

See corresponding article on page 1578.

## Grading nutrition evidence: where to go from here?

Deirdre K Tobias,<sup>1,2</sup> Clemens Wittenbecher,<sup>2,3</sup> and Frank B Hu<sup>2,4</sup>

<sup>1</sup>Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; <sup>2</sup>Department of Nutrition, Harvard TH Chan School of Public Health, Boston, MA, USA; <sup>3</sup>Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany; and <sup>4</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Systematic reviews and meta-analyses are useful tools in summarizing a large body of evidence and informing policy and guidelines. Although findings from systematic reviews and meta-analyses are regarded by some to be the most authoritative form of available evidence, there is great potential for the misuse of the systematic review and meta-analysis methodology (1). The prevalence and magnitude of such misuses in systematic reviews are particularly concerning in nutrition research owing to the proliferation of meta-analyses in the nutrition literature and increased reliance on using systematic reviews to develop food policies and dietary guidelines (2). In their study published in *The American Journal of Clinical Nutrition*, Zeraatkar et al. (3) at McMaster University conducted a comprehensive review and analysis of 150 randomly sampled systematic reviews of nutritional epidemiology studies published between January 2018 and August 2019. Their detailed quality assessment highlighted several common methodological problems in published meta-analyses of nutritional epidemiologic studies.

One of the methodological issues raised by the study was that only a small proportion of the 150 systematic reviews (10%) implemented a formal evaluation of the certainty of the evidence, and “most did not discuss risk of bias.” A careful assessment of certainty of evidence and risk of bias in systematic reviews is critical to evaluate the quality of overall evidence on specific nutrition topics and these are, therefore, important considerations in developing clinical and public health guidelines. To this end, Zeraatkar et al. recommended the GRADE (Grading of Recommendations Assessment, Development and Evaluation) system for rating the certainty of a body of evidence (4). The GRADE system, which was originally developed to evaluate the quality of clinical intervention evidence, relies on a hierarchy of study designs: meta-evidence from randomized trials is automatically considered to be high certainty, whereas meta-evidence from nonrandomized studies is regarded as low certainty owing to potential confounding and selection bias (5). Although the GRADE system is relatively straightforward to implement in assessing the strength of the evidence from randomized double-blinded and placebo-controlled trials, the interrater reliability of GRADE ratings for complex interventions that are not amenable to double-blinding or placebo-controls is only modest (6). The reliability of the GRADE system to evaluate the strength of

observational evidence is likely to be more uncertain given the heterogeneity of observational study designs and different degrees of exposure measurement errors and adjustment for confounding factors. Although we agree with the authors that it is important to “[maintain] consistent standards for evaluating the certainty of evidence across health fields,” the complexity of environmental and behavioral exposures such as diet warrants additional considerations when grading the evidence, and one should not blindly apply the existing GRADE criteria to the development of public health guidelines regarding diet, lifestyle, and environmental factors.

Recently, a series of systematic reviews rated the meta-evidence for the relation between intake of red and processed meats and risk of major chronic disease incidence and mortality as “very low and/or low certainty” using GRADE, and consequently, the authors recommended individuals to continue their red and processed meat consumption habits. These recommendations have caused a great deal of public confusion (7) and raised doubt about the appropriateness of using the GRADE system in developing nutrition recommendations (8). A separate research group has proposed a modified system for rating the certainty of meta-evidence from nutritional studies (NutriGrade). Although NutriGrade shares several scoring components with the GRADE criteria, it does not automatically consider the evidence from observational studies as low certainty. Instead, the assessment of evidence certainty is based on an overall quantitative score of 9 components. Applying NutriGrade to the same body of meta-evidence on red meat intake and chronic disease risk resulted in ratings of “moderate quality” and “high quality” on the associations of red and processed meat intakes with mortality (9) and type 2 diabetes (10), respectively.

The red meat example underscores the challenge in assessing the quality of meta-evidence for diet and lifestyle factors (11). Using the current GRADE criteria, meta-evidence from observational studies (without clear distinction between their different types, e.g., cohort studies compared with case-control

This work is supported by NIH grants HL60712 and DK 46200 (to FBH). Address correspondence to FBH (e-mail: [fhu@hsph.harvard.edu](mailto:fhu@hsph.harvard.edu)). First published online May 8, 2021; doi: <https://doi.org/10.1093/ajcn/nqab124>.

studies) is all rated low certainty owing to lack of randomization, which can be further downgraded to very low level of certainty for additional reasons (e.g., because of unknown confounders and significant heterogeneity) (5, 12). To overcome the problem of excessive downgrading of observational evidence, ROBINS-I (Risk of bias in non-randomized studies of interventions) was developed to assess the risk of bias across multiple domains, rather than simply the lack of randomization, in grading the certainty of observational evidence (13). Although the integration of ROBINS-I with GRADE may provide a more balanced approach to grade observational evidence, the validity of this complex system has not been carefully assessed in the context of nutritional, lifestyle, and environmental exposures.

Methodological problems in assessing the risk of bias in nutrition research are not limited to observational studies. **Currently available risk of bias instruments including GRADE often fail to capture common limitations of dietary intervention trials including poor dietary adherence and high dropout rates.** In addition, because most dietary interventions are focused on food substitutions while maintaining the same total energy intake, the effects of interventions are likely to vary with the types of replacement foods. **For example, the effects of red meat consumption on cardiovascular disease risk may depend on whether red meat is replaced by plant-based protein foods such as legumes and nuts or starchy foods such as bread and potatoes.** The current GRADE system, heavily relying on the clinical intervention paradigm, does not adequately consider these methodological issues when assessing the strength of evidence from dietary intervention studies.

Another issue in applications of GRADE to nutrition research is that dietary interventions are seldom designed to test the effect of the exposure on a hard clinical endpoint, but often rely on intermediate outcomes (e.g., fasting glucose, blood lipids, and blood pressure) as surrogate endpoints. Whereas the Bradford Hill criteria consider experimental evidence from small and mechanistic studies on intermediate endpoints when evaluating causality (14), the GRADE system does not consider the issue of biological plausibility as a domain of certainty. Another important consideration in the Bradford Hill criteria is a dose–response relation from observational studies, which can estimate the effect of dietary exposures across the full distribution of dietary intake of a population to evaluate dose–response and nonlinear relations, whereas dietary intervention studies typically test a fixed dose of dietary exposure.

For these reasons, there is a critical need for the GRADE system and similar metrics to be modified to accommodate unique characteristics of nutritional epidemiology and intervention studies. So where to go from here? The first step is to improve the overall rigor and quality of systematic reviews and meta-analyses by adhering to a set of guidelines as discussed by Zeraatkar et al. including a priori registration of systematic review protocols, **stricter adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement,** and careful assessment of heterogeneity and biases. We should be cognizant that although systematic reviews and meta-analyses are valuable, a meta-analysis is only as good as the studies it pools together. Too often, systematic reviews of nutrition topics are conducted by authors who do not fully appreciate the complexity of the subject or lack content knowledge, which can lead to flawed methodology and misleading conclusions. Moreover, it

is important to recognize the inherent limitations of systematic reviews and consider the conclusions in the broader context of the field (1).

Second, we need to be cautious in applying existing tools to grade the quality or certainty of nutritional evidence. Although the GRADE system was initially developed to assess the strength of evidence from clinical interventions, it has been increasingly used to evaluate the evidence for lifestyle and environmental exposures. However, the infeasibility of conducting large long-term randomized trials for most dietary and lifestyle factors renders the current GRADE criteria inadequate for these exposures. Recent developments such as integrating ROBINS-I for bias assessment with the GRADE system may improve its applicability for nutrition research, but further validation of this combined approach is clearly warranted. Alternative approaches such as those of the World Cancer Research Fund (15), the Hierarchies of Evidence Applied to Lifestyle Medicine (HEALM) (16), and NutriGrade (12) should also be carefully evaluated. Interestingly, a recent systematic review applied GRADE and NutriGrade to the same body of meta-evidence for low-carbohydrate dietary intervention trials and type 2 diabetes remission and found that the 2 metrics concurred for only 53% of the study outcomes with respect to the certainty level of the evidence (17). Considering the pros and cons of different grading systems and the complexity of nutritional studies, there is an urgent need for the nutrition science community to come together and develop a consensus on the appropriate tools for nutrition evidence synthesis and grading.

DKT is the Academic Editor of *The American Journal of Clinical Nutrition*. The authors report no conflicts of interest.

## References

1. Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. *Adv Nutr* 2015;6:5–18.
2. Barnard ND, Willett WC, Ding EL. The misuse of meta-analysis in nutrition research. *JAMA* 2017;318:1435–6.
3. Zeraatkar D, Bhasin A, Morassut RE, Churchill I, Gupta A, Lawson DO, Miroshnychenko A, Sirotych E, Aryal K, Mikhail D, et al. Descriptive analysis of the characteristics and quality of systematic reviews and meta-analyses of observational nutritional epidemiology: a cross-sectional study. *Am J Clin Nutr* 2021;113(6):1578–92.
4. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
5. Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, Morgan RL, Gartlehner G, Kunz R, Katikireddi SV, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol* 2019;111:105–14.
6. Movsisyan A, Melendez-Torres GJ, Montgomery P. Outcomes in systematic reviews of complex interventions never reached “high” GRADE ratings when compared with those of simple interventions. *J Clin Epidemiol* 2016;78:22–33.
7. Neuhauser ML. Red and processed meat: more with less? *Am J Clin Nutr* 2020;111:252–5.
8. Qian F, Riddle MC, Wylie-Rosett J, Hu FB. Red and processed meats and health risks: how strong is the evidence? *Diabetes Care* 2020;43:265–71.
9. Schwingshackl L, Schwedhelm C, Hoffmann G, Lampousi A-M, Knüppel S, Iqbal K, Bechthold A, Schlesinger S, Boeing H. Food groups and risk of all-cause mortality: a systematic review and meta-analysis of prospective studies. *Am J Clin Nutr* 2017;105:1462–73.
10. Neuenschwander M, Ballon A, Weber KS, Norat T, Aune D, Schwingshackl L, Schlesinger S. Role of diet in type 2 diabetes

- incidence: umbrella review of meta-analyses of prospective observational studies. *BMJ* 2019;366:12368.
11. Boon MH, Thomson H, Shaw B, Akl EA, Lhachimi SK, López-Alcalde J, Klugar M, Choi L, Saz-Parkinson Z, Mustafa RA, et al. Challenges in applying the GRADE approach in public health guidelines and systematic reviews: a concept article from the GRADE Public Health Group. *J Clin Epidemiol* 2021;135:42–53.
  12. Schwingshackl L, Knüppel S, Schwedhelm C, Hoffmann G, Missbach B, Stelmach-Mardas M, Dietrich S, Eichelmann F, Kontopantelis E, Iqbal K, et al. Perspective: NutriGrade: a scoring system to assess and judge the meta-evidence of randomized controlled trials and cohort studies in nutrition research. *Adv Nutr* 2016;7: 994–1004.
  13. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
  14. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
  15. Bouvard V, Loomis D, Guyton KZ, Grosse Y, Ghissassi FE, Benbrahim-Tallaa L, Guha N, Mattock H, Straif K, International Agency for Research on Cancer Monograph Working Group. Carcinogenicity of consumption of red and processed meat. *Lancet Oncol* 2015;16:1599–600.
  16. Katz DL, Karlsen MC, Chung M, Shams-White MM, Green LW, Fielding J, Saito A, Willett W. Hierarchies of evidence applied to lifestyle Medicine (HEALM): introduction of a strength-of-evidence approach based on a methodological systematic review. *BMC Med Res Methodol* 2019;19:178.
  17. Goldenberg JZ, Day A, Brinkworth GD, Sato J, Yamada S, Jönsson T, Beardsley J, Johnson JA, Thabane L, Johnston BC. Efficacy and safety of low and very low carbohydrate diets for type 2 diabetes remission: systematic review and meta-analysis of published and unpublished randomized trial data. *BMJ* 2021;372:m4743.