American Journal of Epidemiology, 2025, 194, 1681–1686
https://doi.org/10.1093/aje/kwae332
Advance access publication date August 31, 2024
Practice of Epidemiology

Assessing the certainty of the evidence in systematic reviews: importance, process, and use

Romina Brignardello-Petersen* and Gordon H. Guyatt

Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

*Corresponding author: Romina Brignardello-Petersen, HSC 2C, 1280 Main Street W, Hamilton, ON (brignarr@mcmaster.ca)

Abstract

When interpreting results and drawing conclusions, authors of systematic reviews should consider the limitations of the evidence included in their review. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach provides a framework for the explicit consideration of the limitations of the evidence included in a systematic review, and for incorporating this assessment into the conclusions. Assessments of certainty of evidence are a methodological expectation of systematic reviews. The certainty of the evidence is specific to each outcome in a systematic review and can be rated as high, moderate, low, or very low. Because it will have an important impact, before assessing the certainty of evidence, reviewers must clarify the intent of their question: are they interested in causation or association. Serious concerns regarding limitations in the study design, inconsistency, imprecision, indirectness, and publication bias can decrease the certainty of the evidence. Using an example, this article describes and illustrates the importance and the steps for assessing the certainty of evidence and drawing accurate conclusions in a systematic review.

Key words: systematic reviews; certainty of evidence; quality of evidence; grade approach.

Introduction

A recent systematic review evaluated the association between ultra-processed food intake and all-cause mortality.[1] The authors concluded: "This meta-analysis suggests that high consumption of [ultra-processed food], sugar-sweetened beverages, artificially sweetened beverages, processed meat, and processed red meat might increase all-cause mortality, while breakfast cereals might decrease it." While reporting estimates of association (risk ratios and their corresponding 95% confidence intervals) that showed what they labeled as "significant" associations, when drawing conclusions the authors appropriately considered the limitations of the evidence in their systematic review. In the discussion, they mentioned issues such as the limited number of studies, lack of adjustment of potential confounding factors, and applicability concerns. It is likely that these issues led the authors to use terms such as "suggests" and "might" in their conclusion, which reflects some degree of uncertainty.

As described in our previous article in this series,[2] systematic reviews collate all existing evidence that answer a specific question and use explicit and reproducible methods that make their conclusions more trustworthy. The methods should be explicit and reproducible at all steps of a systematic review, from searching and selecting the studies to include, to drawing conclusions after reviewers perform data analysis. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach provides a framework for the explicit consideration of the limitations of the evidence included in a systematic review (that is, a framework for assessing the certainty of evidence), and for incorporating this assessment into the conclusions. Using the GRADE approach minimizes biases and leads to more trustworthy conclusions.

The certainty of evidence (also known as the quality of evidence) reflects the confidence that the truth lies on one side of a specified threshold or within a specific range.[3] This definition applies to questions regarding association, effect, prevalence, or any question in which systematic reviewers are estimating "the truth" (or, in statistical terms, a parameter). Ratings of certainty of evidence can be high, moderate, low, or very low. On one extreme, high certainty evidence implies that reviewers believe that the evidence is very likely to represent the truth and are confident making statements about what such truth is. On the other extreme, very low certainty evidence implies that reviewers believe that the likelihood that the evidence represents the truth is much lower, and they are not confident making statements about the truth.

Over 110 academic, professional, policy-making, and guideline development organizations around the world have adopted GRADE, including the World Health Organization, the United States Centers for Disease Control and Prevention (CDC), and the world's leading health technology group, the Swedish Agency for Health Technology Assessment.[4] In addition, systematic review authorities, including the Cochrane Collaboration, consider GRADE certainty of evidence assessments a methodological expectation for systematic reviews.[5]

These considerations highlight the importance of systematic reviewers knowing how to conduct these assessments, and of

Received: October 25, 2023. **Accepted:** August 19, 2024

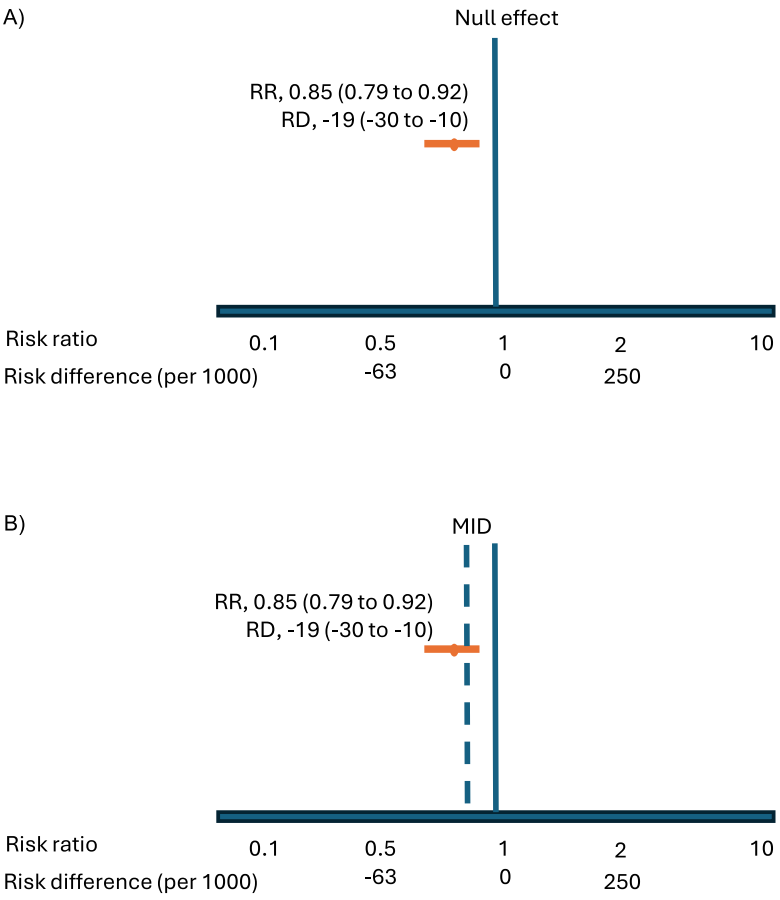

Figure 1. Pooled estimate from a systematic review assessing the association between consumption of breakfast cereals and all-cause mortality. If the authors want to assess their certainty in an association of any magnitude, the confidence interval of the risk ratio (RR) and risk difference (RD) not crossing the null value supports the authors conclusions (A). If, on the other hand, the authors want to assess their certainty on an association of an important magnitude (ie, larger than a minimally important difference), a confidence interval crossing their threshold of importance does not support the authors conclusions (B).

users knowing how to interpret them. This article introduces the GRADE approach for assessing the certainty of the evidence and describes how the assessment is done, how they are incorporated when drawing conclusions from a systematic review, and tools to support systematic reviewers and users faced with certainty of evidence assessments.

Assessing the certainty of evidence

Figure 1(A) illustrates the pooled estimate from the meta-analysis the reviewers conducted (risk ratio of 0.85, with a 95% confidence interval of 0.79-0.92). Based on the language they used, the authors were interested in knowing whether there was an association between breakfast cereals and reduction of mortality, regardless of the magnitude. That is, had they assessed the certainty of the evidence, they would have rated the certainty that the true association was entirely on the range of a reduction in mortality (any value lower than the null effect). Because the confidence interval is entirely in this range, the results support the authors' conclusion.

Alternatively, the authors could have been interested in knowing whether the association between cereals and reduction in mortality was important; that is, that the magnitude of the association is larger than a threshold that separates a range of trivial association (or no important association) from a range of important association (for example, −15 per 1000 deaths). In that case, the authors would have rated the certainty that the true association was entirely below −15 per 1000. The 95% confidence interval of the risk difference of −30 to −10, suggesting the possibility of a trivial association, would have provided more limited support for the authors' conclusion (Figure 1B).

This example illustrates how there may be more certainty or less certainty in the same body of evidence depending on the specific question reviewers are trying to answer, highlighting the importance of being explicit about this question. This example only focused on the statistical uncertainty reflected by the confidence interval around the pooled estimate from the meta-analysis. However, just as the authors correctly pointed out in their discussion, there are other potential limitations in the body of evidence. We later describe the process for assessing the certainty of the evidence using GRADE and how this process addresses these issues.

Step 1: Clarify the intent of the research question

Before assessing the certainty of the evidence, reviewers must clarify the intent of their question.[6] For instance, when addressing the effects of an antiviral drug on symptoms in people with a viral infection, the intent of the researchers is to determine to what extent the drug causes an improvement in such symptoms. If this causal relationship is shown, then medical doctors would consider prescribing the drug to their patients. To address this type of question, assuming they are well designed and conducted,

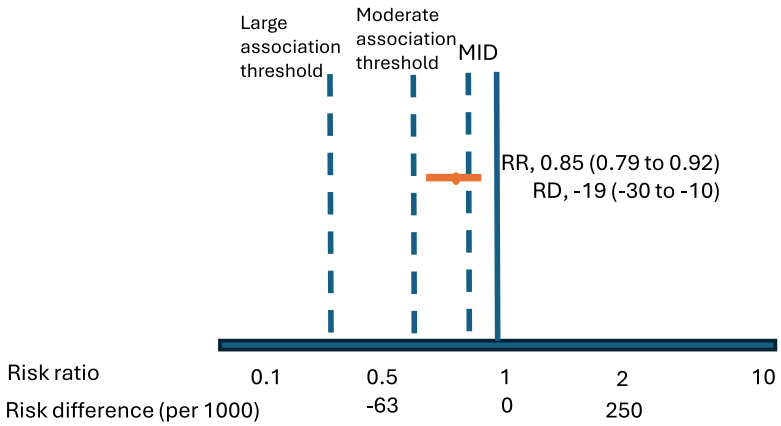

Figure 2. Pooled estimate from a systematic review assessing the association between consumption of breakfast cereals and all-cause mortality. If the authors want to assess their certainty in a small association, they need to consider the relationship between the confidence interval of the pooled estimate and the thresholds for a small association and a moderate association; and they must determine if this pooled estimate falls between them (supporting a conclusion of a small association) or crosses either or both of them (raising uncertainty about the association being small).

randomized clinical trials are the ideal type of study design. Observational studies could also provide some evidence, but the risk of bias due to confounding (or even residual confounding when all known confounders have been accurately measured and used in an adjusted analysis) is high.

In contrast to this focus on causation, often times reviewers aim to address association questions. Although there may be a relationship between an exposure and an outcome, this relationship does not need to be causal because there is no intention to modify the status of the exposure (introduce it when it is not present or remove it when it is) to impact the health outcomes. For instance, review authors often examine the association between age and adverse outcomes. We cannot modify age, and the causal relation is largely between the morbidities that accompany aging rather than age itself (and to that extent age is not itself causal). Here, researchers use this information to identify people at high risk of an outcome, without intervening on the variable associated with the outcome.[6] Observational studies can appropriately address these types of questions.

The relation between nocturia and mortality illustrates the difference between intending to answer an association or a causation question. The authors of a systematic review aimed to assess this relationship . Their meta-analysis of 11 observational studies resulted in a risk ratio of 1.27, with a 95% confidence interval of 1.16-1.40. When assessing the certainty of evidence, they determined that if the intent was to aid those deciding whether they should intervene to reduce nocturia to reduce mortality (ie, a causation question), the certainty of evidence was very low. However, if the intent was to use a history of nocturia to identify people at risk of death (ie, an association question), the certainty of the evidence was moderate.[7]

In the example about breakfast cereal consumption and all-cause mortality, the intent of the question of the authors is not explicit, and it is challenging to infer. Because of that, in illustrating the concepts later on using this example, we will describe both situations: a causation and an association question.

Step 2: Define a target of certainty rating

After clarifying the intent of the research question, reviewers must be explicit about in what specifically they are rating their certainty—we refer to this concept as the "target of the certainty rating."[8] At the beginning of this section, we illustrated 2 possible targets of certainty rating: any association (Figure 1A), or an important association (Figure 1B). Reviewers could also rate their certainty on a particular magnitude of effect or association; that is, they could rate their certainty that the association is trivial, small, moderate, or large by establishing thresholds that delimit these ranges.

For instance, authors could establish a threshold separating the small and the moderate ranges of mortality reduction (ie, a moderate association threshold) of -35 deaths per 1000. The 95% confidence interval of -30 to -10 deaths per 1000 supports the conclusion that the association is smaller than moderate. However, based on the threshold separating the trivial and small association ranges previously established (−15 per 1000), reviewers would not be as confident that the effect is small (there is a possibility that the effect is trivial; Figure 2).

To avoid biases in the systematic review process, reviewers must prespecify which thresholds and ranges of values they will use for assessing the certainty of evidence. Considerations to establish such thresholds and the thresholds themselves may vary depending on whether reviewers are answering an association or a causation question.

The most appropriate target of certainty rating depends on these thresholds, together with the point estimate from the meta-analysis.[8] In the example, the pooled estimate in absolute terms is −19 deaths, with a 95% confidence interval of −30 to −10. If reviewers choose a minimally important difference threshold of 15 fewer deaths per 1000 patients (Figure 1B), because the pooled estimate suggests a larger effect, reviewers should rate their certainty that there is an important reduction in mortality. If instead they choose a threshold of 25 fewer deaths per 1000 patients, because the point estimate suggests a smaller effect, they should rate their certainty that the effect is trivial or not important (as described previously, because the confidence interval crosses the threshold in both cases, reviewers would be less certain about either of these inferences).

For illustrating the concepts we describe later on, we will use the target of certainty implicitly set by the authors of the systematic review; that is, using the null effect as a threshold, we will assess the certainty that: (1) breakfast cereals consumption reduces mortality by any magnitude of effect (causation question), and (2) breakfast cereals are associated with a reduction in mortality of any size (that is, those who consume breakfast cereal have a lower risk of all-cause mortality), but not that eating cereals is actually protective (association question).

Table 1. Assessing the certainty of evidence in the systematic review addressing the relationship between cereal consumption and all-cause mortality. The table illustrates the steps and how the intent of the question impacts the certainty level.

Step 1: intent of the question	Step 2: target of certainty rating	Step 3: starting point	Step 4: assessment of each GRADE domain						Step 5: final level of certainty
			Risk of bias	Inconsistency	Imprecision	Indirectness	Publication bias	Others[a]	
Causation	Reduction of any magnitude	Low	No concerns	Serious concerns	No concerns	No concerns	No concerns	Not present	Very low
Association	Negative association of any magnitude	High	No concerns	Serious concerns	No concerns	No concerns	No concerns	Not present	Moderate

[a]Others: Large magnitude of effect, residual confounding acting in opposite direction, dose–response relationship

Step 3: Consider the design of the studies in the body of evidence

Some study designs are more appropriate than others for answering specific questions. For instance, due to their potential for balancing all known and unknown prognostic factors, randomized trials are less likely to be biased when answering questions about effects of interventions (causation questions) than observational studies. Large case series of representative individuals are, on the other hand, the most appropriate design when answering questions about prevalence. When using GRADE, reviewers must determine which is the most appropriate study design to answer their question and, if that is the design of the studies available, start their assessment of the certainty of the evidence at "high." Otherwise, the assessment of the certainty of evidence starts at "low."[9,10]

If the authors of the systematic review addressing the impact of breakfast cereals on mortality had the intent of answering a causation question, because this question would most appropriately be answered with randomized trials but all the evidence available came from observational studies, the authors would have started their assessment of certainty at "low." For answering an association question (inferring not that cereal results in reduction in mortality but presumably because they covary with factors that are actually causal, their consumption is associated with reduced mortality), observational studies are appropriate; and the assessment of certainty of evidence would therefore start at "high."

Step 4: Evaluate each of the GRADE domains

Because the limitations of the body of evidence may vary depending on which specific studies contribute to an estimate, reviewers should make GRADE assessments for each comparison and each outcome. In the example, there are 6 studies included in the meta-analysis of the relation between breakfast cereals and mortality, and those are the 6 studies that should be considered in the assessment of the certainty of evidence. There are, in contrast, 14 studies included in the meta-analysis addressing the relationship between processed red meat and all-cause mortality, and therefore, those 14 studies should be considered in the assessment of the certainty of evidence.

GRADE classifies the potential limitations of the body of evidence into 5 categories: limitations in the study design, inconsistency, imprecision, indirectness, and publication bias. For each of these categories, reviewers judge whether there are no concerns, serious concerns, or very serious concerns.

Limitations in study design refer to any aspect of the studies design and conduct that would increase the risk of bias, thus decreasing how certain we are in the pooled estimate.[11] In our example, the authors found 6 studies addressing the relationship between breakfast cereal and all-cause mortality. Taking at face value the authors' assessment of methodological quality using the Newcastle-Ottawa scale (range 0-9, higher scores better quality), and not accounting for specific concerns about the tool, most of the evidence (that is, most of the weight of the pooled estimate) came from studies with scores 8 or 9. Therefore, there are no serious concerns in this domain. It is important to highlight, however, that the assessment of this domain depends on the appropriate assessment of risk of bias at the study level, using appropriate tools optimally.

When the studies included in a meta-analysis present similar estimates, we are more certain about the pooled estimate than when the results of the studies vary (and we do not understand why). This is what GRADE refers to as inconsistency.[12,13] In the example, some studies reported a negative association and others a positive one; the confidence intervals of some of the studies did not overlap with the confidence interval of others, and the statistical heterogeneity was high (an I^2 of 77.5%). Therefore, there are serious concerns in this domain.

In addition, as illustrated previously, the confidence interval of the pooled estimate reflects how much statistical uncertainty we have, which GRADE refers to as imprecision.[14,15] When the confidence interval crosses the threshold(s) established, leading to the possibility of different conclusions, we are less certain about the association or effect. As described, because the confidence interval of the pooled estimate is completely on the side of a mortality reduction, there are no serious concerns about imprecision in the example.

Indirectness refers to concerns about the applicability of the evidence to answer the question of the systematic review.[16] Although all studies included in a systematic review meet eligibility criteria, sometimes reviewers find that the evidence available only addresses part of the question of interest (for example, all 6 studies may have addressed a specific type of cereal) and must judge to what extent the evidence they found applies to their general question. Other times, due to lack of evidence, the best available evidence comes from slightly different populations (eg, in rare bleeding disorders, a similar disorder), or the outcomes measured are surrogates for the outcomes in the systematic review. In these situations, reviewers must assess if there are applicability concerns; that is, if they believe that the estimate (of association, effect, prevalence, etc.) would be importantly different in their overall question. In the example, the reviewers cited having applicability concerns but were not specific about these, and thus for this example, there are no serious concerns.

Finally, reviewers should consider to what extent they found all the evidence addressing the question; in other words, whether there are concerns about publication bias.[17] Because publication

is more challenging for smaller studies and studies with "negative" results, when there is a strong suspicion that there may be relevant studies unpublished or that could not be accessed, we are less certain about the evidence. Comprehensive searches in all relevant resources (eg, electronic databases, trials registries, gray literature, etc.) minimize the risk of publication bias. When there are sufficient studies in a meta-analysis, reviewers can use statistical tests or graphical representations (ie, funnel plot) to assess publication bias. In the example, the authors of the systematic review could not assess publication bias for the association between breakfast cereal consumption and all-cause mortality statistically. Because they conducted comprehensive searches, however, there are no serious concerns about publication bias.

In addition to the 5 domains that can decrease the certainty of the evidence, GRADE proposes 3 reasons that can increase it: a large magnitude of effect, residual confounding expected to act in the opposite direction of what is observed, and a dose response relationship between the exposure or intervention and the outcome.[18] These domains are usually considered when the body of evidence comes from observational studies. In the example, there is no evidence that any of those 3 criteria are present.

Step 5: Establish the final certainty of evidence

In the example, there were serious concerns about inconsistency and perhaps indirectness; the authors were not specific with regards to which pieces of evidence their concern referred. As mentioned previously, reviewers choose a starting point for the assessment of certainty of evidence, and after addressing each domain, they decide how many levels they would rate down (when there are serious or very serious concerns) or rate up (when they come across considerations that can increase the certainty of evidence). Many times, the concerns are not serious enough to rate down the certainty of evidence one full level (eg, from moderate to low), and reviewers can rate down only one level because of concerns in 2 different domains. The key is for reviewers to be transparent about their decisions.

For answering the question about causation, the certainty of evidence started as low (due to the inherent limitations of observational studies when addressing questions about the effect of interventions) and could be rated down one level due to inconsistency, resulting in very low certainty. For answering a question about association, the certainty of evidence started as high and could be rated down one level due to inconsistency, resulting in moderate certainty (Table 1).

How the certainty of evidence is formally incorporated into the results when drawing conclusions

The ratings of certainty also allow drawing conclusions that express this certainty in a standardized and reproducible way.[19] For example, for a causation question, had the reviewers judged the certainty of evidence as high, they would have concluded that breakfast cereals decrease mortality. If the certainty of evidence was moderate, they would have concluded that breakfast cereals probably decrease mortality. The language the authors intuitively used is compatible with low certainty evidence (ie, breakfasts cereals may decrease mortality). As described previously, however, the certainty of evidence was very low; therefore, reviewers should have concluded that they were very uncertain about the effects of breakfast cereal on mortality. This conclusion would have represented the evidence and its limitations more accurately than the wording that the authors used. If addressing an association question instead, they should have concluded that those who consume breakfast cereal probably have a lower risk of dying. GRADE provides guidance for creating these narrative statements, incorporating the certainty of evidence and, when applicable (depending on the target of certainty rating), the magnitude of effect.[19]

Tools to facilitate the assessment of the certainty of evidence

There are several resources available for reviewers to make their assessments of the certainty of evidence. The GRADE working group has provided detailed guidance over the years about how to conduct assessments of the certainty of evidence; the key papers are cited throughout this article. In addition, reviewers can benefit from the use of online software, such as GRADEpro (https://gdt.gradepro.org) and MAGICapp (https://app.magicapp.org), which guides them through their assessments and aids in the preparation of GRADE Summary of Findings Tables. Finally, there are many GRADE centers around the world, such as the US GRADE network (https://us.gradeworkinggroup.org) that provides training activities.

Conclusions

Although the first GRADE article was published almost 20 years ago, its uptake in specific fields is just starting. GRADE provides a framework for the systematic assessment of the strengths and limitations of the body of evidence included in a systematic review and the assessment of the certainty of evidence. Because it allows for asystematic and appropriate interpretation of the evidence and accurate conclusions, assessing the certainty of the evidence is a crucial step of data synthesis in any type of systematic review. The appropriate use of GRADE, however, requires clarifying the intent of the question, and therefore reviewers should always be explicit about the inferences they want to make.

It is important to highlight, finally, that GRADE assessments of the certainty of evidence assume that the evidence (ie, studies) is trustworthy; that is, that the data and results have not been fabricated, falsified, or modified in any way. Systematic reviewers should be aware of these potential issues and make all possible efforts to detect problematic studies[20] and not include them in the body of evidence.

Funding

None declared.

Conflict of interest

The authors do not have any financial conflict of interest.

References

1. Taneri PE, Wehrli F, Roa-Díaz ZM, et al. Association between ultra-processed food intake and all-cause mortality: a systematic review and meta-analysis. *Am J Epidemiol*. 2022;191(7):1323-1335. https://doi.org/10.1093/aje/kwac039
2. Brignardello-Petersen RSN, Guyatt GH. Systematic reviews of the literature: an introduction to current methods. *Am J Epidemiol*. 2025;194(2):536-542. https://doi.org/10.1093/aje/kwae232
3. Hultcrantz M, Rind D, Akl EA, et al. The GRADE working group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;87:4-13. https://doi.org/10.1016/j.jclinepi.2017.05.006

4. The GRADE Working Group Organizations. "https://www.gradeworkinggroup.org." Published date not available. Updated 2024; Accessed June 6, 2024.
5. Higgins J, Lasserson T, Chandler J, et al. *Methodological Expectations of Cochrane Intervention Reviews*. London: Cochrane; 2016.
6. Han MA, Leung G, Storman D, et al. Causal language use in systematic reviews of observational studies is often inconsistent with intent: a systematic survey. *J Clin Epidemiol*. 2022;148:65-73. https://doi.org/10.1016/j.jclinepi.2022.04.023
7. Pesonen JS, Cartwright R, Vernooij RWM, et al. The impact of nocturia on mortality: a systematic review and meta-analysis. *J Urol*. 2020;203(3):486-495. https://doi.org/10.1097/JU.0000000000000463
8. Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. *J Clin Epidemiol*. 2021;137:163-175. https://doi.org/10.1016/j.jclinepi.2021.03.026
9. Schunemann HJ, Cuello C, Akl EA, 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;111:105-114. https://doi.org/10.1016/j.jclinepi.2018.01.012
10. Morgan RL, Thayer KA, Santesso N, et al. GRADE working group a risk of bias instrument for non-randomized studies of exposures: a users' guide to its application in the context of GRADE. *Environ Int*. 2019;122:168-184. https://doi.org/10.1016/j.envint.2018.11.004
11. 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(4):407-415. https://doi.org/10.1016/j.jclinepi.2010.07.017
12. Guyatt GH, Oxman AD, Kunz R, et al. GRADE working group GRADE guidelines: 7. Rating the quality of evidence–inconsistency. *J Clin Epidemiol*. 2011;64(12):1294-1302. https://doi.org/10.1016/j.jclinepi.2011.03.017
13. Guyatt G, Zhao Y, Mayer M, et al. GRADE guidance 36: updates to GRADE's approach to addressing inconsistency. *J Clin Epidemiol*. 2023;158:70-83. https://doi.org/10.1016/j.jclinepi.2023.03.003
14. Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach. *J Clin Epidemiol*. 2022;150-224. https://doi.org/10.1016/j.jclinepi.2022.07.014
15. Schünemann HJ, Neumann I, Hultcrantz M, et al. GRADE working group GRADE guidance 35: update on rating imprecision for assessing contextualized certainty of evidence and making decisions. *J Clin Epidemiol*. 2022;150:225-242. https://doi.org/10.1016/j.jclinepi.2022.07.015
16. Guyatt GH, Oxman AD, Kunz R, et al. GRADE working group GRADE guidelines: 8. Rating the quality of evidence–indirectness. *J Clin Epidemiol*. 2011;64(12):1303-1310. https://doi.org/10.1016/j.jclinepi.2011.04.014
17. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence–publication bias. *J Clin Epidemiol*. 2011;64(12):1277-1282. https://doi.org/10.1016/j.jclinepi.2011.01.011
18. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64(12):1311-1316. https://doi.org/10.1016/j.jclinepi.2011.06.004
19. Santesso N, Glenton C, Dahm P, et al. GRADE working group GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol*. 2020;119:126-135. https://doi.org/10.1016/j.jclinepi.2019.10.014
20. Boughton SL, Wilkinson J, Bero L. When beauty is but skin deep: dealing with problematic studies in systematic reviews. *Cochrane Database Syst Rev*. 2021;6(6):Ed000152. https://doi.org/10.1002/14651858.ED000152

American Journal of Epidemiology, 2025, **194**, 1681–1686
https://doi.org/10.1093/aje/kwae332
Practice of Epidemiology