Is GRADE the Right Choice for Clinical Practice Guidelines Developed by the American Academy of Otolaryngology–Head and Neck Surgery Foundation?

Andrés M. Bur, MD¹ and Richard M. Rosenfeld, MD, MPH, MBA²

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Abstract
Clinical practice guidelines (CPGs), developed to inform clinicians, patients, and policy makers about what constitutes optimal clinical care, are one way of increasing implementation of evidence into clinical practice. Many factors must be considered by multidisciplinary guideline panels, including strength of available evidence, limitations of current knowledge, risks/benefits of interventions, patient values, and limited resources. Grading of Recommendations Assessment, Development and Evaluation (GRADE) is a framework for summarizing evidence that has been endorsed by many national and international organizations for developing CPGs. But is GRADE the right choice for CPGs developed by the American Academy of Otolaryngology–Head and Neck Surgery Foundation (AAO-HNSF)? In this commentary, we will introduce GRADE, discuss its strengths and limitations, and address the question of what potential benefits GRADE might offer beyond existing methodology used by the AAO-HNSF in developing CPGs.

Keywords
clinical practice guidelines, evidence, otolaryngology–head and neck surgery

Received March 7, 2019; accepted April 5, 2019.

Grading of Recommendations Assessment, Development and Evaluation (GRADE) is a framework for developing and presenting summaries of evidence that provides a standardized approach for developing clinical practice guidelines (CPGs).¹ In recent years, GRADE has been endorsed by many well-known organizations, including UpToDate, the American College of Physicians, and the World Health Organization. However, the American Academy of Otolaryngology–Head and Neck Surgery Foundation (AAO-HNSF) has not adopted GRADE and instead uses a transparent methodology that is consistent with Institute on Medicine (IOM) and Guideline International Network (G-I-N) standards.² The purpose of this commentary is to introduce GRADE and to discuss what potential benefits GRADE might offer beyond existing methodology used by the AAO-HNSF in developing CPGs.

How Is GRADE Implemented?
The GRADE approach consists of 2 basic determinations. Guideline developers first conduct a systematic literature review to assess the quality of evidence available to address each clinical question. GRADE has 4 levels of evidence (Table 1), and depending on the strengths and limitations of the body of evidence (eg, bias, magnitude of effect), guideline developers may increase or decrease the level of evidence accordingly (Table 2). Evidence is then translated into a strong or weak recommendation using a framework that incorporates not only the quality of evidence but also risks/benefits of the intervention, patient values, and resource considerations.³ Considerable training and experience are required to implement the specific evaluation criteria described in great detail by the GRADE authors.

Strengths of GRADE
There are several strengths of the GRADE framework. First, GRADE provides transparency about how evidence is translated into recommendations and the specific factors underlying each recommendation. By providing a framework that is both explicit and comprehensive for down- or upgrading level of evidence, GRADE attempts to ensure reproducible results. Individuals with extensive experience using GRADE have good intrarater reliability (IRR) when

1. University of Kansas, Kansas City, Kansas, USA
2. SUNY Downstate Medical Center, Brooklyn, New York, USA

The work presented herein was conducted at the University of Kansas, Kansas City, Kansas, USA.

Corresponding Author:
Andrés M. Bur, MD, Department of Otolaryngology–Head and Neck Surgery, University of Kansas School of Medicine, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA.
Email: abur@kumc.edu
assessing the quality of a body of evidence, although IRR diminishes when the evidence is more complex.4

Second, GRADE provides a clear interpretation of strong vs weak recommendations for clinicians, patients, and policy makers. When the desirable effects of an intervention clearly outweigh the undesirable effects, or clearly do not, GRADE advocates a “strong” recommendation. On the other hand, when the trade-off is less certain, either because of low-quality evidence or because of a small effect size, a weak recommendation is made.1

Limitations of GRADE

GRADE methods are not always easily adaptable to the clinical questions posed by guideline developers. GRADE is simplest to apply when high-quality evidence allows quantitative estimates of the effect sizes for each outcome. Unfortunately, this is rarely the case in otolaryngology, in which randomized control trials (RCTs) may not be available and guideline developers may depend on observational studies, for which the GRADE approach may be suboptimal.5 If data are qualitative or if there is significant heterogeneity between studies precluding statistical pooling of estimates, user judgment is needed to assign levels of certainty and to develop recommendations.6 A considerable amount of expert consensus is therefore inevitable using GRADE in this circumstance.

With the detailed analyses that are necessary to adhere to GRADE methodology, guideline development using GRADE is likely to be costlier and more time intensive than other approaches. In addition, adherence to GRADE will result in a guideline product that is more complex to read. Finally, GRADE assigns observational studies a default rating of “low quality,” and users can then modify these evidence ratings according to GRADE criteria. Consequently, recommendations that are not based on RCTs are less likely to be designated as having moderate or high evidence rating.

Is GRADE the Right Choice for AAO-HNSF Guidelines?

Although GRADE draws a bridge between evidence-based health practices and clinical practices, the criteria used to establish the quality of evidence and strength of recommendation are better suited to clinical trials than observational research. In addition, GRADE is not standardized in how a clinician or a patient would evaluate a body of literature, since the judgment would inevitably be based on an individual’s beliefs and opinions.7

There is no supportive evidence that use of GRADE yields better clinical recommendations than other approaches.8 With its added complexity and multistep framework, GRADE is likely to increase the time, cost, and labor involved in developing AAO-HNSF CPGs with uncertain benefit. Although GRADE has been endorsed by numerous high-profile organizations, is adherence to GRADE’s more complex methodology the best use of AAO-HNSF resources?

Conclusions

Presently, AAO-HNSF CPGs are produced using a transparent methodology that adheres to G-I-N and IOM standards. Like GRADE, recommendation strength in AAO-HNSF guidelines depends on the aggregate quality of the supporting evidence and anticipated benefit vs harm if the recommendation is followed. Similarly, the strength of recommendation may be adjusted based on factors that affect our confidence

| Table 1. Quality of Evidence Ratings in GRADE (Grading of Recommendations, Assessment, Development and Evaluations).a |
|---|---|
| Quality of Evidence | Description |
| Very low | The true effect is likely markedly different from the estimated effect. |
| Low | The true effect might be markedly different from the estimated effect. |
| Moderate | The true effect is probably close to the estimated effect. |
| High | The true effect is highly likely to be close to the estimated effect. |


| Table 2. Reasons That Quality of Evidence May Be Downgraded or Upgraded Using GRADE Framework.a |
|---|---|
| Quality of Evidence Can Be Rated Down for | Quality of Evidence Can Be Rated Up for |
| • Risk of bias | • Large magnitude of effect |
| • Imprecision | • Dose-response gradient |
| • Inconsistency | • Residual confounding would increase magnitude of effect |
| • Inconsistency | |
| • Publication bias | |

in the evidence, such as quantity, consistency, precision, and generalizability. Whereas GRADE has only 2 levels of recommendation strength (strong vs weak), the AAO-HNSF process has 3 levels (strong recommendation, recommendation, and option), which may more realistically reflect the choices faced by clinicians.2

Until there is further evidence that GRADE improves outcomes over the existing AAO-HNSF process, we do not believe that the added complexity, cost, or time of implementing GRADE is warranted. We believe that the quality-driven, multidisciplinary approach used by AAO-HNSF, which is well suited to situations where evidence may be limited to observational studies, is most likely to result in efficient, pragmatic guidelines that reflect the daily concerns and issues faced by clinicians and patients.9

Author Contributions
Andrés M. Bur, design, drafting, approval, accountability, Richard M. Rosenfeld, design, drafting, approval, accountability.

Disclosures
Competing interests: Richard M. Rosenfeld, member of the AAO-HNSF Guidelines Task Force, first author of the AAO-HNSF clinical practice guideline manual.
Sponsorships: None.
Funding source: None.

References