

The Eastern Association of the Surgery of Trauma approach to practice management guideline development using Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) methodology

Andrew J. Kerwin, MD, Elliott R. Haut, MD, J. Bracken Burns, DO, John J. Como, MD,
Adil Haider, MD, Nicole Stassen, MD, Philipp Dahm, MD,
and Eastern Association for the Surgery of Trauma Practice Management
Guidelines Ad Hoc Committee

ABSTRACT: The Eastern Association for the Surgery of Trauma (EAST) is a leader in evidence-based medicine and the development of practice management guidelines (PMGs) in trauma and acute care surgery. The previous primer describing EAST's approach for assessing the quality of available evidence and making recommendations for developing PMGs was published in 2000. Since that time, many new systems have been developed in an attempt to overcome previous shortcomings and to devise a methodologically rigorous and transparent approach to the assessment of quality of evidence and development of guidelines. One of these is the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) system. The membership of EAST has determined that the GRADE methodology will be the system used in all future EAST PMGs. The purpose of this article was thus to describe the GRADE methodology. (*J Trauma Acute Care Surg.* 2012;73: S283–S287. Copyright © 2012 by Lippincott Williams & Wilkins)

KEY WORDS: Guideline; GRADE; evidence based medicine; trauma; surgery.

Medical decision making requires the knowledge of a vast amount of information. A busy physician may not be able to critically evaluate all the evidence regarding a specific clinical question. Because of this, the practice of developing clinical practice management guidelines (PMGs) developed. Countless professional societies and national agencies have developed their own systems for evidence synthesis and guideline development to answer specific questions. This evidence-based approach is done in an effort to reduce variation and to improve outcomes. The Eastern Association for the Surgery of Trauma (EAST) is well recognized for its evidenced-based approach to reviewing the literature and developing PMGs, specifically in the field of trauma and acute care surgery.

In recent years, there has been a rapid proliferation of guidelines. The National Guideline Clearinghouse website (www.guidelines.gov) has nearly 2,500 published guidelines available. Unfortunately, this ubiquitous proliferation of PMGs has not always resulted in greater understanding of a

specific disease state and, at times, has only added confusion because different authors and organizations have published conflicting recommendations for specific clinical questions. For example, the guidelines from EAST and the American College of Chest Physicians give very different guidance on the use of prophylactic inferior vena cava filters for the prevention of pulmonary embolism in very high-risk trauma patients.^{1,2} The growth of evidence-based medicine (EBM) and the development of PMGs have resulted in different organizations adopting different approaches to rating evidence and formulating guidelines. In 2002, there were more than 100 of these systems.³ In addition, guideline development has been fraught with numerous problems that have undermined quality and trustworthiness. These problems include the variable quality of studies, lack of transparency, limitations in the systematic review, lack of a multidisciplinary development group, conflicts of interest, and a failure to use a rigorous methodology for evidence review and guideline development. In light of these problems, the Institute of Medicine has recently published standards that authors should use to produce trustworthy guidelines (Table 1).⁴

To combat the wide variation in guideline production, a working group was formed in 2000 with the intent of developing a standardized method to rate a body of evidence and to make recommendations about specific clinical questions. This group developed the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) system. This methodology is a two-part process of, first, grading the evidence for a particular question and, second, making recommendations based on this evidence in combination with other factors important to clinical decision making. The group, now

Submitted: May 24, 2012, Revised: July 26, 2012, Accepted: August 19, 2012.

From the Department of Surgery (A.J.K., J.B.B.), University of Florida College of Medicine, Jacksonville; Department of Urology (P.D.), University of Florida College of Medicine, Gainesville, Florida; Department of Surgery (A.H., E.R.H.), The Johns Hopkins University School of Medicine, Baltimore, Maryland; Department of Surgery (J.J.C.), MetroHealth Medical Center, Cleveland, Ohio; Department of Surgery (N.S.), University of Rochester, Rochester, New York.

Address for reprints: Andrew J. Kerwin, MD, Department of Surgery, Division of Acute Care Surgery, 655 W 8th St, Jacksonville, FL 32209; email: andy.kerwin@jax.ufl.edu.

DOI: 10.1097/TA.0b013e31827013e9

J Trauma Acute Care Surg
Volume 73, Number 5, Supplement 4

S283

TABLE 1. Institute of Medicine Standards for Developing Trustworthy PMGs²

Standards for Developing Trustworthy PMGs
Establish a transparent process
Manage conflict of interest
Use a multidisciplinary guideline development group
Guideline systematic review interaction
Clearly describe summary of evidence and differences of opinion
Clearly articulate the recommendations and how they should be used
Allow for external review from other experts and stakeholders
Keep guidelines up to date

simply known as the GRADE Working Group, has continued to refine the standardized approach to guideline formulation that has resulted in the adoption of this system by more than 60 national and international organizations.

In addition to the guideline development, GRADE can be used for policy development, formulary decisions, purchasing decisions, and quality assessment of organizations. It may be used by patients as well as practitioners. A complete description of the goals of the GRADE Working Group can be found on its website (www.gradeworkinggroup.org). GRADE was initially published in a comprehensive six-part series⁵⁻¹⁰ in 2008 in the *British Medical Journal*. The 2011 to 2012 update is being completed and is being published as a 20-part series¹¹⁻²¹ in the *Journal of Clinical Epidemiology*. As of this writing, 10 parts of the series have been published.

Given the developments in evidence rating systems since EAST first published its original primer on EBM in 2000,²² the PMG Committee and the EAST Board of Directors have decided to update EAST's approach to rating evidence and making guideline recommendations. The objective of this article was to provide a summary of GRADE and how it should be used for the development of future EAST PMGs.

FRAMING THE QUESTION

The first step in using the GRADE methodology is to reformat an "informal question" into a specific question that can be answered. For example, an informal question might be "how do I treat a patient with a blunt splenic injury?" or "should I use angioembolization when managing blunt splenic injury?" The question must then be formatted into the

TABLE 2. Classification of Important Outcomes of Embolization of Blunt Splenic Injuries

Outcome Type	Value	Outcome
Critical outcomes	9	Mortality
	8	Splenic salvage
	7	Need for further transfusions
Important outcomes	6	Arterial access complications
	5	Contrast induced nephropathy
Limited importance outcomes	4	Contrast allergy
	3	Cost, resource use
	2	
	1	

"PICO" format. When formatted correctly, the question must clearly identify the patient population (P), the intervention (I), the comparator or comparators (C), and the outcome (O). A question in this format for our example might read, "In patients with blunt splenic trauma (P), should angioembolization (I) be performed compared to no angioembolization (C) to improve splenic salvage (O) for patients treated with non-operative management?" PICO questions drive the systematic review of the literature search and guideline development. Each informal question may lead to multiple PICO questions, and all possible outcomes (including benefits and harms) should be considered.

DEFINING OUTCOMES

Predefining which outcomes are important is relevant for both the literature search and guideline development process. To use GRADE, the outcome for each PICO is categorized as "critical" for decision making, "important but not critical," or of "limited importance" with respect to decision making. The outcomes can be classified with a numerical value based on a rating scale of 1 to 9 to describe their importance. A rating of 7 to 9 is given for critical outcomes, 4 to 6 for important outcomes, and 1 to 3 for limited importance outcomes. An example, classifying the outcomes in the previous example of angioembolization in the management of patients with blunt splenic injury, is shown in Table 2.

SYSTEMATIC REVIEW OF PUBLISHED LITERATURE

Although this primer on GRADE cannot cover the specific details of how to perform a systematic review in great detail, the importance of reliably identifying all relevant published (and potentially unpublished) data is imperative. Excellent resources for this step are available from both the Cochrane Collaboration (<http://www.cochrane.org>) as well as from the Institute of Medicine.²³ Meta-analysis should be used to combine data from different studies to give an overall point estimate and confidence interval for the effect size that the intervention has on the outcome of interest, if appropriate.

GRADING THE EVIDENCE

The next step is to grade the evidence for each outcome of each PICO question. In the context of guideline development, the quality of evidence relates to the overall body of

TABLE 3. GRADE Methodology Levels for Rating the Quality of Evidence¹⁴

Quality Level	Definition
High (A)	Very confident that the true effect lies close to estimate of effect.
Moderate (B)	Moderate effect; true effect is likely close to estimate of effect but may be substantially different.
Low (C)	Limited confidence; true effect may be substantially different from estimate of effect.
Very low (D)	Little confidence; true effect likely substantially different from estimate of effect.

TABLE 4. GRADE Definition of Strong and Weak Recommendation⁷

	Strong Recommendation	Weak Recommendation
For patients	Most patients would want the recommended course of action.	Most patients would want the recommended course of action, but many would not.
For clinicians	Most patients should receive the recommended course of action.	Different choices will exist for different patients, and clinicians should help patients decide.
For policy makers	Recommended course should be adopted as policy.	Considerable debate and stakeholder involvement needed to make policy.

evidence (usually multiple studies summarized in a systematic review), addressing a given focused PICO question. GRADE describes four levels of quality of the evidence. When rating the quality of evidence, the decision makers must consider the confidence in the estimate of each effect and whether these estimates are likely to be correct. The four levels of quality are “high” (A), “moderate” (B), “low” (C), and “very low” (D). The descriptions of these levels are shown in Table 3. This rating methodology provides a transparent assessment of the quality of evidence and can be applied to either randomized trials or observational studies. As each PICO question is considered, this rating method must be used for each individual outcome—a dramatic change from past methodologies in which individual published studies were rated on quality. Recognizing that the quality of evidence (or confidence in the estimate of effect) may differ by outcome and should therefore be graded separately is a unique GRADE contribution. When grading the quality of evidence, randomized controlled trials (RCTs) are initially considered high-quality evidence (but can be rated down), while observational studies begin as low-quality evidence (but can be rated up).

However, GRADE considers more than simply study design alone when rating the quality of evidence. There are

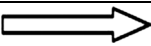
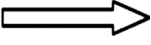
five reasons that the quality of evidence from RCTs should be rated down. These include study limitations (i.e., poor randomization), inconsistency, indirectness, imprecision, and publication bias (Table 4). In addition, for a body of well-designed observational studies, there are three reasons that the quality of evidence should be rated up. These include a large magnitude of effect, the existence of a dose-response gradient, and consideration of the effect of all plausible residual confounders (Table 4). For example, evidence from poorly designed and executed RCTs would be graded down to low- or very low-quality evidence. Contrarily, evidence from multiple well-designed observational trials with no serious flaws might be graded up to moderate or high-quality evidence. In practice, it is more common to rate down the quality of evidence than to rate up.

MAKING RECOMMENDATIONS

Once all of the evidence has been graded and summarized using evidence profiles, the second phase of the process, making recommendations, begins. The evidence profiles are used by the authors to provide a detailed judgment about the quality of evidence for each outcome being considered. These evidence profiles are used to build the summary of findings

TABLE 5. GRADE Approach to Rating Quality of Evidence¹²

A Summary of GRADE’s Approach to Rating Quality of Evidence

Study Design	Initial Quality of a Body of Evidence	Lower If	Higher If	Quality of a Body of Evidence
Randomized trials	High 	Risk of bias -1 Serious -2 Very serious	Large effect +1 Large +2 Very large	High (four pluses: ⊕ ⊕ ⊕ ⊕)
		Inconsistency -1 Serious -2 Very serious	Dose response +1 Evidence of a gradient	Moderate (three pluses: ⊕ ⊕ ⊕ ○)
Observational studies	Low 	Indirectness -1 Serious -2 Very serious	All plausible residual confounding +1 Would reduce a demonstrated effect +1 Would suggest a spurious effect if no effect was observed	Low (two pluses: ⊕ ⊕ ○ ○)
		Imprecision -1 Serious -2 Very serious		Very low (one plus: ⊕ ○ ○ ○)
		Publication bias -1 Likely -2 Very likely		

Reprinted with permission from Elsevier.

tables. The summary of findings tables are meant to provide concise summaries of the key findings that the readers of the guideline can use when making decisions about patient care.

The GRADE methodology differs from other systems in that it makes guideline recommendations relatively simple and transparent. Only two possible recommendations can be made as follows: (1) strong or (2) weak/conditional. A strong recommendation means that most patients should receive the recommended course of medical care. A weak/conditional recommendation means that, although most of the patients would select the recommended action, there are different choices that will be appropriate for different patients depending on their particular situation. The definitions of strong and weak/conditional recommendations as they apply to patients, clinicians, and policy makers are shown in Table 5.⁹ All appropriate evidence must be considered when advising patients about their medical care. GRADE has tried to avoid the confusion of letters and numbers in its recommendations. However, many guideline authors will refer to strong recommendations as "1" and weak/conditional recommendations as "2," and often, some add the quality of evidence (A, B, C, D) resulting in recommendations such as 1A or 2C. A concise schematic summarizing the entire GRADE methodology is shown in Figure 1.

When making final PMG recommendations, it is important to understand the many factors the guideline committee

must consider. Making a strong or weak recommendation is not automatic and simply based on whether an RCT was performed. Rather, the members of the guideline committee must use judgment when formulating recommendations. In addition to weighing the quality of the evidence, they must always consider the ratio of benefits to harms and the patient's values and preferences. Some guideline panels consider the cost of the care involved as well, although this is not required. This phase of the process should be abundantly transparent. That the GRADE methodology strives for this transparency is one of the main reasons it has been widely adopted.

CONCLUSION

The GRADE methodology is rapidly becoming the most widely used, unified methodological framework for rating the quality of evidence and strengths of recommendations. To maintain its leadership role in the development of trauma and acute care surgery PMGs, it is essential that EAST adopt the GRADE methodology for future PMGs. The GRADE Working Group continues to refine this methodology in hopes that all guideline developers will adopt it. The GRADE method offers numerous benefits over EAST's previous approach to guideline development. It provides a clear separation between rating the quality of evidence and making recommendations. There are transparent, explicit, and comprehensive criteria for

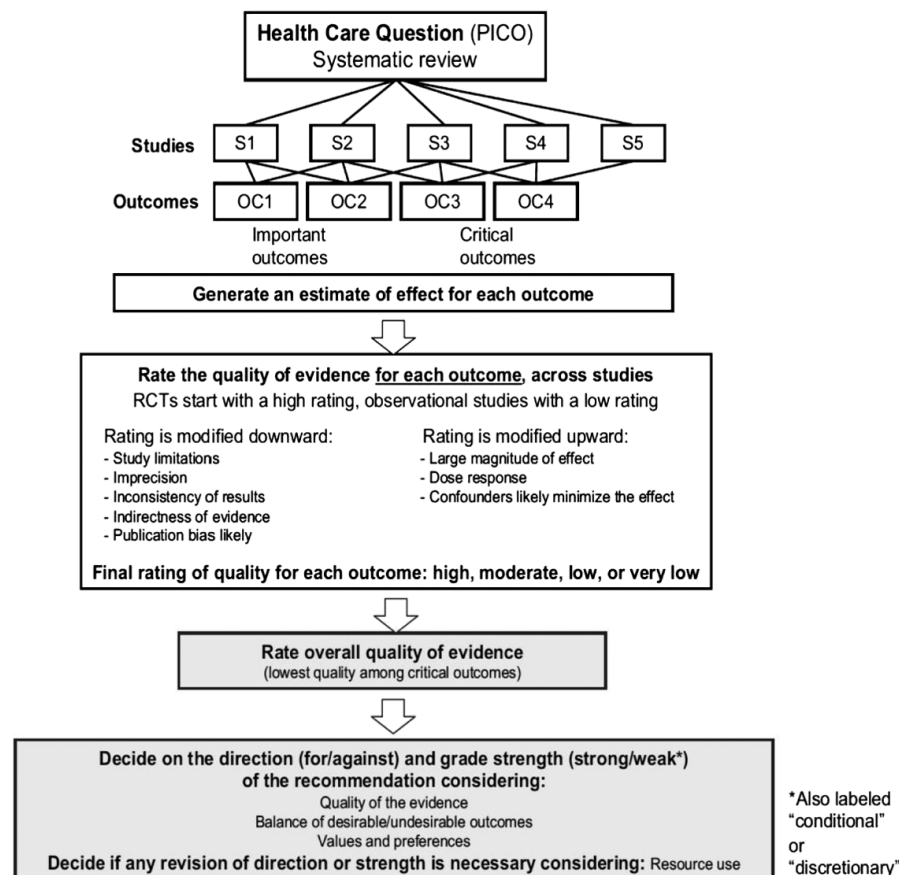


Figure 1. Summary of GRADE methodology.¹² Reprinted with permission from Elsevier.

downgrading or upgrading the quality of evidence. In addition, there are clear definitions of strong and weak/conditional recommendations. Finally, it takes into consideration the importance of patient outcomes and considers the balance between benefit and harm when formulating guidelines.

Because of these clear benefits, the GRADE methodology has been adopted by more than 60 well-recognized national and international organizations such as the Centers for Disease Control, the American College of Chest Physicians, the Infectious Disease Society of America, Up to Date, the Society of Critical Care Medicine, the Surviving Sepsis Campaign, the American Thoracic Society, the World Health Organization, the Cochrane Collaboration, and the Agency for Healthcare Research and Quality. Because it is increasingly used in EBM by numerous other societies and because of benefits described throughout this primer, the leaders of EAST and its PMG committee have decided that GRADE will be used for all future EAST PMGs. Adoption of GRADE will benefit EAST, its members, clinicians, and patients worldwide.

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. Guyatt GH, Akl EA, Crowther M, Gutterman DD, Schunemann HJ. American College of Chest Physicians Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(Suppl 2):7S–47S.
2. Rogers FB, Cipolle MD, Velmahos G, Rozycki G, Luchette FA. Practice management guidelines for the prevention of venous thromboembolism in trauma patients: the EAST practice management guidelines work group. *J Trauma*. 2006;53:142–164.
3. Straus S, Shepperd S. Challenges in guideline methodology. *J Clin Epidemiol*. 2011;64:347–348.
4. Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E (eds). Committee on Standards for Developing Trustworthy Clinical Practice Guidelines; Institute of Medicine. *Clinical Practice Guidelines We Can Trust*. Washington, DC: The National Academies Press; 2011.
5. Guyatt GH, Oxman AD, Vist G, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924–926.
6. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Schunemann HJ. GRADE: what is “quality of evidence” and why is it important to clinicians? *BMJ*. 2008;336:995–998.
7. Schunemann HJ, Oxman AD, Brozek J, et al. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336:1106–1110.
8. Guyatt GH, Oxman AD, Kunz R, et al. GRADE: incorporating considerations of resource use into grading recommendations. *BMJ*. 2008;336:1170–1173.
9. Guyatt GH, Oxman AD, Kunz R, et al. GRADE: going from evidence to recommendations. *BMJ*. 2008;336:1049–1051.
10. Jaeschke R, Guyatt GH, Dellinger P, et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ*. 2008;337:744.
11. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *J Clin Epidemiol*. 2011;64:380–382.
12. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64:383–394.
13. Guyatt GH, D. OA, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important questions. *J Clin Epidemiol*. 2011;64:395–400.
14. Balshe H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64:401–406.
15. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol*. 2011;64:407–415.
16. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol*. 2011;64:1277–1282.
17. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guideline: 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol*. 2011;64:1283–1293.
18. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol*. 2011;64:1294–1302.
19. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol*. 2011;64:1303–1310.
20. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64:1311–1316.
21. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol*. Epub. 27 April 2012..
22. Eastern Association for the Surgery of Trauma Ad Hoc Committee on Practice Management Guideline Development. *Utilizing evidence based medicine outcome measures to develop practice management guidelines: a primer*. EAST; 2000. Available at: www.east.org/content/documents/east_pmg_primer.pdf.
23. Eden J, Levit L, Berg A, Morton S. Committee on Standards for Systematic Reviews of Comparative Effectiveness; Institute of Medicine. *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, D.C.: The National Academies Press; 2012.