

Grading the Strength of a Body of Evidence When Comparing Medical Interventions—Agency for Healthcare Research and Quality and the Effective Health-Care Program

Owens, D.K., **Lohr, K.N.**, Atkins, D., Treadwell, J.R., Reston, J.T., Bass, E.B., et al. (2010) AHRQ Series Paper 5: Grading the strength of a body of evidence when comparing medical interventions—Agency for Healthcare Research and Quality and the Effective Health-Care Program. *Journal of Clinical Epidemiology* 63 (5):513-523.



Kathleen Lohr

In the fields of evidence-based practice, comparative effectiveness, quality of care, practice guidelines, and related areas, systematic reviews are a crucial tool for assembling and synthesizing evidence published in peer-reviewed journals and other authoritative sites. For more than 15 years, the U.S. Agency for Healthcare Research and Quality (AHRQ) has supported production of systematic reviews through its program of Evidence-based Practice Centers (EPCs). The RTI-University of North Carolina EPC is one such center. A key part of conducting such rigorous reviews is grading the strength of the overall bodies of evidence amassed to answer key clinical or policy questions. The aim is to provide readers—clinicians, policymakers, patients or their representatives, third-party payers, and others—with a straightforward way to understand the confidence that systematic review authors have (or do not have) in their findings, so that users have a sound basis for making clinical, administrative, or policy decisions.

In this paper, a cross-EPC workgroup set out authoritative guidance for EPCs on how best to grade strength of evidence. It drew from international approaches from the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group, with adaptations for the U.S. setting and the AHRQ program. Using ratings on several domains relating to the quality of published studies and the consistency and other characteristics of findings from those studies, reviewers grade the strength of evidence in one of four categories: high, moderate, low, or insufficient. These grades reflect how strongly reviewers believe the findings (and the likelihood that they will change); high is very robust evidence, and insufficient typically means either no or inconsistent and uninterpretable results.

Guidance updating this paper, which will be more detailed and more statistically advanced, will appear by 2014. The critical importance to decision makers of clear, transparent strength of evidence grades remains, nonetheless, central to the validity and usefulness of systematic reviews of all clinical and policy issues.

Link: <http://dx.doi.org/10.1016/j.jclinepi.2009.03.009>