



Korean Translation of the GRADE Series Published in the *BMJ*, ‘GRADE: What Is “Quality of Evidence” and Why Is It Important to Clinicians?’ (A Secondary Publication)

Translated by: Ho Won Kang, Jae Hung Jung^{1,2}, Do Kyung Kim³, Ja Yoon Ku⁴, Hyun Jin Jung⁵, Hong Wook Kim⁶, Eu Chang Hwang⁷; Guideline Development Committee in the Korean Association of Urogenital Tract Infection and Inflammation

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

This article is second translation of a GRADE series published in the *BMJ* to create a highly structured, transparent, and informative system for rating quality of evidence for developing recommendations. The process to develop a guideline, we should formulate a clear question with specification of all outcomes of importance to patients. Grading of Recommendations, Assessment, Development and Evaluation (GRADE) offers four levels of evidence quality: high, moderate, low, and very low for these patient-important outcomes. Randomized trials begin as high quality evidence and observational studies as low quality evidence. Although randomized trials begin as high quality evidence, quality may be downgraded as a result of study limitations (risk of bias), inconsistency (variability in results), indirectness, imprecision (wide confidence intervals), or publication bias. While the quality of evidence derived from observational studies starts at ‘low’ but may be upgraded based on a very large magnitude of effect, a dose-response gradient, and if all plausible biases would reduce an apparent treatment effect.

Received: 30 August, 2019

Revised: 30 August, 2019

Accepted: 30 August, 2019

Correspondence to: Eu Chang Hwang

<https://orcid.org/0000-0002-2031-124X>

Department of Urology, Chonnam National University Hwasun Hospital, Chonnam National University Medical School, 322 Seoyang-ro, Hwasun-eup, Hwasun 58128, Korea

Tel: +82-61-379-7747, Fax: +82-61-379-7750

E-mail: urohwan@gmail.com

This article is the secondary publication (complete translation in Korean) of the article originally published in the *BMJ* in English (GRADE: What is “quality of evidence” and why is it important to clinicians? 2008;336:995-8). 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.

GRADE (Grading of Recommendations, Assessment, Development and Evaluation): “근거수준(Quality of Evidence)”이란 무엇이며, 임상 의사에게 왜 중요한가?

진료지침 개발자들은 권고안의 바탕이 되는 근거수준 (quality of evidence)을 평가하기 위해 다양한 방법론을 사용합니다. 그러나 일부의 방법론은 복잡하고 일관되지 못한 것이 현실입니다.

2004년 Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group은 진료지침 개발을 위해 새로운 방법론을 제안하였습니다[1].

본 논문은 진료지침을 개발을 위한 GRADE 방법론에 대한 두 번째 시리즈로, 체계적이고 투명한 근거수준 평가를 위한 방법으로서 GRADE를 알아볼 것입니다[2].

진료지침은 핵심 질문을 명확하게 정의하여야 한다.

진료지침의 핵심 질문은 인구집단(patients), 중재(intervention), 대조군(comparison) 및 건강 결과(outcomes of interest) 등의 4가지 요소로 구성됩니다[3]. 예를 들어 췌장암 환자의 수술방법으로 유문부 보존술식(pylorus preserving)을 표준 광범위 절제술(standard Whipple pancreatico-

duodenectomy)과 비교할 때, 단기 및 장기 사망률, 수혈량, 담즙 누출, 입원기간 및 위 배출(gastric emptying)에 어떤 영향을 미치는지 고려되어야 합니다.

진료지침 개발자는 건강 결과의 상대적 중요도를 다루어야만 한다.

진료지침 개발자들은 권고안 개발 과정에서 환자에게 중요한 모든 건강 결과를 명시하고, 건강 결과에 대하여 핵심적인 것(critical)과 중요하지만 핵심적이지 않은 것(important but not critical)으로 분류하여야 합니다[4].

GRADE는 중요도를 판단하기 위한 9단계의 척도를 제시합니다. 척도의 상단 7-9는 의사결정(decision making)에 있어 핵심적인 건강 결과이며 4-6등급은 의사결정에 있어 중요하지만 핵심적이지 않은 건강 결과, 의사결정과정에서 1-3등급은 덜 중요한 항목입니다(Fig. 1) [2].

근거수준의 결정은 의료 환경의 고려가 필요하다.

GRADE는 권고안을 만드는 과정에서 근거수준에 대한 정의를 제시합니다. 근거수준이란 효과에 대한 평가가 권고안을 뒷받침하는 데 충분한지를 반영합니다. 이러한 정의는 두 가지 중요한 의미를 가집니다. 첫째, 진료지침 개발 위원회는 권고안의 근거수준을 결정할 때 권고안이 적용되는 의료 환경을 고려하여 판단을 내려야 합니다. 둘째, 근거수준의 의미는 체계적 문헌고찰(systematic review)과 진료지침에서 다를 수 있습니다.

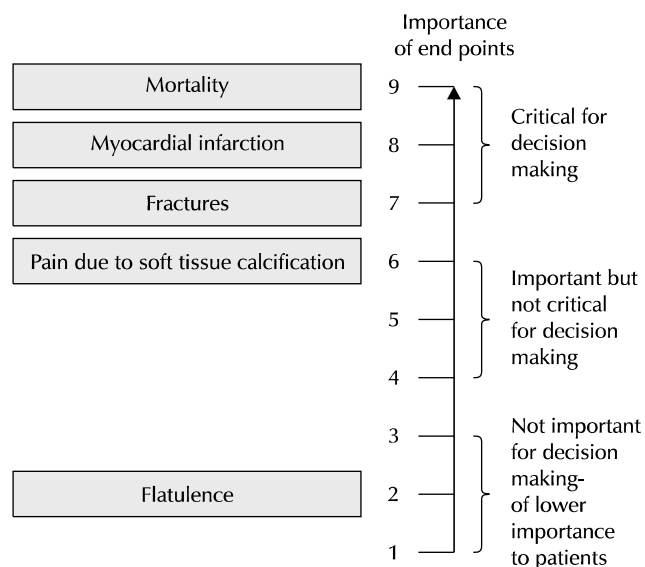


Fig. 1. Hierarchy of outcomes according to importance to patients to assess effect of phosphate lowering drugs in patients with renal failure and hyperphosphataemia. Adapted from the article of Guyatt et al. BMJ 2008;336:995-8 [2].

니다. 체계적 문헌고찰에서 근거수준은 치료 효과의 추정치(estimate of effect)에 대한 신뢰도(confidence)를 의미합니다.

연구 설계(Study Design)는 근거수준을 결정하는 데 중요하다.

근거수준을 평가하는 이전의 방법론과 마찬가지로[5], GRADE는 연구 설계에 대한 고려로부터 시작됩니다. 일반적으로 무작위 연구(randomized trials)가 관찰 연구(observational studies)보다 강력한 근거를 제공하며 엄격한 관찰 연구(rigorous observational studies)는 통제되지 않은 환자군 연구(uncontrolled case series)보다 강력한 근거가 될 수 있습니다. GRADE는 중대한 결함이 없는 무작위 연구를 근거수준 '높음'으로, 반면 관찰 연구를 '낮음'으로 평가합니다.

근거수준을 낮출 수 있는 5가지 고려요소

GRADE는 개개의 환자에게 중요한 건강 결과의 근거수준을 평가하기 위해 5가지 요소를 고려합니다(Table 1 and Appendix 1) [2,6]. 이러한 5가지 평가 요소는 무작위 대조 임상연구(randomized controlled trials)뿐만 아니라 관찰 연구에도 동일하게 적용됩니다.

1. 비풀림 위험(Study Limitations)

만약 진료지침 개발에 포함된 연구가 비풀림 위험을 가지고 있다면 치료 효과에 대한 신뢰도가 낮아질 것입니다[7]. 이러한 비풀림 위험에는 무작위배정 은폐(allocation concealment) 또는 눈가림(blinding)과 관련된 비풀림(건강 결과에 측정에 주관적 평가가 필요한 경우 비풀림 위험도가 더욱 높음); 높은 추적관찰 실패율(large losses to follow-up); 치료의향에 따른 분석 원칙 위배(failure to adhere to an intention to treat analysis); 연구 조기중단(stopping early for bene-

Table 1. Factors in deciding on quality of evidence

Factors that might decrease quality of evidence

- Study limitations
- Inconsistency of results
- Indirectness of evidence
- Imprecision
- Publication bias

Factors that might increase quality of evidence

- Large magnitude of effect
- Plausible confounding, which would reduce a demonstrated effect
- Dose-response gradient

Adapted from the article of Guyatt et al. BMJ 2008;336:995-8 [2]. See Appendix 1 (complete translate in Korean).

fit) [8], 또는 선택적 보고(일반적으로 중재 효과가 관찰되지 않은 경우) 등이 있습니다.

예를 들면, 췌장암에 대한 유문부 보존술식과 표준 광범위 절제술에 대한 대부분의 무작위 비교연구는 무작위배정 은폐나 눈가림의 결함, 그리고 높은 추적관찰 실패율로 인해 비뚤림 위험을 가지고 있어 각각의 중요한 건강 결과에 대한 근거수준은 '중등도' 이상으로 평가할 수 없을 것입니다(Table 2 and Appendix 2) [2,4].

2. 비일관성(Inconsistent Results)

여러 연구 간에 치료 효과의 추정치가 크게 다른 경우는(결과의 이질성 또는 변이도) 근본적인 치료 효과의 차이가 있음을 시사합니다. 변이도(variability)는 대상 인구집단의 차이(예: 환자의 중증도에 따른 효과 차이), 중재(예: 약제 용량 차이로 인한 효과 차이) 또는 결과(예: 시간 경과에 따른 치료 효과 감소)로 인해 발생할 수 있습니다. 설명할 수 없는 이질성이 존재할 경우 근거수준은 낮아지게 됩니다.

예를 들어, 췌장암 환자의 수술방법에 대한 무작위 연구들은

위 배출에 대한 효과 추정치가 각각 크게 상이하여 근거수준은 더 낮아지게 됩니다(Fig. 2) [2].

3. 비직접성(Indirectness of Evidence)

근거의 비직접성은 두 가지 유형이 있습니다. 첫째, 둘 이상의 중재 약물들을 직접 비교한 무작위 연구(A vs. B)는 없으나 각각의 중재 약물을 위약과 비교한 무작위 연구(A vs. placebo or B vs. placebo)가 존재하는 경우입니다. 이러한 연구들은 두 중재 약물의 효과를 간접적으로(indirect) 비교할 수 있으나, 근거수준은 두 약물을 직접 비교(head to head comparisons)보다 낮을 것입니다. 두 번째 유형의 비직접성은 인구집단의 차이, 중재법의 차이, 대조군의 차이, 건강 결과의 차이에 의해 발생합니다(Table 3 and Appendix 3) [2].

4. 비정밀성(Imprecision)

연구에 포함된 대상수가 적고 사건이 드물면 추정치의 신뢰구간이 넓어지며, 진료지침 개발위원회는 근거수준을 낮출 수 있습니다. 예를 들어, 대부분의 건강 결과에 대한 췌장암

Table 2. GRADE evidence profile for impact of surgical alternatives for pancreatic cancer from systematic review and meta-analysis of randomised controlled trials in inpatient hospitals of pylorus preserving versus standard Whipple pancreaticoduodenectomy for pancreatic or periampullary cancer by Karanicolas et al. [9]

No of studies (no. of participants)	Quality assessment					Summary of findings			
	Study limitations ^{a)}	Consistency	Directness	Precision	Publication bias	Relative effect (95% CI) ^{b)}	Best estimate of Whipple group risk	Absolute effect (95% CI)	Quality
Five year mortality:									
3 (229)	Serious limitations (-1)	No important inconsistency	Direct	No important imprecision	Unlikely	0.98 (0.87 to 1.11)	82.5%	20 less/1,000; 120 less to 80 more	+++ , moderate
In-hospital mortality:									
6 (490)	Serious limitations (-1)	No important inconsistency	Direct	Imprecision (-1) ^{c)}	Unlikely	0.40 (0.14 to 1.13)	4.9%	20 less/1,000; (50 less to 10 more)	++ , low
Blood transfusions (unit):									
5 (320)	Serious limitations (-1)	No important inconsistency	Direct	No important imprecision	Unlikely	-	2.45 units	-0.66 (-1.06 to -0.25); favours pylorus preservation	+++ ; moderate
Biliary leaks:									
3 (268)	Serious limitations (-1)	No important inconsistency	Direct	Imprecision (-1) ^{c)}	Unlikely	4.77 (0.23 to 97.96)	0	20 more/1,000 20 less to 50 more	++ , low
Hospital stay (d):									
5 (446)	Serious limitations (-1)	No important inconsistency	Direct	Imprecision (-1) ^{c)}	Unlikely	-	19.17 days	-1.45 (-3.28 to 0.38); favours pylorus preservation	++ , low
Delayed gastric emptying:									
5 (442)	Serious limitations (-1)	Unexplained heterogeneity (-1) ^{d)}	Direct	Imprecision (-1) ^{c)}	Unlikely	1.52 (0.74 to 3.14)	25.5%	110 more/1,000; 80 less to 290 more	+ , very low

^{a)}Unclear allocation concealment in all studies, patients blinded in only one study, outcome assessors not blinded in any study, >20% loss to follow-up in three studies, not analysed using intention to treat in one study. ^{b)}Relative risks (95% confidence intervals) are based on random effect models.

^{c)}Confidence interval (CI) includes possible benefit from both surgical approaches. ^{d)} $I^2 = 72.6\%$, $p = 0.006$.

Adapted from the article of Guyatt et al. BMJ 2008;336:995-8 [2].

See Appendix 2 (complete translate in Korean).

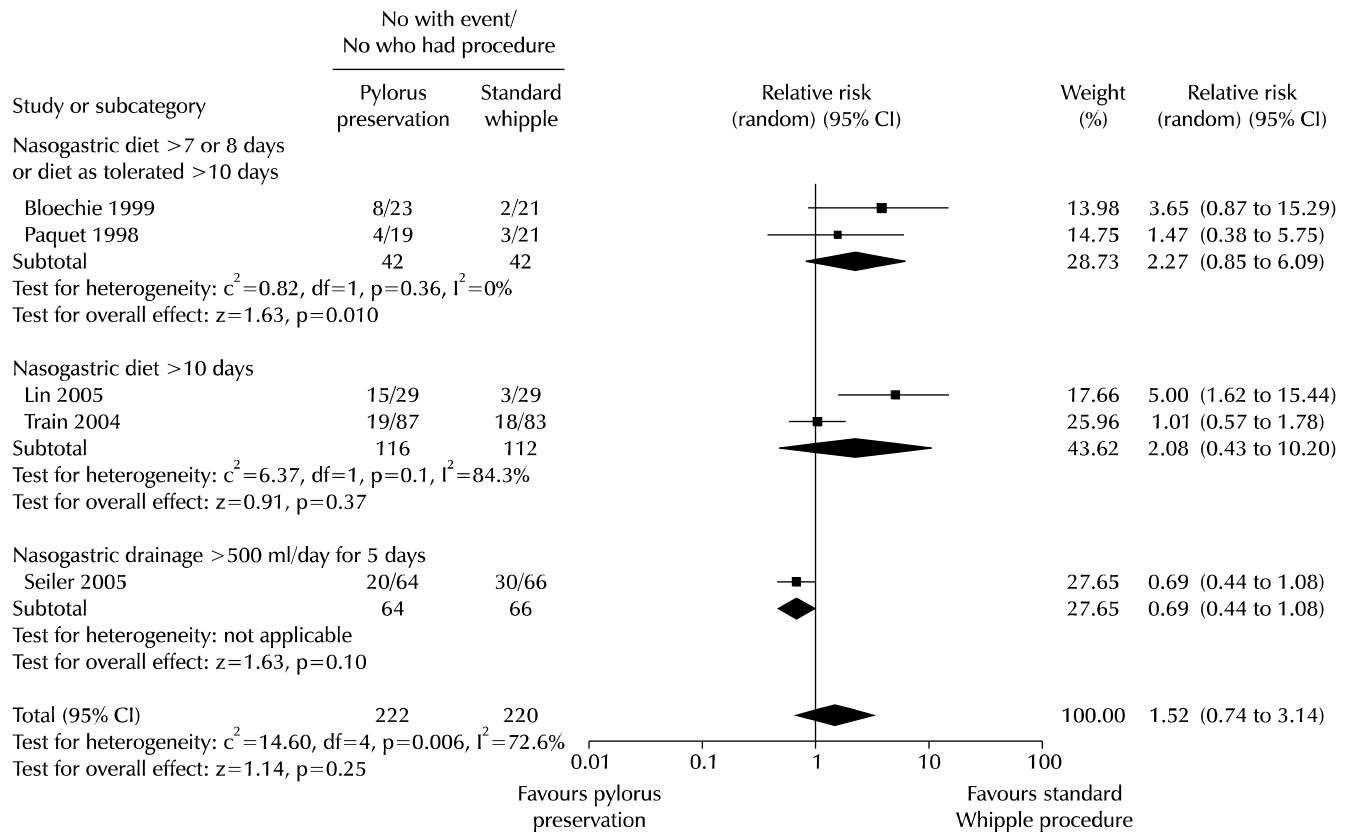


Fig. 2. Effect on delayed gastric emptying of pylorus preserving pancreaticoduodenectomy compared with standard Whipple procedure for pancreatic adenocarcinoma. CI: confidence interval. Adapted from the article of Guyatt et al. BMJ 2008;336:995-8 [2].

Table 3. Quality of evidence is weaker if comparisons in trials are indirect

Question of interest	Source of indirectness
Relative effectiveness of alendronate and risedronate in osteoporosis	Indirect comparison: randomized trials have compared alendronate with placebo and risedronate with placebo, but trials comparing alendronate with risedronate are unavailable
Oseltamivir for prophylaxis of avian flu caused by influenza A (H5N1) virus	Differences in population: randomised trials of oseltamivir are available for seasonal influenza, but not for avian flu
Sigmoidoscopic screening for prevention of mortality from colon cancer	Differences in intervention: randomized trials of faecal occult blood screening provide indirect evidence, bearing on potential effectiveness of sigmoidoscopy
Choice of drug for schizophrenia	Differences in comparator: series of trials comparing newer generation neuroleptic agents with fixed doses of haloperidol 20 mg provide indirect evidence of how newer agents would compare with lower, flexible doses of haloperidol that clinicians typically use
Rosiglitazone for prevention of diabetic complications in patients at high risk of diabetes	Differences in outcome: randomized trial shows delay in development of biochemical diabetes with rosiglitazone but was underpowered to tackle diabetic complications

Adapted from the article of Guyatt et al. BMJ 2008;336:995-8 [2].
See Appendix 3 (complete translate in Korean).

환자의 수술방법에 대한 무작위 연구들의 신뢰구간은 ‘효과 있음(important effect)’뿐만 아니라 ‘효과 없음(no effect)’을 모두 포함하고 있습니다(wide confidence interval, Table 2 and Appendix 2) [2,9].

5. 출판비뮌(Publication Bias)

연구자가 선택적으로 자신의 연구를 보고하지 않는 경우 (reporting bias) 근거수준은 낮아질 수 있습니다. 이러한

선택적 결과보고는 연구수가 적은 경우 또는 기업에서 연구비를 후원한 경우에 전형적으로 발생합니다.

근거수준을 높일 수 있는 3가지 고려요소

관찰 연구는 비록 잘 수행되었다고 할지라도 근거수준 ‘낮음’으로 평가되지만, 특수한 상황에서는 근거수준 ‘중등도’ 또는 ‘높음’으로 평가되는 경우가 존재합니다(Table 1 and

Appendix 1) [2,10]. 방법론적으로 엄격히 수행된 관찰 연구에서 치료 효과의 크기가 클 때 우리는 그 결과를 매우 신뢰할 수 있을 것입니다. 비록 이러한 경우, 관찰 연구가 실제 효과보다 치료 효과를 과대 평가할 수 있으나, 치료의 명백한 이득(benefit)은 연구 설계에 따른 단점을 상쇄할 수 있을 것입니다.

예를 들어, 관찰 연구를 이용한 메타분석에서 자전거 헬멧은 충돌 사고에 참여한 자전거 운전자의 두경부 손상 위험을 크게 감소시켰습니다(교차비 0.31, 95% 신뢰구간 0.26-0.37) [11]. 이렇게 큰 치료 효과는 근거수준 '중등도'에 해당할 것입니다. 마찬가지로, 심장 판막 치환술 환자에서 warfarin 사용은 혈전색전증의 발생을 크게 감소시킨다는 것을 관찰 연구를 이용한 메타분석에서 확인할 수 있었습니다(상대 위험도 0.17, 95% 신뢰구간 0.13-0.24) [12]. 이렇게 충분히 큰 치료 효과는 근거수준 '높음'에 해당할 것입니다.

양-반응 관계(dose-response gradient)가 있거나 교란변수가 추정 효과의 크기를 낮출 때(plausible confounding, which would reduce a demonstrated effect) 역시 근거수준을 높일 수 있습니다.

핵심적 건강 결과(Critical Outcome)이 근거수준을 결정한다

진료지침은 환자에게 중요한 모든 건강 결과에 대한 근거와 근거수준에 기반하여야 합니다. 근거수준이 각각의 건강 결과에 따라 상이한 경우 GRADE는 종합적 근거수준을 평가할 때 건강 결과들의 근거수준 중에서 가장 낮은 근거수준에 따르도록 권고하고 있습니다. 예를 들어, 췌장암 환자의 수술방법에 대한 연구는 6개의 개별 건강 결과에 대하여 '중등도'부터 '매우 낮음'까지 다양한 근거수준을 보여주고 있습니다(Table 2 and Appendix 2) [2,9]. 만약 위 배출 지연이 핵심적 건강 결과라면 종합적 근거수준은 '매우 낮음'으로 생각할 수 있을 것입니다. 그러나 위 배출 지연이 중요하지만 핵심적이지 않은 건강 결과라면 근거수준은 비록 5년 사망률의 근거수준이 '중등도'임에도 불구하고 병원 내 사망률의 근거수준에 따라 '낮음'으로 결정될 것입니다(5년 사망률 및 병원내 사망률은 핵심적인 건강 결과[critical outcome]).

근거요약표

GRADE는 임상 의사의 근거에 기반한 의사결정을 돕기 위해 간결하고 명료한 요약표를 제공합니다(Table 3 and Appendix 3) [2].

결론

GRADE는 진료지침 개발 각 단계의 의사결정과정에 필요한 철저한 검토와 토론을 통하여 이를 뒷받침하는 근거수준을 평가하고 요약하기 위한 명확하고 포괄적인 방법론을 제공합니다.

CONFLICT OF INTEREST

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

AUTHOR CONTRIBUTIONS

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

ORCID

Ho Won Kang, <https://orcid.org/0000-0002-8164-4427>
Jae Hung Jung, <https://orcid.org/0000-0002-4990-7098>
Do Kyung Kim, <https://orcid.org/0000-0002-3696-8756>
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>
Eu Chang Hwang, <https://orcid.org/0000-0002-2031-124X>

REFERENCES

1. 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;328:1490.
2. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ; GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? BMJ 2008;336:995-8.
3. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. CMAJ 1988;138:697-703.
4. Schunemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 10. integrating

- values and consumer involvement. *Health Res Policy Syst* 2006;4:22.
5. Fletcher SW, Spitzer WO. Approach of the Canadian task force to the periodic health examination. *Ann Intern Med* 1980;92: 253-4.
 6. Schunemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, et al.; ATS Documents Development and Implementation Committee. 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;174:605-14.
 7. Guyatt G, Cook D, Devereaux PJ, Meade M, Straus S. Therapy. In: Guyatt G, Rennie D, editors. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago: American Medical Association; 2002.
 8. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203-9.
 9. Karanickolas PJ, Davies E, Kunz R, Briel M, Koka HP, Payne DM, et al. The pylorus: take it or leave it? Systematic review and meta-analysis of pylorus-preserving versus standard whipple pancreaticoduodenectomy for pancreatic or periampullary cancer. *Ann Surg Oncol* 2007;14:1825-34.
 10. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007;334:349-51.
 11. Thompson DC, Rivara FP, Thompson R. Helmets for preventing head and facial injuries in bicyclists. *Cochrane Database Syst Rev* 2000;(2):CD001855.
 12. Cannegieter SC, Rosendaal FR, Briet E. Thromboembolic and bleeding complications in patients with mechanical heart valve prostheses. *Circulation* 1994;89:635-41.

Appendix 1. 근거수준을 결정하는 요소

근거수준을 낮출 수 있는 요소

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

근거수준을 높일 수 있는 요소

- 효과크기가 클 때(large magnitude of effect)
- 교란변수가 추정 효과의 크기를 낮출 때(plausible confounding, which would reduce a demonstrated effect)
- 양-반응 관계(dose-response gradient)

Appendix 2. 췌장 또는 췌대부주위 암에 대한 표준 Whipple 이자생창자절제술과 유문 보존 이자생창자절제술의 체계적 고찰, 메타분석을 통해 췌장암 환자의 수술적 대안의 영향에 대한 GRADE 근거요약표

연구수 (대상자수)	근거수준 평가					결과요약			
	비뚤림 위험 ^{a)}	일관성	직접성	정밀성	출판비뚤림	상대 효과 (95% CI) ^{b)}	표준 Whipple 이자생창자절 제술 추정치	절대 효과 (95% CI)	수준
5년 사망률:									
3 (229)	심각함(-1)	비일관성 없음	직접적	비정밀성 없음	가능성 적음	0.98 (0.87 to 1.11)	82.5%	1,000명당 20명 적음; 120명 적음에서 80명 많음	+++, 중등도
병원내 사망률:									
6 (490)	심각함(-1)	비일관성 없음	직접적	부정확 (-1) ^{c)}	가능성 적음	0.40 (0.14 to 1.13)	4.9%	1,000명당 20명 적음; 50명 적음에서 10명 많음	++, 낮음
수혈(유닛):									
5 (320)	심각함(-1)	비일관성 없음	직접적	비정밀성 없음	가능성 적음	-	2.45 유닛	-0.66 (-1.06 to -0.25); 유문 보존술식 선호	+++, 중등도
담즙액 누출:									
3 (268)	심각함(-1)	비일관성 없음	직접적	부정확 (-1) ^{c)}	가능성 적음	4.77 (0.23 to 97.96)	0	1,000명당 20명 많음; 20명 적음에서 50명 많음	++, 낮음
입원기간(일):									
5 (446)	심각함(-1)	비일관성 없음	직접적	부정확 (-1) ^{c)}	가능성 적음	-	19.17일	-1.45 (-3.28 to 0.38); 유문보존술식 선호	++, 낮음
위 배출 지연:									
5 (442)	심각함(-1)	설명되지 않는 이질성(-1) ^{d)}	직접적	부정확 (-1) ^{c)}	가능성 적음	1.52 (0.74 to 3.14)	25.5%	1,000명당 110명 많음; 80명 적음에서 290명 많음	+, 매우 낮음

CI: confidence interval.

^{a)} 모든 연구에서 불명확한 무작위 배정은 폐, 한 연구에서만 환자 눈가림이 이루어짐, 결과평가자 눈가림 이루어지지 않음, 새 연구에서 20% 이상의 추적관찰 실패율, 한 연구에서는 치료의향에 따른 분석 원칙이 지켜지지 않음. ^{b)} 상대 위험 (95 % 신뢰 구간)은 변량효과모형을 기반으로 함. ^{c)} 신뢰 구간에는 두 가지 수술 방법의 추정 효과를 포함. ^{d)} $I^2=72.6\%$, $p=0.006$.

Appendix 3. 비직접성의 유형(연구의 비교가 간접적 인 경우 증거의 질이 약하다)

핵심 질문	비직접성의 원인
골다공증에서 alendronate와 risedronate의 상대적 효과	간접 비교: 무작위 연구는 alendronate와 위약을 사용한 군 및 risedronate와 위약을 사용한 군을 비교했지만 alendronate와 risedronate를 직접 비교하는 연구는 수행되지 않았다.
인플루엔자 A (H5N1) 바이러스에 의한 조류 인플루엔자 예방을 위한 Oseltamivir 결장암으로 인한 사망을 감소시키기 위한 S자 결장경 선별검사	인구 집단의 차이: Oseltamivir의 무작위 연구는 계절성 인플루엔자에서 수행되었지만 조류 독감을 대상으로 이루어지지 않았다.
정신 분열증 치료제 선택	중재법의 차이: 대변 잠혈 검사의 무작위 연구는 S자 결장경 검사의 잠재적 효과에 대한 간접적인 증거를 제공한다.
고위험 당뇨병 환자에서 당뇨병 합병증 예방을 위한 Rosiglitazone	대조군의 차이: 새롭게 개발된 항정신병제제 투여군과 haloperidol 20 mg의 지속 투여한 군을 비교하는 일련의 임상 시험은 임상적으로 사용되는 용량의 haloperidol (20 mg 미만)에 비하여 새로운 항정신병약제의 효용성에 대하여 간접적인 증거를 제공한다.
	건강 결과의 차이: 무작위 임상 시험에서 Rosiglitazone은 생화학적 당뇨병 발병을 지연 시켰으나 당뇨병 합병증을 줄이지 못하였다.