

B. Rochwerg  
W. Alhazzani  
Roman Jaeschke

## Clinical meaning of the GRADE rules

---

Received: 13 March 2014  
Accepted: 15 March 2014  
Published online: 26 March 2014  
© Springer-Verlag Berlin Heidelberg and ESICM 2014

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s00134-014-3273-0](https://doi.org/10.1007/s00134-014-3273-0)) contains supplementary material, which is available to authorized users.

---

B. Rochwerg (✉) · W. Alhazzani · R. Jaeschke  
Department of Medicine, McMaster University,  
Hamilton, ON, Canada  
e-mail: bram.rochwerg@medportal.ca

R. Jaeschke  
e-mail: jaeschke@mcmaster.ca

W. Alhazzani  
Department of Critical Care, Prince Sultan Military Medical City,  
Riyadh, Saudi Arabia

R. Jaeschke  
Department of Clinical Epidemiology and Biostatistics,  
McMaster University, Hamilton, ON, Canada

Clinical practice guidelines (CPGs) are usually presented with a large number of recommendations and a long text. This format can be intimidating and challenging in terms of interpretation and application to patients. Although the GRADE rules are used frequently in critical care CPGs, these rules may not be familiar to most clinicians. Understanding the implications of the GRADE rules and how to interpret the different combinations of symbols, numbers, and letters is essential.

Often the biggest question facing both guideline authors and guideline readers is how confident they are that a recommendation should be followed. Sometimes it is clear that one course of action is better than the alternative (say, giving aspirin in the setting of myocardial

infarction rather than not giving it); sometimes the choice is less clear (a shorter course of antibiotics for the treatment of ventilator-associated pneumonia compared to a longer course). The GRADE system asks the authors of each recommendation to express their confidence that following a specific recommended course of action (rather than a potentially plausible alternative) will result in more benefits than downsides to patients. The level of such confidence (in 'GRADE language,' strength of recommendation) is first expressed in the assertiveness of the wording. For example, a strong recommendation will say, "we recommend rapid delivery of antibiotics in patients with sepsis," here indicating most patients in this situation should be managed this way. Other examples of strong wording used in this high confidence situation are 'clinicians should' or even more simply 'do this.' Alternatively, if the authors are less clear on the best course of action in a specific clinical scenario, they may express their lower confidence through a weak recommendation (more of a suggestion). Here, the choice of wording is different, for example, "we suggest that sedation strategies using non-benzodiazepine sedatives may be preferred over sedation with benzodiazepines." The alternative phrase is 'clinicians might' (as opposed to 'clinicians should'). The strength of recommendations has implications for patients, clinicians and policy makers [1] (Table 1).

Now, what makes CPG authors more or less confident in their recommendations? What determines whether recommendations are strong versus weak?

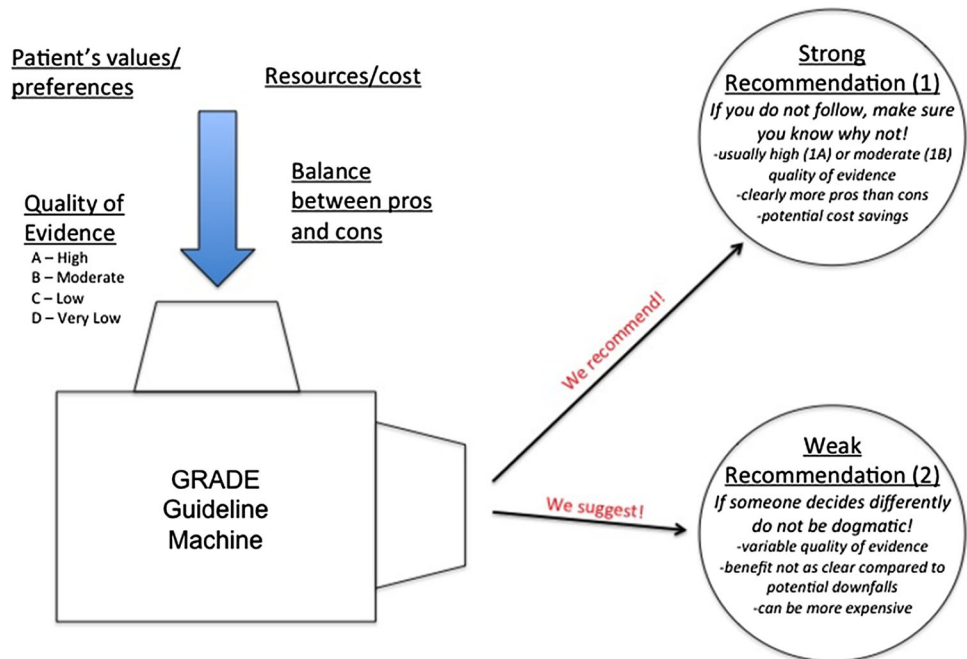
The first factor is the confidence that the true effects of the intervention are already known. In other words, how confident/certain are the authors that the true effect lies close to that of the available estimate? The GRADE term for this concept is quality of evidence (QoE). Intuitively, the more confident authors are about treatment effects, the more likely they are to issue a strong recommendation—a good example is the evolution of information regarding

**Table 1** Interpretation of strong and weak recommendations for different stakeholders

Implications	Strong recommendation	Conditional (weak) recommendation
For patients	Most individuals in this situation would want the recommended course of action and only a small proportion would not	The majority of individuals in this situation, if fully informed, would choose the suggested course of action, but some would not
For clinicians	Most individuals should receive the intervention. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator	Recognize that different choices may be appropriate for individual patients or that you are less confident in the advantages of using one course of action versus alternative. When relevant, the pros and cons should be explained to or discussed with the patient
For policy makers	The recommendation can be adapted as policy in most situations	Policy making will require substantial debate and involvement of various stakeholders

Modified from table found in [1]

**Fig. 1** Summary of the GRADE process



starches in the resuscitation of septic shock patients over the last few years [2, 3]. The GRADE approach incorporates several key factors into judging QoE [4], labeling this quality (confidence) from high, or moderate through to low or even very low (usually, from A to D sequentially). In general, we are more confident in evidence from randomized controlled trials, which start as high quality, as opposed to observational studies, which start as low quality. QoE may then be lowered if the estimates are not consistent from study to study (remember the activated protein C story? [5]). Or it may be lowered if the data from the studies are not directly reflective of our specific clinical situation (for example, applying evidence from adult studies to a pediatric patient population). In GRADE language, the above phenomena are labeled as inconsistency (heterogeneity) and indirectness, respectively. Indirectness may also apply in terms of the outcome.

GRADE recommendations should be made based on patient-important factors such as mortality or quality of life and QoE may be downgraded when the available literature focuses on only substitute outcomes (e.g., FEV1 or cardiac output). QoE from observational studies can also be raised under strict rules if there is stronger inclination to believe the data (e.g., a large treatment effect or major dose response) [6]. In order to make a strong recommendation, in effect confidently saying most clinicians should follow this course of action, usually a high (or moderate) QoE is needed [7].

On the other hand, a high QoE does not always mandate a strong recommendation—consider a situation in which the comparative effects of two alternatives are confidently known, but those effects are small and balanced in terms of risks and benefits to patients (e.g., warfarin for low-risk atrial fibrillation). This is an

example of the second factor influencing strength of recommendation, the balance between desirable and undesirable effects (in this case, a small decrease in stroke versus a small increase in bleeding risk). Overall, the larger the difference that exists between the potential benefits of intervention versus the potential downfalls to patients, the higher the likelihood that a strong recommendation will be issued.

The third determinant of strength of recommendation is patient variability in underlying values and preferences. This concept is relatively new to guideline developers and attempts to incorporate the understanding that acceptance of some treatments or tests is patient-dependent. The final consideration is cost and resource allocation. This is especially important to guideline panelists and policy makers on a population level. In general, assuming other factors discussed above are equal (confidence in the accuracy of the data, size of the probable effects, uniformity of preferences), the higher the cost of a particular intervention, the less likely it is to result in a strong recommendation [8].

The guideline panel decides on the final strength of recommendation after considering these four factors: QoE, estimated size of effect, preferences, and cost (see Fig. 1). How should the clinician behave in the face of such recommendations? First, no practice guideline recommendation should trump common sense or strongly held (patient's) personal beliefs. In this sense, every decision based on guidelines has to be re-interpreted in

the context of a specific clinical situation. On the other hand, clinicians should feel reassured that in cases of strong recommendations, a reasonable panel of expert clinicians (stakeholders) considered all (or most) pros and cons and concluded that a particular intervention should be used under most circumstances and for most patients. Knowing the reasons that lead to a weak recommendation can also be helpful. Are we issuing a weak recommendation because the estimated effects are close in the balance between potential benefits and downsides to patients? Or is it because the intervention is very costly? Or is it because we have low-quality data and we simply are not confident enough to be secure in what we recommend.

GRADE helps authors to be more specific about such issues and helps interested readers to follow the rationale for specific recommendations. As a minimum, clinical decisions not following strong recommendations should be accompanied by explicit reasoning. A strong recommendation should rarely (if ever) be based on low or very low QoE. On the other hand, clinical decisions based on weak recommendations should not be dogmatic. It also follows that no guideline is meant to be absolute, and pragmatically every clinician must interpret the recommendations and make clinical decisions that seem most reasonable for their individual patients.

**Conflicts of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A et al (2008) Going from evidence to recommendations. *BMJ* 336(7652):1049–1051
- Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R et al (2008) Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2008. *Intensive Care Med* 34(1):17–60
- Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM et al (2013) Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med* 39(2):165–228
- Balslem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J et al (2011) GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 64(4):401–406
- Angus DC (2012) Drotrecogin alfa (activated)—a sad final fizzle to a roller-coaster party. *Crit Care* 16(1):107
- Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P et al (2011) GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 64(12):1311–1316
- Alexander PE, Bero L, Montori VM, Brito JP, Stoltzfus R, Djulbegovic B et al (2014) World Health Organization recommendations are often strong based on low confidence in effect estimates. *J Clin Epidemiol*
- Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A et al (2008) Incorporating considerations of resources use into grading recommendations. *BMJ* 336(7654):1170–1173