Quality Guidelines for Corona Virus Disease 2019 with AGREE II Instrument.

Yao Song¹, Qirui Liu¹, Jianjun Ren¹, Wenjie Yang¹, Ke Qiu¹, Jianqing Qiu², Wendu Pang¹, Yijun Dong¹, Deying Kang¹, Yu Zhao¹, and Wei Zhang¹

¹Sichuan University West China Hospital
²Sichuan University West China School of Public Health Department of Epidemiology and Health Statistics

April 29, 2020

Abstract

Objective: Numerous clinical practice guidelines (CPGs) for the corona virus disease 2019 (COVID-19) have been developed since its outbreak. To date, however, the methodological quality of these guidelines has not been fully evaluated. Quality evaluation of the guidelines is to facilitate the clinic practitioners in diagnosis and treatment of the COVID-19. Methods: The evaluation was conducted by searching seven databases and government organizations websites (December 2019 to March 2020). Four assessors assessed the quality of the CPGs independently with the Appraisal of Guidelines for Research & Evaluation II (AGREE II) instrument. For domain scores (0-100%), >60% were considered of sufficient quality, and >80% of good quality. Results: Of 471 retrieved records, 20 guidelines were included. The median score for the domain of scope and purpose was 56.9% (range 40.3–90.3%) and nine guidelines scored >60%, among which three guidelines scored >80%. The median score for the stakeholder involvement domain was 7% (range 0–65.3%) and only two guidelines scored >60%. Nine guidelines scored 0% in this domain. The median score for the rigour of development domain was 0% (range 0–91.7%) and fourteen guidelines scored 0% in this domain. The median score for the clarity of presentation domain was 30.6% (range 13.9–91.7%) and six guidelines scored >60%, among which three guidelines scored >80%. The median score for the applicability domain was 0% (0–57.3%) and eleven guidelines scored 0% in this domain. The median score for the editorial independence domain was 0% (0–100%): four guidelines scored >80%. Discussions: Three guidelines were recommendable, and four were recommendable with modification in the stakeholder involvement, the rigour of development, and the applicability domains. Our results could contribute to improve development of future guidelines, and affect the reasonable selection and use of guidelines in clinical practice.

1 Introduction

The corona virus disease 2019 (COVID-19), caused by new virus that first reported in Wuhan, China, in mid-December 2019, has so far infected more than 2.4 million people and spread to nearly 211 countries and areas, causing huge losses to public health and property. For standardizing the diagnosis and treatment of COVID-19 infections, the Chinese government, the World Health Organization (WHO) and clinical experts in relevant disciplines around the world have published numerous clinical practice guidelines (CPGs) for COVID-19.

CPGs are regarded as systematically developed statements to assist clinical practitioners in making decisions regarding appropriate health care for specific clinical circumstance¹. However, CPGs drawn up by different groups to respond for same clinic diseases may result in quite difference, even conflicting recommendations, making it difficult for clinic practitioners to choose the superior one. AGREE II (the Appraisal of Guidelines for Research & Evaluation II) is the new (2010) international tool to assess the quality and reporting
of practice guidelines\(^2\). It can be used to critically appraise the comprehensiveness, rigor, clarity, and applicability, etc., of CPGs.

Furthermore, due to the urgency and damage of COVID-19, plenty of CPGs have been drafted quickly. Therefore, it is necessary and meaningful to evaluate and compare their methodologic quality, which, to date, hasn’t been performed. The objective of our present study was to evaluate the quality of currently available COVID-19 guidelines using the AGREE II instrument, so as to assist clinicians in choosing the most appropriate guideline.

2 Methods and Materials

2.1 Information Sources and Eligibility Criteria

The search for the management of COVID-19 CPGs was carried out on March 30, 2020, using documents issued by regional or international groups or government organizations in English and Chinese, through the following search engines using appropriate keywords: Embase; Medline; UpToDate; Cochrane Library; Database for Chinese Technical Periodicals (VIP); China National Knowledge Infrastructure (CNKI); the websites of China national, provincial, local Health and Fitness Commission; the website of WHO. In this study the following keywords were used: ‘guideline’, ‘novel/ new coronavirus’, ‘COVID-19’. The search was limited to December 2019 and not earlier.

For included CPGs in this research, they must meet the inclusion criteria as follows: (1) the CPGs provide systematically recommendations or strategies in any population with COVID-19 on epidemiology, screening, diagnosis of disease, preventive measures, therapeutic interventions that assists clinical practitioners to make appropriate decisions under specific circumstances; (2) the CPGs was conducted under the concerted effort of public health organizations, professional societies, medical associations, or government agencies at the national, provincial, or local level. Since the role in the assessment of AGREE II of health technology assessments has not been formally evaluated\(^2\), the exclusion criteria are as follow: (1) guidance documents or handbook that address health care organizational issues, or guidelines were not issued on a regional level; (2) documents that were not guidelines; (3) guidelines that were concerning mainly about traditional Chinese medicine.

2.2 Assessment of Guideline Quality

Four assessors, two were experienced clinicians and the other two were public health fellows with experiences in developing and evaluating guidelines, completed the online overview tutorial and practice exercise recommended by the AGREE collaboration before evaluation\(^2\).

The assessors independently responded to a total of 23 questions in six domains using the AGREE II instrument: (1) scope and purpose of the guideline, (2) stakeholder involvement, (3) rigour of development, (4) clarity of presentation, (5) applicability, (6) editorial independence\(^2, 3\). Each item was rated on a scale of 1 for “strongly disagree” to 7 for “strongly agree”\(^2, 3\). On evaluating the 23 items, each appraiser provided an overall assessment of each guideline, and decide if the guideline is recommendable. The decision was based on the personal judgement of assessors and domain scores\(^4\). In order to reduce discrepancies among four assessor, we referred to a previous method\(^4, 5\): if the scores assigned by four appraisers differed by 1 point, the lower was kept; if the scores differed by 2 points, they were averaged; and if the scores differed by 3 points, an agreement was reached after discussion.

According to the AGREE II methodology\(^2, 4\), domain scores were calculated as follows: \(\frac{\text{obtained score-minimum possible score}}{\text{maximum possible score-minimum possible}} \times 100\%\), while the obtained score was defined as the sum of four assessors’ scores of each item. Then, as reported in previous researches\(^3, 4, 6\), a value >60% was considered as sufficient and a value >80% as good. A median score across all six domains was calculated for each guideline.

3 Results

3.1 Guideline Selection
After screening for relevance in databases, 471 records were initially identified and 43 records were excluded for duplicates. After screening the title and abstract of records, 375 records were excluded because they were about animals, sanitary technical guidance or irrelevant. 53 remaining records were retrieved for full text and 33 records were excluded because they were guidance documents on health work. Finally, 20 CPGs were selected (Figure 1). Table 1 presented the basic information of the included guidelines,

3.2 AGREE II Scores

Four researchers participated in the evaluation of each guidelines and the standardized scores in six domains were shown in Table 2.

3.3 Scope and Purpose

The median score for the domain of scope and purpose was 56.9% (range 40.3-90.3%). Guideline 2,3,4,5,10,14,16,18,19,20 scored above 60% for this domain. The other guidelines did not clearly describe their overall objectives, health questions and target populations.

3.4 Stakeholder Involvement

The median score for stakeholder involvement domain was 7% (range 0%-65.3%). Only Guideline 2 and 19 scored above 60% in this domain as it clearly defined the target users of the guideline and included individuals from most relevant professional groups. No guidelines considered the views and preferences of the target population.

3.5 Rigour of Development

The median score for rigour of development domain was 0% (range 0-91.7%). Guideline 19 had the highest score (91.7%) in this domain because it described systematic method they applied to search for evidence, the strengths and limitations of the body of evidence, the methods for formulating the recommendations, updating procedure, reviewed by experts prior to its publication and considered the health benefits, side effects and risks when developing the guideline. Other guidelines all performed poor in this domain.

3.6 Clarity of Presentation

The median score for clarity of presentation domain was 30.6% (range 13.9-91.7%). Guideline 1,2,3,8,18,19,20 scored above 60% for this domain. Other guidelines did not provide specific, unambiguous and easily identifiable recommendations.

3.7 Applicability

The median score for applicability domain was 0% (range 0-57.3%). Guideline 2 had the highest score (57.3%) in this domain because it described the facilitators and barriers to its application, provided advices and tools on the practice of the recommendations. Guideline 3, as well as the Guideline 2,6,17,18,11,19,18,13,19,20 also presented monitoring and auditing criteria.

3.8 Editorial independence

The median score for editorial independence domain was 0% (range 0-100%). The scores of the included guidelines in this domain was polarized: Guideline 1,2,3 and 19 scored >80%, and 18 scored 45.8%, while the others all scored 0%

4 Discussion

No relevant research was found in an extended search on April 23th, this was the first study to systematically evaluate the quality of COVID-19 guidelines with the AGREE II instrument. The overall quality of the 20 guidelines for COVID-19 was highly variable, and significant variability can be seen across domains within guidelines.

Domain scores of scope and purpose in most guidelines were below 60%, indicating that the inclusion and exclusion criteria of target populations and health question remained unclear for early published guidelines.
Stakeholder involvement domains were poorly described, making those guidelines less professional. The involvement of patients in making decision might improve clinical outcomes and increase patients’ guideline adherence\(^2\). Therefore, collecting the views and preferences of target populations is a necessary part of standard guideline, and needs a period of time to carry out. However, considering over 60000 people died, and COVID-19 continues to spread worldwide, no guideline described related information in this item. More attention are ought to paid on the composition of the guideline development team and target population preferences by local and regional guideline developers\(^20\).

Systematic reviews are expected to form the basis of high quality of CPGs\(^21\). The importance had been demonstrated by the guideline manuals published by WHO\(^22\). No guideline involved review of outside methodological or health economic experts, which explained all guidelines scored below 60% in the rigor of development domain and suggested a fundamental methodology problem with these CPGs.

Only six guidelines had moderate to high scores in the clarity of the recommendations. This might result from the unclear audience and low scores in stakeholder and rigour of development domain in documents issued by government organizations. CPGs developers should provide a concrete and precise description of different options in different situations, as informed by the body of evidence, and make recommendations easily identifiable to the first-line clinicians audience.

In the applicability domain, most guidelines scored 0%, and did not describe the facilitators and barriers to their application and definite audit criteria. Some CPGs even lacked potential resource and educational tools of applying the recommendations, which might have a great influence on the speed and spread of adoption of guidelines. This finding is inversely proportional to the need of developing guidelines for user-friendliness and clarity suggested in some studies\(^4, 23\) and should be taken into more consideration when developing new guidelines or updating new versions.

Information on the domain of editorial independence was also neglected in most guidelines. This is particularly important given that influences of the founding body and conflicts of interest are the most common source of bias in guideline development\(^23, 24\). Perhaps some guideline developers did not realize the significance of editorial independence disclosures and management. Studies showed that financial conflicts of interest were prevalent among CPGs in a variety of clinical areas\(^24\), and some evidence suggested that such financial conflicts of interest might have an influence on guideline recommendations\(^25\). Therefore, the guideline developers can’t emphasize editorial independence domain enough.

Guideline 27,38 and 1914 can be recommended to guide clinical practice, while 116, 49, 1813, 2015 were recommendable with modification. In view of the urgency of the COVID-19 outbreak, the lower quality and limitations of the early published guidelines are understandable and acceptable. But they are expected to update more evidence-based medical recommendations and modify nonstandard methodology of guidelines to increase their credibility and applicability promptly. And guidelines developers are supposed to focus on not only high speed but also rigorous quality, especially in stakeholder involvement, rigour of development, applicability, and editorial independence domain.

This study had limitation as follows, first, the AGREE II instrument established an appraisal system for methodological quality of guidelines, but the evaluation of guideline recommendations was not stated. Second, the excluded guidelines concerning mainly about traditional Chinese medicine, the next research to work on in our team, might cause this study not to be representative of all CPGs.

In conclusion, the overall quality of CPGs for COVID-19 was uneven. Further research is needed for the appraisal of guideline recommendations. The results of our study could contribute to improve development of future guidelines, and affect the reasonable selection and use of guidelines in clinical practice.

5 Reference:


16. Expert Consensus on Diagnosis and Treatment of New Pediatric Coronavirus Pneumonia in Guangdong Province. Guangdong Medicine. [https://doi.org/10.13820/j.cnki.gydx.20200287](https://doi.org/10.13820/j.cnki.gydx.20200287)


Declarations

Funding: This work was supported by The Scientific and Technological Projects for COVID-19 of Sichuan Province [grant number No.2020YFS0007]; COVID-19 Emergency Project of Research Institute,[grant number No. 0082604151010]

Conflict of Interest: The authors declare that they have no conflicts of interest.

Availability of data and material: all data and materials support published claims and comply with field standards.

Figure 1 Flow chart of searching and selecting guidelines
Hosted file