

GRADE guidelines—an introduction to the 10th–13th articles in the series

1. Introduction

The 10th–13th of a series of 22 articles providing detailed guidance for using Grading of Recommendations Assessment, Development, and Evaluation (GRADE) appear in this issue of *Journal of Clinical Epidemiology*. This article provides an updated overview of the series (Box 1) and an introduction to the four articles in this issue. We also explain our use of terminology, particularly “confidence in estimates of effect” (quality of evidence).

The first nine articles in the series describe the GRADE approach of rating quality of evidence in systematic reviews, health technology assessments, and clinical practice guidelines. Table 1 summarizes this approach.

The 10th article in the series, appearing in this issue of the journal, addresses judgments about how much confidence to place in estimates of resource use or cost. The GRADE approach to making these judgments is the same as for other outcomes. However, judgments about cost present special challenges. Cost is far more variable over time and across jurisdictions than are other outcomes and guideline developers deal differently with costs. For example, sometimes they ignore them and sometimes they require explicit consideration of cost-effectiveness.

After considering the five reasons for rating down and the three reasons for rating up the confidence in estimates of effect, systematic review authors and guideline developers must decide how much confidence to place in the estimate of effect for each outcome across those considerations. Moreover, guideline developers must look across outcomes and decide on a category for the overall quality of evidence. The 11th article in the series describes the GRADE approach to making these judgments.

The final two articles appearing in this issue of the journal deal with the challenges of creating evidence profiles and summary of findings tables. For binary outcomes, these include choosing between competing bodies of evidence (e.g., randomized trials and observational studies), and

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the *Journal of Clinical Epidemiology* website.

Box 1 GRADE *Journal of Clinical Epidemiology* series—list of articles

Introductory articles

1. Introduction and summary of findings tables
2. Framing the question and deciding on the importance of outcomes
3. Rating the quality of evidence—introduction

Rating confidence in estimates of effect (quality of evidence)

4. Rating the quality of evidence—risk of bias
5. Rating quality of evidence—publication bias
6. Rating the quality of evidence—imprecision (random error)
7. Rating the quality of evidence—inconsistency
8. Rating the quality of evidence—indirectness
9. Rating up the quality of evidence
10. Rating the quality of evidence for resource use

Summarizing the evidence

11. Summarizing the quality of evidence for individual outcomes and across outcomes
12. Preparing summary of findings tables—binary outcomes
13. Preparing summary of findings tables—continuous outcomes

Diagnostic tests

14. Applying GRADE to diagnostic tests—summarizing the evidence

Making recommendations

15. Going from evidence to recommendations—the meaning of strong and weak recommendations
16. Going from evidence to recommendations—determinants of a recommendation’s direction and strength
17. Going from evidence to recommendations—resource use
18. Going from evidence to recommendations—diagnostic tests

Observational studies

- 19. Special challenges in using observational studies
- 20. GRADE for public health and health policymaking

Concluding articles

- 21. Group processes, variations of GRADE, and further developments of GRADE part 1
- 22. Group processes, variations of GRADE, and further developments of GRADE part 2

Abbreviation: GRADE, Grading of Recommendations Assessment, Development, and Evaluation.

generating both relative and absolute estimates of effect. Additional complexities (e.g., conveying the magnitude of effect in an interpretable way) arise in generating summary of findings tables for continuous variables, which therefore warrants its own article in the series.

2. GRADE’s approach to terminology

Readers of this series may have noticed a shift in the terminology that we use from “quality of evidence” to “confidence in estimates of effect.” Although we strive to be clear in our use of language, we sometimes have found it challenging to find suitable terms that can be translated to different languages and have the same meaning to different audiences. Consequently, the GRADE approach is flexible with regard to the choice of words.

Nonetheless, poorly chosen words can be confusing. “Quality of evidence” is an example of this. In some contexts, it is easily understood and conveys the intended meaning, which is confidence in an estimate of effect. However, it sometimes is confused with “risk of bias,” which is one of the several considerations that might decrease confidence

in an estimate of effect. It also is sometimes interpreted as a derogatory comment about the efforts of researchers, rather than a judgment about confidence in an estimate of effect derived from research. Despite the best efforts of researchers, estimates of effect derived from well-designed and implemented research can sometimes warrant only a low level of confidence.

Similarly, although “confidence in estimates of effect” is easily understood in many contexts and is more likely to convey the intended meaning than “quality of evidence”, it might also be misunderstood or confusing. For example, it may be confused with “confidence intervals” (imprecision), which is also one of the several considerations that might decrease confidence in estimates of effect. It also might be misinterpreted as someone’s subjective feeling, rather than a judgment of how much confidence is warranted based on an explicit consideration of the issues in the table.

Thus, we are continuing to discuss and evaluate other alternatives, such as “certainty of the anticipated effect.” We invite advice and feedback on this terminology and we will be testing different audiences’ understanding of alternative terms.

Other examples of challenging terminology include “study limitations” (which we now call “risk of bias”) [1], “directness” (which groups together considerations of applicability, surrogate outcomes, and indirect comparisons) [2], and “weak recommendation” (which we will address in article 15 in this series). Our approach to challenges such as these is to:

- Shift our use of words when we discover problems while trying to avoid confusion by not making unnecessary changes and by using more than one term, at least for a period of time, when we make changes;
- Invite wide input from people with different languages and backgrounds and discuss potential changes over a period of time before disseminating them;
- Test understanding among different audiences when there is uncertainty or disagreement about what term to use; and

Table 1
GRADE’s approach to rating confidence in effect estimates (quality of evidence)

1. Establish initial level of confidence		2. Consider lowering or raising level of confidence		3. Final level of confidence rating
Study design	Initial confidence in an estimate of effect	Reasons for considering lowering or raising confidence		Confidence in an estimate of effect across those considerations
		↓ Lower if	↑ Higher if	
Randomized trials →	High confidence	Risk of Bias Inconsistency Indirectness Imprecision Publication bias	Large effect Dose response All plausible residual confounding and bias • Would reduce a demonstrated effect or • Would suggest a spurious effect if no effect was observed	High ⊕⊕⊕⊕
Observational studies →	Low confidence			Moderate ⊕⊕⊕○
				Low ⊕⊕○○
				Very low ⊕○○○

Abbreviation: GRADE, Grading of Recommendations Assessment, Development, and Evaluation.

- Offer a choice of terms when it is not possible to find a term that is optimal in all situations.

Gordon H. Guyatt

Andrew D. Oxman

Holger J. Schünemann

E-mail address: guyatt@mcmaster.ca (G.H. Guyatt)

References

- [1] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
- [2] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011;64:1303–10.