



# Moving towards Evidence-Based Clinical Practice Guidelines

Jae Hung Jung<sup>1,2</sup>, Juan V A Franco<sup>3,4</sup>, Philipp Dahm<sup>5,6</sup>

<sup>1</sup>Department of Urology, <sup>2</sup>Institute of Evidence-Based Medicine, Yonsei University Wonju College of Medicine, Wonju, Korea, <sup>3</sup>Argentine Cochrane Centre, Instituto Universitario Hospital Italiano, Buenos Aires, <sup>4</sup>Family and Community Medicine Service, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina, <sup>5</sup>Urology Section, Minneapolis Veterans Healthcare System, Minneapolis, <sup>6</sup>Department of Urology, University of Minnesota, Minneapolis, MN, USA

The Institute of Medicine in its report “Clinical Practice Guidelines we can trust” defined standards for clinical practice guidelines. However, many guidelines continue to rely on expert opinion and lack a formal framework for moving from evidence to recommendations. These guidelines may or may not be labeled as “consensus statements” and do not meet contemporary standards for guideline documents we would refer to as “evidence-based”. Therefore, the Grading of Recommendations Assessment, Development and Evaluation working group developed a novel, rigorous and transparent approach to grading certainty (quality) of evidence. In addition, it created a system for “moving from evidence to decisions”, for example for the development of evidence-based guidelines. In this article, we aim to introduce this approach to appraising the certainty of relevant evidence and estimate the benefits and detriments of health care interventions within the larger context of evidence-based medicine.

**Keywords:** Evidence-based medicine; Guideline

**Received:** 4 October, 2018

**Revised:** 4 November, 2018

**Accepted:** 4 December, 2018

Copyright © 2018, 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.

**Correspondence to:** Philipp Dahm

<https://orcid.org/0000-0003-2819-2553>

Urology Section, Minneapolis Veterans Healthcare System, 112D, 1 Veterans Drive, Minneapolis, MN 55417, USA

Tel: +1-612-467-3532, Fax: +1-612-467-2232

E-mail: [pdahm@umn.edu](mailto:pdahm@umn.edu)

## INTRODUCTION

We applaud the publication issued by Korean Centers for Disease Control and Prevention entitled ‘Guidelines for the antibiotics use in urinary tract infection’, which has been endorsed by the Korean Association of Urogenital Tract Infection and Inflammation [1]. This guideline meets an important prerequisite of an evidence-based approach as it links recommendations to supporting evidence. However, this link is primarily based on study design, which is based on the hierarchy of evidence that we associate with the Center for Evidence-Based Medicine in Oxford. Integral to this framework is that it places “weaker” study designs (pre-clinical studies and case series) at the lowest level and follows this with case-control and cohort studies,

and finally with randomized controlled trials (RCTs) and systematic reviews, which are placed at the very top of the hierarchy. However, this hierarchy of evidence has been challenged because we cannot place the same confidence in different studies at the same hierarchical “level”. We are all familiar with examples such as randomized controlled studies labeled as “level I evidence” with critical limitations unrelated to study design that only provide low quality evidence. Murad et al. [2] therefore developed an alternative hierarchy in which the straight lines separating study designs were replaced with wavy lines (Fig. 1) [2]. This modified version signals that evidence from poorly conducted RCTs may be less reliable than that drawn from methodologically rigorous cohort studies.

In recognition of the shortcoming that such hierarchies

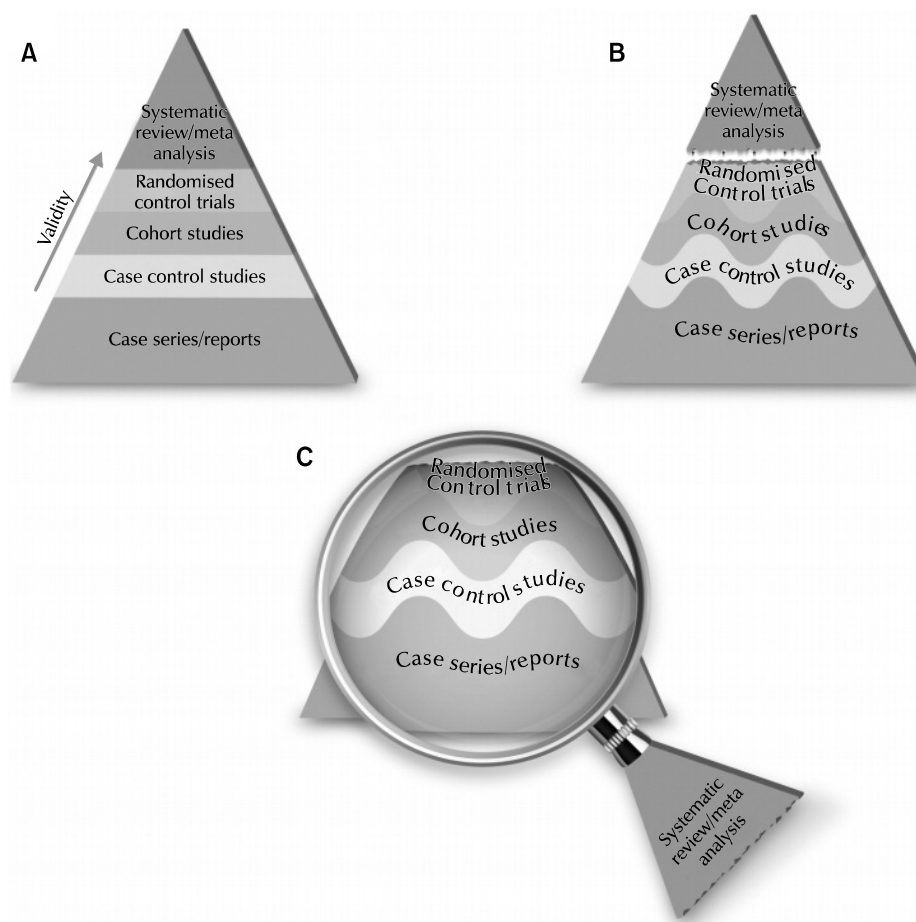


Fig. 1. The traditional pyramid (A), new evidence-based medicine pyramid (B, C). (B) Wavy lines separating study designs and systematic reviews separated from hierarchy. (C) Critical appraising process based on systematic reviews. Adapted from the article of Murad et al. *Evid Based Med* 2016;21:125-7 [2].

of evidence place too much emphasis on study design, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group, which has been active since the early 2000s, developed a novel, rigorous, transparent approach to grading the certainty (quality) of evidence that incorporates additional domains that affect confidence. In addition, it created a system for “moving from evidence to decisions”, for example for the development of evidence-based guidelines [3].

In this article, we introduce this approach to the appraisal of the certainty of relevant evidence and estimate the benefits and detriments of health care interventions within the larger context of evidence-based medicine.

## EVIDENCE-BASED MEDICINE

Evidence does not only include RCTs and systematic

reviews with or without meta-analysis, but also any empirical observations regardless of whether they were systematically collected [4]. For example, proper follow-up studies, such as cohort studies, are needed to estimate the natural course and determinants of a disease. Evidence may even be provided by the basic sciences such as genetics or immunology.

In 1991, Guyatt [5] from McMaster Medical School coined the term ‘evidence-based medicine’ in a short editorial for the ACP Journal Club. Evidence-based medicine was defined as the conscientious, explicit, and judicious use of current best evidence when making clinical decisions about the care of individual patients [6], whereas most clinicians have been taught to refer to authority (e.g., textbooks and expert opinions) to resolve issues of patient management. Therefore, Guyatt [5] suggested a new strategy, which included finding relevant research evidence, critically

appraising it, and applying its results to patient care. This strategy requires clinicians regularly consult the medical literature to answer clinical questions and make independent assessments of evidence, and thus, evaluate the credibilities of opinions offered by experts [7]. This definition was further expanded to include patient values and preferences when determining the best course of action for a given patient [8]. As a result, even the existence of high-quality evidence does not automatically imply that a given treatment or diagnostic test should or should not be performed or used. Instead, clinicians need to consider patients' values and preferences and individual circumstances in addition to available evidence; "evidence alone is never enough".

## EVIDENCE-BASED CLINICAL PRACTICE GUIDELINES

Many guidelines are developed by panels of experts with specific content expertise who arrive at recommendations without systematically reviewing evidence. These guidelines may or may not be labeled as "consensus statements" but they do not meet contemporary standards for "evidence-based" guideline documents. Recognizing this issue, the Institute of Medicine in its report "Clinical Practice Guidelines we can trust" defined standards for clinical practice guidelines, and one of these refers to the need for systematic review of evidence, which includes the rating of certainty. Other important aspects relate to stakeholder representation, management of conflict of interest, and the use of a transparent approach when moving from evidence to recommendations [9]. Several tools can be used to assess the quality of the guideline development process [10], and

the most comprehensive and well-known is the AGREE instrument (<https://www.agreetrust.org>).

### 1. Summarizing the Evidence: Systematic Reviews

The number of RCTs published in MEDLINE expanded from 5,000 per year in 1978-1985 to 25,000 per year in 1994-2001, and thus, clinicians can no longer keep up with the rapidly expanding knowledge base and are at risk of being overwhelmed by vast volumes of evidence of uncertain value [9,11]. Systematic reviews have become necessary tools as a first step of evidence-based medicine approaches, and thus, the development of high-quality guidelines critically depends on the availability of reliable systematic reviews [9].

Unlike narrative reviews, which provide broad overviews of clinical conditions, systematic reviews present a summary of research evidence that addresses a specific clinical question in a systematic, reproducible manner [12]. Furthermore, because systematic reviews serve a vital role in clinical decision making, clinicians should expect that clinical questions be addressed using consistent and unbiased methodological standards (Table 1) [13]. Systematic flaws in the design or conduct of a review may introduce bias in any stage of the review process, and explicit methodological guidelines have been introduced on how to conduct systematic reviews. In particular, Cochrane formally adopted the MECIR (Methodological Expectations of Cochrane Intervention Review) guidelines and the US Institute of Medicine has recommended standards for conducting high-quality systematic reviews [9,12]. In addition, the development and adoption of the PRISMA (Preferred Reporting Items for Systematic Reviews and

**Table 1.** Methodological standards for conducting rigorous systematic reviews

Review process	Method
Defining scope of questions	Focused clinical question with PICO components
Defining methods	A priori written protocol
Search methods used to identify studies	Comprehensive, transparent, reproducible search of diverse database including trial/study registries and grey literature without any restriction of language or publication status
Screening and selecting studies	Based on predefined criteria (PICO)/study selection in duplicate
Assessing risk of bias	Explicit quality assessment (e.g., Cochrane Collaboration's tool for assessing risk of bias)/assessment of risk of bias in duplicate
Data extraction	Continuous or dichotomous statistical values/data extraction in duplicate
Data synthesis	Quantitative summary (e.g., meta-analysis)
Sensitivity and subgroup analyses	Based on a prior written protocol
Data quality	Scientific quality of evidence interpreting/discussing the results of the review (e.g., GRADE)

PICO: participants, intervention, comparator, outcomes, GRADE: Grading of Recommendations Assessment, Development and Evaluation. Adapted from the article of Shea et al. *BMJ* 2017;358:j4008 [13].

Meta-Analyses) statement has led to improvements in the reporting of systematic reviews [14], and AMSTAR 2 (A MeaSurement Tool to Assess systematic Reviews 2) and ROBIS (Risk Of Bias In Systematic reviews) were devised to provide critical appraisal and quality assessments of systematic reviews at all stages of the review process [13,15].

However, systematic reviews containing overlapping, redundant, and misleading information with little value in terms of informed clinical decision-making or health policy are being increasingly published [11,16]. Accordingly, clinical practice guideline developers should make efforts to prevent potential biases when including the results of such reviews.

## 2. Grading of Recommendations Assessment, Development and Evaluation Approach

Evidence summaries constitute the first step toward developing evidence-based practice guidelines [17]. The GRADE Working Group (<http://www.gradeworkinggroup.org/>) was formed in 2000 by individuals interested in addressing the shortcomings of healthcare grading systems. Systematic reviews of the effects of healthcare have become components of guideline developments but are not sufficient for making well-informed decisions since they do not adequately integrate the certainties of findings [18]. However, judgments about certainties of evidence and strengths of recommendations in healthcare are complex, and thus, GRADE developed a common, sensible, and transparent approach to their grading. Furthermore, Guyatt et al. [18] and Alonso-Coello et al. [19] recommended evidence summaries should be added to systematic reviews to provide bases for judgments about evidence certainties and strengths of recommendations. In 2017, over 100

organizations including the World Health Organization, Cochrane Collaboration, National Institute of Health and Clinical Excellence, and UpToDate have endorsed GRADE as the standard for guideline development [18,20].

### 1) Rating certainties of evidence

Clinical practice guidelines fundamentally depend on appraisals of the quality of relevant evidence related to clinically important outcomes. GRADE defines certainty of evidence as the extent to which confidence in an estimate of an effect influences critical and important outcomes (e.g. mortality, quality of life, adverse events) and reliably supports a specific recommendation (Table 2) [3,21]. The GRADE approach involves a four-tiered rating system of high, moderate, low, and very low, which reflect a gradient of confidence in estimates of treatment effect. Although initially the highest quality rating was assigned to RCT evidence, guideline developers may downgrade randomized trial evidence to moderate, low, or even very low certainty, based on considerations of five factors, that is, risk of bias, inconsistency, indirectness, imprecision, and publication bias. The certainty of evidence derived from observational studies starts at 'low' but may be re-graded. Furthermore, there are settings in which the certainties of evidence of observational studies can be upgraded based on; considerations of a large effect size, demonstration of a dose-response gradient, and when all plausible confounding would reduce a demonstrated effect or suggest a spurious effect [3]. The most common reason for upgrading is a large or very large effect size. Table 3 listed the five upgrade and three downgrade factors. The most comprehensive series describing the GRADE approach was published in the *Journal of Clinical Epidemiology* [20].

**Table 2.** GRADE Working Group grades of evidence

Levels of certainty	Underlying methodology	Definition of certainty of evidence
High	Randomized trials; or double-upgraded observational studies	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	Downgraded randomized trials; or upgraded observational studies	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Double-downgraded randomized trials; or observational studies	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	Triple-downgraded randomized trials; or downgraded observational studies; or case series/case reports	Any estimate of effect is very uncertain

GRADE: Grading of Recommendations Assessment, Development and Evaluation. Adapted from the article of Guyatt et al. *J Thromb Haemost* 2013;11:1603-8 [21].

**Table 3.** Factors that may affect certainty level of a body of evidence

Factors for downgrade	Factors for upgrade <sup>a)</sup>
Study limitations (risk of bias)	Large magnitude of effect
Inconsistency (unexplained heterogeneity)	All plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results show no effect
Indirectness (indirect PICO - external validity)	Dose-response gradient
Imprecision (wide confidence interval)	
Publication bias (omission of the studies that show no effect)	

PICO: participants, intervention, comparator, outcomes.

<sup>a)</sup>Upgrade cannot be applied to a randomized controlled trial and upgrade may not be possible if the evidence has been downgraded by some of the factors in the left column.

Adapted from the article of Guyatt et al. J Thromb Haemost 2013;11:1603-8 [21].

## 2) Rating strength of recommendations

The strength of a recommendation reflects the extent to which we can be confident that the desirable effects (e.g., reductions in morbidity and mortality) of an intervention outweigh its undesirable effects (e.g., adverse effects) [19,22,23].

GRADE classifies recommendations using only two categories of strength, that is, strong or weak. Briefly, strong recommendations are made when a guideline panel is confident that the desirable effects of adherence to a recommendation clearly outweigh its undesirable effects. When a panel is less confident, they can make a weak recommendation that indicate the desirable effects of adherence probably outweigh undesirable effects [19,23]. This binary classification provides clear direction to patients, clinicians, and policy makers. Strong recommendations imply most people with the same clinical condition would choose the recommended management and that clinicians should guide their patients to accept recommendation. For policymakers, a strong recommendation can be adopted as policy in most situations. However, if the strength of a recommendation is weak, clinicians should discuss the merits and demerits of recommended management with patients and compare these with those of alternative management strategies to ensure adequate implementation of shared-decision making [24].

Whereas the certainty of evidence is important, guideline developers should consider other key factors when making recommendations such as the balance between desirable and undesirable effects, patients' values and preferences, healthcare resource utilization and equity implications, and the feasibility and acceptability of interventions as defined in the *Evidence to Decision Framework* [19]. A high certainty of evidence does not necessarily imply strong recom-

mendations as strong recommendations can arise from low quality evidence [22]. For example, patients' values and preferences may be age-dependent. Younger patients may place greater value on the prolongation of life, whereas older patients may consider quality of life to be more important. Recently, the BMJ published a GRADE-based, weak recommendation for prostate-specific antigen-based prostate cancer screening [25], which was justified by moderate quality evidence of small benefit but greater harm and considerable variation of its value with respect to the outcomes of screening [26].

## CONCLUSIONS

Many guidelines are developed with undue reliance on expert opinion and no formal framework for moving from evidence to recommendations. Guidelines devised in this manner do not meet minimal standards for "evidence-based guidelines" and should, therefore, be abandoned in favor of an approach such as that offered by GRADE. The defining features of GRADE include reliance on high quality systematic reviews, multidimensional assessments of certainty of evidence that go well beyond study design, and the use of a formal evidence-to-decision framework for moving from evidence to recommendations. These elements will result in guidelines that inspire trust by patients, clinicians, and policy makers and ultimately, improve patient outcomes.

## CONFLICT OF INTEREST

Philipp Dahm serves as Coordinating Editor of Cochrane Urology for the International Cochrane Collaboration. He is also a member of GRADE Working Group and the US

GRADE Network, Jae Hung Jung and Juan V A Franco are Cochrane Urology Contact Editors.

## REFERENCES

1. Korea Centers for Disease Control. Guidelines for the antibiotic use in urinary tract infections [Internet]. Cheongju: Korea Centers for Disease Control; 2018 [cited 2018 Sep 24]. Available from: <http://www.cdc.go.kr/CDC/together/CdcKrTogether0302.jsp?menuIds=HOME006-MNU2804-MNU3027-MNU2979&cid=138017>.
2. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med* 2016;21:125-7.
3. 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.
4. Zimerman AL. Evidence-based medicine: a short history of a modern medical movement. *Virtual Mentor* 2013;15:71-6.
5. Guyatt GH. Evidence-based medicine. *ACP J Club* 1991;114:A16.
6. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. 1996. *Clin Orthop Relat Res* 2007;455:3-5.
7. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
8. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 2nd ed. Edinburgh: Churchill Livingstone; 2000.
9. Robin Graham; Institute of Medicine. Clinical practice guidelines we can trust. Washington, DC: National Academies Press; 2011.
10. Siering U, Eikermann M, Hausner E, Hoffmann-Eßer W, Neugebauer EA. Appraisal tools for clinical practice guidelines: a systematic review. *PLoS One* 2013;8:e82915.
11. Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q* 2016;94:485-514.
12. Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological Expectations of Cochrane Intervention Reviews [Internet]. London: Cochrane; 2018 [cited 2018 Sep 24]. Available from: <https://community.cochrane.org/mecir-manual/introduction-key-points>.
13. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
14. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
15. Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225-34.
16. Han JL, Gandhi S, Bockoven CG, Narayan VM, Dahm P. The landscape of systematic reviews in urology (1998 to 2015): an assessment of methodological quality. *BJU Int* 2017;119:638-49.
17. Guyatt G, Meade M. What is evidence-based medicine? In: Guyatt G, Rennie D, Meade M, Cook D, editors. *Users' guides to the medical literature. Essentials of evidence-based clinical practice*. 3rd ed. New York: McGraw-Hill education; 2015. p. 16-8.
18. 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.
19. Alonso-Coello P, Schunemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: introduction. *BMJ* 2016;353:i2016.
20. Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *J Clin Epidemiol* 2011;64:380-2.
21. Guyatt G, Eikelboom JW, Akl EA, Crowther M, Gutterman D, Kahn SR, et al. A guide to GRADE guidelines for the readers of JTH. *J Thromb Haemost* 2013;11:1603-8.
22. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al.; GRADE Working Group. Going from evidence to recommendations. *BMJ* 2008;336:1049-51.
23. Schunemann HJ, Mustafa R, Brozek J, Santesso N, Alonso-Coello P, Guyatt G, et al.; GRADE Working Group. GRADE guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *J Clin Epidemiol* 2016;76:89-98.
24. McCormack J, Elwyn G. Shared decision is the only outcome that matters when it comes to evaluating evidence-based practice. *BMJ Evid Based Med* 2018;23:137-9.
25. Tikkinen KAO, Dahm P, Lytvyn L, Heen AF, Vernooij RWM, Siemieniuk RAC, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a clinical practice guideline. *BMJ* 2018;362:k3581.
26. Vernooij RWM, Lytvyn L, Pardo-Hernandez H, Albarqouni L, Canelo-Aybar C, Campbell K, et al. Values and preferences of men for undergoing prostate-specific antigen screening for prostate cancer: a systematic review. *BMJ Open* 2018;8:e025470.