

**The AHRQ National Guideline Clearinghouse (NGC, [guideline.gov](http://guideline.gov)) Web site will not be available after July 16, 2018 because federal funding through AHRQ will no longer be available to support the NGC as of that date. For additional information, read our [full announcement](#).**

EXPERT COMMENTARY JANUARY 04, 2010

## Using GRADE for International Guidelines on Kidney Disease

By: *Ethan M. Balk, MD, MPH and Katrin Uhlig, MD, MS*

To enhance transparency and accountability in guideline development and to facilitate guideline adaptation, the quality of supporting evidence and the strength of recommendations are commonly graded. According to the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach, the grade of a

<https://www.guideline.gov/expert/expert-commentary/164> following the recommendation will do more good than harm (1,2).  
7 captures 30 Sep 2016 12 Jul 2018

At the Tufts Center for Kidney Disease Guideline Development and Implementation at Tufts University, we have used the GRADE approach to develop guidelines for Kidney Diseases: Improving Global Outcomes (KDIGO; [www.kdigo.org](http://www.kdigo.org)), an international guideline-development initiative. Here we describe our experience with and challenges implementing the GRADE approach (3, 4). While our application is in the area of kidney disease, the issues we discuss are not domain specific.

The Tufts Center is an academically-based, independent group of methodologists and nephrologists under contract with the National Kidney Foundation. Each guideline (16 since 2000, 4 in progress) takes 12 to 24 months to complete, includes up to dozens of specific topics with systematic reviews of up to hundreds of studies, and is developed by a dedicated staff of physicians and other methodologists, a workgroup of 10 to 20 domain experts (primarily nephrologists), and dedicated supporting staff at the National Kidney Foundation. Initially, guideline development was under the auspices of the US-based Kidney Diseases Outcomes Quality Initiative (NKF KDOQI™; <http://www.kidney.org/professionals/kdoqi/>). When KDIGO was founded in 2003, we reviewed existing grading systems to standardize our methods. We adopted a modification of the GRADE approach.

(corresponding to strong, moderate, and weak). However, based on feedback from members of the GRADE working group and difficulties with expressing and translating the nuances of the 3 grading levels in English and other languages, the approach was again revised in December 2008 to more closely adhere to GRADE, with 2 levels of recommendations: Strong — "Do (or don't) do it"; and Weak — "Probably do (or don't do) it" (1). We have modified these to: Level 1, in which the recommendation includes "We recommend"; and Level 2, in which the recommendation includes "We suggest." In addition, as is described further below, the KDIGO approach allows "ungraded statements."

A recurring concern among KDIGO guideline workgroups is what to do when the evidence reviewed for a particular question provides an answer that is inconclusive. According to GRADE, where there are unclear trade-offs (between benefits and harms) or lack of agreement, "It may not be appropriate to make a recommendation" (1). However, a majority of KDIGO workgroup members and its Board of Directors have a different philosophy: the primary purpose of guidelines is to offer guidance to clinicians, patients, and other stakeholders on best or most appropriate care, even in the face of limited evidence. KDIGO encourages its workgroups to make recommendations based on the overwhelming consensus of its experts' judgments. It is therefore common for KDIGO guidelines to have Level 2 recommendations based on low or very low quality evidence, which may not have been made if a more rigid philosophy were used that requires higher quality or more definitive evidence to make a recommendation (like the U.S. Preventive Services Task Force [USPSTF] recommendations) (5).

Another common issue regards what to do when a topic does not lend itself to a formal evidence review because the question is not specific enough or the recommendation is a reminder of the obvious. Common examples include recommendations about frequency of testing, referral to specialists, and routine medical care. The KDIGO leadership has agreed that in order for the guidelines to be of greatest value to users, "clinical pearls" or "common sense" recommendations may be included as ungraded statements (6). While we strive to minimize the use of ungraded statements, they remain common; for example, 20% (49 of 247) of the recommendations in the transplantation guidelines are ungraded (7).

Beyond these modifications to GRADE, the KDIGO experience with applying GRADE has highlighted several challenges in implementation, including the effort required to complete the many grading steps and the remaining subjectivity in grading and determining net benefit. To strictly apply GRADE, for each research question related to each recommendation one must compile a list of outcomes of interest, rank the importance of these outcomes, systematically review the evidence for each outcome, and assess the quality of the relevant studies, the consistency across the studies, the "directness" of the evidence and other

regarding the consistency in the "direction of effect [and] the size of the differences in effect," as well as determining the "estimated size of the effect [and] the confidence limits around those estimates" (1). However, with the expansive scope of our guidelines, performing meta-analysis for each outcome is not feasible. Our transplant guideline has almost 250 specific recommendations. We restricted the number of topics with full evidence profiles, but the guideline still includes 21 evidence profiles, which combined would contain over 200 potential meta-analyses (one per intervention-outcome pair). For any systematic review, even if unlimited time and resources were available, meta-analysis is often inappropriate due to limitations of the evidence or clinical heterogeneity of the studies. With few exceptions, we do not perform *de novo* meta-analyses for the workgroups, but without the meta-analyses, judging consistency and estimating overall effect sizes and variability becomes a more subjective process.

To assess the quality of a body of evidence, it is necessary to grade the quality of the individual studies for each outcome. The GRADE Working Group provides little guidance on how to grade the quality of individual studies, which is reasonable given the lack of consensus on the best approach to take. We use a 3-level ranking of study quality (good, fair, poor) based on a range of standard and topic-specific features that may result in biased or inaccurate results. This approach has also been adopted for Agency for Healthcare Research and Quality's (AHRQ's) Comparative Effectiveness Reviews (8). Whether using this or score-based systems to grade study quality, the process is intrinsically subjective. Further subjectivity and arbitrary decision-making occurs when aggregating the quality grades from individual studies into a summary grade to express whether there are serious, very serious, or no limitations to the methodological quality of a body of evidence. For example, it is problematic to summarize the overall methodological quality of one large, good quality study, one small, fair quality study, and three poor quality studies. There is no simple system to derive a grade for methodological quality of a body of evidence; thus, we aim to achieve consensus among the methodologists and guideline workgroup members.

In the GRADE system, outcomes are ranked by clinical importance to weigh the individual effects on each outcome when deriving the overall net benefit. This ranking also impacts on the overall quality of the evidence, since the quality grades of important outcomes carry more weight. Our workgroups have had few problems with ranking the clinical outcomes of interest. However, difficulties have arisen when applying these rankings to decision making. This has been the case particularly when the "crucial" outcome is of lesser clinical importance than critical outcomes. While death will always be a critically important outcome, it may not be crucial when deciding whether to use an intervention. For example, when evaluating the use of human growth hormone in children with kidney transplants, the poor

workgroup to come to consensus about the relative meaning of different outcomes, the corresponding effect sizes, and the net effect. This is compounded in the global KDIGO guidelines since there is no particular point of reference for calibrating judgments about values, preferences, and costs.

In summary, evidence appraisal requires judgments and thus is susceptible to bias at several steps of synthesis and appraisal. Despite these issues, the Tufts Center and KDIGO have found that the GRADE system provides a structured and transparent approach to determine the quality of evidence, summarize effects, and grade the strength of recommendations. A key advantage to GRADE is that it prompts guideline developers to consider all relevant outcomes and to follow a consistent and comprehensive process in evaluating the quality of bodies of evidence. Furthermore, it has prompted our workgroups to be more complete, explicit, and transparent in discussing how they weigh the evidence and incorporate values and preferences to arrive at the recommendations.

While the GRADE system does not remove subjectivity or potential bias from the guideline development process, nor the need for ad hoc adaptations to fit different topics and different types of evidence, our experience using GRADE for guidelines on kidney disease has resulted in greater methodological rigor on the part of the Tufts Center methodologists, a better understanding of the guideline development steps by the workgroup domain experts, and a simpler nomenclature and convention for rating recommendations.

## **Authors**

**Ethan M. Balk, MD MPH**  
Boston, MA

**Katrin Uhlig, MD, MS**  
Boston, MA

## **Disclaimer**

The views and opinions expressed are those of the author and do not necessarily state or reflect those of the National Guideline Clearinghouse™ (NGC), the Agency for Healthcare Research and Quality (AHRQ), or its contractor, ECRI Institute.

## **Potential Conflicts of Interest**

Dr. Balk notes personal financial interests in QuantRx and Echo Therapeutics. He has worked on the following clinical practice guideline or quality measure development projects: Kidney Disease: Improving Global Outcomes guidelines (various on chronic kidney disease); Kidney

a member of the Core Editorial Board for the National Guideline Clearinghouse and the National Quality Measures Clearinghouse.

Dr. Uhlig discloses that she is paid by the National Kidney Foundation for her work in the development of Kidney Disease Improving Global Outcomes (KDIGO) clinical practice guidelines. She also serves as a paid consultant to the Society for Gynecological Surgeons Systematic Review Group in the development of systematic reviews and clinical practice guidelines.

## References

1. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490.
2. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ et al. What is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008;336:995-98.
3. Uhlig K, Balk EM, Lau J, Levey AS. Clinical practice guidelines in nephrology—for worse or for better. *Nephrol Dialysis Transplant*. 2006;21:1145-53.
4. Uhlig K, Macleod A, Craig J, Lau J, Levey AS, Levin A et al. Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int*. 2006;70:2058-65.
5. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D, U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Ann Intern Med*. 2007;47(12):871-5.
6. Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Work Group. KDIGO clinical practice guideline for the diagnosis, evaluation, prevention, and treatment of chronic kidney disease-mineral and bone disorder (CKD-MBD). *Kidney Int*. 2009;76:S1-S130.
7. Kidney Disease: Improving Global Outcomes (KDIGO) Transplant Work Group. KDIGO clinical practice guideline for the care of kidney transplant recipients. *Am J Transplant* 2009;9(Suppl 3):S1-S157.
8. Agency for Healthcare Research and Quality. Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews, Version 1.0 [Draft posted Oct. 2007] Available at: [http://effectivehealthcare.ahrq.gov/repFiles/2007\\_10DraftMethodsGuide.pdf](http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf) (PDF Help). 2008. Rockville, MD.

