

METHODOLOGY

An instrument for evaluating the clinical applicability of guidelines

Zeng Linan^{1,2} | Yi Qiusha^{1,2,3}  | Zhang Chuan^{1,2} | Huang Chao⁴ | Li Hailong^{1,2} | Yang Chunsong^{1,2} | Lin Mao^{1,2,5} | Huang Liang^{1,2,5} | Liu Dan^{1,2} | Kang Deying⁶  | Liu Guanjian⁶ | Liu Qiaolan⁷ | Zhao Rongsheng^{8,9} | Zhang Junhua¹⁰ | Li Youping⁶ | Liu Hanmin^{2,11} | Wang Qiang⁴ | Zhang Lingli^{1,2,9}

¹ Department of Pharmacy/Evidence-Based Pharmacy Centre, West China Second University Hospital, Sichuan University, Chengdu, P.R. China

² Key Laboratory of Birth Defects and Related Diseases of Women and Children, Ministry of Education, Chengdu, P.R. China

³ West China School of Medicine, Sichuan University, Chengdu, P.R. China

⁴ Medical Management Service Guidance Center, National Health Commission of the People's Republic of China, Beijing, P.R. China

⁵ West China School of Pharmacy, Sichuan University, Chengdu, P.R. China

⁶ Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, Chengdu, P.R. China

⁷ West China School of Public Health, Sichuan University, Chengdu, P.R. China

⁸ Department of Pharmacy, Peking University Third Hospital, Beijing, P.R. China

⁹ Evidence-Based Pharmacy Committee, Chinese Pharmaceutical Association, Beijing, P.R. China

¹⁰ Centre for Evidence-Based Medicine, Tianjin University of TCM, Tianjin, P.R. China

¹¹ Department of Pediatrics, West China Second University Hospital, Sichuan University, Chengdu, P.R. China

Correspondence

Wang Qiang, Medical Management Service Guidance Center, National Health Commission of the People's Republic of China, Beijing 100044, China.

Email: jason2019@sina.cn

Abstract

Objective: To establish an instrument for evaluating the clinical applicability of guidelines from the guideline-users' perspective.

Methods: We established this instrument through forming a working group, forming an initial list of items based on a qualitative systematic review, establishing initial instrument via two rounds of modified Delphi surveys, and external review the initial instrument.

Results: The results of modified Delphi surveys establishing appraisal aspects, appraisal items, general information of the evaluator met the preset requirements. The instrument includes three parts: general information of the evaluator (12 items), evaluation of clinical applicability (12 items, including items on the availability, readability, acceptability, feasibility, and overall applicability of guideline), and scoring scheme.

Conclusions: The instrument for evaluating the clinical applicability of guidelines from the guideline-users' perspective provides criteria and methods for improving the clinical applicability of guidelines during development and updating.

KEYWORDS

clinical applicability, clinical practice guideline, evaluation instrument

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Journal of Evidence-Based Medicine* published by Chinese Cochrane Center, West China Hospital of Sichuan University and John Wiley & Sons Australia, Ltd

Zhang Lingli, Department of Pharmacy/Evidence-Based Pharmacy Centre, West China Second University Hospital, Sichuan University, No. 20, Third Section, Renmin Nan Lu, Chengdu 610041, China.
Email: zhanglingli@scu.edu.cn

Zeng Linan and Yi Qiusha contributed equally to this study.

Funding information

This research has received funding from the National Health Commission of the People's Republic of China.

1 | INTRODUCTION

Clinical practice guidelines (CPGs) are statements that include recommendations intended to optimize patient care that is informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options.¹ Clear, explicit, and transparent clinical guidelines have a positive effect on both patients and health care practitioners² by supporting decision-making processes during patient care, improving the quality of health care, leading to better patient outcomes, and decreasing medical costs. To this end, CPGs play an important role in optimizing health care.^{1,3}

As health care practitioners continue to improve their awareness of evidence-based decision-making, CPGs have become more and more popular. The number of global CPGs has grown rapidly in recent years. As of September 2019, more than 25 000 CPGs have been included in the PubMed, an increase of 12 500% from 1990. To enable doctors and other health care stakeholders to obtain high-quality guidelines and promote CPGs dissemination, implementation, and use,⁴ several countries and institutions have established guideline databases, including the National Guideline Clearinghouse (NGC) (closed July 2018), the National Institute for Health and Care Excellence (NICE), and the Scottish Intercollegiate Guidelines Network (SIGN), etc. However, despite CPGs widespread use and popularity, they are not consistently successful in improving health care. Indeed, some result in an extensive waste of resources.^{5,6}

The lack of clinical applicability is an important reason for the poor results of the promotion and implementation of CPGs.⁷ Clinical applicability is defined as the extent to which the users can apply a recommendation in practice.⁸ The research of Melissa Brouwers, the leader of developing the Appraisal of Guidelines for Research and Evaluation (AGREE) instrument, pointed out that high AGREE scores may not be a determinant of CPGs use, so only improving the quality of CPGs may not improve the clinical applicability of CPGs.⁹ Evaluating the clinical applicability of CPGs is helpful in understanding the extent to which CPGs are accepted and used by medical practitioners, and to identifying CPGs with high clinical applicability and promoting their application. Appropriate clinical applicability instrument can provide key indicators and information for CPGs developers to optimize and improve CPGs, help to improve the clinical applicability of CPGs continuously, and truly make the CPGs available and useful.

However, there are additional shortcomings in the instruments available to evaluate CPGs, meaning that it may be hard to determine which are effective health care instruments. This suggests that an adequate system is needed for evaluating clinical applicability of CPGs that considers guideline-users as the target evaluators. To this end, in this article, we developed such an instrument for evaluating the clinical applicability of guidelines.

2 | METHODS

This project was funded by Medical Management Service Guidance Center of National Health Commission of the People's Republic of China and carried out by the Department of Pharmacy/the Evidence-Based Pharmacy Center, West China Second Hospital of Sichuan University.

In the first phase of our study, we formed a working group with experts from the Medical Administration Center of the National Health and Health Commission, West China Second Hospital, and the Chinese Evidence-Based Medicine Center. We conducted a qualitative systematic review of instruments for evaluating clinical applicability of CPGs.¹⁰ We then drafted a list of items to evaluate the clinical applicability of CPGs, including 5 primary items, 15 secondary items, and 59 tertiary items (Table S1). Based on these results, the working group then held a face-to-face meeting to aggregate all potential items and remove duplicates. After further discussion, we developed an initial list of items, including clinical applicability evaluation domains, evaluation items, scoring scheme, and general information of the evaluator items.

In the second phase, the working group invited 89 assessors (methodological experts, clinicians, pharmacists, nurses and hospital administrators) to select the evaluation items and draw up the scoring scheme by two rounds of modified Delphi surveys. In the first round, the assessors scored each item in importance and understandability using a 5-point Likert scale ranging from 1 ("completely unimportant"/"completely can't understand") to 5 ("very important"/"easy to understand").¹¹ At the same time, we asked assessors to clarify the importance, familiarity, and judgment of each item in availability, readability, acceptability, feasibility and overall applicability domains. Content validity, reliability, authority coefficient, coefficient of variation (CV), and average value were calculated according to

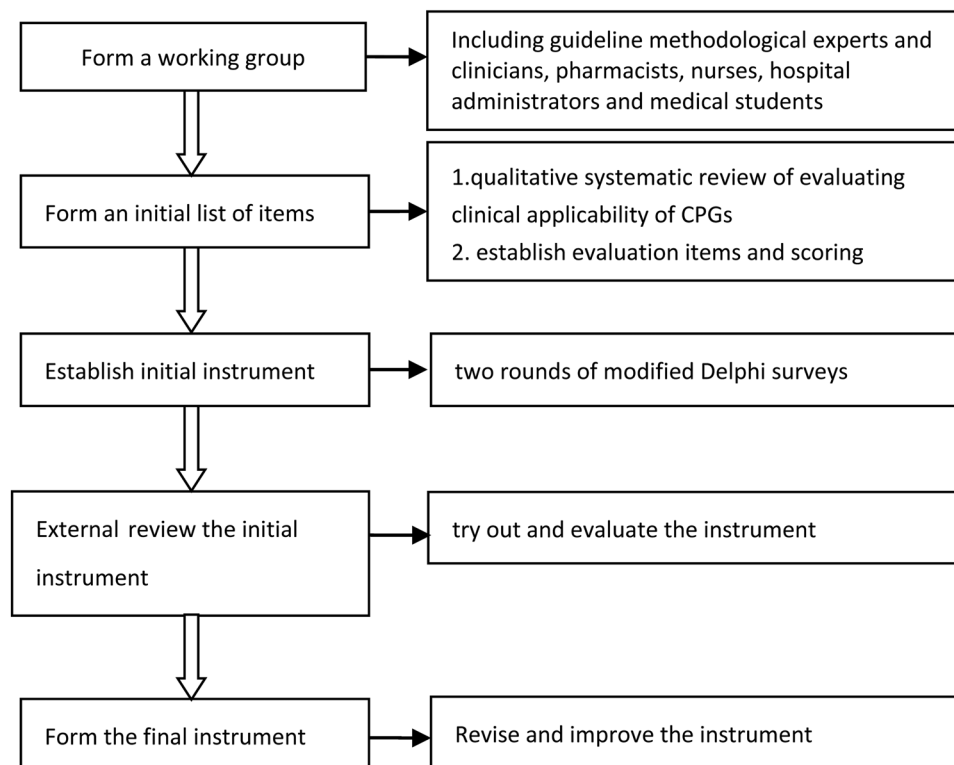


FIGURE 1 The process of establish the instrument

the assessors' evaluation results. In the second round, the assessors evaluated general information of the evaluator, evaluation items, and scoring scheme by the degree of recognition. We then developed an initial instrument based on the results of this survey.

In the last phase, the working group set up an external peer-review group, including methodological experts and guideline-users, to evaluate the instrument in five rounds of an on-site consultation. This process allowed us to improve and optimize the evaluation instrument (Figure 1).

3 | RESULTS

3.1 | Result of modified Delphi surveys

3.1.1 | Appraisal aspects

Initially, four clinical applicability appraisal aspects (availability, readability, acceptability, and feasibility) and overall evaluation were developed. The results of modified Delphi surveys suggested that the degree of disagreement in appraisal aspects was low; the degree of authority was high (Table 1).

3.1.2 | Appraisal items

The results indicated that 77% item-level content validity index, reliability index, 100% coefficient of variation, and 84.6% average value met

the standards, suggesting that the degree of disagreement in appraisal aspects was low, the consistency of expert opinions was good, and items can reflect the clinical applicability of the CPGs. After modifying 4 items and removing 12 items according to experts' opinions, the degree of recognition of all items was $\geq 60\%$, suggesting that the items were reasonable (Table 2).

3.1.3 | General information of the evaluator

The recognition degree of all general information items was $\geq 60\%$, suggesting that the set of items was basically reasonable (Table 3). Based on the opinions of experts, two new items were added: "Age" and "Type of your medical institution."

3.2 | Results of external review

The external review experts read the instrument and put forward suggestions for optimizing the instrument to make the instrument more in line with the evaluator's reading habits. For example, change "Do you think you have the professional skills required to implement this guideline?" to "How well do you have the professional skills required to implement this guideline? (for content related to your work)", "Is there any barrier to implement this guideline (in your medical institution)?" to "Is there any barrier to implement this guideline in your medical institution?" At the same time, adjust item 9 to "acceptability." According to the opinions of the external review experts, we

TABLE 1 The results of Modified Delphi surveys on applicability appraisal aspects

Appraisal aspects	Average value of importance	Coefficient of variation	Authority coefficient
Availability	4.62	0.141	0.809
Readability	4.63	0.136	0.826
Acceptability	4.60	0.126	0.827
Feasibility	4.58	0.138	0.825
Overall evaluation	4.59	0.126	0.811

developed the instrument for evaluating the clinical applicability of guidelines.

3.3 | The instrument for evaluating the clinical applicability of guidelines

The evaluating instrument for the clinical applicability of guidelines is mainly divided into three parts: (1) general information of the evaluator, (2) evaluation items, and (3) scoring scheme. The first section, basic evaluator information, includes 12 items that collect data on the evaluator's age, geographic location, education, occupation, professional title, type and level of the medical institution, familiarity with the guideline, clinical speciality, years of working, and declaration of interest. This section is used to investigate the evaluator's characteristics. The second section, evaluation items, includes 12 items that address the CPG's availability (Items 1-2), readability (Items 3-4), acceptability (Items 5-7), feasibility (Items 8-10), and overall evaluation (Items 11-12). Nine of these items are evaluated using a 5-point Likert scale, two are multiple-choice questions, and one requires free form answers. Finally, the third section is the scoring scheme, which is a standardized grading system across fields. The entire guideline evaluation instrument can be seen in Supporting Information (Table S2).

4 | DISCUSSION

We found many problems when our working group reviewed the available instruments for evaluating the clinical applicability of CPGs. For example, evaluators were not usually users of the actual guidelines, which led to the inability to accurately evaluate the clinical applicability of CPGs due to the lack of pertinence. In addition, the necessary high skill level and long reporting time meant that the compliance of evaluation instruments was poor and promotion difficult, making it difficult to use directly.

An effective CPG is easy to access, understand and accept, and results in practical recommendations. To evaluate a CPG, therefore, one must rank it according to the following: availability or how easy it is to access; readability, or how easy it is to comprehend; acceptability, or whether guideline-users agree with its recommendations; and feasibility, or how easy it is to implement its suggestions. These four areas rep-

resent different dimensions of clinical applicability and cover all stages from access to use.

4.1 | Comparison with similar instruments

In 2005, Shiffman et al¹² established the "The Guideline Implementability Appraisal (GLIA)" standard. The target evaluators are two or more people, who score each recommendation separately, and at least one subject matter expert and one implementation expert should participate. In 2011, Gagliardi et al¹³ established a "conceptual framework of implementability," and the users are guideline developers. Both instruments focus on the feasibility of developing CPG evaluation guidelines. In this study, our instrument builds on these instruments by considering not only the feasibility of a CPG but also its availability, readability, and acceptability.

In 2009, the AGREE Next Steps Consortium published the "AGREE-II instrument," which is recognized as the golden standard for international evaluation guidelines, and the potential users are health care providers, guideline developers, policymakers, and educators. However, only one area of this instrument relates to CPG's clinical applicability. Evaluation items include the following: "This guideline provides advice and/or instruments for putting recommendations into practice"; "The guideline describes factors that foster and impede its application"; "The potential resource implications of applying the recommendations have been considered"; and "The guideline presents monitoring and/or auditing criteria."¹⁴

In 2018, Li et al¹⁵ established a new scale for the evaluation of clinical practice guidelines applicability, and the users are clinicians. This instrument used items based on a 4-point Likert scale to improve an instrument's operability. However, it only collected a small amount of basic information about users: institution, major, professional title, and length of employment. It didn't consider factors such as educational background and interests, which may affect a users' use of a CPG. In addition, our instrument uses a 5-point Likert scale, unlike the 4-point scale, which increases the likelihood of a neutral score.

4.2 | Strengths and weakness

Unlike other clinical applicability evaluation instruments, this instrument uses a 5-point Likert scale. Studies have shown that after a simple

TABLE 2 The results of modified Delphi surveys on applicability evaluation items

Appraisal aspects	Evaluation items	Content validity				Importance			Understandability		
		I-CVI	K*	S-CVI/Ave	Reliability	Average value	Coefficient of variation	Average value	Coefficient of variation	Degree of recognition (%)	
Availability	Item1	0.828	0.828	0.816	0.964	4.16	0.234	4.48	0.153	98.7	
	Item2	0.750	0.750			4.22	0.190	4.28	0.182	98.7	
	Item3	0.763	0.763			3.98	0.209	4.19	0.189	98.7	
Readability	Item4	0.828	0.828			4.31	0.173	4.28	0.173	97.4	
	Item5	0.859	0.859			4.34	0.174	4.34	0.169	97.4	
Acceptability	Item6	0.859	0.859			4.38	0.174	4.39	0.164	97.4	
	Item7	0.875	0.875			4.53	0.151	4.53	0.140	96.2	
Feasibility	Item8	0.813	0.813			4.13	0.212	4.13	0.216	96.2	
	Item9	0.762	0.762			3.97	0.250	4.10	0.225	96.2	
	Item10	0.825	0.825			4.24	0.161	4.27	0.163	97.4	
	Item11	0.828	0.828			4.31	0.168	4.42	0.143	98.7	
Overall evaluation	Item12	0.825	0.825			4.30	0.174	4.42	0.165	98.7	
	Item13	0.792	0.792			4.19	0.213	4.10	0.245	98.7	

Abbreviations: I-CVI, item-level content validity index; K*, kappa; S-CVI/Ave, scale-level content validity index/average.

TABLE 3 Degree of recognition of general information items

Number	Degree of recognition (%)	Number	Degree of recognition (%)
Item 1	84.6	Item6	96.2
Item 2	100.0	Item7	98.7
Item 3	98.7	Item8	100.0
Item 4	100.0	Item9	98.7
Item 5	97.4	Item10	93.6

data conversion, data collected using different point rating scales are similar in average, variation, skewness, and kurtosis.¹⁶ However, the 5-point scale makes it easier to give each score an accurate meaning, thereby improving evaluators' understanding as well as the response rate and quality.¹⁷

We consider the overall evaluation separately instead of developing a general score using the sums of the scores in each area. There are two reasons. First, the weights of each dimension may be different: direct summation may not accurately reflect this weighted influence. Second, this allows us to verify the relevance of the evaluation field and the clinical applicability of items to determine whether the instrument is perfect.¹⁸

This instrument adopts standardized score processing for scores in each field. When the number of evaluation indicators in each field is different, then the field scores can be standardized to the same level (percentage system), similar to AGREE II.¹⁴ The instrument can therefore simultaneously compare scores in different areas of the same guidelines, scores in the same field of different guidelines, and overall clinical applicability scores for different guidelines, allowing for a more intuitive comparison of the same indicator across different guidelines and areas.

There are some weaknesses in our research. First, the qualitative systematic review of instruments for evaluating CPGs clinical applicability only included evidence published in Chinese and English. Second, considering the maneuverability, the scoring part of the guideline didn't use the centesimal system, and a certain amount of data accuracy is lost. Third, due to time and labor cost constraints, the study failed to incorporate more guidelines for empirical research.

5 | CONCLUSIONS

In this article, we have developed an instrument for guideline-users with reference to existing international standards and methods, which can be used to evaluate the clinical applicability of CPGs. The instrument collects data on evaluator information, the CPGs, and scoring to offer an accurate and transparent way of evaluating these guidelines. This instrument remains an initial version: we will revise and improve the instrument every 3–5 years based on feedback from guideline-users, CPG developers, and data statisticians. We foresee that this instrument will be used to evaluate multiple guidelines, analyze the reliability and validity of CPGs, verify the feasibility of the CPG system, and establish a model for data analysis and evaluation.

ACKNOWLEDGMENT

The authors thank the persons who responded to the modified Delphi surveys for their thoughtful comments.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ORCID

Yi Qiusha  <https://orcid.org/0000-0001-8096-0228>

Kang Deying  <https://orcid.org/0000-0003-1488-7047>

REFERENCES

- Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E. *Clinical Practice Guidelines We Can Trust*. Washington, D.C.: National Academies Press; 2011.
- Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 14. Reporting guidelines. *Health Res Policy Syst*. 2006;4:26.
- The National Institute for Health and Clinical Excellence (NICE). The guidelines manual. Available at: <https://www.nice.org.uk/process/pmg6/chapter/introduction> (accessed 12 May 2019).
- Agency for Healthcare Research and Quality. About NGC and NQMC. Available at: <https://www.ahrq.gov/gam/about/index.html> (accessed 12 May 2019).
- Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis. *BMJ*. 2014;348:g37225.
- Pronovost PJ. Enhancing physicians' use of clinical guidelines. *JAMA*. 2013;310:2501-2502.
- Mehndiratta A, Sharma S, Gupta NP, et al. Adapting clinical guidelines in India—a pragmatic approach. *BMJ*. 2017;359:j5147.
- ADAPTE. Guideline adaptation: a resource toolkit, 2010. Available at: <https://www.g-i-n.net/document-store/working-groupsdocuments/adaptation/adapte-resource-toolkit-guidelineadaptation-2-Opdf/view> (accessed 29 Jan 2018).
- Gagliardi AR, Brouwers MC. Do guidelines offer implementation advice to target users? A systematic review of guideline applicability. *BMJ Open*. 2015;5:e007047.
- Lin M, Zeng LN, Li JL, et al. Global guideline clinical applicability evaluation tool: a systematic review. *Chin J Evid-Based Med*. 2020;20(4):437-443.
- Geng YX, Zhao LY, Wang Y, et al. Competency model for dentists in China: results of a Delphi study. *PLoS One*. 2018;13(3):e0194411.
- Shiffman RN, Dixon J, Brandt C, et al. The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. *BMC Med Inform Decis Mak*. 2005;5(1):23.
- Gagliardi AR, Brouwers MC, Palda VA, et al. How can we improve guideline use? A conceptual framework of implementability. *Implement Sci*. 2011;6(1):26.

14. Agree Collaboration. Appraisal of guidelines for research and evaluation in Europe II (AGREE II) instrument. 2009. Available from: <http://www.agreetrust.org/> (accessed 12 May 2019).
15. Li H, Xie R, Wang Y, et al. A new scale for the evaluation of clinical practice guidelines applicability: development and appraisal. *Implement Sci.* 2018;13(1):61.
16. Dawes JG. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *Int J Market Res.* 2008;50(50):61-77.
17. Babakus E, Mangold WG. Adapting the SERVQUAL scale to hospital services: an empirical investigation. *Health Serv Res.* 1992;26(6):767.
18. The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical prac-

tice guidelines: the AGREE project. *Qual Saf Health Care.* 2003;12:18-23.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Linan Z, Qiusha Y, Chuan Z, et al. An instrument for evaluating the clinical applicability of guidelines. *J Evid Based Med.* 2021;14:75–81.
<https://doi.org/10.1111/jebm.12416>