

Developing a Scoring Guide for the Appraisal of Guidelines for Research and Evaluation II Instrument in Korea: A Modified Delphi Consensus Process

You Kyoung Lee,¹ Ein Soon Shin,²
Jae-Yong Shim,³ Kyung Joon Min,⁴
Jun-Mo Kim,⁵ Sun Hee Lee,⁶
On behalf of the Executive Committee
for CPGs, The Korean Academy of
Medical Sciences

¹Department of Laboratory Medicine and Genetics, Soonchunhyang University College of Medicine, Bucheon; ²Department of Preventive Medicine & Public Health, Ajou University School of Medicine, Suwon; ³Department of Family Medicine, Gangnam Severance Hospital, Yonsei University Health System, Seoul; ⁴Department of Psychiatry, Chung-Ang University College of Medicine, Seoul; ⁵Department of Urology, Soonchunhyang University College of Medicine, Bucheon; ⁶Department of Preventive Medicine, Ewha Womans University School of Medicine, Seoul, Korea

Received: 2 October 2012

Accepted: 7 December 2012

Address for Correspondence:

Ein Soon Shin, PhD

Department of Preventive Medicine & Public Health,
Ajou University School of Medicine, 164 Worldcup-ro,
Yeongtong-gu, Suwon 443-721, Korea

Tel: +82.31-219-5147, Fax: +82.31-219-5084

E-mail: shin2738@ajou.ac.kr

This study was supported by a 2011 research grant for health policy from the Ministry of Health and Welfare, Republic of Korea.

INTRODUCTION

The development of clinical practice guidelines (CPGs) in Western countries started in the 1980s and increased rapidly in the 1990s. The number of publications related to CPGs recently exceeded 1,000 per year (1). In Korea, an increased interest and emphasis on the development of CPGs has been presented in the healthcare community (2). About 54 CPGs has been developed in Korea according to the survey of the Korean Academy of Medical Sciences (KAMS) in 2006 (3). However, only 33 CPGs offer full-text service at this point in time on Korean Medical Guideline Information Center (3). And the methodology and content of Korean CPGs are not evaluated as systematically as are those in Western countries (2, 4). A survey of the recognition and use of CPGs found that 73.4% of respondents recognized

Korea has a relatively short history in the development and use of clinical practice guidelines (CPGs). Additionally, it has been difficult to employ the Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument due to the lack of consensus and the presence of differences in Korean medical settings and in the Korean socio-cultural environment. An AGREE II scoring guide was therefore developed to reduce differences among evaluators using the same tool. In consideration of the importance of using a quantitative measure of satisfaction with the elements described in the AGREE II manual, a final draft was developed through a *Delphi* consensus process. Ninety-two draft scoring guides for anchor points 1, 3, 5, and 7 (full score) in 23 items were developed. Consensus was defined as agreement among at least 70% of the raters. Agreement on 88 draft scoring guidelines was reached in the first *Delphi* round, and agreement for the remaining four was achieved in the second round. The development of an AGREE II scoring guide in this study is expected to contribute to improving the CPG environment.

Key Words: Clinical Practice Guideline; Peer Review

the existence of CPGs, but only 53.3% used the CPGs in actual practice (5). These results reflect the short history of CPGs in Korea and the absence of a consensus about the development and use of these CPGs.

The Appraisal of Guidelines for Research and Evaluation (AGREE) was first introduced in 2003, and the supplemented the AGREE II instrument (6), a 23-item instrument addressing six quality domains, was published in 2009; it is the most widely used quality evaluation tool for CPGs (7). In Korea, the original AGREE instrument, which was a translation, was distributed in 2006 (8); the AGREE II instrument was translated into Korean and introduced to Korean medical societies in 2010 (9). This evaluation tool is used not only to evaluate the quality of medical practice guidelines but also to contribute to the development of a method for refining CPGs and providing information rele-

vant to the content and wording of CPGs (6).

We do not have as much experience as do those in Western countries with the development of CPGs or the use of the AGREE instrument. One reason for the difficulties in the application of the AGREE instrument is lack of consensus, and a report issued by the KAMS in 2009 noted that efforts to reduce misunderstandings about AGREE items among evaluators are needed (5). The KAMS Executive Committee for CPGs (ECC) discussed the possibility of developing an AGREE II instrument scoring guide to resolve ambiguities. Accordingly, the ECC sought to develop Korean-style criteria that considered differences in the medical environments and cultures of Korea and the West while not sacrificing the need for “how-to” information regarding rating each item on the AGREE II. The ECC attempted to propose a methodology for and provide information about optimal editing practices as well as to facilitate achievement of the original goals of AGREE II with respect to evaluating the quality of CPGs.

MATERIALS AND METHODS

The ECC is an expert panel established to develop CPGs and evaluate guidelines using the AGREE instrument; this body is responsible for developing a domestic CPG evaluation system, providing an educational program for CPG developers, and disseminating information about the methodology used to develop CPGs. Each item on the AGREE II specifies a goal, a method of description, and the elements required to meet that goal. These are presented under the following headings: “User’s Manual Description,” “Where to Look,” and “How to Rate,” respectively (6). Those charged with evaluating the guidelines should thoroughly understand all content before grading the CPGs. These individuals rate the guidelines from 7 to 1 according to their satisfaction with the requirements. Based on the sections addressing “Where to Look” and “How to Rate,” the ECC defined a rating of “7” as “strongly agree” and proposed four anchor points (1, 3, 5, and 7); ratings of 2, 4, and 6 could be given at the discretion of the peer reviewer. In terms of the three anchor points and the full score, a draft that addressed the importance of and quantitatively measured satisfaction with the elements and the final document emerged from a *Delphi* consensus process.

Development of the draft scoring guide

A focus group consisting of an internist, a urologist, a clinical pathologist, a psychiatrist, a family physician, and an evidence-based medicine (EBM) methodologist, all of whom had great interest in and experience with CPGs and the AGREE instrument, was established. This group generated the draft scoring guide for scores of 1, 3, 5, and 7 for each of the AGREE II items.

Modified *Delphi* consensus process

We used a modified *Delphi* consensus process to avoid domi-

nation by individual views in open discussions. Based on a structured questionnaire, levels of agreement and disagreement during the *Delphi* consensus process were expressed in terms of a nine-point Likert scale. Agreement was rated as 7 or more and disagreement as 3 or less. Consensus was defined as more than 70% agreement.

Experts on the modified *Delphi* consensus process were selected from among ECC members, clinical practitioners who had participated in the development of CPGs during the past 1-2 yr, and experts in EBM methodology. A total of 13 people participated in the process.

In the first round, the draft scoring guide developed by the focus group was distributed to all participants via email. Participants were presented with a copy of the Korean version of the AGREE II instrument, which included the 92 draft scoring guides. They were asked to review the document and, using the questionnaire, to indicate their level of agreement on a scale ranging from strongly disagree 1 to strongly agree 9. Participants also had the opportunity to provide written feedback.

After collecting the completed survey sheets, we analyzed data for each scoring guide separately to determine whether consensus had been reached. Scoring guides were considered complete when consensus had been reached. In the absence of consensus in the first round, a second round was executed. In the second round, participants received structured *Delphi* questionnaires targeted to the particular foci of this round. Participants reviewed the original draft scoring guides, the data from the first consensus round, including information about their own data relative to those from other participants, and the modified scoring guides. Participants were then asked to rate the modified scoring guides using a nine-point Likert scale. *Delphi* rounds were repeated until consensus was reached.

RESULTS

Round 1

In the first round, all 13 experts returned the survey sheets, yielding a response rate of 100%. Consensus was reached on 88 of the 92 scoring guides for the 23 items (Table 1). The experts participating in the consensus process were asked to propose modifications to the scoring guides when they did not agree, and the focus group revised the scoring guides that did not achieve consensus during the first round in the light of these modifications. These were then reviewed by the focus group during the second round (Table 2).

Round 2

Thirteen experts participated in the second round, and the response rate was 100%. Agreement was reached with regard to the four scoring guides that did not achieve consensus in the first round. The 92 scoring guides developed through the modi-

Table 1. First-round agreement on each anchor point of the AGREE items. Consensus was defined as more than 70% agreement. The anchor point 3 on items 2, 5, 11, and 12 failed to reach consensus (bold italic font)

AGREE item No.	Anchor point	Agree (7-9)	AGREE item No.	Anchor point	Agree (7-9)	AGREE item No.	Anchor point	Agree (7-9)
1	1	13 (100)	9	1	13 (100)	17	1	13 (100)
	3	12 (92.3)		3	13 (100)		3	11 (84.6)
	5	13 (100)		5	13 (100)		5	12 (92.3)
	7	13 (100)		7	13 (100)		7	13 (100)
2	1	10 (76.9)	10	1	13 (100)	18	1	12 (92.3)
	3	8 (61.5)		3	11 (84.6)		3	12 (92.3)
	5	12 (92.3)		5	12 (92.3)		5	12 (92.3)
	7	12 (92.3)		7	13 (100)		7	12 (92.3)
3	1	11 (84.6)	11	1	13 (100)	19	1	12 (92.3)
	3	10 (76.9)		3	9 (69.2)		3	12 (92.3)
	5	13 (100)		5	11 (84.6)		5	12 (92.3)
	7	12 (92.3)		7	13 (100)		7	13 (100)
4	1	13 (100)	12	1	10 (76.9)	20	1	11 (84.6)
	3	11 (84.6)		3	9 (69.2)		3	10 (76.9)
	5	12 (92.3)		5	11 (84.6)		5	10 (76.9)
	7	13 (100)		7	12 (92.3)		7	11 (84.6)
5	1	11 (84.6)	13	1	13 (100)	21	1	11 (84.6)
	3	9 (69.2)		3	13 (100)		3	11 (84.6)
	5	11 (84.6)		5	13 (100)		5	11 (84.6)
	7	11 (84.6)		7	13 (100)		7	11 (84.6)
6	1	13 (100)	14	1	13 (100)	22	1	10 (76.9)
	3	10 (76.9)		3	11 (84.6)		3	12 (92.3)
	5	11 (84.6)		5	12 (92.3)		5	12 (92.3)
	7	11 (84.6)		7	13 (100)		7	12 (92.3)
7	1	12 (92.3)	15	1	12 (92.3)	23	1	12 (92.3)
	3	13 (100)		3	13 (100)		3	11 (84.6)
	5	12 (92.3)		5	11 (84.6)		5	12 (92.3)
	7	13 (100)		7	13 (100)		7	12 (92.3)
8	1	13 (100)	16	1	11 (84.6)			
	3	10 (76.9)		3	12 (92.3)			
	5	11 (84.6)		5	11 (84.6)			
	7	12 (92.3)		7	12 (92.3)			

Table 2. First and revised drafts of the four scoring guidelines that failed to reach consensus in the first Delphi round

AGREE item	Score	Scoring guide
2	First Draft of Point 3	If it can be inferred that this is a clinical question although the subtitle consists of keywords only.
	Revised Draft of Point 3	If the question is presented with minimal information (e.g., subtitle etc.)
5	First Draft of Point 3	When the perspective (experiences and expectations) and the preference of the target group are described or can be estimated based on the subjective experiences of the developer.
	Revised Draft of Point 3	Although the perspective (experiences and expectations) and the preference of the target group are described, the survey method is not described.
11	First Draft of Point 3	When evidence or detailed data on benefits, side effects, and factors that are hazardous to health are not presented, and they are not sufficiently addressed in the recommendations.
	Revised Draft Point 3	When evidence or detailed data on benefits, side effects, and factors hazardous to health are not presented, and only some aspects are addressed in the recommendations.
12	First Draft of Point 3	When the connection between the recommendations and the supporting evidence is insufficient.
	Revised Draft of Point 3	When only some of the recommendations are related to the supporting evidence.

fied *Delphi* process are available on the website of the Korean Medical Guideline Information Center (<http://www.guideline.or.kr>).

DISCUSSION

The AGREE II evaluates the quality of CPGs, addresses how and what to present in published guidelines, and provides a methodology for the development of CPGs (6). It is an important tool

for education regarding the development of CPGs; it also allows developers to understand the strengths and weaknesses of their guidelines when evaluating their own or others' guidelines and to update their guidelines accordingly.

Although the history of CPGs in Korea is short, demand for the development of CPGs has increased. Indeed, nearly 120 guidelines are being planned by professional member societies of the KAMS (10, 11). Presently, CPG developers in Korea typically use an "adaptation process." Differences in the interpreta-

tions of evaluators during the process of applying the AGREE instrument in evaluations of the quality of existing CPGs is among the most difficult problems faced by CPG developers. For example, an AGREE II evaluation was performed on the 10 draft guidelines developed with regard to the CPGs for stomach cancer in Korea in 2011. Significant deviations were observed during this process, such as differences of 2 points or more on 6.6 ± 3.5 (mean \pm SD) items (unpublished data), which reflect problems in the use of the AGREE instrument in a Korean situation. The reasons that Korean CPG developers and evaluators experience relatively greater difficulty using the AGREE tool include the absence of consensus among medical professionals regarding the relative levels of importance of the elements listed in the “how-to-rate” guidelines in the AGREE II manual, differences among medical environments, socio-cultural differences between Korea and Western nations, and subjective interpretations of each of the questions. Although AGREE is a widely accepted tool in the field of CPGs evaluation, it has the disadvantage that it can be affected by the subjective perceptions of evaluators (12-14). In 2010, Dans and Dans (15) noted that AGREE II items demand that activities be “described well” rather than be “be performed well,” which causes confusion about the purpose of the evaluation and, ultimately, about the grading. As the Korean medical community does not yet have sufficient experience with CPGs, differing interpretations and understandings among evaluators constitute major obstacles. In Korea, almost every participant in the development and evaluation of CPGs is a medical doctor who has majored in medicine and has experience in clinical practice. Thus, majority of evaluators considered the quality of expected performance of recommendations in addition to the quality of description in the evaluation process. This may be another reason for the major differences among evaluators using the AGREE II instrument. Thus, it is expected that provision of an AGREE II scoring guide will facilitate the achievement of consensus about the purpose of and the approach to the development of CPGs.

ECC attempted to provide anchor points for scores of 1, 3, 5, and 7 based on importance and a quantitative measure of satisfaction after agreeing on the standards for a seven point scale. These were based on the “How to Rate” instructions for the AGREE II instrument because these provided a good description of the purpose and the content of each item. In the first round of the *Delphi* consensus process, however, we could not reach consensus on the anchor point 3 for AGREE II item numbers 2, 5, 11, and 12 (Table 1). Thus, it was difficult to identify four phased anchor points based on the “How to Rate” guidelines alone. In these cases, we tried to provide standards that considered being “be performed well.”

To identify problems, the guide was applied to several recently developed domestic CPGs. Although no major differences among evaluators were observed, the data are not reported here

because too few guidelines were involved. This scoring guide will be further organized and modified in the process of actually applying it to diverse CPGs, and it will be revised to reflect further developments in CPGs and medical environments in Korea. The scoring guide for the AGREE II instrument proposed herein can be used to evaluate previously developed CPGs in Korea and is a useful foundation for the creation of new CPGs in the future.

ACKNOWLEDGMENTS

We thank the ECC members who participated in the *Delphi* process and provided valuable comments about the Korean scoring guide for the AGREE II instrument. This undoubtedly improved the quality of our work and we thank YK Song, HT Kang, HK Jung, JM Park, KJ Lee, W Joo, SM Lim, and KH Seo. The authors have no conflicts of interest to disclose.

Author contributions

YK Lee, SH Lee, and ES Shin conceived the idea for this research and designed the study with J Shim, K Min, and J Kim. All authors participated in the planning and processing of the *Delphi* consensus process and in the interpretation of data. YK Lee wrote the first draft of the report, and J Shim, J Kim, and ES Shin made important intellectual contributions. All other authors commented on the draft and approved the final version.

REFERENCES

- Alonso-Coello P, Irfan A, Solà I, Gich I, Delgado-Noguera M, Rigau D, Tort S, Bonfill X, Burgers J, Schunemann H. *The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. Qual Saf Health Care* 2010; 19: e58.
- Ahn HS, Kim HJ. *Development and implementation of clinical practice guidelines: current status in Korea. J Korean Med Sci* 2012; 27: S55-60.
- Korean Medical Guideline Information Center. Available at <http://www.guideline.or.kr> [accessed on 16 November 2012].
- Ministry of Health & Welfare, The Korean Academy of Medical Sciences. *A process model for developing clinical practice guidelines in Korea and its practical application for sample CPGs. Available at http://www.guideline.or.kr/about/board/list.php?code=pds2* [accessed on 16 August 2012].
- Ministry of Health & Welfare, The Korean Academy of Medical Sciences. *Clinical practice guideline development and establishment of informative systems. Available at http://www.guideline.or.kr/about/board/list.php?code=pds2* [accessed on 25 September 2012].
- AGREE Next Steps Consortium. *The AGREE II Instrument: Electronic version. Available at http://www.agreetrust.org.* [accessed on 4 May 2012].
- Oxman AD, Schünemann HJ, Fretheim A. *Improving the use of research evidence in guideline development: 16. Evaluation. Health Res Policy Syst* 2006; 4: 28.
- Ahn HS, Kim SY, Kim NS, Park MH. *Appraisal of Guidelines for Research and Evaluation (Korean version). Available at http://www.guideline.*

- [or.kr/about/board/viewbody.php?code=pds2&page=1&number=31&keyfield=&key=](http://www.guideline.or.kr/about/board/viewbody.php?code=pds2&page=1&number=31&keyfield=&key=) [accessed on 25 September 2012].
9. Lee SH, Lee YK, Shin IS. *Appraisal of guidelines for research and evaluation II (Korean version)*. Available at <http://www.agreetrust.org/resource-centre/agree-ii-translations/> [accessed on 16 February 2012].
 10. Ministry of Health & Welfare, The Korean Academy of Medical Sciences. *Korean Clinical Practice Guidelines: Designing a model system for development, education, propagation and a web-based AGREE II evaluation system*. Available at <http://www.guideline.or.kr/about/board/viewbody.php?code=pds2&page=1&number=29&keyfield=&key=> [accessed on 14 September 2012]
 11. Ministry of Health & Welfare, The Korean Academy of Medical Sciences. *Study on the propagation of clinical practice guidelines and a plan to invigorate the KOMGI*. Available at <http://www.guideline.or.kr/about/board/list.php?code=pds> [accessed on 14 September 2012].
 12. De Hert M, Vancampfort D, Correll CU, Mercken V, Peuskens J, Sweers K, van Winkel R, Mitchell AJ. *Guidelines for screening and monitoring of cardiometabolic risk in schizophrenia: systematic evaluation*. *Br J Psychiatry* 2011; 199: 99-105.
 13. Stone MA, Wilkinson JC, Charpentier G, Clochard N, Grassi G, Lindblad U, Müller UA, Nolan J, Rutten GE, Khunti K; GUIDANCE Study Group. *Evaluation and comparison of guidelines for the management of people with type 2 diabetes from eight European countries*. *Diabetes Res Clin Pract* 2010; 87: 252-60.
 14. Delgado-Noguera M, Tort S, Bonfill X, Gich I, Alonso-Coello P. *Quality assessment of clinical practice guidelines for the prevention and treatment of childhood overweight and obesity*. *Eur J Pediatr* 2009; 168: 789-99.
 15. Dans AL, Dans LF. *Appraising a tool for guideline appraisal (the AGREE II instrument)*. *J Clin Epidemiol* 2010; 63: 1281-2.