Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

BJA

British Journal of Anaesthesia, xxx (xxx): xxx (xxxx)

doi: 10.1016/j.bja.2022.09.008
Advance Access Publication Date: xxx
Review Article

REVIEW ARTICLE

Evaluation of the quality of COVID-19 guidance documents in anaesthesia using the Appraisal of Guidelines for Research and Evaluation II instrument

Sinead M. O'Shaughnessy[1,*], Arnaldo Dimagli[2], Bessie Kachulis[1], Mohamed Rahouma[2], Michelle Demetres[3], Nicolas Govea[1] and Lisa Q. Rong[1]

[1]Department of Anesthesiology, New York, NY, USA, [2]Department of Cardiothoracic Surgery, New York, NY, USA and [3]Information, Technology and Services, Weill Cornell Medicine, New York, NY, USA

*Corresponding author. E-mail: oshaugs@gmail.com

Abstract

Background: Guidance documents are a valuable resource to clinicians to guide evidenced-based decision making. The quality of guidelines in anaesthesia and across other specialties has been demonstrated to be poor. COVID-19 presented an urgent need for immediate guidance for anaesthetists as frontline clinicians. The aim of this study was to evaluate the quality of COVID-19 guidance documents using the internationally validated Appraisal of Guidelines for Research & Evaluation (AGREE) II tool.

Methods: A search was conducted in Ovid EMBASE and Ovid MEDLINE to identify all COVID-19 anaesthesia guidance documents from 2020-2021. Thirty-eight guidance documents were selected for analysis by 4 independent appraisers using the AGREE II instrument, across its 6 domains and 23 items. A scoring threshold for high quality was agreed by the working group via consensus.

Results: Overall, the body of COVID-19 guidance documents achieved poor scores using AGREE II. Only 5% of documents met the high-quality criteria. Markers of quality included international and multi-institutional collaboration. Document title ('guideline' vs 'consensus statement'/ 'recommendations') did not yield any differences in domain scores and overall quality ratings. Compared with recent general anaesthesia guidelines, COVID-19 guidelines performed significantly worse.

Conclusions: COVID-19 guidance documents published during the first two years of the pandemic lacked rigour and appropriate quality. This raises concern about their trustworthiness for use in clinical practice. Enhanced systems are required to ensure the integrity of rapidly formulated guidance.

Keywords: anaesthesia; consensus statements; guidelines; COVID-19; recommendations

Editor's key points

- COVID-19 guidance documents have been demonstrated to be of poor quality in diverse clinical settings.
- No large study had been performed to date in anaesthesia.
- The authors examined such documents published in the first 2 yr of the COVID-19 pandemic in the field of anaesthesia using a standardised tool. These guidance documents were found to be of low quality.
- Markers of quality included appraisal tool use and multi-institutional and international collaboration. Guidance documents were labelled incorrectly.
- Living guidelines, a prospective register, and journal input have potential to improve the quality and accuracy of COVID-19 guidance.

Received: 1 July 2022; **Accepted:** 3 September 2022

A vast array of guidance documents are available to assist evidence-based decision-making in clinical practice in anaesthesia.[1] Clinical practice guidelines (CPGs) are increasingly important to anaesthetists, as they navigate progressively complex patient populations in an era characterised by the exponential evolution and dissemination of scientific information.[2] The COVID-19 pandemic presented an urgent need for immediate guidance especially for anaesthetists as key frontline providers.[3,4] A CPG is defined as a 'statement that includes recommendations intended to optimise patient care that are informed by systematic review of evidence and as assessment of benefits and harms of alternative care options'.[5] Typically, CPGs are regarded as more robust than consensus statements, which are based on expert opinion in the absence of comprehensive evidence.[6] Other guidance documents that are less well-defined include position statements and recommendations. Despite the clear differences between these documents, these terms are often used interchangeably, which can be confusing for clinicians and potentially translate into use of less evidence-based sources in patient care.[7]

The Appraisal of Guidelines for Research and Evaluation II (AGREE II) is the gold-standard appraisal tool exclusively designed for clinical guidance documents, which has been validated across multiple specialties.[8–12] It consists of six domains, with Domain 3 (Rigour of Development) widely regarded as the most important for guideline quality (Table 1).[13–15] Despite the availability of AGREE II and other useful tools, concerns regarding the quality and integrity of guidelines in anaesthesia have existed for some time.[16] This has been amplified during the COVID-19 era because of lack of an existing evidence base, purported absence of rigour in some guideline development strategies, use of non-standard document formats, and use of confusing terminology.[17,18] Several publications have detailed the quality of anaesthesia guidelines and the quality of COVID-19 recommendations in other fields of medicine.[15,17,19] One small anaesthesia study offered an early indication as to the standard of COVID-19 guidance.[20] The current study is the largest to date to evaluate the quality of COVID-19 guidance documents in anaesthesia.

The main aim of this study was to critically evaluate the quality and methodology of COVID-19 guidance documents using the AGREE II instrument, considered the gold-standard appraisal tool. Documents labelled as guidelines, consensus statements, and those providing recommendations predominantly aimed at anaesthesia practitioners were assessed with this tool. Additional objectives included (i) the identification of characteristics linked to quality within these documents, (ii) a comparison of the transparency and methodological rigour of COVID-19 guidelines with consensus statements and recommendations, and (iii) a comparison between the methodological quality of COVID-19 guidelines and recent general anaesthesia guidelines.

The authors hypothesised that the methodological quality of the COVID-19 guidance documents in anaesthesia would be poor and significantly worse than recently published anaesthesia guidelines. This was based on knowledge of an existing poor baseline quality of anaesthesia guidelines and an acknowledgement of the unique challenges for guidance

Table 1 Appraisal of Guidelines for Research and Evaluation II: domain number, domain title, and specific domain items.

Domain number	Domain title	Domain items
1	Scope and Purpose	Overall objective specifically described
		Health questions specifically described
		Population in question specifically described
2	Stakeholder Involvement	Development group includes all relevant professional groups
		Views and preferences of target population sought
		Target users clearly defined
3	Rigour of Development	Systematic methods used to search
		Criteria for selecting the evidence clearly described
		Strengths and limitations of the body of evidence are clearly described
		Methods for formulating the evidence are clearly described
		Health benefits, side-effects, and risks have been considered in formulating the recommendations
		Explicit link between the recommendations and supporting evidence
		External review by experts before publication
		Procedure for updating the guideline is provided
4	Clarity of Presentation	Recommendations are specific and unambiguous
		Different options for management of the health condition are clearly presented
		Key recommendations are easily identifiable
5	Applicability	Guideline describes facilitators and barriers to application
		Guideline provides advice and tools to put recommendations into practice, or both
		Potential resource implications of applying the recommendations have been considered
		Guideline provides monitoring, monitoring criteria, or both
6	Editorial Independence	Views of the funding body have not influenced the content of the guideline
		Competing interests of guideline development group members have been recorded and addressed

synthesis posed during the pandemic. The authors also proposed that documents entitled 'guidelines' may have been defined inaccurately, and thus would not be superior in quality to those labelled as 'consensus statements' and 'recommendations'. Possible markers of quality were identified at the outset as use of a recognised appraisal tool, professional society involvement, and a high number of institutions and countries involved in development.

Methods

A comprehensive literature search was performed by a medical librarian (MD) to identify COVID-19 guidance documents for anaesthetists. Two major scientific databases were searched in June 2021: Ovid MEDLINE (all; 1946 to present) and Ovid EMBASE (1974 to present). In addition, the World Federation of Societies of Anesthesiologists' COVID-19 guidance page was evaluated.[21] The search strategy included all the appropriate controlled vocabulary and keywords for the concepts of 'anaesthesia' and 'guidelines' and limited to COVID-19 literature using Ovid Expert Searches COVID-19 filter. The full search strategies for all databases are available in the Supplementary appendix. To limit publication bias, there were no language restrictions at the time of the initial search strategy (n=352).

Inclusion criteria of articles selected from the initial cohort for analysis were as follows (n=38): (i) written in the English language; (ii) relating to all anaesthesia perioperative subspecialties, including chronic pain; and (iii) documents labelled as guidelines, consensus statements, or those providing recommendations aimed at anaesthetists for COVID-19 care. There was no limitation by journal type or impact factor, type of working group involved, or document methodology used. Excluded papers were those relating to critically ill patients with COVID-19 and COVID-19 care outside of the perioperative period. Articles including recommendations outside of anaesthesia and review articles providing no formal recommendations were also excluded. The perioperative period was defined as 24 h before and 24 h after a surgical procedure.

Thirty-eight guidance documents were selected for further analysis after completion of the search strategy and application of the inclusion/exclusion criteria. This was performed by four independent appraisers (SMOS, NG, MD, and BK) using the AGREE II instrument. Three of the appraisers had considerable experience in guideline synthesis and assessment, and they comprised two consultant anaesthetists (BK and SMOS) and a medical librarian (MD). NG, an anaesthesia resident, utilised the AGREE II manual, designed to orient and instruct the novice user. All participants had previously published in the area of guideline appraisal using AGREE II.

The included articles were each assessed across the six domains and 23 items using a 7-point Likert scale: Domain 1 (Scope and Purpose; Items 1–3), Domain 2 (Stakeholder Involvement; Items 4–6), Domain 3 (Rigour of Development; Items 7–14), Domain 4 (Clarity of Presentation; Items 15–17), Domain 5 (Applicability; Items 18–21), and Domain 6 (Editorial Independence; Items 22–23). In addition to the six domains, an overall rating (1–7) and an overall recommendation ('yes', 'yes with modifications', and 'no') were performed by each appraiser. After all 38 guidelines were assessed, a review was performed to identify any discrepancies in scores of 3 or greater between appraisers. These items were revisited, and a consensus was reached between all appraisers before proceeding to analysis. A senior author, who was not involved in appraisals, was available for consultation to resolve any persistent discrepancies in scoring.

As described in the AGREE II user manual, the individual domain scores were calculated as a scaled domain score according to the following formula: (score obtained−minimum possible score)/(maximum possible score−minimum possible score) × 100.[22] Mean domain scores (with standard deviations [SDs]) were calculated for each domain. The intra-class correlation coefficient, based on a 95% confidence interval, was used to assess inter-appraiser agreement as follows: excellent at >0.9, good at 0.75–0.89, moderate at 0.5–0.74, and poor at <0.5. Two quality scores were calculated: a subjective quality score and an objective quality score. The subjective quality score/mean recommendation for each guideline was calculated based on the overall rating and recommendation. An average subjective rating of 6 or 7 was recommended as 'yes', a rating of 4 or 5 as 'yes with modifications', and 3 or below as 'no'. The objective quality score remains undefined by AGREE II and is designed by each appraiser group according to their own unique context. There was little precedent in literature relating to AGREE II quality scores for COVID-19 guidelines; therefore, the authors agreed the quality threshold based on previous experience examining anaesthesia guidelines with this instrument and typical practice in the literature. A document with a score of >50% for Domains 1, 3, 4, and 6 was deemed high quality. Finally, documents entitled 'guidelines' (n=9) were compared with a data set of guidelines in anaesthesia from 2016 to 2020 (n=51). The study used for comparison of quality previously assessed 51 anaesthesia guidelines, across all subspecialties, published in the top 10 anaesthesia journals over a 5 yr period.[15] For this data set, guidelines were defined as 'statements that include recommendations intended to optimise patient care that are informed by systematic review of evidence and an assessment of benefits and harms of alternate care options'.[5] Both studies used similar methodologies facilitating direct comparison.

Categorical data were presented as frequency count and percentage and compared between groups using Fisher's exact test or χ^2 test, as appropriate. After testing for normal distribution, continuous data were presented as mean (SD) and median (inter-quartile range [IQR]), and the Kruskal–Wallis rank-sum test was applied for comparison. Individual domain performance, overall rating, and subjective and objective quality scores over four time periods (January–June 2020, July–December 2020, January–June 2021, and July–December 2021) were assessed *via* linear regression. The impact of several variables on domain performance, overall rating, and quality scores was also examined, including use of an appraisal tool *vs* no appraisal tool, society *vs* no society involvement, single *vs* multiple institutions, and single *vs* multiple countries. Documents were divided according to the title of 'guidelines' (n=9) *vs* 'consensus statements'/'recommendations' (n=28) according to the same parameters. Alpha level was set at 0.05 with P-value of <0.05 indicative of statistical significance. All analyses were done using R (version 4.0.5) within RStudio (RStudio Inc., Boston, MA, USA).

Results

Thirty-eight articles over the study period of March 2020 to December 2021 were included for final analysis, with the majority published during 2020 (79%) (Fig 1). A diverse range of journals were included, totalling 20 in number. Two journals accounted for 27% of articles analysed, *Anaesthesia* and

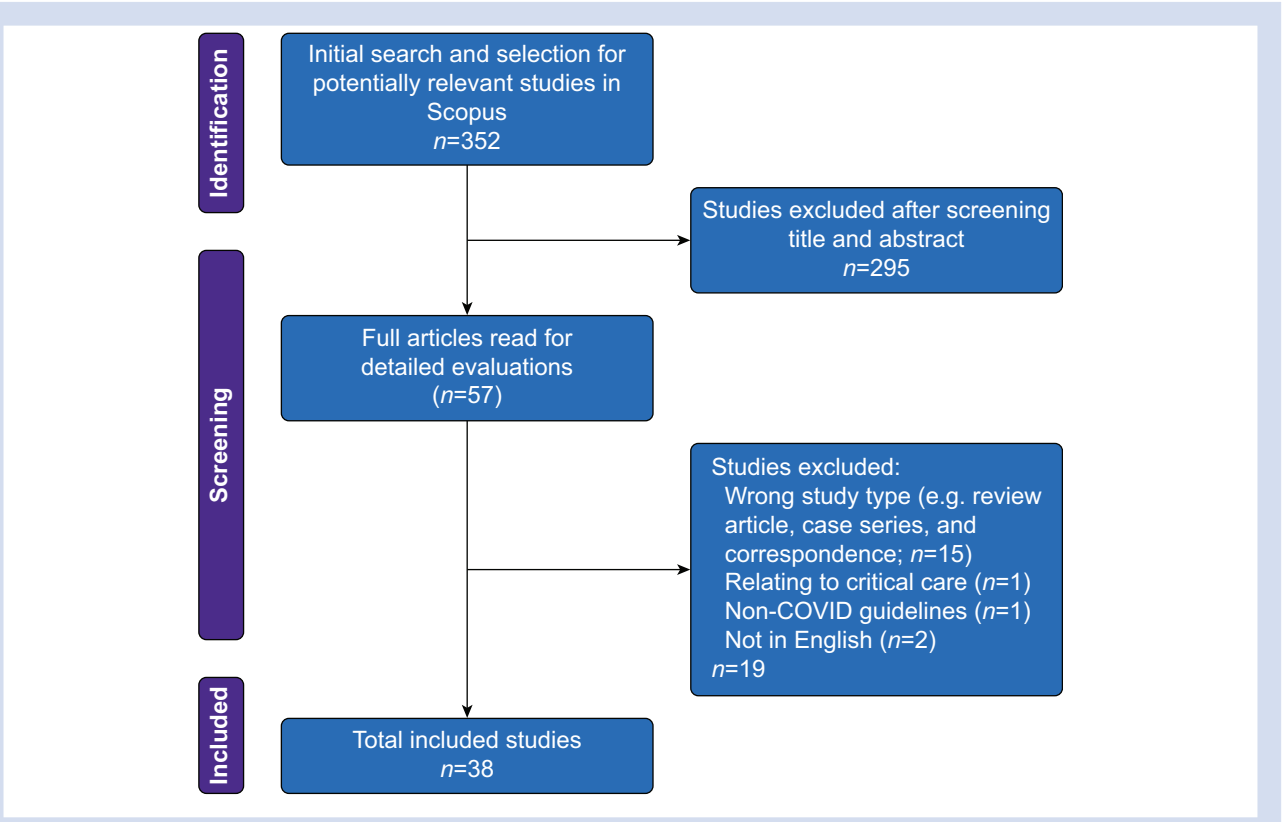

Fig 1. Consolidated Standards of Reporting Trials diagram.

Anesthesia & Analgesia. The breakdown per anaesthesia subspecialty was as follows: perioperative 63% (cardiothoracic 24%, general 18%, regional 8%, neuro 5%, orthopaedics 3%, ophthalmology 3%, non-operating theatre anaesthesia 3%, airway 13%, pain 13%, obstetrics 5%, and paediatrics 5%. Cross-regional cooperation was evident in the geographic origin of the articles with 39% involving authors from more than one continent. Of the remaining articles, 29% were from Asia, 16% were from North America, 7.9% were from the UK, 5.3% were from Europe, and 2.6% were from Australia. The median altimetric score was 8 (IQR: 0–24), and the median Scopus citation score was 9 (IQR: 1–20). Only three (8%) articles used an appraisal tool during formulation. Of these, AGREE II, Grading of Recommendations Assessment, Development and Evaluation (GRADE), and CAse REports (CARE)/Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) accounted for one use each. A descriptive summary of the articles appraised is shown in Supplementary Table 1.

Anaesthesia societies were involved in the formulation of 71% (n=27) of the guidance documents; 47%, including one society; and 24%, including multiple societies (median of 4; IQR: 2–70). Most guidelines were multi-institutional in origin (79%; median 7 [IQR: 3–37]) but confined to a single country (61%). Of the guidelines developed through international collaboration, the median number of countries involved was 3 (IQR: 2–16).

The individual domain scores, documented in Fig 2a and b, were highest for Domain 1 (Scope and Purpose; median 92 [IQR: 86–94]; mean 89 [SD 8]), Domain 6 (Editorial Independence; median 100 [IQR: 50–100]; mean 80 [SD 27]), and Domain 4 (Clarity of Presentation; median 78 [IQR: 69–85]; mean 77 [13]). Domain 3 (Rigour of Development) achieved the lowest score (median 16 [IQR: 10–31]; mean 21 [SD 15]), followed by Domain 5 (Applicability; median 20 [IQR: 15–29]; mean 23 [SD 15]) and Domain 2 (Stakeholder Involvement; median 44 [IQR: 39–50]; mean 44 [SD 6]). The intra-class correlation coefficient was excellent for four domains (2, 3, 5, and 6) and good for the remaining two domains (1 and 4); see Supplementary Table 2. Most of the guidance documents (63%) were assigned an overall subjective quality rating of 'yes with modifications' (overall rating of 4–5) with a minority (16%) recommended without modifications (overall rating of 6–7). Over one-fifth of the papers reviewed were not recommended (21%) in their current format (overall rating 1–3). Only two guidance documents (5%), both labelled as 'recommendations', met the objective quality criteria of >50% in Domains 1, 3, 4, and 6. Both of these guidelines correlated with the highest subjective score[6,7] and were recommended for use.

Table 2 reports the trend of scores over time, characterised by four 6 month time intervals (from January 2019 to December 2021). There was no significant difference in the individual domain scores, overall rating, and subjective or objective quality scores across the study period. The use of an appraisal tool compared with no appraisal tool was associated with significantly higher scores in Domain 3 score (P=0.028) and for the overall rating (P=0.024); see Table 3. Similarly,

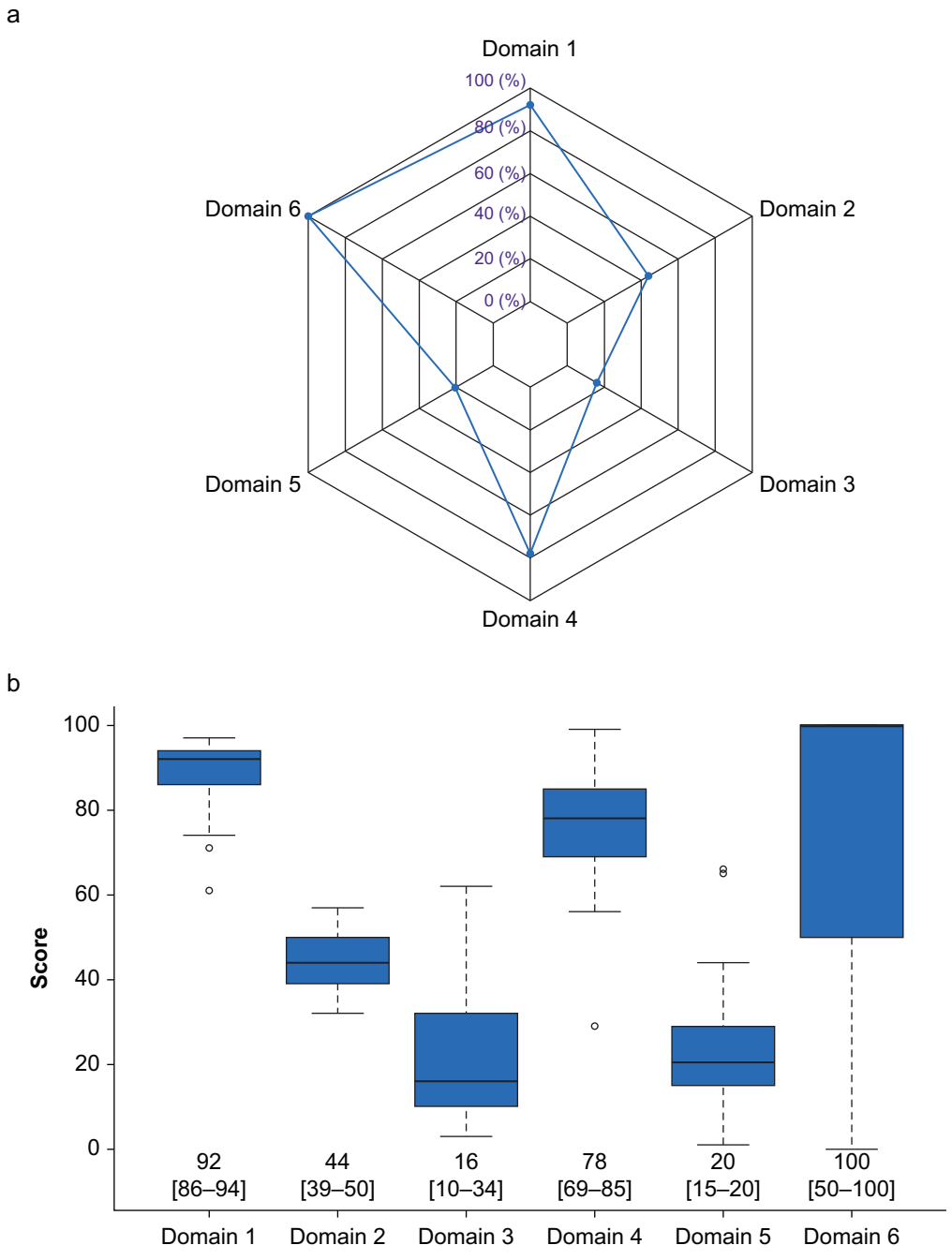

Fig 2. (a) Radar plot detailing the median domain scores; (b) boxplots detailing the median (black) and interquartile range (blue) domain scores.

multi-institutional cooperation improved the Rigour of Development (Domain 3) of the articles (P=0.043); see Supplementary Table 3. Society involvement did not improve ratings for individual domain scores, overall rating, and subjective or objective quality scores (Supplementary Table 4). An international perspective and participation from multiple countries (Table 4) emerged as the most influential property in COVID-19 guidance document development yielding significantly higher scores in Domain 2 (P=0.023), Domain 3 (P=0.007), overall rating (P=0.005), and subjective quality score (P=0.001).

Articles entitled 'guidelines' were not scored higher in any of the categories than those self-labelled as 'consensus recommendations' or 'recommendations' (Supplementary Table 5). There were also no differences noted in appraisal tool use, society involvement (none *vs* single *vs* multiple), or

Table 2 Trend over time of individual domain scores (mean/median), mean rating, and subjective quality and objective quality. *Mean (standard deviation); median (25th–75th quartiles); n (%). †Kruskal–Wallis rank-sum test; Fisher's exact test.

Characteristic	Overall, n=38*	January–June 2020, n=15*	July–December 2020, n=18*	January–July 2021, n=3*	July–December 2021, n=2*	P-value†
Domain 1	89 (8); 92 (86–94)	91 (5); 93 (90–94)	87 (9); 88 (86–93)	91 (5); 94 (90–94)	82 (15); 82 (76–87)	0.3
Domain 2	45 (6); 44 (39–50)	45 (5); 44 (42–49)	43 (6); 44 (38–49)	49 (6); 51 (47–53)	48 (13); 48 (44–53)	0.3
Domain 3	21 (15); 16 (10–31)	22 (15); 16 (12–30)	18 (15); 14 (8–25)	34 (13); 33 (28–40)	20 (10); 20 (16–24)	0.2
Domain 4	77 (13); 78 (69–85)	76 (17); 78 (66–88)	77 (9); 78 (70–82)	80 (4); 78 (78–82)	76 (29); 76 (66–87)	>0.9
Domain 5	23 (15); 20 (15–29)	26 (16); 23 (17–37)	20 (15); 18 (12–26)	22 (6); 22 (20–25)	24 (6); 24 (22–27)	0.4
Domain 6	80 (27); 100 (50–100)	73 (32); 100 (50–100)	89 (21); 100 (100–100)	67 (29); 50 (50–75)	75 (35); 75 (62–88)	0.3
Overall rating	4 (1); 4 (4–5)	4 (1); 4 (4–5)	4 (1); 4 (3–5)	5 (1); 5 (5–6)	5 (2); 5 (4–5)	0.4
Subjective quality, n (%)						0.2
No	8 (21)	2 (13)	5 (28)	0 (0)	1 (50)	
Yes	6 (16)	3 (20)	1 (6)	1 (33)	1 (50)	
Yes with modifications	24 (63)	10 (67)	12 (67)	2 (67)	0 (0)	
Objective quality, n (%)						>0.9
No	36 (95)	14 (93)	17 (94)	3 (100)	2 (100)	
Yes	2 (5)	1 (7)	1 (6)	0 (0)	0 (0)	

multi-institutional (single *vs* multiple) or international collaboration (single *vs* multiple) between these two groups (Supplementary Table 6).

When COVID-19 anaesthesia guidelines were compared with a comprehensive review of recent general anaesthesia guidelines, several noteworthy findings were revealed. These included statistically worse performances by COVID-19 guidelines in (i) Domain 3 (Rigour of Development; prior study: median 51 [IQR: 27–70] and mean 51 [SD 23] *vs* current study: median 10 [IQR: 7–26] and mean 16 [SD 13]; P<0.001), (ii) Domain 4 (Clarity of Presentation; prior study: median 89 [IQR: 82–94] and mean 87 [SD 10] *vs* current study: median 71 [IQR: 67–85] and mean 73 [SD 20]; P=0.021), (iii) Domain 5 (Applicability; prior study: median 41 [IQR: 29–63] and mean 44 [SD 21] *vs* current study: median 18 [IQR: 4–38] and mean 22 [SD 23]; P=0.01), (iv) overall rating (prior study: median 5 [IQR: 4–6] and mean 5 [SD 1] *vs* current study: median 4 [IQR: 3–5] and mean 4 [SD 1]; P=0.016), and (v) subjective quality score (P=0.003). These results are detailed in Supplementary Table 7.

Discussion

This is the most comprehensive study assessing COVID-19 guidance documents in anaesthesia using the AGREE II instrument to date. Several studies with small numbers of anaesthesia-specific documents have raised serious concerns regarding overall quality and rigour of the guidance documents produced during the COVID-19 pandemic.[17,20] All of the main anaesthesia subspecialties, exclusive of ICU, were represented amongst the 38 guidance documents studied. The analysis revealed poor methodological quality of these documents and an interchangeability between the terms of guidelines, consensus statements, and recommendations. This is reflected in the very low scores achieved in Domain 3

Table 3 Comparison of individual domain scores (mean/median), mean rating, subjective quality, and objective quality between guidance documents with and without appraisal tool use. *Mean (standard deviation); median (25th–75th quartiles); n (%). †Wilcoxon rank-sum test; Fisher's exact test.

Characteristic	No appraisal tool, n=35*	Appraisal tool used, n=3*	P-value†
Domain 1	89 (8); 92 (87–94)	89 (3); 89 (88–90)	0.4
Domain 2	44 (6); 44 (39–50)	49 (9); 49 (45–53)	0.3
Domain 3	19 (13); 15 (10–26)	44 (16); 46 (36–52)	0.028
Domain 4	76 (13); 78 (68–85)	87 (10); 86 (82–92)	0.15
Domain 5	23 (15); 20 (14–28)	24 (5); 24 (22–26)	0.5
Domain 6	81 (27); 100 (50–100)	67 (29); 50 (50–75)	0.3
Overall rating	4 (1); 4 (4–5)	6 (1); 6 (6)	0.024
Subjective quality, n (%)			0.09
No	8 (23)	0 (0)	
Yes	4 (11)	2 (67)	
Yes with modifications	23 (66)	1 (33)	
Objective quality, n (%)			0.2
No	34 (97)	2 (67)	
Yes	1 (3)	1 (33)	

Table 4 Comparison of individual domain scores (mean/median), mean rating, subjective quality, and objective quality between guidance documents involving one *vs* multiple countries. *Mean (standard deviation); median (25th–75th quartiles); *n* (%). †Wilcoxon rank-sum test; Fisher's exact test.

Characteristic	Single country, *n*=23*	Multiple countries, *n*=15*	*P*-value†
Domain 1	90 (6); 92 (88–94)	87 (10); 92 (84–94)	0.7
Domain 2	43 (6); 42 (39–46)	47 (6); 49 (44–51)	0.023
Domain 3	16 (11); 13 (8–21)	29 (17); 27 (16–36)	0.007
Domain 4	75 (14); 78 (68–82)	80 (12); 85 (72–90)	0.3
Domain 5	20 (14); 19 (10–25)	27 (14); 28 (18–30)	0.14
Domain 6	85 (28); 100 (75–100)	73 (26); 50 (50–100)	0.13
Overall rating	4 (1); 4 (3–5)	5 (1); 5 (4–6)	0.005
Subjective quality, n (%)			0.001
No	7 (30)	1 (6.7)	
Yes	0 (0)	6 (40)	
Yes with modifications	16 (70)	8 (53)	
Objective quality, n (%)			0.15
No	23 (100)	13 (87)	
Yes	0 (0)	2 (13)	

(Rigour of Development), widely regarded as the most important domain and most reflective of potential bias and evidence-based document formulation.[14] Only 16% of papers met the subjective quality threshold graded in our study to recommend their publication without modification. Only two papers scored >50% in Domains 1, 3, 4, and 6, meeting the objective quality criteria. These findings demonstrate that recommendations during the early COVID era were of similar quality and based on low levels of evidence, regardless of their title. In the absence of other useful resources, these documents have often been applied clinically. However, their limitations should be considered, and their suggestions should be reassessed or updated when new information arises.

The clinical and informational challenge of the COVID-19 pandemic was unprecedented in the modern era, creating enormous challenges for anaesthetists to provide the best care possible whilst operating in an evidence vacuum.[4,23,24] Urgent guidance with rapid dissemination was required, necessitating a move away from traditional, time-consuming approaches.[25] Development of CPGs can take many months or potentially years from the conception phase to publication, and even longer for translation into the clinical environment.[26,27] In response to this, novel types of guidance documents, such as rapid statements and living CPGs, have more recently emerged, gaining additional prominence during COVID-19.[28] Whilst case reports/case series and anecdotal/expert experience are regarded as the lowest form of data in the Oxford 'Levels of Evidence', they constituted most of the early publications across all specialties during the pandemic.[29–31] Whilst this may be expected early in the pandemic, more robust evidence did not appear to underpin the guidance over time from early 2019 to late 2020. This is not entirely surprising, as the overall time frame in question of less than 2 yr is short for sequential medical evidence generation and subsequent application in guidance development.[32–34]

Only 8% of documents utilised appraisal tools in their formulation. We found that the use of an appraisal tool was associated with a higher Domain 3 score and mean overall rating. The types of appraisal tools used were AGREE II, GRADE, and CARE. AGREE II is the most multifunctional and well-validated of all the tools available and can be used in guideline development, reporting, and appraisal.[22,35] GRADE provides additional granularity in the assessment of the level of evidence and is increasingly being used in guideline appraisal, often in conjunction with AGREE II.[36,37] CARE is a tool used to guide and assess a low level of evidence, specifically case reports.[38] The use of CARE as an appraisal tool for guidance documents is unusual, and its use in this study likely reflects a poor underlying evidence base. Given the association of appraisal tool use with improved scores in Domain 3 and in the overall rating, guidance document working groups should be encouraged to use these instruments to improve rigour.

International collaboration emerged as the feature most aligned to document quality compared with multi-institutional or society involvement. Although participation from multiple institutions demonstrated certain benefits (higher Domain 3 scores), this was exaggerated by participation across countries (higher Domains 2 and 3 scores, overall rating, and subjective quality). By definition, international cooperation reflects a diversity of institutions, which may reduce duplication of effort, reduce cost, and improve uptake of guidance in practice.[39] Society involvement had no impact on quality, a finding that has been replicated elsewhere.[40,41] Single societies were represented to a greater extent than multiple societies, which may indicate a smaller range of expertise available and a more limited structure for guidance document development. In the context of COVID-19, with clinicians calling for guidance on best practice, societies might have felt the need to rapidly produce guidance documents; however, we demonstrate that international collaboration and resulting knowledge-sharing produce improved quality.

The AGREE II domains that achieved the highest scores were Domain 1 (Scope and Purpose), Domain 4 (Clarity of Presentation), and Domain 6 (Editorial Independence). These domains appear to be the easiest to maximise, both in anaesthesia and in other specialties.[42,43] Domain 6, assessing conflicts of interest and sources of controversy in funding, has been suggested to possess equal importance to Domain 3.[14] It is possible that this domain is over generously scored. Whilst the AGREE II user manual prompts description of the influence of competing interests in guideline development, less detail is required about the nature of any financial support.

Furthermore, the scale of industry–physician relationships is not reflected in the rate of declaration observed in these guidance documents.[44]

Domain 2 (Stakeholder Involvement) and Domain 5 (Applicability) achieved the lowest scores. The lack of stakeholder involvement, particularly during the early stages of the pandemic, is not unexpected. The COVID-19 era necessitated innovative forms of communication, which created obvious challenges for the involvement of patients and the wider public.[45,46] Full compliance with these two domains is also resource-intensive in terms of personnel, time, and finances, all of which were limited during COVID-19. For these reasons, the objective quality score created by this working group did not include these two domains, which are less likely to be as reflective of quality in the context of COVID-19.

The nomenclature used to describe guidance documents is not standardised, which can lead to misunderstanding regarding the quality and trustworthiness of documents produced. There were three main types of document titles identified in this study. The title of an article as either a 'guideline', 'consensus statement', or set of 'recommendations' has important implications for the reader in terms of strength of evidence, rigour of methodology, and overall trustworthiness. During the pandemic, this study found that many of the documents labelled as guidelines were incorrectly titled and were not based on systematic review, a risk/benefit analysis, and assessment of alternative therapeutic options as per the definition. This is reflected in extremely low Domain 3 scores for the nine 'guidelines' included (median score of 10 and mean of 16). Moreover, when the 'guidelines' group was compared with the group of 'consensus statements' and 'recommendations', there were no differences in scores for any of the AGREE II categories studied. Other markers of quality identified in this study, such as appraisal tool use, multi-institution involvement, and international collaboration, were also similar between groups. We demonstrate that the title of CPGs was inconsistent, not matched to quality or known definitions, which could potentially create a source for confusion for practitioners.

A comparison of the quality of this body of COVID-19 guidance documents with 51 recent general anaesthesia guidelines from the top 10 anaesthesia journals by impact factor revealed significant disparities in performance as per AGREE II was performed.[15] The latter study revealed poor quality in anaesthesia guidelines overall but generated an encouraging trend of improvement in Domain 3 from 2016 to 2020. Despite this, it established a need for increased transparency and rigour in anaesthesia guideline development. In comparison, COVID-19 guidance documents had statistically worse scores in Domains 3, 4, and 5 overall rating and subjective quality score. Interestingly, when compared with COVID-19 guidance documents from other specialties, the raw AGREE II domain scores for anaesthesia documented in this study were almost identical.[17] This consistency not only supports our findings but indicates that anaesthesia was not an outlier in the quality of its COVID-19 guidance documents.

In the face of knowledge gaps and clinical equipoise during the pandemic, poor quality guidance based on low-level evidence may incur harm. Perhaps a defined scoring system or standardised rating should be applied by journals to indicate the quality of guidance documents would allow clinicians to make more informed decisions about their appropriateness and applicability in the clinical environment.[28] Living guidelines potentially play a role, particularly in a viral pandemic such as COVID-19, because of the risk of resurgence or the emergence of new information.[30,47] A living guideline is defined as an optimisation of the guideline development process to allow updating of recommendations as new evidence emerges.[48] Living guidelines, introduced by the WHO in 2018, combine the processes of continuous literature surveillance, repeated systematic review, and virtual expert panel discussions to disseminate the latest evidence most efficiently [28,49] They have been used in the treatment and prevention of COVID-19 but not yet within anaesthesia.[50,51] Use of living guidelines in anaesthesia requires consideration of the rate of new evidence available, whether this should influence practice and the risk of confusing readers with frequent changes. Another potential improvement to guidance document development is a prospective register that could reduce duplication and foster greater collaboration.[18] This study demonstrates that whilst the general quality of guidelines in anaesthesia is recognised as poor, the guidance documents produced during COVID-19 offered an even lower standard, despite readily available aids, such as appraisal tools. Based on the findings of this study, a suggested ideal COVID-19 guidance document in anaesthesia would utilise an appraisal tool during development; take a collaborative approach with multi-institutional and international involvement; and adhere to the requirements of AGREE II, particularly Domains 1, 3, 4, and 6.

This study has several strengths. To the authors' knowledge, it is the largest study of COVID-19 guidance documents pertaining to anaesthesia. To capture all relevant articles, a rigorous search methodology was used. There was no limitation by journal type, impact factor, or country/region. Only articles pertaining to ICU were excluded, as the scope of practice in relation to COVID-19 disease was much broader in this setting and assessment of these guidelines would warrant its own separate study. Four independent appraisers assessed each article, which is the maximum number cited and encouraged by the AGREE II user manual. Three of these appraisers had expertise in guideline methodology and in the application of AGREE II, specific to anaesthesia, and have previously published experience with AGREE II. To contextualise the body of COVID-19 guidelines studied, a data set of 51 anaesthesia guidelines from 2016 to 2020 analysed by AGREE II was available and used for comparative purposes.

There were also a number of limitations to this study. Although the study included all guidance documents published at the height of the COVID-19 pandemic and in the immediate aftermath, articles after the search date were not assessed. The articles examined were limited by language to English only. The fourth time period (July–December 2021) may have had an incomplete number of guidance documents because of the search date. Papers with a broad multispecialty focus, inclusive of but not uniquely pertaining to anaesthesia, were not selected for analysis. A potentially useful comparison between COVID-19 guidance documents and those of past pandemics was not feasible because of lack of publications in the perioperative settings for the latter. Despite COVID-19 guidelines being developed in different circumstances than most anaesthesia guidelines, the authors felt that a comparison between the two was appropriate. Prior data on general anaesthesia guidelines provided a validated baseline. It was not expected that COVID-19 guidelines would perform as well as the general anaesthesia guidelines, but the comparison provided valuable insights into quality differences particularly in specific domains and certain quality scores. The quality of the evidence underpinning each recommendation in each guidance document was not examined. AGREE II,

although the gold standard for guideline assessment and appraisal, is subjective and limited by user application. Furthermore, individual domains are not weighted by importance and are considered equal by the tool developers despite clear discrimination applied by users in practice. The 'objective' quality score, which often places a heavier weighting to Domain 3, is not standardised and is open to interpretation by different guideline assessment groups. Finally, AGREE II is a methodological guide only and does not measure the clinical intervention performed or the clinical context within each guidance document.

Conclusions

The COVID-19 pandemic necessitated urgent development and dissemination of guidance for anaesthetists. The guidance documents produced were of universally poor quality, as assessed by AGREE II, regardless of their title. Furthermore, they were of significantly worse quality when compared with recent non-COVID-related general anaesthesia guidelines. Markers of quality included appraisal tool use and multi-institution involvement with international collaboration emerging as the most influential characteristic. Journals may have a role in screening and signposting guidance document quality. Furthermore, the establishment of a prospective registry and the use of living guidelines may be helpful in enhancing guidance document quality.

Authors' contributions

Study conception/design: SMOS, BK, LQR.
Data acquisition: SMOS, MD, NG, BK, LQR.
Data analysis: SMOS, AD, MR, LQR.
Article writing: SMOS, BK, LQR.
Revision of article: all authors.
All authors approved the final version of the article for submission and have agreed to be accountable for all aspects of this work.

Declarations of interest

LQR is an associate editorial board member of the *British Journal of Anaesthesia*.

Funding

US National Institutes of Health (K23 HL153836) to LQR.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bja.2022.09.008.

References

1. Allen D, Harkins K. Too much guidance? *Lancet* 2005; **365**: 1768
2. Fortunato S, Bergstrom CT, Börner K, et al. Science of science. *Science* 2018; **359**, eaao0185
3. McCartney CJ, Mariano ER. COVID-19: bringing out the best in anesthesiologists and looking toward the future. *Reg Anesth Pain Med* 2020; **45**: 586–8
4. Roberts T, Hirst R, Sammut-Powell C, et al. Psychological distress and trauma during the COVID-19 pandemic: survey of doctors practising anaesthesia, intensive care medicine, and emergency medicine in the United Kingdom and Republic of Ireland. *Br J Anaesth* 2021; **127**: e78–80
5. Steinberg E, Greenfield S, Wolman DM, Mancher M, Graham R. *Clinical practice guidelines we can trust*. Washington, DC: National Academies Press; 2011
6. Bakaeen FG, Svensson LG, Mitchell JD, Keshavjee S, Patterson GA, Weisel RD. The American Association for Thoracic Surgery/Society of Thoracic Surgeons position statement on developing clinical practice documents. *Ann Thorac Surg* 2017; **103**: 1350–6
7. Kredo T, Bernhardsson S, Machingaidze S, et al. Guide to clinical practice guidelines: the current state of play. *Int J Qual Health Care* 2016; **28**: 122–8
8. Smith CA, Toupin-April K, Jutai JW, et al. A systematic critical appraisal of clinical practice guidelines in juvenile idiopathic arthritis using the Appraisal of Guidelines for Research and Evaluation II (AGREE II) instrument. *PLoS One* 2015; **10**, e0137180
9. MacQueen G, Santaguida P, Keshavarz H, et al. Systematic review of clinical practice guidelines for failed antidepressant treatment response in major depressive disorder, dysthymia, and subthreshold depression in adults. *Can J Psychiatry* 2017; **62**: 11–23
10. Barrette L-X, Xu K, Suresh N, et al. A systematic quality appraisal of clinical practice guidelines for Ménière's disease using the AGREE II instrument. *Eur Arch Otorhinolaryngol* 2022; **279**: 3439–47
11. Zhang X, Zhao K, Bai Z, Yu J, Bai F. Clinical practice guidelines for hypertension: evaluation of quality using the AGREE II instrument. *Am J Cardiovasc Drugs* 2016; **16**: 439–51
12. Nygaard CC, Tsiapakidou S, Pape J, et al. Appraisal of clinical practice guidelines on the management of obstetric perineal lacerations and care using the AGREE II instrument. *Eur J Obstet Gynecol Reprod Biol* 2020; **247**: 66–72
13. Corp N, Mansell G, Stynes S, et al. Evidence-based treatment recommendations for neck and low back pain across Europe: a systematic review of guidelines. *Eur J Pain* 2021; **25**: 275–95
14. Hoffmann-Eßer W, Siering U, Neugebauer EA, Brockhaus AC, McGauran N, Eikermann M. Guideline appraisal with AGREE II: online survey of the potential influence of AGREE II items on overall assessment of guideline quality and recommendation for use. *BMC Health Serv Res* 2018; **18**: 1–43
15. O'Shaughnessy SM, Lee JY, Rong LQ, et al. Quality of recent clinical practice guidelines in anaesthesia publications using the Appraisal of Guidelines for Research and Evaluation II instrument. *Br J Anaesth* 2022; **128**: 655–63
16. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines?: the methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; **281**: 1900–5
17. Stamm TA, Andrews MR, Mosor E, et al. The methodological quality is insufficient in clinical practice guidelines in the context of COVID-19: systematic review. *J Clin Epidemiol* 2021; **135**: 125–35
18. Zhao S, Lu S, Wu S, et al. Analysis of COVID-19 guideline quality and change of recommendations: a systematic review. *Health Data Sci* 2021; **2021**: 9806173

19. Mai HT, Croxford D, Kendall MC, De Oliveira G. An appraisal of published clinical guidelines in anesthesiology practice using the AGREE II instrument. *Can J Anaesth* 2021; **68**: 1038–44
20. Ong S, Lim WY, Ong J, Kam P. Anesthesia guidelines for COVID-19 patients: a narrative review and appraisal. *Korean J Anesthesiol* 2020; **73**: 486–502
21. World Federation of Societies of Anaesthesiologist. *COVID-19 guidance for anaesthesia and perioperative care providers* 2020. Available from: https://wfsahq.org/resources/COVID-19/. [Accessed 27 September 2022].
22. Brouwers MC, Kho ME, Browman GP, et al. Agree II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010; **182**: E839–42
23. Ortega R, Chen R. Beyond the operating room: the roles of anaesthesiologists in pandemics. *Br J Anaesth* 2020; **125**: 444–7
24. Afshari A, Disma N, von Ungern-Sternberg BS, Matava C. COVID-19 implications for pediatric anesthesia: lessons learnt and how to prepare for the next pandemic. *Pediatr Anesth* 2022; **32**: 385–90
25. Alam M, Harikumar V, Ibrahim SA, et al. Principles for developing and adapting clinical practice guidelines and guidance for pandemics, wars, shortages, and other crises and emergencies: the PAGE criteria. *Arch Dermatol Res* 2022; **314**: 393–8
26. Practice advisory for the prevention, diagnosis, and management of infectious complications associated with neuraxial techniques: an updated report by the American society of anesthesiologists task force on infectious complications associated with neuraxial techniques and the American society of regional anesthesia and pain medicine. *Anesthesiology* 2017; **126**: 585–601
27. Moppett IK, Gardiner D, Harvey DJ. Guidance in an uncertain world. *Br J Anaesth* 2020; **125**: 7–9
28. Rong LQ, Audisio K, O'Shaughnessy SM. Guidelines and evidence-based recommendations in anaesthesia: where do we stand? *Br J Anaesth* 2022; **128**: 903–8
29. The Centre for Evidence-Based Medicine. *The Oxford 2011 levels of evidence 2011*. Available from: https://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf. [Accessed 11 May 2022]
30. Dagens A, Sigfrid L, Cai E, et al. Scope, quality, and inclusivity of clinical guidelines produced early in the covid-19 pandemic: rapid review. *BMJ* 2020; **369**: m1936
31. Zhao S, Cao J, Shi Q, et al. A quality evaluation of guidelines on five different viruses causing public health emergencies of international concern. *Ann Transl Med* 2020; **8**: 500
32. Rosenfeld RM, Shiffman RN. Clinical practice guideline development manual: a quality-driven approach for translating evidence into action. *Otolaryngol Head Neck Surg* 2009; **140**: S1–43
33. Powell K. Does it take too long to publish research? *Nature* 2016; **530**: 148–51
34. Ross JS, Mocanu M, Lampropulos JF, Tse T, Krumholz HM. Time to publication among completed clinical trials. *JAMA Int Med* 2013; **173**: 825–8
35. Brouwers MC, Kerkvliet K, Spithoff K, Consortium ANS. The AGREE reporting checklist: a tool to improve reporting of clinical practice guidelines. *BMJ* 2016; **352**: i1152
36. Thornton J, Alderson P, Tan T, et al. Introducing GRADE across the NICE clinical guideline program. *J Clin Epidemiol* 2013; **66**: 124–31
37. Hazlewood GS, Akhavan P, Schieir O, et al. Adding a "GRADE" to the quality appraisal of rheumatoid arthritis guidelines identifies limitations beyond AGREE-II. *J Clin Epidemiol* 2014; **67**: 1274–85
38. Gagnier JJ, Kienle G, Altman DG, Moher D, Sox H, Riley D. The CARE guidelines: consensus-based clinical case reporting guideline development. *J Med Case Rep* 2013; **7**: 223
39. Philip T, Fervers B, Haugh M, Otter R, Browman G. European cooperation for clinical practice guidelines in cancer. *Br J Cancer* 2003; **89**: S1–3
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**: 148–57
41. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000; **355**: 103–6
42. Merchan-Galvis AM, Caicedo JP, Valencia-Payán CJ, Calvache JA. Methodological quality and transparency of clinical practice guidelines for difficult airway management using the appraisal of guidelines research & evaluation II instrument: a systematic review. *Eur J Anaesthesiol* 2020; **37**: 451–6
43. Andrade R, Pereira R, van Cingel R, Staal JB, Espregueira-Mendes J. How should clinicians rehabilitate patients after ACL reconstruction? A systematic review of clinical practice guidelines (CPGs) with a focus on quality appraisal (AGREE II). *Br J Sports Med* 2020; **54**: 512–9
44. Tringale KR, Marshall D, Mackey TK, Connor M, Murphy JD, Hattangadi-Gluth JA. Types and distribution of payments from industry to physicians in 2015. *JAMA* 2017; **317**: 1774–84
45. Kudchadkar SR, Carroll CL. Using social media for rapid information dissemination in a pandemic:# PedsICU and coronavirus disease 2019. *Pediatr Crit Care Med* 2020; **21**: e538–46
46. Fan Z, Yin W, Zhang H, et al. COVID-19 information dissemination using the WeChat communication index: retrospective analysis study. *J Med Internet Res* 2021; **23**, e28563
47. Munn Z, Twaddle S, Service D, et al. Developing guidelines before, during, and after the COVID-19 pandemic. *Ann Intern Med* 2020; **173**: 1012–4. Available from, https://www.acpjournals.org/doi/10.7326/M20-4907. [Accessed 11 May 2022]
48. Akl EA, Meerpohl JJ, Elliott J, et al. Living systematic reviews: 4. living guideline recommendations. *J Clin Epidemiol* 2017; **91**: 47–53
49. World Health Organization. *Keeping global recommendations up-to-date: a 'living guidelines' approach to maternal and perinatal health* 2019. Available from: https://www.who.int/news/item/19-08-2019-keeping-global-recommendations-up-to-date-a-living-guidelines-approach-to-maternal-and-perinatal-health. [Accessed 11 May 2022]
50. Agarwal A, Rochwerg B, Lamontagne F, et al. A living WHO guideline on drugs for covid-19. *BMJ* 2020; **370**: m3379
51. Lamontagne F, Agoritsas T, Siemieniuk R, et al. A living WHO guideline on drugs to prevent covid-19. *BMJ* 2021; **372**: n526

Handling editor: Jonathan Hardman