characteristics, which should not be restricted to populations' characteristics (e.g., sex, gender, race, genetic profile), and consider behavioral and social aspects and treatment costs. The involvement of other stakeholders, such as health policymakers and government representatives, is paramount for the continuous development and practical implementation of a guideline [24, 38, 39]. For this to occur, communication between CPG's working parties, including a wide range of representatives (e.g., professional organizations, regional and local offices, and relevant national bodies), is essential. Nonetheless, there may be several barriers hampering involvement, including the identification and recruitment of appropriate stakeholders, the definition of roles and responsibilities, and disclaiming the potential conflicts of interest and resources, which should be clearly stated in the guideline (either on the main document or supplementary materials) for data transparency. Moreover, if applicability issues, including types of facilitators and barriers surrounding guideline implementation, are not adequately addressed, CPG may be ineffective in clinical practice and increase the waste of resources. O'Connor et al. [40] reported that less than 20% of the published guidelines on antenatal management of twin pregnancies were considered in implementing recommendations. The costs and budget concerns of applying the findings are equally poorly communicated by guidelines in PAH or other areas (including mental health, viral infections, and oncology), limiting the extent to which health systems can disseminate and use the evidence [41, 42].

Although there has been a significant increase in the number of CPG publications in the past years, including for PAH [43, 44], which is probably due to the approval of new treatments and the intention to improve the quality of care for patients while informing healthcare costs and reducing variability in clinical practice [43], their overall methodological

quality (AGREE II scores) has not improved, reinforcing the need to strictly develop and report these documents to avoid suboptimal decision-making processes [30, 31, 44]. From a global perspective, although some countries have initiated national protocols regarding developing, appraising, and implementing high-quality CPG, standardized approaches for their production are still lacking [43]. This may partially justify the heterogeneity among PAH guidelines regarding the evidence-based methodological approaches used to build the reports. Although a panel of experts and GRADE collaboration tools are widely recommended strategies, some guidelines use alternative approaches, which may be the cause of low results due to poor external validity and reproducibility of findings. Conversely, we observed a slight improvement in the quality of recommendations reported by CPG published after 2018, which may be due to greater use of the AGREE and GRADE instruments as a form of critical assessment of evidence and more careful elaboration of clinical guidelines.

Like Seow et al. [45] analysis of insomnia guidelines, we reported that the most productive regions of PAH guidelines were North America, Asia, and Europe, following the global geographical distribution of researchers, available technologies, science funding, and international collaboration [46]. However, although documents from Europe and some from North America scored slightly better in the AGREE II and AGREE REX tools, no significant differences were found in the overall quality of these guidelines. Likewise, guidelines' scores did not differ according to the type of developer (society or government), which is consistent with the findings by De Melo et al. [41], who reported higher quality documents from professional societies and government institutions compared to independent researchers/universities. This may occur due to the different financial and human resources available in these organizations, especially considering CPG development is a time-consuming and costly process, as well as their scope and purposes.

Regarding the apparent lack of relationship between recommendations on PAH target therapy and AGREE II and AGREE REX quality, it is important to consider that there is no conclusive evidence comparing monotherapy and combination therapy or classes and drugs within classes. Therefore, the expert's opinion is decisive on this matter, and thus, the influence of one guideline on another is notable, regardless of its quality.

Overall, future studies should prioritize methodological rigor (e.g., by means of systematic evaluations of existing guidelines to highlight areas where evidence is scarce or of poor quality), stakeholders' engagement (e.g., qualitative research with peers, patients, caregivers, healthcare providers, and policymakers on their values and preferences can provide valuable insights regarding the real-world needs and challenges in PAH management), and clinical applicability

and implementability (e.g., quality improvement initiatives including assessing barriers and facilitators aimed at testing the cost-effectiveness, feasibility, and adherence of treatment recommendations in clinical settings) [47, 48]. Moreover, although creating and updating high-quality guidelines every 2–5 years, as recommended, may be a costly process, especially for a small national professional society or governmental body, further strategies, including international consortiums, gathering external funding, and adaptations of high-quality published documents, are considered efficient. The use of evidence-to-decision frameworks can also provide better documents with more credible and transparent recommendations [1, 48].

Our review has some limitations. We only included freely available guidelines published after 2012, the year of the last update of the PAH treatment algorithm, which addressed pharmacological treatments; thus, results and conclusions cannot be generalized to other conditions. Nonetheless, a systematic and replicable methodology was performed following international recommendations and using four databases. Given the high heterogeneity among guidelines (i.e., year of publication, countries, target populations, and different methods to assess the level/certainty of evidence and propose recommendations), comparisons between guideline content should be avoided [27]. AGREE tools rely on inherent subjectivity when scoring their domains and may not fully account for the level of evidence attributed to recommendations, which can be subject to debate (e.g., perceptions of the tools and their measurement properties). Nonetheless, these instruments provide important frameworks to constantly evaluate and critically assess clinical practice guidelines. Although overall agreement across reviewers was considered substantial in our study, the evaluation of some items was arbitrary, leading to low consensus rates; however, this may reflect the opinions of a multidisciplinary team with different clinical experiences.

Conclusions

Although some slight improvements in the report of recommendations from CPG on PAH pharmacological treatment occurred in the past years, methodological weaknesses are still common in the documents addressing this rare disease, especially regarding scientific rigor, values and preferences, and implementability. Thus, particular attention should be given to stakeholders' perspectives and clinical applicability for developing future guidelines in this field.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10557-024-07605-w>.

Author Contribution All authors contributed to the conception and design of the study. RP, FST, and APOV planned and designed the

study. FST and APOV developed search strategies. FD and APOV selected potential studies and extracted data from the included studies. APOV, FD, FFM, and YMSP assessed the quality of the guideline. FST and RCL performed the statistical analyses. FST and RCL conducted the arbitration under disagreement. FST, RP, RCL, and APOV interpret the results. APOV and FST wrote the manuscript with editing and reviewing contributions from all authors.

Funding Open access funding provided by FCTIFCCN (b-on). This study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability All data generated or analyzed during this study are included in this published article (and its supplementary information files). Further information are available from the corresponding author on reasonable request.

Code Availability Not applicable.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Maron BA. Revised definition of pulmonary hypertension and approach to management: a clinical primer. *J Am Heart Assoc.* 2023;12(8):e029024. <https://doi.org/10.1161/JAHA.122.029024>.
2. Simonneau G, Montani D, Celermajer DS, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J.* 2019;53(1):1801913. <https://doi.org/10.1183/13993003.01913-2018>.
3. Humbert M, Kovacs G, Hoeper MM, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J.* 2022;43(38):3618–731. <https://doi.org/10.1093/eurh/rtj/ehac237>.
4. Cullivan S, Higgins M, Gaine S. Diagnosis and management of pulmonary arterial hypertension. *Breathe (Sheff).* 2022;18(4):220168. <https://doi.org/10.1183/20734735.0168-2022>.

5. Maron BA, Galiè N. Diagnosis, treatment, and clinical management of pulmonary arterial hypertension in the Contemporary Era: a review. *JAMA Cardiol.* 2016;1(9):1056–65. <https://doi.org/10.1001/jamacardio.2016.4471>.
6. Leber L, Beaudet A, Muller A. Epidemiology of pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension: identification of the most accurate estimates from a systematic literature review. *Pulm Circ.* 2021;11(1):2045894020977300. <https://doi.org/10.1177/2045894020977300>.
7. Ranchoux B, Harvey LD, Ayon RJ, et al. Endothelial dysfunction in pulmonary arterial hypertension: an evolving landscape (2017 Grover Conference Series). *Pulm Circ.* 2018;8(1):2045893217752912. <https://doi.org/10.1177/2045893217752912>.
8. Maron BA, Abman SH, Elliott CG, et al. Pulmonary arterial hypertension: diagnosis, treatment, and novel advances. *Am J Respir Crit Care Med.* 2021;203(12):1472–87. <https://doi.org/10.1164/rccm.202012-4317SO>.
9. Bourgeois A, Omura J, Habbout K, Bonnet S, Boucherat O. Pulmonary arterial hypertension: new pathophysiological insights and emerging therapeutic targets. *Int J Biochem Cell Biol.* 2018;104:9–13. <https://doi.org/10.1016/j.biocel.2018.08.015>.
10. Humbert M, Sitbon O, Guignabert C, et al. Treatment of pulmonary arterial hypertension: recent progress and a look to the future. *Lancet Respir Med.* 2023;11(9):804–19. [https://doi.org/10.1016/s2213-2600\(23\)00264-3](https://doi.org/10.1016/s2213-2600(23)00264-3).
11. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice G. Clinical practice guidelines we can trust. 2011; <https://doi.org/10.17226/13058>.
12. Shiffman RN, Shekelle P, Overhage JM, et al. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. *Ann Intern Med.* 2003;139(6):493–8. <https://doi.org/10.7326/0003-4819-139-6-200309160-00013>.
13. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med.* 2010;7(2):e1000217. <https://doi.org/10.1371/journal.pmed.1000217>.
14. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet (London, England).* 2000;355(9198):103–6. [https://doi.org/10.1016/s0140-6736\(99\)02171-6](https://doi.org/10.1016/s0140-6736(99)02171-6).
15. Brouwers MC, Kho ME, Browman GP, et al. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. *CMAJ Can Med Assoc J J Assoc Med Can.* 2010;182(10):1045–52. <https://doi.org/10.1503/cmaj.091714>.
16. 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 Can Med Assoc J J Assoc Med Can.* 2010;182(10):E472–8. <https://doi.org/10.1503/cmaj.091716>.
17. Brouwers MC, Spithoff K, Kerkvliet K, et al. Development and validation of a tool to assess the quality of clinical practice guideline recommendations. *JAMA Netw Open.* 2020;3(5):e205535. <https://doi.org/10.1001/jamanetworkopen.2020.5535>.
18. Brouwers MC, Spithoff K, Lavis J, et al. What to do with all the AGREEs? The AGREE portfolio of tools to support the guideline enterprise. *J Clin Epidemiol.* 2020;125:191–7. <https://doi.org/10.1016/j.jclinepi.2020.05.025>.
19. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality & safety in health care.* 2003;12(1):18–23. <https://doi.org/10.1136/qhc.12.1.18>.
20. Florez ID, Brouwers MC, Kerkvliet K, 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 IS.* 2020;15(1):79. <https://doi.org/10.1186/s13012-020-01036-5>.
21. Niu M, Gao Y, Yang M, et al. The quality and clinical applicability of recommendations in anxiety disorders guidelines: a systematic review of seventeen guidelines from seven countries. *J Affect Disord.* 2021;295:1301–9. <https://doi.org/10.1016/j.jad.2021.08.103>.
22. Rethnam V, Hayward KS, Bernhardt J, Churilov L. Early mobilization after stroke: do clinical practice guidelines support clinicians' decision-making? *Front Neurol.* 2021;12:606525. <https://doi.org/10.3389/fneur.2021.606525>.
23. Steeb T, Wessely A, Drexler K et al. The quality of practice guidelines for melanoma: a methodologic appraisal with the AGREE II and AGREE-REX instruments. *Cancers.* 2020;12(6). <https://doi.org/10.3390/cancers12061613>.
24. Wessely A, Steeb T, Heppt F, et al. A critical appraisal of evidence- and consensus-based guidelines for actinic keratosis. *Curr Oncol (Toronto, Ont).* 2021;28(1):950–60. <https://doi.org/10.3390/currenconcol28010093>.
25. Aromataris E M. JBI Manual for Evidence Synthesis. JBI database of systematic reviews and implementation reports. 2020. <https://doi.org/10.46658/jbimes-20-01>.
26. Higgins JT J, Chandler, et al. *Cochrane handbook for systematic reviews of interventions.* Chichester: John Wiley and Sons. 2019;2. <https://doi.org/10.1002/9781119536604>.
27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clin Res ed).* 2021;372:n71. <https://doi.org/10.1136/bmj.n71>.
28. 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>.
29. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–63. <https://doi.org/10.1016/j.jcm.2016.02.012>.
30. Guerra-Farfan E, Garcia-Sanchez Y, Jornet-Gibert M, et al. Clinical practice guidelines: the good, the bad, and the ugly. *Injury.* 2023;54(Suppl 3):S26–9. <https://doi.org/10.1016/j.injury.2022.01.047>.
31. Schünemann HJ, Wiercioch W, Etzeandía I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ Can Med Assoc J J Assoc Med Can.* 2014;186(3):E123–42. <https://doi.org/10.1503/cmaj.131237>.
32. Noyahr JK, Tatuco-Babet OA, Chapple LS et al. Methodological rigor and transparency in clinical practice guidelines for nutrition care in critically ill adults: a systematic review using the AGREE II and AGREE-REX tools. *Nutrients.* 2022;14(13). <https://doi.org/10.3390/nu14132603>.
33. Santero M, de Mas J, Rifà B, et al. Assessing the methodological strengths and limitations of the Spanish Society of Medical Oncology (SEOM) guidelines: a critical appraisal using AGREE II and AGREE-REX tool. *Clin Trans Oncol Off Publ Fed Span Oncol Soc Natl Cancer Inst Mexico.* 2023. <https://doi.org/10.1007/s12094-023-03219-0>.
34. Cui N, Zhang Y, Li Q, et al. Quality appraisal of guidelines on physical restraints in intensive care units: a systematic review. *Intensive Crit Care Nurs.* 2022;70:103193. <https://doi.org/10.1016/j.iccn.2021.103193>.
35. Zhou H, Zhang S, Sun X, et al. Lipid management for coronary heart disease patients: an appraisal of updated international guidelines applying Appraisal of Guidelines for Research and Evaluation II-clinical practice guideline appraisal for lipid management in coronary heart disease. *J Thorac Dis.* 2019;11(8):3534–46. <https://doi.org/10.21037/jtd.2019.07.71>.

36. Santos RD. III Brazilian Guidelines on Dyslipidemias and Guideline of Atherosclerosis Prevention from Atherosclerosis Department of Sociedade Brasileira de Cardiologia. *Arq Bras Cardiol.* 2001;77(Suppl 3):1–48. <https://doi.org/10.1590/S0066-782X2001001500001>.
37. Frank L, Basch E, Selby JV. The PCORI perspective on patient-centered outcomes research. *JAMA.* 2014;312(15):1513–4. <https://doi.org/10.1001/jama.2014.11100>.
38. Gillespie BM, Latimer S, Walker RM, et al. The quality and clinical applicability of recommendations in pressure injury guidelines: a systematic review of clinical practice guidelines. *Int J Nurs Stud.* 2021;115:103857. <https://doi.org/10.1016/j.ijnurstu.2020.103857>.
39. Steeb T, Hayani KM, Forster P, et al. Guidelines for uveal melanoma: a critical appraisal of systematically identified guidelines using the AGREE II and AGREE-REX instrument. *J Cancer Res Clin Oncol.* 2020;146(4):1079–88. <https://doi.org/10.1007/s00432-020-03141-w>.
40. O'Connor C, O'Connor E, Leitao S, Barrett S, O'Donoghue K. Clinical practice guidelines for the antenatal management of dichorionic diamniotic twin pregnancies: a systematic review. *BMC Pregnancy Childbirth.* 2023;23(1):347. <https://doi.org/10.1186/s12884-023-05652-z>.
41. Gabriel FC, De Melo DO, Stein AT et al. Factors associated with higher quality of clinical practice guidelines and their recommendations for the pharmacological treatment of depression: a systematic review. *BMJ Open.* 2023;13(3). <https://doi.org/10.1136/bmjopen-2022-067390>.
42. Zheng Q, Gao Y, Xiong L, et al. Chinese herbal medicine and COVID-19: quality evaluation of clinical guidelines and expert consensus and analysis of key recommendations. *Acupunct Herb Med.* 2022;2(3):152–61. <https://doi.org/10.1097/hm9.0000000000000043>.
43. De Leo A, Bloxsome D, Bayes S. Approaches to clinical guideline development in healthcare: a scoping review and document analysis. *BMC Health Serv Res.* 2023;23(1):37. <https://doi.org/10.1186/s12913-022-08975-3>.
44. Kretser A, Murphy D, Bertuzzi S, et al. Scientific integrity principles and best practices: recommendations from a Scientific Integrity Consortium. *Sci Eng Ethics.* 2019;25(2):327–55. <https://doi.org/10.1007/s11948-019-00094-3>.
45. Seow SY, Kwok KFV, Tay KH, et al. Systematic review of clinical practice guidelines for insomnia disorder. *J Psychiatr Pract.* 2022;28(6):465–77. <https://doi.org/10.1097/prs.0000000000000670>.
46. Tebala GD. What is the future of biomedical research? *Med Hypotheses.* 2015;85(4):488–90. <https://doi.org/10.1016/j.mehy.2015.07.003>.
47. Mc Allister M, Florez ID, Stoker S, McCaul M. Advancing guideline quality through country-wide and regional quality assessment of CPGs using AGREE: a scoping review. *BMC Med Res Methodol.* 2023;23(1):283. <https://doi.org/10.1186/s12874-023-02101-5>.
48. Meneses-Echavez JF, Bidonde J, Montesinos-Guevara C, et al. Using evidence to decision frameworks led to guidelines of better quality and more credible and transparent recommendations. *J Clin Epidemiol.* 2023;162:38–46. <https://doi.org/10.1016/j.jclinepi.2023.07.013>.

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