ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 380–382

Journal of Clinical Epidemiology

GRADE SERIES - GUEST EDITORS, SHARON STRAUS AND SASHA SHEPPERD

GRADE guidelines: A new series of articles in the *Journal of Clinical Epidemiology*

Gordon H. Guyatt[a,b,*], Andrew D. Oxman[c], Holger J. Schünemann[a,b], Peter Tugwell[d], Andre Knottnerus[e]

[a]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[b]*Department of Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[c]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs plass, 0130 Oslo, Norway*
[d]*Centre for Global Health, Institute of Population Health, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada*
[e]*Department of General Practice, PO Box 616, 6200 MD Maastricht, The Netherlands*

Accepted 29 September 2010

Abstract

The "Grades of Recommendation, Assessment, Development, and Evaluation" (GRADE) approach provides guidance for rating quality of evidence and grading strength of recommendations in health care. It has important implications for those summarizing evidence for systematic reviews, health technology assessment, and clinical practice guidelines. GRADE provides a systematic and transparent framework for clarifying questions, determining the outcomes of interest, summarizing the evidence that addresses a question, and moving from the evidence to a recommendation or decision. Wide dissemination and use of the GRADE approach, with endorsement from more than 50 organizations worldwide, many highly influential (http://www.gradeworkinggroup.org/), attests to the importance of this work. This article introduces a 20-part series providing guidance for the use of GRADE methodology that will appear in the *Journal of Clinical Epidemiology*. © 2011 Elsevier Inc. All rights reserved.

Keywords: GRADE; Evidence quality; Strength of recommendations; Risk of bias; Health technology assessment; Clinical practice guidelines

1. Introduction

The Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) Working Group is a group of health professionals, researchers, and guideline developers worldwide who, in 2000, began to work together to develop an optimal system of rating quality of evidence and determining strength of recommendations for clinical practice guidelines. The group now includes more than 200 members and continues, after a decade of work, to meet to refine and extend its methods. The group's more than 25 one- to two-day meetings thus far and uncountable e-mail discussions have become a laboratory for the development and refinement of the methodology of interpreting research evidence for clinical practice and health care decisions, and for optimally presenting that evidence to clinicians, patients, and policymakers.

The published literature includes a number of articles describing the GRADE approach, of which the most comprehensive is a six-part series published in 2008 in the *BMJ* [1–6]. The audience for these articles is, however, the clinician and policy-making users of GRADE's output, which includes evidence profiles, summary of findings tables, and graded recommendations (all facilitated by a computer program, GRADEpro, that the working group has produced [7] and an associated help file [8]).

What previous articles fail to do is provide detailed guidance for those responsible for using GRADE to produce this output: systematic review and health technology assessment authors and the guideline panelists and methodologists who provide support for guideline panels. A series of articles, the first four of which are included in this issue of *JCE*, address this deficiency.

This series, which provides guidance for each step in the application of GRADE, will include 20 articles (Table 1). The first introduces GRADE and its use in systematic reviews, guidelines, and health technology assessment, as well as presenting the final product of the GRADE approach to collecting and summarizing evidence: the evidence profile and summary of findings table. The second

* Corresponding author. CLARITY Research Group, Department of Clinical Epidemiology & Biostatistics, Room 2C12, 1200 Main Street West Hamilton, Ontario L8N 3Z5, Canada. Tel.: +905-527-4322; fax: +905-523-8781.

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

0895-4356/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi: 10.1016/j.jclinepi.2010.09.011

Table 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 the quality of evidence
4. Rating the quality of evidence—risk of bias
5. Rating the 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
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
GRADE and observational studies
18. Special challenges in using observational studies
Concluding articles
19. Group processes, variations of GRADE, and further developments of GRADE part 1
20. Group processes, variations of GRADE, and further developments of GRADE part 2

Abbreviation: GRADE, Grades of Recommendation, Assessment, Development, and Evaluation.

shows how GRADE uses the patient/intervention/comparator/outcome framework for structuring a clinical question and its approach to defining critical, important, and less important outcomes. The last of the three introductory articles presents GRADE's definition of quality of evidence (confidence in effect estimates). This third article provides the rationale for randomized trials beginning as high-quality evidence, and observational studies as low quality in GRADE's four-category system or quality rating (high, moderate, low, and very low). It also introduces five categories of reasons for rating down quality of evidence and three categories of reasons for rating up quality of evidence.

The subsequent five articles—the fourth to the eighth in the series—address the five categories of issues that may result in rating down the quality of evidence. The fourth article deals with risk of bias, presenting an approach similar to the Cochrane risk of bias tool. The fifth article is devoted to the other type of bias—publication bias—that can lower the quality of the evidence. The sixth article presents GRADE's approach to considering imprecision, an approach that focuses on the consideration of confidence intervals around point estimates associated with each outcome.

The series' seventh article explains the fourth reason for rating down quality, inconsistency, and outlines three relevant considerations: similarity of point estimates, the extent to which confidence intervals overlap, and the available statistical tests related to heterogeneity between study results. The eighth article presents the final category of rating down: indirectness. This refers first to differences between the population, intervention, and outcome addressed in the available studies and those of interest to systematic review authors and guideline developers. Second, it refers to indirect comparisons in which one is interested in recommending between two agents that have each been tested against a third comparator, but not directly against each other.

The ninth article deals with possibilities of rating up quality of evidence from observational studies. It presents the most common reason for rating up (a large effect) and two less common reasons (a dose–response gradient; and a conclusion that plausible residual confounding would further support inferences regarding treatment effect). The 10th article deals with special considerations in assessing risk of bias when the outcome is resource use (cost).

The 11th to 13th articles deal with issues in summarizing the evidence. Every body of evidence has limitations, and when to rate down quality for a particular outcome, and how much, is a major challenge. Furthermore, because the GRADE approach rates quality of evidence separately for each outcome, it is frequently the case that quality differs across outcomes. Deciding on an overall quality of evidence across outcomes is therefore challenging. The 11th article in the series addresses these issues. The 12th and 13th articles address details regarding the production of evidence profiles and summary of findings tables, the 12th dealing with binary end points and the 13th with continuous variables.

The 14th article addresses a particular challenge that the working group has faced: how to rate quality of evidence for diagnostic tests within the GRADE framework. The 15th and 16th articles deal with moving from evidence to recommendations and whether to classify recommendations as strong or weak (alternative terms for the latter are weak, discretionary, or contingent). These two articles explain four issues relevant to deciding on the strength of recommendations: the trade-off between desirable and undesirable consequences of the alternative management strategies, the quality of evidence, the extent of variability in values and preferences, and resource use considerations.

The current plan for the final articles in the series includes one dealing with the special challenges that observational studies present and two presenting the GRADE working group's perspective on group process, variations of GRADE, and possible developments of GRADE in the future. With respect to the future, there should be no expectation that the methodology presented in this series will remain the static definitive guide to applying GRADE. It

will not. This series provides suggestions for approaching a host of methodological issues. Some of the approaches are innovative: innovations include how to deal with surrogate end points; criteria for judging limitations as a result of imprecision; criteria for evaluating the credibility of a subgroup analyses; judging quality of evidence for diagnostic tests; and summarizing the magnitude of effect for continuous variables.

There is inherent instability in innovative approaches—refinements are inevitable, not necessarily for all but certainly for some. Moreover, there will in the future be methodological advances and refinements not only of innovations but also of established concepts. We hope that users of GRADE will not be dismayed by our inability to present a set of immutable criteria for applying the system. They can be reassured by the knowledge that this series presents a broad and comprehensive foundation that will stand them in good stead applying GRADE to systematic reviews, guidelines, and health technology assessments for now and in the future. The editors are proud to be presenting the series; we are confident that it will ultimately be seen as a milestone in the development of clinical epidemiology.

References

[1] 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;336:924–6.

[2] Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ. What is "quality of evidence" and why is it important to clinicians? BMJ 2008;336:995–8.

[3] Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Going from evidence to recommendations. BMJ 2008;336:1049–51.

[4] Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 2008;336:1106–10.

[5] Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A, et al. Incorporating considerations of resources use into grading recommendations. BMJ 2008;336:1170–3.

[6] Jaeschke R, Guyatt GH, Dellinger P, Schunemann H, Levy MM, Kunz R, et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ 2008;337. a744.

[7] Brozek J, Oxman AD, Schünemann H. GRADEpro [computer program]. Version 3.2 for Windows. Available at http://www.cc-ims.net/gradepro or http://mcmaster.flintbox.com/technology.asp?page=3993. 2008.

[8] Schünemann H, Brozek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendation; 2010.