

Original Article
Medicine General &
Social Medicine



Determinants Used to Justify the Strength of Recommendations among Korean Clinical Practice Guidelines

Ein-Soon Shin ,^{1*} Ji-Eun Jang ,^{2*} Ji-Yun Yeon ,¹ Da-sol Kim ,¹
Sung-Goo Chang ,³ and Yoon-Seong Lee ⁴

¹Research Agency for Clinical Practice Guidelines, Research Center, Korean Academy of Medical Sciences, Seoul, Korea

²Department of Medical Sciences, Ewha Womans University Graduate School, Seoul, Korea

³Department of Urology, Kyung Hee University School of Medicine, Seoul, Korea

⁴Department of Forensic Medicine, Seoul National University College of Medicine, Seoul, Korea

OPEN ACCESS

Received: Oct 23, 2017

Accepted: Dec 26, 2017

Address for Correspondence:

Ji-Eun Jang, PhD

Department of Medical Sciences, Ewha
Womans University Graduate School,
1071 Anyangcheon-ro, Yangcheon-gu,
Seoul 07985, Korea.

E-mail: jieun88@ewhain.net

*Ein-Soon Shin and Ji-Eun Jang contributed
equally to this work.

© 2018 The Korean Academy of Medical
Sciences.

This is an Open Access article distributed
under the terms of the Creative Commons
Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>)
which permits unrestricted non-commercial
use, distribution, and reproduction in any
medium, provided the original work is properly
cited.

ORCID iDs

Ein-Soon Shin
<https://orcid.org/0000-0002-5086-6086>

Ji-Eun Jang
<https://orcid.org/0000-0002-0691-0289>

Ji-Yun Yeon
<https://orcid.org/0000-0002-6029-910X>

Da-sol Kim
<https://orcid.org/0000-0002-2498-699X>

Sung-Goo Chang
<https://orcid.org/0000-0003-0733-7266>

Yoon-Seong Lee
<https://orcid.org/0000-0002-1899-3632>

ABSTRACT

Background: A standardized systematic approach to grade evidence and the strength of recommendations is important for guideline users to minimize bias and help interpret the most suitable decisions at the point of care. The study aims to identify and classify determinants used to make judgement for the strength of recommendations among 56 Korean clinical practice guidelines (CPGs), and explore strong recommendations based on low quality of evidence.

Methods: Determinants used in the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach among 34 CPGs which have reported both strength of recommendations and level of evidence were reviewed.

Results: Five of 34 CPGs (14.7%) considered quality of evidence, benefits and harms, patients' values and preferences, and costs. And 24 of 34 CPGs (70.6%) considered both magnitude of effect and feasibility as additional determinants. Judgement table was not widely provided for use to translate evidence into recommendations. Eighty-two of 121 recommendations (67.8%, ranged 20.0% to 100.0%) among 11 CPGs using the same judgement scheme showed 'strong' strength of recommendations based on low or very low quality of evidence. Among 5 paradigmatic situations that justify strong recommendations based on low or very low evidence, situation classified as 'potential equivalence, one option clearly less risky or costly' was 87.8% for 82 strong recommendations. Situation classified as 'uncertain benefit, certain harm' was 4.9%.

Conclusion: There is a need to introduce and systematize an evidence-based grading system. Using judgement table to justify the strength of recommendations and applying the 5 paradigmatic situations mentioned above is also recommended in the near future.

Keywords: Clinical Practice Guidelines; GRADE Approach; Strength of Recommendations; Quality of Evidence; Judgement Table

INTRODUCTION

To minimize bias and help interpret the most suitable decisions at the point of care for users of clinical practice guidelines (CPGs), a standardized systematic approach to grade evidence and the strength of recommendations is important.¹ However, because several grading systems have

Disclosure

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Shin ES, Chang SG, Lee YS.
Data curation: Shin ES, Jang JE, Yeon JY, Kim DS. Formal analysis: Shin ES, Jang JE, Yeon JY, Kim DS. Writing - original draft: Shin ES, Jang JE. Writing - review & editing: Shin ES, Jang JE, Chang SG, Lee YS.

shown inconsistency and wide variations to grade quality of evidence and recommendations,^{2,3} it is difficult for guideline users to understand the message. Different grading systems include Grading of Recommendations Assessment, Development and Evaluation (GRADE), National Institute for Health and Clinical Excellence (NICE), Scottish Intercollegiate Guideline Network (SIGN), and Centre for Evidence-Based Medicine, Oxford (CEBM) are used to give a clinical recommendation in a CPG which have an influence on the clinician's point of view and can significantly change clinician's decisions and clinical outcomes.¹

The GRADE is an emerging consensus on rating quality of evidence and strength of recommendations. GRADE provides an explicit, comprehensive, transparent, and pragmatic approach⁴⁻⁶ and is increasingly being adopted by organizations worldwide.⁷⁻⁹ Separating the judgements regarding the quality of evidence from judgements about the strength of recommendations is a critical and defining feature of this new grading system.^{10,11} The GRADE approach also provides a framework to move from evidence to the recommendation.¹²⁻¹⁵ The strength of recommendations¹⁶⁻¹⁸ depends on estimates of magnitude of effect,¹⁹⁻²¹ estimates of values and preference and their variability,²² confidence in each of these estimates, and resource use^{23,24} considerations.

GRADE guidance allows five situations that justify strong recommendations in the face of low or very low confidence in estimates for critical outcomes.^{7,25} Despite GRADE guidance's warning against the transparent approach, World Health Organization (WHO) guidelines showed that approximately one-half of the recommendations were based on low or very low confidence.²⁶⁻²⁸ It raises concerns about whether GRADE is being optimally applied in the WHO guideline development process.

In Korea, few CPGs are trying to adopt GRADE.²⁹⁻³³ They considered little bit different components to grade recommendations and utilize varied grading systems. Some guidelines considered only an evidence and there was lack of information on the resource use and patients' value and preferences mostly. There has been no review of factors or grading systems used to grade recommendations in Korea. Therefore, it is necessary to obtain a clear rationale for the adoption of the GRADE approach which is currently accepted worldwide through the analysis of the current situation, particularly the situation where strong recommendations are based on low evidence.

The study aims to identify and classify determinants used to make judgement for the strength of recommendations among 56 Korean CPGs, and explore strong recommendations based on low quality of evidence.

METHODS

Data source

To identify determinants used to make judgement for the strength of recommendations in Korea, 56 CPGs developed based on the academic societies that are listed in the Korean Medical Guideline Information Center (<http://www.guideline.or.kr/>) were included. Of the 56 CPGs, 27 (48.2%) were developed between 2001 and 2010, and 29 (51.8%) were developed between 2011 and the end of June 2017. Fifty-six CPGs were developed in a total of 33 academic societies; 22 of 33 academic societies (66.7%) developed 1 CPG, 6 of 33 (18.2%) developed 2 CPGs, and 5 of 33 (15.1%) developed 3 or more CPGs.

Data abstraction process

We conducted 3-step process. First, 56 Korean CPGs were classified by reporting level of evidence and/or the strength of recommendations. Second, determinants used to make judgement for the strength of recommendations among 34 CPGs were identified (**Appendix 1**). The following determinants used GRADE approach worldwide is included: quality of evidence, balance between benefits and harms, patients' values and preferences, and costs (resource use). And additional determinants include absolute magnitude of effect and feasibility (**Appendix 2**). To collect data on feasibility, we considered two aspects, feasible to implement and whether to consider barriers. Third, to explore the strength of recommendations based on low or very low quality of evidence, 11 CPGs using the same judgement scheme (strong, weak) were reviewed.²⁹⁻³⁹ We will compare a summary on the strength of recommendations by low or very low quality of evidence with WHO CPGs.^{7,26} Fourth, classification by 5 paradigmatic situations that justify strong recommendations based on low or very low certainty of the evidence in GRADE approach was performed. In case of 'life-threatening situations and potential equivalence' and 'one option is clearly less risky or costly,' the strength of the recommendation in favor of the intervention can be given a strong. In case of 'uncertain benefit with certain harm,' 'potential catastrophic harm,' and 'high similar benefits, one option potentially more risk or costly,' the strength of the recommendation against the intervention can be given a strong (**Appendix 3**).

RESULTS

Classification on reporting for the strength of recommendations and/or level of evidence among 56 Korean CPGs were as follows: 18 of 56 CPGs (32.1%) did not report both strength of recommendations and level of evidence, 2 CPG (3.6%) reported level of evidence only, 2 CPG (3.6%) reported strength of recommendations only, and 34 CPGs (60.7%) reported both strength of recommendations and level of evidence (**Table 1**).

Determinants used in the GRADE approach and judgement table used to justify the strength of recommendations among 34 CPGs which have reported both strength of recommendations and level of evidence were as follows: 12 of 34 CPGs (35.3%) considered only quality of evidence, and 12 CPGs (35.3%) considered quality of evidence and benefits and harms. Five of 34 CPGs (14.7%) considered quality of evidence, benefits and harms, patients' values and preferences, and costs such as GRADE approach. And 24 of 34 CPGs (70.6%) considered both magnitude of effect and feasibility as additional determinants (**Table 2**). Judgement table was not widely provided for use to translate evidence into recommendations. There was inconsistency between information in method part and information in background on the determinants for making judgement the strength of recommendations (**Appendix 1**).

Table 1. Classification of reporting for the strength of recommendations and/or level of evidence among 56 Korean CPGs

Reporting items	Number of CPGs (%)
Reported quality or level of evidence only	2 (3.6)
Reported strength of recommendations only	2 (3.6)
Reported neither strength of recommendations nor quality or level of evidence	18 (32.1)
Reported both strength of recommendations and quality or level of evidence	34 (60.7)
Total	56 (100.0)

CPG = clinical practice guideline.

Table 2. Determinants and judgement table used to justify the strength of recommendations

Considered determinants by GRADE approach ^a	Additional determinants ^a	Number of CPGs (%)	Provide judgement table to justify the strength of recommendations
Quality or level of evidence only (n = 12)	None	1 (2.9)	No
	Magnitude of effect	6 (17.7)	No
	Magnitude of effect and feasibility	5 (14.7)	No
Quality or level of evidence and benefits and harms (n = 12)	Magnitude of effect	2 (5.9)	No
	Magnitude of effect and feasibility	10 (29.4)	No
Quality or level of evidence, benefits and harms, and patients' value and preferences (n = 3)	Magnitude of effect and feasibility	3 (8.9)	No
Quality or level of evidence, benefits and harms, and cost (n = 2)	Magnitude of effect	1 (2.9)	No
	Magnitude of effect and feasibility	1 (2.9)	No
Quality or level of evidence, benefits and harms, patients' value and preferences, and cost (n = 5)	Magnitude of effect and feasibility	5 (14.7)	No
Total		34 (100.0)	

GRADE = Grading of Recommendations Assessment, Development and Evaluation, CPG = clinical practice guideline.

^aSee **Appendix 1**.

Table 3. Strength of recommendations based on low or very low quality of evidence among 11 Korean CPGs using the same judgement scheme (strong, weak or conditional)

Guidelines (n = 11)	Strength of recommendations based on low or very low quality of evidence		
	Strong	Weak (or conditional)	Total
CPG 1 (2010) ²⁹	7 (58.3)	5 (41.7)	12 (100.0)
CPG 2 (2011) ³⁴	20 (95.2)	1 (4.8)	21 (100.0)
CPG 3 (2012) ³⁵	3 (42.9)	4 (57.1)	7 (100.0)
CPG 4 (2012) ³⁶	3 (20.0)	12 (80.0)	15 (100.0)
CPG 5 (2012) ³⁰	13 (86.7)	2 (13.3)	15 (100.0)
CPG 6 (2012) ³¹	10 (71.4)	4 (28.6)	14 (100.0)
CPG 7 (2013) ³⁷	7 (100.0)	0 (0.0)	7 (100.0)
CPG 8 (2014) ³²	11 (73.3)	4 (26.7)	15 (100.0)
CPG 9 (2014) ³³	1 (25.0)	3 (75.0)	4 (100.0)
CPG 10 (2014) ³⁸	2 (33.3)	4 (66.7)	6 (100.0)
CPG 11 (2015) ³⁹	5 (100.0)	0 (0.0)	5 (100.0)
Total	82 (67.8)	39 (32.2)	121 (100.0)

Values are presented as number (%).

CPG = clinical practice guideline.

Strength of recommendations based on low quality of evidence among 11 Korean CPGs using the same judgement scheme (strong, weak, or conditional) were as follows: 82 of 121 recommendations (67.8%, ranged 20.0% to 100.0%) showed 'strong' strength of recommendations based on low or very low quality of evidence. Six of 11 CPGs (54.5%) made judgement 'strong' for more than 70% of recommendations based on low or very low quality of evidence (**Table 3**). The determinants used to make judgement varied among 11 CPGs (**Appendix 1**). **Table 4** showed comparison a summary on the strength of recommendations by low or very low quality of evidence. Sixty-seven point eight % of recommendations was given 'strong' recommendations based on low or very low quality of evidence in 11 Korean CPGs, follows 67.4% for 44 CPGs of 20 countries, 53.0% for 43 CPGs of WHO, and 48.4% for 8 CPGs (human immunodeficiency virus [HIV] and tuberculosis [TB] only) of WHO. It was ranged from 57.8% to 74.5% for low quality of evidence and ranged from 27.6% to 57.8% for very low quality of evidence.

Classification by 5 paradigmatic situations that justify strong recommendations based on low or very low certainty of the evidence is shown in **Table 5**. Situation classified as 'potential equivalence, one option clearly less risky or costly' was 87.8% for 82 strong recommendations; 86.8% (66 of 76) in low quality of evidence and 100% (6 of 6) in very low quality of evidence. Situation classified as 'uncertain benefit and certain harm' was 4.9% (4 of 82 recommendations), and 7.3% (6 of 82) was classified as other situation including no intervention.

Table 4. Comparison a summary on the strength of recommendations by low or very low quality of evidence

Nation	Number of guidelines	Quality of evidence	Strength of recommendations		
			Strong	Weak (or conditional)	Total
Korea	11	Low	76 (74.5)	26 (25.5)	102 (100.0)
		Very low	6 (31.6)	13 (68.4)	19 (100.0)
		Total	82 (67.8)	39 (32.2)	121 (100.0)
20 countries ⁷	44	Low	505 (74.4)	174 (25.6)	679 (100.0)
		Very low	283 (57.8)	207 (42.2)	490 (100.0)
		Total	788 (67.4)	381 (32.6)	1,169 (100.0)
WHO A ²⁶	43	Low	95 (60.1)	63 (39.9)	158 (100.0)
		Very low	65 (45.1)	79 (54.9)	144 (100.0)
		Total	160 (53.0)	142 (47.0)	302 (100.0)
WHO B ⁷	8 (HIV/TB)	Low	37 (57.8)	27 (42.2)	64 (100.0)
		Very low	8 (27.6)	21 (74.4)	29 (100.0)
		Total	45 (48.4)	48 (51.6)	93 (100.0)

Values are presented as number (%).

WHO = World Health Organization, HIV = human immunodeficiency virus, TB = tuberculosis.

Table 5. Classification by 5 paradigmatic situations that justify strong recommendations based on low or very low certainty of the evidence in GRADE approach^a

Quality of evidence	Strong recommendations in favor of the intervention		Strong recommendations against the intervention			Other situation (no intervention)	Total
	Life-threatening situation ^b	Potential equivalence, one option clearly less risky or costly ^c	Uncertain benefit, certain harm ^d	High similar benefits, one option potentially more risky or costly ^e	Potential catastrophic harm ^f		
Low	-	66 (86.8)	4 (5.3)	-	-	6 (7.9)	76 (100.0)
Very low	-	6 (100.0)	-	-	-	-	6 (100.0)
Total	0 (0.0)	72 (87.8)	4 (4.9)	0 (0.0)	0 (0.0)	6 (7.3)	82 (100.0)

Values are presented as number (%).

GRADE = Grading of Recommendations Assessment, Development and Evaluation.

^aSee **Appendix 3**; ^bUncertain (low or very low) benefit, immaterial (very low to high) harm, high mortality, absence of effective alternative; ^cUncertain (low or very low) benefit, certain (high or moderate) harm, similar rates of complete response in comparison with the alternatives, but high confidence of less harm, morbidity, and cost; ^dUncertain (low or very low) benefit, certain (high or moderate) harm, intervention is associated with a substantial established harm; ^eCertain (high/moderate) benefit, uncertain (low or very low) harm, potential greater harm; ^fImmaterial (very low to high) benefit, uncertain (low or very low) harm, potential important harm of the intervention.

DISCUSSION

Determinants used to justify the strength of recommendations among Korean CPGs were identified and classified. Five different grading systems each considered different factors; 1) only quality of evidence was considered, 2) quality of evidence and benefits and harms were considered, 3) quality of evidence, benefits and harms, and patients' values and preferences were considered, 4) quality of evidence, benefits and harms, and cost were considered, and 5) quality of evidence, benefits and harms, patients' values and preferences, and cost were considered. We also considered additional factors including absolute magnitude of effect and feasibility. GRADE approach has been adopted by 5 of 34 CPGs (14.7%). And 24 of 34 CPGs (70.6%) considered both magnitude of effect and feasibility as additional determinants (**Table 2**). However, we could not find judgement table with transparent explanation and evidence to justify the strength of recommendations among Korean CPGs. Template and judgement table by recommendation has been provided by WHO (**Appendix 2**). It is a summary table for translating evidence of effectiveness into recommendations. It will be helpful to the guideline end users at the point of care if guideline developers can present a judgement table to clearly explain what factors are considered important.

If CPGs do not provide a structured judgement table, it would require laborious review of the method and/or background for every recommendation listed to discover which determinants were considered. Also, there were some discrepancies between information in method part

and information in background on the determinants for making judgement the strength of recommendations (**Table 2**). To overcome these shortcomings, we should adopt a systematic and transparent GRADE approach which is used by many organizations and associations worldwide. We also reviewed current situation regarding strong recommendations based on low or very low quality of evidence among Korean CPGs. Lower percentage of 'strong' recommendations in very low quality of evidence (31.6% vs. 57.8% in 20 countries) is shown in **Table 4**.

Alexander et al.^{27,28} explained the reason why WHO guideline developers make strong recommendations inconsistent with GRADE guidance. The main reason was limitations in their understanding and optimal application of GRADE. To utilize GRADE approach consistent with GRADE guidance, it requires training of guideline developers or panelists and formal processes to maximize adherence to GRADE principles.⁷ GRADE guidance present five paradigmatic situations that justify strong recommendation based on low or very low certainty of the evidence (**Appendix 3**).^{25,27} Among 5 paradigmatic situations, situation classified as 'potential equivalence, one option clearly less risky or costly' was 87.8% for 82 strong recommendations. Situation classified as 'uncertain benefit and certain harm' was 4.9% (4 of 82). There was no recommendation classified as 'life-threatening situation' on the strong recommendations in favor of the intervention, and there was no recommendation classified as 'potential catastrophic harm,' and 'high similar benefits, one option potentially more risk or costly' on the strong recommendations against the intervention. Especially, when guideline developers have to make a decision the strong recommendation based on low evidence, it will be better to consider the five paradigmatic situations mentioned above.

Limitations for carrying out this study are as follow: First, since only 56 CPGs in the Korean Medical Guideline Information Center are analyzed among all Korean CPGs, there was lack of representative. At the end of 2013, 115 CPGs were reported to have been developed by 45 societies and organizations.⁴⁰ Second, 34 Korean CPGs were using different grading systems. Therefore, we performed data abstraction on the 11 CPGs using the same judgement scheme (strong, weak or conditional), because we could not explore judgement of strong recommendation based on low or very low level of evidence for all CPGs included in the study. Third, when extracting the factors considered in the recommendation grading, we classified them based on the information provided in the method part. However, there were many cases in which the background part of the recommendation actually provided information that considers other factors. We summarized the two kinds of information in **Appendix 1** because there are differences according to whether classification of factors considered in grading recommendation is based on information of method part or information of background part.

In conclusion, there were several grading systems with wide variations to grade quality of evidence and strength of recommendations, and strong recommendations based on low or very low quality of evidence were very common among Korean CPGs. There is a need to introduce and systematize an evidence-based grading system. It is also necessary to aggressively review, apply, and disseminate the worldwide GRADE approach that grades recommendations in consideration of important factors including quality of evidence, benefits and harms, patients' value and preferences, and cost in Korea. Before applying and disseminating the GRADE approach on rating quality of evidence and strength of recommendations in Korea, it is necessary to study what external and internal barriers are to use this grading system in advance. Judgement table was not widely provided for use to translate evidence into recommendations, and the five paradigmatic situations were not used against low evidences to decide strength of recommendations to be made.

REFERENCES

1. Cuello García CA, Pacheco Alvarado KP, Pérez Gaxiola G. Grading recommendations in clinical practice guidelines: randomised experimental evaluation of four different systems. *Arch Dis Child* 2011;96(8):723-8.
[PUBMED](#) | [CROSSREF](#)
2. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004;4(1):38.
[PUBMED](#) | [CROSSREF](#)
3. Seo KH, Lee SH, Shin ES, Lim SM, Jang JE, Jung YM, et al. Trend analysis of grading systems for level of evidence and strength of recommendation. *J Korean Med Assoc* 2011;54(7):758-68.
[CROSSREF](#)
4. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.
[PUBMED](#) | [CROSSREF](#)
5. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, et al. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005;5(1):25.
[PUBMED](#) | [CROSSREF](#)
6. Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66(7):736-42.
[PUBMED](#) | [CROSSREF](#)
7. Nasser SM, Cooke G, Kranzer K, Norris SL, Olliaro P, Ford N. Strength of recommendations in WHO guidelines using GRADE was associated with uptake in national policy. *J Clin Epidemiol* 2015;68(6):703-7.
[PUBMED](#) | [CROSSREF](#)
8. Thornton J, Alderson P, Tan T, Turner C, Latchem S, Shaw E, et al. Introducing GRADE across the NICE clinical guideline program. *J Clin Epidemiol* 2013;66(2):124-31.
[PUBMED](#) | [CROSSREF](#)
9. Al-Hameed F, Al-Dorzi HM, Shamy A, Qadi A, Bakhsh E, Aboelnazar E, et al. The Saudi clinical practice guideline for the diagnosis of the first deep venous thrombosis of the lower extremity. *Ann Thorac Med* 2015;10(1):3-15.
[PUBMED](#) | [CROSSREF](#)
10. Kavanagh BP. The GRADE system for rating clinical guidelines. *PLoS Med* 2009;6(9):e1000094.
[PUBMED](#) | [CROSSREF](#)
11. Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, et al. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006;174(5):605-14.
[PUBMED](#) | [CROSSREF](#)
12. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Going from evidence to recommendations. *BMJ* 2008;336(7652):1049-51.
[PUBMED](#) | [CROSSREF](#)
13. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 2013;66(7):719-25.
[PUBMED](#) | [CROSSREF](#)
14. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66(7):726-35.
[PUBMED](#) | [CROSSREF](#)
15. Brozek JL, Akl EA, Compalati E, Kreis J, Terracciano L, Fiocchi A, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines part 3 of 3. The GRADE approach to developing recommendations. *Allergy* 2011;66(5):588-95.
[PUBMED](#) | [CROSSREF](#)
16. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328(7454):1490.
[PUBMED](#) | [CROSSREF](#)
17. Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. *Allergy* 2009;64(5):669-77.
[PUBMED](#) | [CROSSREF](#)

18. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64(4):383-94.
[PUBMED](#) | [CROSSREF](#)
19. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013;66(2):151-7.
[PUBMED](#) | [CROSSREF](#)
20. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64(4):401-6.
[PUBMED](#) | [CROSSREF](#)
21. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ, et al. What is “quality of evidence” and why is it important to clinicians? *BMJ* 2008;336(7651):995-8.
[PUBMED](#) | [CROSSREF](#)
22. Fraenkel L. Incorporating patients' preferences into medical decision making. *Med Care Res Rev* 2013;70(1 Suppl):80S-93S.
[PUBMED](#) | [CROSSREF](#)
23. Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. *J Clin Epidemiol* 2013;66(2):140-50.
[PUBMED](#) | [CROSSREF](#)
24. Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A, et al. Incorporating considerations of resources use into grading recommendations. *BMJ* 2008;336(7654):1170-3.
[PUBMED](#) | [CROSSREF](#)
25. Neumann I, Santesso N, Akl EA, Rind DM, Vandvik PO, Alonso-Coello P, et al. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. *J Clin Epidemiol* 2016;72:45-55.
[PUBMED](#) | [CROSSREF](#)
26. Alexander PE, Bero L, Montori VM, Brito JP, Stoltzfus R, Djulbegovic B, et al. World Health Organization recommendations are often strong based on low confidence in effect estimates. *J Clin Epidemiol* 2014;67(6):629-34.
[PUBMED](#) | [CROSSREF](#)
27. Alexander PE, Brito JP, Neumann I, Gionfriddo MR, Bero L, Djulbegovic B, et al. World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance. *J Clin Epidemiol* 2016;72:98-106.
[PUBMED](#) | [CROSSREF](#)
28. Alexander PE, Gionfriddo MR, Li SA, Bero L, Stoltzfus RJ, Neumann I, et al. A number of factors explain why WHO guideline developers make strong recommendations inconsistent with GRADE guidance. *J Clin Epidemiol* 2016;70:111-22.
[PUBMED](#) | [CROSSREF](#)
29. The Korean Society of Gynecologic Oncology; Korean Academy of Medical Sciences. *Practice Guideline for Gynecologic Cancer Version 2.0*. Seoul, Korea: The Korean Society of Gynecologic Oncology; 2010.
30. The Korean Academy of Tuberculosis and Respiratory Diseases. *Revision of the Guidelines for Chronic Obstructive Pulmonary Disease 2012*. Seoul, Korea: The Korean Academy of Tuberculosis and Respiratory Diseases; 2012.
31. Korean Academy of Medical Sciences. *Korean Clinical Practice Guideline for Colon and Rectal Cancer v.1.0*. Seoul, Korea: Korean Academy of Medical Sciences; 2012.
32. The Korean Academy of Tuberculosis and Respiratory Diseases. *Revision of the Guidelines for Chronic Obstructive Pulmonary Disease 2014*. Seoul, Korea: The Korean Academy of Tuberculosis and Respiratory Diseases; 2014.
33. The Korean Academy of Tuberculosis and Respiratory Diseases. *Korean Guidelines for Asthma, Revised 2014*. Seoul, Korea: The Korean Academy of Tuberculosis and Respiratory Diseases; 2014.
34. The Korean Association for the Study of the Liver. *KASL Clinical Practice Guidelines: Management of Chronic Hepatitis B*. Seoul, Korea: The Korean Association for the Study of the Liver; 2011.
35. The Korean Society of Neurogastroenterology and Motility. *Korean Guideline for the Management of Gastroesophageal Reflux Disease*. Seoul, Korea: The Korean Society of Neurogastroenterology and Motility; 2012.
36. Korean Academy of Medical Sciences. *Korean Clinical Practice Guideline for Gastric Cancer*. Seoul, Korea: Korean Academy of Medical Sciences; 2012.
37. Korean Pancreatobiliary Association. *Korean Clinical Practice Guidelines for Common Bile Duct Stones*. Seoul, Korea: Korean Pancreatobiliary Association; 2013.

38. The Korean Academy of Tuberculosis and Respiratory Diseases. *Cough Guideline*. Seoul, Korea: The Korean Academy of Tuberculosis and Respiratory Diseases; 2014.
39. The Korean Urological Association; The Korean Academy of Family Medicine; Korean Continence Society. *Korean Clinical Practice Guideline for Benign Prostate Hyperplasia*. Seoul, Korea: The Korean Urological Association; 2015.
40. Korean Academy of Medical Sciences; Korea Centers for Disease Control and Prevention. *Development and Dissemination of CPGs for Chronic Diseases*. Seoul, Korea: Korean Academy of Medical Sciences; 2013.

Appendix 1. Data abstraction on the determinants used to make judgement for the strength of recommendations

Guidelines (n = 34)	Four GRADE components						Two additional components			
	Quality of evidence		Benefits and harms		Patients' values and preferences		Cost (resource use)		Feasibility	
	Part of method	Part of background	Part of method	Part of background	Part of method	Part of background	Part of method	Part of background	Part of method	Part of background
A. Determinants information provided in a methodology part: ① Quality or level of evidence only (n = 12, 35.3%)										
• Treatment Guideline for Diabetes (2007)	○	★	-	★	-	★	-	★	-	△
• Korea Breast Cancer Society Practice Recommendations of Breast Cancer (2008)	○	★	-	★	-	★	-	★	-	△
• Diagnostic Guideline of Ulcerative Colitis (2009)	○	★	-	-	-	-	-	-	-	-
• Diagnostic Guideline of Crohn's Disease (2009)	○	★	-	-	-	-	-	-	-	-
• Clinical Practice Guideline for Stroke Rehabilitation in Korea 2009 (2009)	○	★	-	★	-	★	-	-	-	△
• Practice Guidelines for Management of Gallbladder Polyps (2010)	○	★	-	-	-	-	-	-	-	-
• Intravenous Infusion Nursing Practice Guideline (2012)	○	★	-	★	-	★	-	-	-	△
• Clinical Practice Guideline for Stroke Rehabilitation in Korea (2012)	○	★	-	★	-	★	-	★	-	△
• Practical Guidelines for the Surgical treatment of Gallbladder Cancer (2014)	○	★	-	★	-	★	-	★	-	△
• HIV/AIDS Treatment Guidelines (2014)	○	★	-	★	-	★	-	★	-	△
• Clinical Practice Guideline for the prevention and management of diabetes in Korea (2015)	○	★	-	★	-	★	-	★	-	△
• Korean Guideline for Sexually Transmitted Infections (2016)	○	★	-	★	-	★	-	-	-	△
B. Determinants information provided in methodology part: ② Quality or level of evidence and benefits and harms (n = 12, 35.3%)										
• Evidence Based Medicine Guideline for Post-traumatic Stress Disorder (2008)	○	★	○	★	-	-	-	-	-	△
• Acute Inflammatory Facial Nerve Paralysis (2010)	○	★	○	★	-	-	-	-	-	△
• Korean Guideline for the Management of Gastroesophageal Reflux Disease (2012)35	○	★	○	★	-	-	-	-	-	△
• Clinical Practice Guideline for CAPD Peritonitis (2012)	○	★	○	★	-	-	-	-	-	△
• Korean Clinical Practice Guideline: Otitis Media in Children (2014)	○	★	○	★	-	★	-	★	-	△
• Guideline for Intervention Recanalization of Lower Extremity Artery (2014)	○	★	○	★	-	★	-	-	-	△
• Guidelines for initial management and transfer of burns at emergency department (2015)	○	★	○	★	-	★	-	-	-	△
• Evidence-based Recommendations for Hypertension in Primary Care (2016)	○	★	○	★	-	★	-	-	-	△
• Evidence-based Recommendations for Type 2 Diabetes in Primary Care (2016)	○	★	○	★	-	★	-	-	-	△
• Diagnosis and Treatment of Lower Extremity Deep Vein Thrombosis: Korean Practical Guideline (2016)	○	★	○	★	-	★	-	★	-	△
• Evidence-based Clinical Imaging Guidelines (2016)	○	★	○	★	-	★	-	★	-	△
• Evidence-based Recommendations for Dyslipidemia in Primary Care (2017)	○	★	○	★	-	★	-	-	-	△

(continued to the next page)

Appendix 1. (Continued) Data abstraction on the determinants used to make judgement for the strength of recommendations

Guidelines (n = 34)	Four GRADE components				Two additional components			
	Quality of evidence		Benefits and harms		Patients' values and preferences		Cost (resource use)	
	Part of method	Part of background	Part of method	Part of background	Part of method	Part of background	Part of method	Part of background
C. Determinants information provided in methodology part: ③ Quality or level of evidence, benefits and harms, and patient's values and preferences (n = 3, 8.8%)								
• Korean Clinical Practice Guideline for Gastric Cancer (2012) ³⁶	○	★	○	★	○	-	-	△
• Korean Clinical Practice Guidelines for Common Bile Duct Stones (2013) ³⁷	○	★	○	★	○	★	★	△
• Korean Clinical Practice Guideline for Benign Prostate Hyperplasia (2015) ³⁹	○	★	○	★	○	★	★	○
D. Determinants information provided in methodology part: ④ Quality or level of evidence, benefits and harms, and cost (n = 2, 5.9%)								
• KASL Clinical Practice Guidelines: Management of Chronic Hepatitis B (2011) ³⁴	○	★	○	★	-	★	○	△
• Cough Guideline 2014. (2014) ³⁸	○	★	○	★	-	-	○	○
E. Determinants information provided in methodology part: ⑤ Quality or level of evidence, benefits and harms, patient's values and preferences, and cost (n = 5, 14.7%)								
• Practice Guideline for Gynecologic Cancer Version 2.0 (2010) ²⁹	○	★	○	★	○	★	○	○
• Revision of the Guidelines for Chronic Obstructive Pulmonary Disease (2012) ³⁰	○	★	○	★	○	★	★	○
• Korean Clinical Practice Guideline for Colon and Rectal Cancer v1.0 (2012) ³¹	○	★	○	★	○	★	○	○
• Revision of the Guidelines for Chronic Obstructive Pulmonary Disease (2014) ³²	○	★	○	★	○	★	★	○
• Korean Guidelines for Asthma, revised (2014) ³³	○	★	○	★	○	★	★	○

GRADE = Grading of Recommendations Assessment, Development and Evaluation.

^aMarked Δ: cited existing meta-analysis or systematic review, ○: analyzed, performed meta-analysis or systematic review.

Appendix 2. A template and sample of judgement table to justify the strength of recommendations

Judgement sample		
Intervention	What is the intervention?	Vitamin D supplementation
Quality of the evidence	The higher the quality of the evidence, the stronger the recommendation. However, when “low” or “very-low” quality, consider more carefully the other criteria below in deciding the strength of the recommendation.	<input type="checkbox"/> High <input type="checkbox"/> Moderate <input checked="" type="checkbox"/> Low <input type="checkbox"/> Very low
Values and preferences	This refers to values placed by health workers, policy-makers, patients and other stakeholders on the intended outcomes of interventions. If there is wide variability between values and preferences of various stakeholders, it is less likely to have a strong recommendation.	<input checked="" type="checkbox"/> No significant variability <input type="checkbox"/> Significant variability
Absolute magnitude of effect	This refers to the potential of the intervention to have large effects. The effects can be enhanced by combining with other interventions. Consider what are the possible associations (or “bundles”) that will enhance the effect. The larger the potential effects and for longer periods of time, the more likely to have a strong recommendation.	<input type="checkbox"/> Large effect in the long term <input checked="" type="checkbox"/> Small effect for short duration
Balance of benefits versus disadvantages	Benefits should consider the intended effects of the intervention. Disadvantages should consider the potentially negative effects of the intervention, as well as the unintended effects. The less potentially negative effects, the more likely to have a strong recommendation.	<input type="checkbox"/> Benefits clearly outweigh disadvantages <input checked="" type="checkbox"/> Benefits and disadvantages are balanced <input type="checkbox"/> Disadvantages clearly outweigh benefits
Resource use	The resource needed for implementing the recommendation may comprise financial resources, human resources, and infrastructure or equipment. Ideally, the benefits of the intervention should come at reasonable, affordable and sustainable costs. One should consider that capital costs, such as for infrastructure development, even if initially high, may yield benefits in the long run. The higher the incremental or recurrent costs, all other things being equal, the less likely it is to have a strong recommendation.	<input checked="" type="checkbox"/> Less resource intensive <input type="checkbox"/> More resource intensive
Feasibility	All interventions require political commitment and wide stakeholder engagement as a prerequisite. In addition, “technical” feasibility requires functional organizational and institutional structures necessary to manage, follow through, and monitor the implementation of the recommendation. The elements of technical feasibility vary widely by country or context, but if these elements are likely to be functional in a wide variety of settings, the more likely is to have a strong recommendation.	<input checked="" type="checkbox"/> Yes, globally <input type="checkbox"/> Yes, conditionally
Overall ranking	Strength of the recommendation.	<input checked="" type="checkbox"/> Strong recommendation [‡] <input type="checkbox"/> Weak recommendation
Conclusion about recommendation direction		<input type="checkbox"/> In favour of the intervention <input checked="" type="checkbox"/> Against the intervention

[‡]This recommendation was made strong against the intervention despite of the low quality of evidence due to the fact that some participants expressed concerns about the limited evidence on safety of vitamin D supplementation during pregnancy. The guideline development group also noted that several studies were ongoing on this topic which may lead to a change in the evidence base in the future.

Ref. World Health Organization. WHO Recommendations for Prevention and Treatment of Pre-eclampsia and Eclampsia. Geneva, Switzerland: World Health Organization; 2011.

Appendix 3. Five situations that justify strong recommendations based on low or very low certainty of the evidence in GRADE approach^{25,27}

Paradigmatic situation	Certainty of the evidence for health outcomes (quality of evidence)		Balance of benefits and harms	Values and preferences	Resource considerations	Recommendation
	Benefits	Harms				
Life-threatening situation	Low or very low	Immaterial (very low to high)	Intervention may reduce mortality in a life threatening situation; adverse events not prohibitive	A very high value is placed on an uncertain but potentially life preserving benefit	Small incremental cost (or resource use) relative to the benefits justify the intervention	Strong recommendation in favor
	Example 1.	Indirect evidence from seasonal influenza suggests that patients with avian influenza may benefit from the use of oseltamivir (low certainty of the evidence). Given the high mortality of the disease and the absence of effective alternatives, the WHO made a strong recommendation in favor of the use of oseltamivir rather than no treatment in patients with avian influenza.				
	Example 2.	Fresh frozen plasma or vitamin K in a patient receiving warfarin with elevated INR and an intracranial bleed. Only low-quality evidence supports the benefits of limiting the extent of the bleeding.				
Uncertain benefit, Low or very low certain harm	High or moderate	Possible but uncertain benefit; substantial established harm	A much higher value is placed on the adverse events in which we are confident than in the benefit, which is uncertain	High incremental cost (or resource use) relative to the benefits may not justify the intervention	Strong recommendation against	
	Example 1.	In patients with idiopathic pulmonary fibrosis, treatment with azathioprine plus prednisone offers a possible but uncertain benefit in comparison with no treatment. The intervention, however, is associated with a substantial established harm. An international guideline made a recommendation against the combination of corticosteroids plus azathioprine in patients with idiopathic pulmonary fibrosis.				
	Example 2.	Head-to-toe CT/MRI screening for cancer. Low-quality evidence of benefit of early detection but high-quality evidence of possible harm and/or high cost (strong recommendation against this strategy).				
Potential equivalence, one option clearly less risky or costly	Low or very low	High or moderate	Magnitude of benefit apparently similar, though uncertain for alternatives; we are confident in less harm or cost for one of the competing alternatives	A high value is placed on the reduction in harm	High incremental cost (or resource use) relative to the benefits may not justify one of the alternatives	Strong recommendation for less harmful/less expensive
	Example 1.	Low-certainty evidence suggests that initial <i>Helicobacter pylori</i> eradication in patients with early-stage extranodal marginal zone (MALT) B-cell lymphoma results in similar rates of complete response in comparison with the alternatives of radiation therapy or gastrectomy, but with high confidence of less harm, morbidity, and cost. Consequently, UpToDate made a strong recommendation in favor of <i>H. pylori</i> eradication rather than radiotherapy in patients with MALT lymphoma.				
	Example 2.	<i>H. pylori</i> eradication in patients with early stage gastric MALT lymphoma with <i>H. pylori</i> positive. Low-quality evidence suggests that initial <i>H. pylori</i> eradication results in similar rates of complete response compared with the alternatives of radiation therapy or gastrectomy; high-quality evidence suggests less harm/morbidity.				
High similar benefits, one option potentially more risky or costly	High or moderate	Low or very low	Established that magnitude of benefit is similar for alternative management strategies; best (though uncertain) estimate is that one alternative has appreciably greater harm	A high value is placed on avoiding the potential increase in harm	High incremental cost (or resource use) relative to the benefits may not justify one of the alternatives	Strong recommendation against the intervention with possible greater harm
	Example 1.	In women requiring anticoagulation and planning conception or in pregnancy, high certainty of the evidence suggests similar effects of different anticoagulants. However, indirect evidence (low certainty of the evidence) suggests potential harm to the unborn infant with oral direct thrombin (e.g., dabigatran) and factor Xa inhibitors (e.g., rivaroxaban, apixaban). The AT9 guidelines recommended against the use of such anticoagulants in women planning conception or in pregnancy.				
	Example 2.	Hypertension in women planning conception and in pregnancy. Strong recommendations for labetalol and nifedipine and strong recommendations against ACE inhibitors and ARBs; all agents have high-quality evidence of equivalent beneficial outcomes, with low-quality evidence for greater adverse effects with ACE inhibitors and ARBs.				
Potential catastrophic harm	Immaterial (very low to high)	Low or very low	Potential important harm of the intervention, magnitude of benefit is variable	A high value is placed on avoiding potential increase in harm	High incremental cost (or resource use) relative to the benefits, may not justify the intervention	Strong recommendation against the intervention
	Example 1.	In men with androgen deficiency, testosterone supplementation likely improves quality of life. Low-certainty evidence suggests that testosterone increases cancer spread in patients with prostate cancer. The US Endocrine Society made a recommendation against testosterone supplementation in patients with prostate cancer.				
	Example 2.	Testosterone in males with or at risk of prostate cancer. High-quality evidence for moderate benefits of testosterone treatment in men with symptomatic androgen deficiency to improve bone mineral density and muscle strength. Low-quality evidence for harm in patients with or at risk of prostate cancer.				

GRADE = Grading of Recommendations Assessment, Development and Evaluation, WHO = World Health Organization, INR = international normalized ratio, CT/MRI = computed tomography/magnetic resonance imaging, MALT = mucosa-associated lymphoid tissue, ACE = angiotensin-converting enzyme, ARB = angiotensin receptor blocker.