



## Korean Translation of the GRADE Series Published in the *BMJ*, 'GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations' (A Secondary Publication)

Translated by: Do Kyung Kim, Eu Chang Hwang<sup>1</sup>, Ho Won Kang<sup>2</sup>, Ja Yoon Ku<sup>3</sup>, Hyun Jin Jung<sup>4</sup>, Hong Wook Kim<sup>5</sup>, Jae Hung Jung<sup>6,7</sup>; Guideline Development Committee in the Korean Association of Urogenital Tract Infection and Inflammation

Department of Urology, Soonchunhyang University Seoul Hospital, Soonchunhyang University College of Medicine, Seoul, <sup>1</sup>Department of Urology, Chonnam National University Medical School, Gwangju, <sup>2</sup>Department of Urology, Chungbuk National University College of Medicine, Cheongju, <sup>3</sup>Department of Urology, Pusan National University Hospital, Busan, <sup>4</sup>Department of Urology, Daegu Catholic University School of Medicine, Daegu, <sup>5</sup>Department of Urology, College of Medicine, Konyang University, Daejeon, <sup>6</sup>Department of Urology, <sup>7</sup>Institute of Evidence Based Medicine, Yonsei University Wonju College of Medicine, Wonju, Korea

Clinical practice guidelines are statements that include recommendations intended to optimize patient care based on a systematic review of the evidence assessing the benefits and harm of alternative care options. Guideline developers should use an explicit, judicious, and transparent methodology to make trustworthy guidelines. Although there are a variety of frameworks that can help translate enormous medical knowledge into recommendations, the most widely adopted tool for grading the quality of evidence and making recommendations is GRADE (Grading of Recommendations, Assessment, Development and Evaluations). This article is the first translation of a series published in the *BMJ* with regard to the GRADE Approach for Evidence Based Clinical Practice Guideline Development to provide informative knowledge for moving from evidence to recommendations to Korean guideline developers.

Received: 1 April, 2019

Revised: 9 April, 2019

Accepted: 9 April, 2019

Correspondence to: Jae Hung Jung

<https://orcid.org/0000-0002-4990-7098>

Department of Urology, Yonsei University Wonju College of Medicine, 20 Ilsan-ro, Wonju 26426, Korea

Tel: +82-33-741-0652, Fax: +82-33-741-1930

E-mail: geneuro95@yonsei.ac.kr

This article is the secondary publication (complete translation in Korean) of the article originally published in the *BMJ* in English (GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. 2008;336:924-6). The Editor-in-Chief of *Urogenital Tract Infection* decided to publish this secondary publication for the reader's sake, and it was approved by *BMJ*. *BMJ* Publishing Group take no responsibility for the accuracy of the translation from the published English language original and are not liable for any errors that may occur.

Copyright © 2019, Korean Association of Urogenital Tract Infection and Inflammation. All rights reserved.



This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://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.

진료지침의 근거수준과 권고등급을 평가하는 방법은 일관성이 부족한 것이 현실입니다. 이 논문에서는 전 세계적으로 점차 사용이 확대되고 있는 GRADE 방법론에 대해 논의하고자 합니다.

2006년부터 *BMJ* (<https://www.bmj.com/>)는 “Instructions to Authors”에 임상 진료지침을 투고할 때 증거의 등급을 평가하기 위해 GRADE 방법론을 이용하는 것이 바람직하다고 권고하였습니다.

이 논문은 진료지침 개발에 있어 근거수준 및 권고등급 평가를 위한 통일된 방법론의 필요성 및 임상적 중요성에 대하여 GRADE 방법론을 중심으로 설명하고자 합니다.

GRADE 방법론은 여러 다른 방법론에 비하여 장점을 가지고 있습니다(Table 1 and Appendix 1) [1,2].

## "근거수준"이란 무엇이며 왜 중요한가?

환자 진료의 최적화를 위해서 임상 의사와 환자는 치료의 원하는 효과(advantage)와 원하지 않는 효과(disadvantage)를 반드시 가늠해보아야 합니다. 치료의 원하는/원하지 않는 효과의 추정치(best estimate of expected advantage and disadvantage)뿐만 아니라 추정치에 대한 확신도(근거수준)는 의사 결정에 영향을 미치게 됩니다. 다시 말하면 중재 효과의 추정치 크기(estimate of the magnitude of intervention)의 임상적 유용성은 그 추정치에 대한 우리의 근거수준에 달려있습니다.

보건의료 관련 전문가 및 전문학회 등은 근거수준을 고려하지 않는 우를 범하였습니다[3]. 예를 들면, 과거 폐경기 여성의

**Table 1.** Advantages of GRADE over other systems

- Developed by a widely representative group of international guideline developers
- Clear separation between the quality of evidence and strength of recommendations
- Explicit evaluation of the importance of outcomes of alternative management strategies
- Explicit, comprehensive criteria for downgrading and upgrading the quality of evidence ratings
- Transparent process of moving from evidence to recommendations
- Explicit acknowledgment of values and preferences
- Clear, pragmatic interpretation of strong versus weak recommendations for clinicians, patients, and policy makers
- Useful for systematic reviews and health technology assessments, as well as guidelines

Adapted from the article of Guyatt et al. BMJ 2008;336:924-6 [1].  
See also Appendix 1 (complete translate in Korean).

심혈관계 위험을 감소시킬 수 있다는 믿음으로 인해 호르몬 대체 요법이 권고되어 왔습니다[4].

그러나, 이는 일관되지 않은 결과를 보고한 관찰 연구 (observational studies)에 기반한 결과로 이에 대한 근거수준은 ‘매우 낮음(very low quality of evidence)’임이 간과되었습니다[5]. 결론적으로, 무작위 대조 임상연구(randomized controlled trials)에서 호르몬 대체 요법이 심혈관계 위험을 감소시키지 못하고 증가시킬 수도 있다는 것이 밝혀졌습니다[6,7]. 또 다른 예로, 미국 식품의약국(US Food and Drug Administration)은 돌연사(sudden death)를 초래할 수 있는 무증상의 심실성 부정맥(asymptomatic ventricular arrhythmia) 치료제인 encainide와 flecainide를 허가하였습니다. 그러나, 이 결정 역시 그 약물들의 효과에 대한 낮은 근거수준(low quality of evidence)을 고려하지 않았으며 마찬가지로, 무작위 대조군 연구에서 두 약물이 급사의 위험을 증가시키는 것으로 보고되었습니다[8]. 만약, 낮은 근거수준에 대한 주의가 있었다면 수많은 생명을 구할 수 있었을 것입니다.

높은 질의 근거수준을 간과하는 것 역시 같은 문제를 야기할 수 있습니다. 엄격히 수행된 무작위 대조군 연구들은 혈전 용해 요법이 심근경색으로 인한 사망률을 감소시킬 수 있음을 보고하였으나 과거 임상진료지침은 10년 이상 이를 반영하지 못하였습니다[9].

이러한 실수를 반복하지 않기 위하여 근거수준에 대하여 이해하는 것이 도움이 될 것입니다.

## 진료지침 개발자들은 임상 의사들에게 어떻게 근거수준을 알려야 하는가?

GRADE 방법론은 근거수준을 4가지 등급(높음[high], 중등도[moderate], 낮음[low], 매우 낮음[very low])으로 범주화함으로써 임상 의사가 근거수준을 명확히 이해할 수 있도록 도울 수 있을 것입니다. 그러나 근거수준은 이렇게 이분법적으로

나눌 수 없는 경우도 있을 수 있습니다. 하지만 이러한 근거수준 분류의 장점은 단순함(simplicity), 투명성(transparency), 선명성(vividness)이라고 할 수 있겠습니다.

## "권고등급"은 무엇이며 왜 중요한가?

진료지침은 적은 비용(cost) 및 환자불편감(inconvenience)과 더불어 중재의 원하는 효과(benefits)와 원하지 않는 효과(side effects)의 차이를 증명한 무작위 대조군 연구에 기반할 수 있을 것입니다. 예를 들어, 천식이 악화된 환자에서 단기간의 경구용 스테로이드의 유용성이 이런 경우에 해당할 것입니다. 임상 의사는 이러한 치료를 아무런 망설임 없이 시행할 수 있을 것입니다.

또는, 진료지침은 관찰 연구에 기반할 수도 있으며 상당한 중재 부작용(harms) 또는 비용을 수반할 수도 있습니다. 인공 심장 판막을 가진 임산부에게 항혈전 치료를 결정하기 위해서 판막 혈전증의 발생 감소 정도와 환자의 불편감, 치료 비용 및 태아 기형 발생 위험을 모두 고려하여 심사숙고하여야 합니다. 임상 의사는 환자가 자신의 가치와 선호도에 따라 치료의 효과와 부작용에 대해 신중하게 가늠할 수 있도록 도와야 합니다.

그러므로 진료지침은 1) 근거수준 ‘높음’ 및 원하는 효과와 원하지 않는 효과의 차이가 큰 경우, 2) 원하는 효과와 원하지 않는 효과의 차이가 우열을 가리기 힘들거나 불확실한 경우인지를 명시해야 합니다. 간단하고 투명한 권고 등급은 이를 효과적으로 전달할 수 있을 것입니다.

GRADE 방법론은 권고등급을 ‘강한 권고(strong recommendation)’ 또는 ‘약한 권고(weak recommendation)’로 나누고 있으나 근거수준과 마찬가지로 권고등급 역시 이분법적으로 나누기 어려운 점이 있을 수 있습니다. 그러나 명확한 권고등급은 진료지침에 있어 단점보다는 장점이 클 것으로 생각합니다.

## 우수한 등급 시스템에 필수적인 것은 무엇인가?

항상 근거수준과 권고등급이 일치하는 것은 아닙니다. 높은 근거수준(high quality of evidence)이 반드시 강한 권고를 의미하는 것은 아니며, 강한 권고 역시 낮은 근거수준에 기반할 수 있습니다.

예를 들어 명백한 유발 요인이 없이 심부정맥혈전증을 처음으로 경험한 환자는 항응고 치료 시행 1개월 후에 와파린의 장기복용 여부를 결정해야 합니다. 무작위 대조군 연구에 따르면 장기적인 와파린 복용은 혈전증의 재발 위험을 줄이지만 출혈의 위험과 환자의 불편이 증가한다는 단점이 있습니다. 환자들은 스스로의 가치와 선호도에 따라 치료를 결정할 수

있기 때문에, 임상진료지침 개발자는 높은 근거수준에도 불구하고(무작위 대조군 연구에 기반한) 약한 권고를 내릴 수도 있을 것입니다.

“전문가 의견(expert opinion)”을 근거(evidence)로서 분류하는 것 역시 혼란을 야기합니다. 증거의 질이 높거나 낮은 것에 상관없이 모든 근거의 해석에는 판단(judgement)이 필요합니다. 전문가 의견은 사례 보고 및 적절하게 수행되지 않은 관찰 연구와 마찬가지로 매우 낮은 근거수준으로 평가되어야 할 것입니다.

근거수준과 권고등급의 단순한 분류 시스템(grading system)은 환자, 임상 의사 및 정책 결정자에게 사용을 용이하게 만들 수 있으며 근거수준 및 권고등급의 상세하고 명쾌한 기준은 진료지침을 보다 투명하게 만들 것입니다[2].

지금까지 많은 분류 시스템이 사용되어 왔으나 시스템의 복잡성으로 인해 이를 임상 진료현장에 사용하기에는 많은 제약이 있습니다[2]. 또한 다양한 시스템을 임상의사가 모두 이해하는 것은 효율적이지 않을뿐더러 실제로 임상의사는 이것을 위해 할애할 시간이 없습니다. GRADE 방법론은 현재 세계보건기구(World Health Organization), 미국내과학회(American College of Physicians), 미국흉부학회(American Thoracic Society), UpToDate (www.uptodate.com) 및 코크란 연합(Cochrane Collaboration) 등 전 세계 유수의 기관에서

진료지침 개발에 사용되고 있습니다.

## GRADE 방법론은 어떻게 근거수준을 분류하는가?

GRADE 방법론은 근거수준을 ‘높음’, ‘중등도’, ‘낮음’ 및 ‘매우 낮음’으로 단순하고 투명하게 분류합니다(Table 2 and Appendix 2) [1]. 무작위 대조군 연구는 근거수준 ‘높음’에서 출발하여 다음의 이유로 중등도, 낮음, 매우 낮음 순으로 낮아질 수 있습니다.

- 비뚤림 위험(risk of bias)
- 비일관성(inconsistency)
- 비직접성(indirectness)
- 비정밀성(imprecision)
- 출판비뚤림(publication bias)

관찰 연구(예: 코호트[cohort study] 및 환자 대조군 연구[case control study])는 증거의 질이 “낮음”으로 출발하지만 만일 중재 효과의 크기가 크거나(very large effect), 교란변수가 효과의 크기를 낮추거나(all plausible biases would decrease the magnitude of an apparent treatment effect), 양-반응 관계(dose-response relation)가 있다면 근거수준의 등급을 높일 수 있습니다.

## GRADE 방법론은 어떻게 권고등급을 분류하는가?

GRADE 방법론은 “강함(strong)” 및 “약함(weak)” 두 가지 등급의 권고 사항을 제공합니다(진료지침 개발 패널은 “약함” 대신 “조건부[conditional]” 또는 “임의[discretionary]”와 같은 용어를 선호할 수 있음). 중재의 효과가 위해보다 명확히 크거나 작은 경우 진료지침 개발 패널은 강한 권고를 제시하는 반면, 만약 낮은 근거수준이거나 효과가 위해보다 거의 비슷한

Table 2. Quality of evidence and definitions

- High quality—Further research is unlikely to change our confidence in the estimate of effect
- Moderate quality—Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- Low quality—Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- Very low quality—Any estimate of effect is very uncertain

Adapted from the article of Guyatt et al. BMJ 2008;336:924-6 [1]. See also Appendix 2 (complete translate in Korean).

Table 3. Factors that affect the strength of a recommendation

Factor	Examples of strong recommendations	Examples of weak recommendations
Quality of evidence	Many high quality randomized trials have shown the benefit of inhaled steroids in asthma	Only case series have examined the utility of pleurodesis in pneumothorax
Uncertainty about the balance between desirable and undesirable effects	Aspirin in myocardial infarction reduces mortality with minimal toxicity, inconvenience, and cost	Warfarin in low risk patients with atrial fibrillation results in small stroke reduction but increases the bleeding risk and causes substantial inconvenience
Uncertainty or variability in values and preferences	Young patients with lymphoma will invariably place a higher value on the life prolonging effects of chemotherapy than on the treatment toxicity	Older patients with lymphoma may not place a higher value on the life prolonging effects of chemotherapy than on the treatment toxicity
Uncertainty about whether the intervention represents a wise use of resources	The low cost of aspirin as a prophylaxis against stroke in patients with transient ischemic attacks	The high cost of clopidogrel and of combination dipyridamole and aspirin as prophylaxis against stroke in patients with transient ischemic attacks

Adapted from the article of Guyatt et al. BMJ 2008;336:924-6 [1]. See also Appendix 3 (complete translate in Korean).

경우 약한 권고를 제시합니다. 근거수준 이외에도 몇 가지 다른 요인들이 권고의 등급에 영향을 미칠 수 있습니다(Table 3 and Appendix 3) [1].

## CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

## AUTHOR CONTRIBUTIONS

D.K.K.: translating the article, and drafting the manuscript, E.C.H.: providing clinical and methodological advices in translation, H.W.K.: helping to translate and draft the manuscript, J.Y.K., H.J.J., and H.W.K.: providing clinical advices in translation, J.H.J.: contacting BMJ editorial office to get the approval, helping to translate and draft the manuscript, and final approval.

## ORCID

Do Kyung Kim, <https://orcid.org/0000-0002-3696-8756>  
 Eu Chang Hwang, <https://orcid.org/0000-0002-2031-124X>  
 Ho Won Kang, <https://orcid.org/0000-0002-8164-4427>  
 Ja Yoon Ku, <https://orcid.org/0000-0003-3460-9386>  
 Hyun Jin Jung, <https://orcid.org/0000-0002-1895-7180>  
 Hong Wook Kim, <https://orcid.org/0000-0002-3847-1401>  
 Jae Hung Jung, <https://orcid.org/0000-0002-4990-7098>

## REFERENCES

1. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al.; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-6.
2. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al.; GRADE Working Group. 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;4:38.
3. Lacchetti C, Guyatt G. Therapy and validity: surprising results of randomized controlled trials. In: Guyatt G, Rennie D, editors. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago: American Medical Association; 2002.
4. American College of Physicians. Guidelines for counseling postmenopausal women about preventive hormone therapy. *Ann Intern Med* 1992;117:1038-41.
5. Humphrey LL, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 2002;137:273-84.
6. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 1998;280:605-13.
7. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.
8. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med* 1991;324:781-8.
9. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992;268:240-8.

## Appendix 1. GRADE의 장점

- 전 세계적으로 대표되는 국제 진료지침 개발자 그룹에 의해 개발
- 근거수준(quality of evidence)과 권고등급(strength of recommendation)의 명확한 분리
- 대안적 치료 전략의 결과의 중요성에 대한 명확한 평가
- 근거수준 등급을 낮추거나 높이기 위한 명확하고 포괄적인 기준
- 근거(evidence)에서 권고(recommendation)로의 투명한 이행 과정
- 환자의 가치(value) 및 선호도(preference)의 명확한 제공
- 임상 의사, 환자 및 정책 입안자에 대한 권고 등급의 명확하고 실용적인 해석
- 체계적 문헌 고찰, 의료 기술 평가 및 진료지침 개발에 유용성

## Appendix 2. 근거수준(quality of evidence)

- 높음(high)—추가 연구가 효과 추정치(estimate of effect)에 대한 우리의 신뢰도(confidence)를 바꿀 가능성이 매우 낮음
- 중등도(moderate)—추가 연구가 효과 추정치에 대한 우리의 신뢰에 중요한 영향을 미칠 수 있으며 추정치를 변경시킬 수도 있음
- 낮음(low)—추가 연구가 효과 추정치에 대한 우리의 신뢰에 중요한 영향을 미칠 가능성이 매우 높으며 추정치를 변경시킬 것으로 예상됨
- 매우 낮음(very low)—효과 추정치는 매우 불확실함

## Appendix 3. 권고등급에 영향을 미치는 요인

요인	강한 권고	약한 권고
근거수준(quality of evidence)	천식 환자에서 흡입 스테로이드의 효과에 대한 많은 무작위 대조군 연구	폐기흉환자에서 흉막유착술(pleurodesis)의 효과에 대한 사례보고만이 존재
원하는 효과(desirable effects)와 원하지 않는 효과(undesirable effects)	심근 경색증에서 aspirin의 효과(사망률을 낮추면서 약제 독성, 불편 및 비용이 적음)	심방 세동의 위험이 낮은 환자에서 warfarin은 뇌졸중 감소효과는 적으나 출혈 위험을 높일 수 있으며 환자 불편을 유발함
가치(values)와 선호(preferences)	젊은 림프종 환자들은 치료 독성보다는 화학 요법의 수명연장 효과에 더 높은 가치를 둘 수 있음	고령의 림프종 환자는 화학 요법의 수명 연장 효과보다는 치료 독성에 더 높은 가치를 둘 수 있음
비용(costs or resources)	일시적인 뇌졸중 환자에서 aspirin 사용은 비용이 적게 들	일시적인 뇌졸중 환자에서 clopidogrel과 dipyridamole 및 aspirin 병용 요법은 비용이 많이 들