

NUQUEST—NUTRITION QUALITY EVALUATION STRENGTHENING TOOLS: DEVELOPMENT OF TOOLS FOR THE EVALUATION OF RISK OF BIAS IN NUTRITION STUDIES

Shannon E Kelly,¹ Linda S Greene-Finestone,² Elizabeth A Yetley,³ Karima Benkhedda,⁴ Stephen PJ Brooks,^{4,5} George A Wells,¹ and Amanda J MacFarlane^{4,5}

¹Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada; ²Applied Research Division, Public Health Agency of Canada, Ottawa, Ontario, Canada; ³US NIH (retired), Bethesda, MD, USA; ⁴Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada; and ⁵Department of Biology, Carleton University, Ottawa, Ontario, Canada

ABSTRACT

Background: Dietary exposure assessments are a critical issue in evaluating human nutrition studies; however, nutrition-specific criteria are not consistently included in existing bias assessment tools.

Objectives: Our objective was to develop a set of risk of bias (RoB) tools that integrated nutrition-specific criteria into validated generic assessment tools to address RoB issues, including those specific to dietary exposure assessment.

Methods: The Nutrition Quality Evaluation Strengthening Tools (NUQUEST) development and validation process included 8 steps. The first steps identified 1) a development strategy; 2) generic assessment tools with demonstrated validity; and 3) nutrition-specific appraisal issues. This was followed by 4) generation of nutrition-specific items and 5) development of guidance to aid users of NUQUEST. The final steps used established ratings of selected studies and feedback from independent raters to 6) assess reliability and validity; 7) assess formatting and usability; and 8) finalize NUQUEST.

Results: NUQUEST is based on the Scottish Intercollegiate Guidelines Network checklists for randomized controlled trials, cohort studies, and case-control studies. Using a purposive sample of 45 studies representing the 3 study designs, interrater reliability was high (Cohen's κ : 0.73; 95% CI: 0.52, 0.93) across all tools and at least moderate for individual tools (range: 0.57–1.00). The use of a worksheet improved usability and consistency of overall interrater agreement across all study designs (40% without worksheet, 80%–100% with worksheet). When compared to published ratings, NUQUEST ratings for evaluated studies demonstrated high concurrent validity (93% perfect or near-perfect agreement). Where there was disagreement, the nutrition-specific component was a contributing factor in discerning exposure methodological issues.

Conclusions: NUQUEST integrates nutrition-specific criteria with generic criteria from assessment tools with demonstrated reliability and validity. NUQUEST represents a consistent and transparent approach for evaluating RoB issues related to dietary exposure

assessment commonly encountered in human nutrition studies. *Am J Clin Nutr* 2022;115:256–271.

Keywords: quality assessment instrument, nutrition, randomized controlled trial, cohort study, case-control study, risk of bias tool

Introduction

Scientific consensus committees tasked with evaluating diet and health–disease relations for policy, clinical, and educational purposes are increasingly applying evidence-based methodologies to the evaluation of nutrition studies (1). However, a universal observation is that although the concepts and methods of evidence-based medicine are applicable to nutrition topics, diet-related challenges in nutrition studies require consideration (1–7). These include the difficulty in obtaining accurate and comprehensive exposure estimates (i.e., intake or status estimates), the potential for confounding and interactive effects of food

Supported by Health Canada.

AJM is an Associate Editor for *The American Journal of Clinical Nutrition* and played no role in the Journal's evaluation of the manuscript.

Supplemental Files 1–7 are available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at <https://academic.oup.com/ajcn/>.

GAW and AJM contributed equally to this work as senior authors.

Address correspondence to AJM (e-mail: amanda.macfarlane@hc-sc.gc.ca).

Abbreviations used: AHRQ, Agency for Healthcare Research and Quality; NUQUEST, Nutrition Quality Evaluation Strengthening Tools; PICO, Population, Intervention (or exposure), Comparison, and Outcome; QAI, quality assessment instrument; RCT, randomized controlled trial; RoB, risk of bias; SIGN, Scottish Intercollegiate Guidelines Network; SR, systematic review.

Received March 7, 2021. Accepted for publication September 29, 2021.

First published online November 23, 2021; doi: <https://doi.org/10.1093/ajcn/nqab335>.

substances, and the impact of participant baseline nutritional status on the relation between exposure to a particular food substance and health outcome. If not properly accounted for, these challenges can lead to bias, misleading interpretation, and erroneous conclusions. Thus, the development of an assessment approach that integrates nutrition-specific concepts with accepted general concepts of study design and conduct is of paramount importance to users of nutrition evidence (8–10).

Our objective was to develop a set of reliable and valid tools—Nutrition QUality Evaluation Strengthening Tools (NUQUEST)—to aid in evaluating the types of study designs commonly used for nutrition studies while retaining generic assessment components from existing tools. We found inconsistencies among publications regarding the appropriate terminology as to whether our goal was best described by the term “quality” or “risk of bias” (RoB) (11). These terms, often applied interchangeably, describe conditions related to study design and conduct associated with the validity of study results (12). Quality assessment instruments (QAIs) assess the quality of a study from conception to interpretation, whereas RoB tools assess the accuracy of estimates of benefit and risk. A QAI, an older term than RoB, includes the assessment of precision and generalizability, as well as RoB. Existing QAIs and RoB tools (assessment instruments) are usually generic and focus on the evaluation of general study design and conduct issues that are universally important. Examples of generic tools for assessing RoB previously used in the nutrition literature include Cochrane RoB for randomized controlled trials (RCTs) and ROBINS-I for observational studies (10, 13). However, they do not address specific nutrition-related challenges that contribute to uncertainty in nutrition studies, but, on occasion, are adapted to address them. For example, the Agency for Healthcare Research and Quality (AHRQ) added several nutrient-specific questions to the Cochrane RoB tool for RCTs and the Newcastle-Ottawa Scale for observational studies to address uncertainties associated with dietary assessment measures (10, 14–16). Although modifications can improve the sensitivity of these assessment instruments for nutrition studies, inconsistent adaptation and application limit the validity and comparability of review outcomes.

Here we present the development and validation of NUQUEST, a suite of RoB tools for evaluating nutrition RCTs, cohort studies, and case-control studies. They combine RoB components from an existing assessment instrument with nutrition-specific criteria, detailed guidance, and worksheets that address RoB issues related to dietary exposure assessment. We intended NUQUEST to be helpful to both nonnutrition scientists (e.g., research methodologists, epidemiologists) who may be unfamiliar with nutritional issues, and nutrition scientists who may lack in-depth expertise in human study design and conduct. NUQUEST complements nutrition reporting guidelines, guidelines for conducting nutrition systematic reviews (SRs), and guidelines for grading nutrition evidence, and it can be used in conjunction with these tools when appropriate.

Methods

The coauthors constituted a multidisciplinary working group of 7 US and Canadian experts in the areas of nutrition science

and epidemiology with the goal to develop a set of nutrition-specific instruments to be based on validated generic instruments for the assessment of RoB in nutrition research studies. For this work, the term “nutrition studies” would broadly apply to studies in which the focus is on the intake of a food substance (nutrient or other food component), food, dietary pattern, or biomarkers of nutritional intake/status. The development of the nutrition-specific RoB tools for RCTs, cohort studies, and case-control studies followed an 8-step approach: 1) identify a development strategy; 2) identify existing and appropriate assessment instruments for adaptation; 3) identify nutrition-specific appraisal issues; 4) generate nutrition-specific items; 5) adapt and/or develop guidance for use; 6) assess reliability and validity; 7) assess formatting and usability; and 8) finalize NUQUEST (Figure 1).

Identify a development strategy

The group’s overall strategy was to develop a nutrition-specific RoB tool that captured the generic RoB concepts addressed in existing instruments while adding specificity for assessing nutrition issues in nutrition studies. The working group defined the goals, focus, and general approach for the project.

Identify existing and appropriate assessment instruments for adaptation

A rapid, systematic process was used to identify instruments from systematic or comparative review articles that assessed existing QAIs and RoB tools that could serve as a starting point for developing nutrition-specific RoB tools. The rationale for starting with existing instruments was to leverage their demonstrated properties (i.e., validity) and experience of use in the field. We performed a literature search in December 2018 of PubMed, MEDLINE, and Google Scholar using a combination of keywords and Medical Subject Headings syntax relevant to critical appraisal, quality assessment, or RoB to locate potentially relevant articles (Supplemental File 1). We included reviews that formally evaluated the appraisal domains and properties of different types of study designs (e.g., RCTs or observational studies). A working group member screened titles and abstracts and then the full text of potentially relevant articles. Included reviews were vetted by the working group until agreement on inclusion status was achieved. In addition, we searched bibliographic databases to identify instruments published after the search date of the most recent review. Individual generic instruments were identified from recommendations made by included reviews, and generic instruments were assessed by the working group using a process that considered the value and feasibility of adding nutrition-specific items in the context of the existing RoB items.

Identify nutrition-specific appraisal issues

In order to develop appropriate nutrition-specific items to add to the generic instruments, the working group identified nutrition-specific appraisal issues. Nutrition-specific issues were defined as those related to nutrition exposure measurement and the interpretation of nutrition data. We considered “dietary exposure assessment” broadly to encompass multiple components

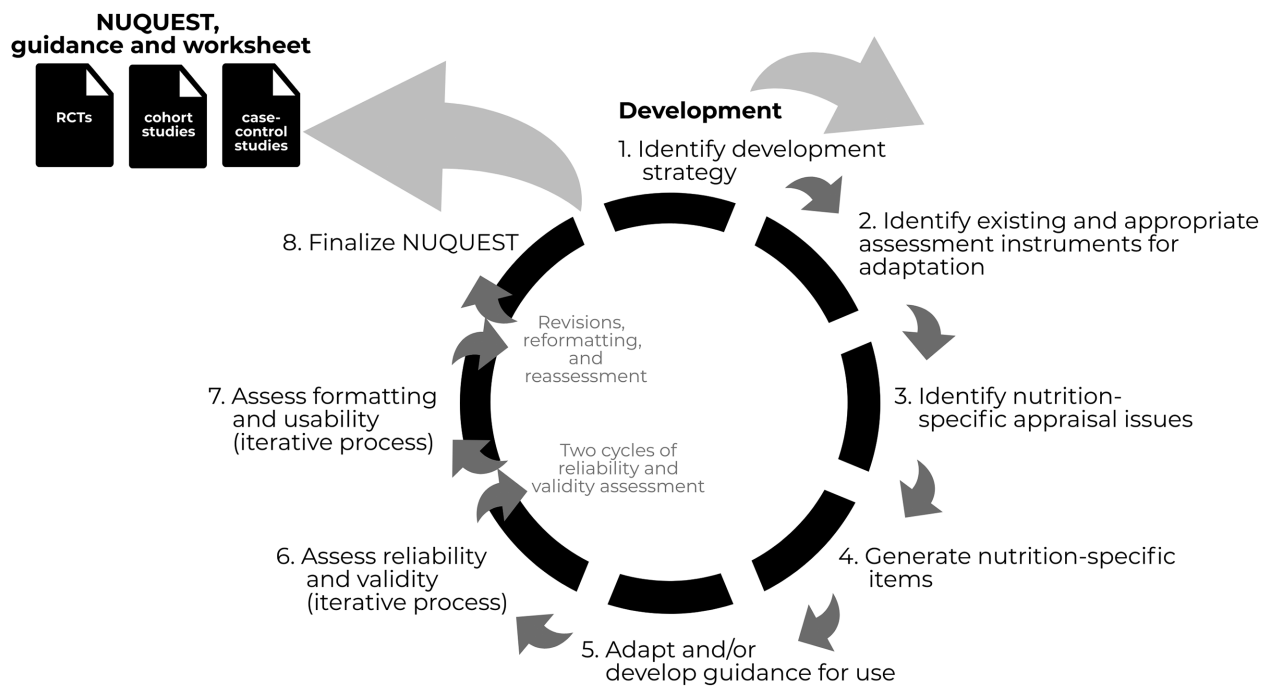


FIGURE 1 Approach for the development of NUQUEST, guidance and worksheet. NUQUEST, NUtrition QUality Evaluation Strengthening Tools; RCT, randomized controlled trial.

of exposure assessment including accuracy and completeness of exposure assessment methodologies, assurances of adherence to interventions, and useful context information for interpreting dietary exposure data. We drew on published literature describing the experiences of scientists who had either conducted or used nutrition-based SRs (7–10), recommendations of expert groups (4, 5, 17–21), articles that had identified nutrition-specific appraisal issues (2, 3), and considered the experiences among the working group members in evaluating nutrition studies. Preliminary nutrition-specific appraisal issues were compiled from all available sources and then vetted within the working group.

Generate nutrition-specific items

The working group drafted a set of nutrition-specific RoB items for each study type (RCTs, cohort studies, case-control studies) based on the identified nutrition-specific appraisal issues with the intention of “bolting” them onto the selected generic instruments as an additional checklist section.

Adapt and/or develop guidance for use

The working group reviewed the existing item-specific guidance for the generic RoB items and edited or revised the text as needed to include context and examples relevant to nutrition studies and the specific study design. We developed detailed guidance de novo for each nutrition-specific item using a structured, iterative process. Two additional experts (non-working group members) in clinical and population nutrition reviewed NUQUEST, including the generic and nutrition-specific items and guidance. The proposed guidance was refined and finalized using an iterative approach.

Assess reliability and validity

To assess the impact of the nutrition-specific items on overall study rating, we used a purposive sample of 45 nutrition-based studies representing 3 study designs (**Supplemental File 2**). Studies were selected based on 2 criteria. First, the study had to have been included in an SR that had examined the association between a nutrition exposure and a health outcome in humans and assigned an overall study rating based on an assessment instrument. Second, the studies as a whole had to cover a diverse range of study ratings, as determined in the original SR, which allowed for comparisons between the original SR overall rating and the NUQUEST overall rating across a range of ratings. The sample consisted of 15 RCTs, 15 cohort studies, and 15 case-control studies on a range of topics. Reliability testing was performed in 2 rounds; the first round consisted of 10 studies for each study design and the second consisted of 5 additional studies for each design. Two raters (not members of the working group), with knowledge of research methodology and critical appraisal or nutrition science, independently applied NUQUEST to the selected studies for each set of studies/study designs.

Reliability.

We assessed reliability using Cohen’s κ statistic for interrater agreement, comparing the overall rating of “poor” with “neutral” or “good.” A κ value was determined for each checklist and for all checklists combined (22). The working group regarded κ values ≤ 0 as less than chance agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement (23).

In the first round of reliability testing, the working group assessed the proportion of agreement between raters across 10 studies for each of the 3 types of study design. Interrater reliability was assessed using the proportion of agreement between raters (i.e., raters' assessments of good, neutral, or poor were exactly the same) and the proportion of near agreement (i.e., raters' assessments of good, neutral, or poor either were exactly the same or differed by only 1 category of good, neutral, or poor). After assessing the first 10 studies for each checklist, raters met with the working group to discuss sources of disagreement and strategies for improving agreement. The working group made modifications, as needed, before raters completed the remaining 5 studies for each checklist. Consensus for the 45 sets of NUQUEST ratings was reached by discussion between the 2 raters with adjudication by a third rater.

Validity.

Because there is no accepted "gold-standard" nutrition-specific assessment instrument, it was not possible to assess criterion validity (i.e., how well NUQUEST is in alignment with a gold-standard comparator). A degree of validity was established by identifying a generic tool on which to build NUQUEST that had a long history of application and has performed well in identifying studies of different RoB to the satisfaction of authors and users (Step 2). In addition, content and concurrent validity of NUQUEST was assessed. Content validity was established by including issues specifically related to dietary exposure and other key aspects needed when evaluating nutrition studies (Steps 3 and 4), which provided specific content that was in alignment with the important issues related to nutrition studies, over and above the general methods items of the selected generic tool on which NUQUEST was based. Concurrent validity, how a new tool compares to an existing tool, was assessed by comparing the NUQUEST overall rating for each study (poor, neutral, and good) to the overall rating reported in the original SR from which the study was selected. The ratings from the original SR assessment were recoded into categories of poor, neutral, and good to allow for comparison to NUQUEST.

Assess formatting and usability

Verbal and written feedback was sought from the raters pertaining to the usability of NUQUEST, clarity of the appraisal items and corresponding guidance, and their overall experience applying the tools. They also tracked the time it took to complete each NUQUEST assessment, and documented instances where they perceived scoring to be challenging or guidance unclear.

Finalize NUQUEST

Taking into consideration the final feedback from raters, the wording of the NUQUEST guidance was edited for clarity and made consistent across the 3 study design-specific checklists.

Results

Identify a development strategy

The working group identified their initial goals for the development of NUQUEST: 1) to integrate accepted generic criteria for assessing study design and conduct with nutrition-specific methodological issues relating primarily to nutritional intake and exposure; and 2) to have the resultant NUQUEST be useful both to epidemiological experts who may have limited knowledge of nutritional methodologies and to nutritional scientists who may have limited knowledge of human study design and conduct issues. To achieve these goals, the group decided to adapt an existing generic assessment instrument to maintain the scientific concepts and qualities associated with the items included in the instrument, with some limited modifications to enhance the usability for nutrition contexts. In addition, the group decided that "bolt-on" nutrition-specific items could be added to the generic assessment instruments when those concepts were not addressed by items in the generic instrument. The group agreed that NUQUEST should initially focus on common study designs used in nutrition science.

Identify existing and appropriate assessment instruments for adaptation

The rapid review (24, 25) located 6 SRs (26–31) comparing various assessment instruments after a review of 367 titles and abstracts and 14 full-text articles. Six potentially relevant records (5, 9, 32–35) identified through additional top-up searching were also considered. One unpublished environmental scan was identified (36, 37), but eventually excluded because it identified, but did not evaluate, a number of available assessment instruments.

The working group used the most recent, comprehensive, and rigorously conducted SR by Bai et al. (26) to select an assessment instrument for adaptation. The Bai et al. (26) SR is an agency-sponsored report completed by the Canadian Agency for Drugs and Technologies in Health that updated the findings of a comprehensive 2002 AHRQ report (38) that identified and evaluated assessment instruments. Briefly, a total of 267 unique assessment instruments (57 for SRs, 94 for RCTs, 99 for observational studies, 17 for multiple designs, and 60 evidence grading systems) were included after a comprehensive multidatabase search of published and unpublished sources and applying a rigorous approach to comparisons and evaluation. Each assessment instrument was assessed independently by ≥ 2 review authors and external content experts based on its adequacy in addressing 5 study-level domains important for generic assessment instruments: comparability of subjects, exposure ascertainment/intervention, outcome measure/ascertainment, statistical analysis, and funding. Bai et al. noted that a variety of instruments for appraising studies exist; yet, no gold standard has been identified. They recommended 3 generic assessment instruments produced by the Scottish Intercollegiate Guidelines Network (SIGN) (39, 40) (1 checklist each for RCTs, cohort studies, and case-control studies) based on internal validity and the potential for feasible, consistent, and systematic application of the checklists during evidence appraisal.

There are many assessment instruments but only a few that have been developed to assess with consistency a range of study designs including RCTs and observational studies (cohort studies and case-control studies), such as those developed by the SIGN (39, 40) and the Joanna Briggs Institute (41–43). Other advantages of the SIGN instruments are that they are simple to use, include key criteria for RoB and clear guidance for their application and interpretation, and have been used extensively. Also, SIGN instruments are not outcome-specific so they can be used for a wide range of health outcomes. The working group concluded that the advantages of the SIGN approach supported its use as the basis for the concurrent development of nutrition-specific RoB tools across 3 commonly used nutrition study designs (i.e., RCTs, cohort studies, and case-control studies) for which a suite of SIGN tools were available. Because a SIGN checklist for cross-sectional studies was not available, a *de novo* process will be required for the future development of NUQUEST for cross-sectional studies.

Identify nutrition-specific appraisal issues

Based on findings from a broad literature review (1–5, 17, 19, 34, 44–47) and building on the experiential knowledge of the working group members in performing assessments of nutrition studies, the group identified 5 essential nutrition-specific issues to consider when evaluating studies. These issues are explained in detail in **Table 1** and broadly relate to 1) accuracy of exposure estimates; 2) baseline exposure; 3) exposure throughout the study; 4) consideration of the dietary context; and 5) the duration of exposure relative to outcome.

The first issue identified was the challenge involved in obtaining accurate estimates of exposure and status—an issue that has received considerable attention (1, 4, 5, 17, 44, 45, 47). Several authoritative reports have offered guidance on how to deal with these issues in reporting and evaluating study RoB, and in analyzing study results (4, 5, 17, 19, 34). The second issue—the need for accurate baseline exposure or status measurements—has been recognized as an important issue for cohort studies (34) but is also important for all nutrition study designs (2, 3, 46). The third issue—the need to assess exposure/status throughout the study—affects all study designs, and is especially relevant to the issue of adherence to study protocols in intervention studies (46). The fourth issue—the need to consider dietary context when evaluating the effect of specific food substances or dietary patterns on health outcomes—is important because it can affect the exposure–outcome relation. Diets are complex mixtures containing multiple biologically active components that affect food substance bioavailability and metabolic utilization, and any food substance of interest may be correlated with other food substances in the diet (1, 3, 4). Energy intakes are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, for identifying potential errors in dietary intake reporting (e.g., under- or overreporting of dietary intake) (44, 45, 48). The various dietary and biological complexities and their effects on food substance and outcome relations of interest may vary systematically between subpopulation groups. The last nutrition issue—the duration of exposure relative to outcome—is important because the duration of exposure must be sufficient to reasonably expect the full effect of the exposure

on development of the outcome from both an effectiveness and a safety perspective (46, 49–51).

Generate nutrition-specific items

Reflecting on the identified nutrition-specific issues from Step 3, the working group developed nutrition-specific items to address the issues when generic items of the existing assessment instrument addressed them either insufficiently or not at all. In the case of the former, the nutrition-specific items were designed to build on concepts addressed by generic RoB items, thereby emphasizing specific important details related to nutrition exposure. The items were focused on measurement of exposure, assessment of baseline exposure, exposure assessment throughout the study, the dietary context of the exposure and potential for confounding, and the duration of exposure relative to the outcome. They were “bolted-on” to the SIGN checklists, as appropriate for each study design. An iterative process among the working group and raters was used to refine the wording of the nutrition-specific items. The final nutrition-specific items vary in number and wording according to the study design and consider methodological rigor, complexity, and user-friendliness (**Table 2**). We added 5 nutrition-specific items to the RCT checklist and 4 to the cohort checklist. For the case-control checklist, we retained the original SIGN item for exposure assessment and added 2 additional nutrition-specific items.

Adapt and/or develop guidance for use

Key nutrition concepts applied to both generic and nutrition-specific items. As such, we added nutrition-specific guidance to the existing guidance for the generic SIGN items, including contextual examples pertinent to nutrition studies (**Table 3** presents examples of nutrition-specific guidance for generic items included in NUQUEST for cohorts). We also developed *de novo* guidance for each nutrition-specific item (**Table 4** presents examples of guidance for nutrition-specific items included in NUQUEST for cohorts). Although exposure was assessed by generic items, we emphasized the importance of accurate exposure assessment in the context of nutrition studies (NUQUEST for RCTs, cohort studies, and case-control studies with complete guidance are in **Supplemental Files 3, 4, and 5**, respectively). In NUQUEST for case-control studies, the generic item on exposure assessment received greatly enhanced nutrition-specific guidance, similar to that of NUQUEST for RCTs and cohort studies.

Guidance pertaining to the assessment of exposure and intake was expanded with each round of validation to include discussion of accuracy and validity in a nutrition context. Topics covered include baseline assessments, adherence to intervention, self-report, supplement and food product composition, dietary patterns or exposures, biomarkers and appropriate documentation, and the appropriateness of dietary assessment methodologies for different research purposes. Guidance was edited to include examples from nutrition studies to illustrate specific issues and facilitate tool application and decision-making.

Assess reliability and validity

Raters completed the assessment of 45 nutrition studies (i.e., 15 RCTs, 15 cohort studies, and 15 case-control studies)

TABLE 1 Nutrition-specific issues¹

Nutrition-specific issue	Rationale for consideration of the issue in the NUQUEST checklists
Accuracy of exposure estimate	<p>This issue refers to the ability to assess nutritional exposures accurately and with minimum error:</p> <ul style="list-style-type: none"> • The 2 main issues are measurement error and validation of the assessment methodology. • Measurement error is the difference between the measured value and the true value. It has 2 forms: random error and systematic bias. • Random error (day-to-day or within-person variation) is related to an individual's measured intake on a specific day of testing vs. an individual's long-term average intake ("usual" intake) based on multiple administrations of the same instrument: <ul style="list-style-type: none"> ○ These data are imprecise but not biased. ○ With repeated measures on a subsample of a population, statistical modeling can be used to adjust for these effects (i.e., to estimate "usual" intakes). ○ Failure to deal with random error results in overestimates of tail probabilities, attenuated relations, and loss of power. • Systematic error (bias) results in measurements that depart from the "true intake" in the same direction: <ul style="list-style-type: none"> ○ This can result from under- or overreporting or misreporting intakes and may also be related to individual characteristics (e.g., BMI). Some biomarker methodologies can also exhibit bias. ○ This type of error cannot be reduced or eliminated by taking repeated measurements. ○ Bias can result in loss of power to detect diet–health relations or in errors in establishing accurate intake–response curves. ○ The magnitude of bias varies by the type of methodology used, among different nutrients, and with participant characteristics. • Validation studies are conducted to determine how accurately exposure instruments measure true exposures: <ul style="list-style-type: none"> ○ One type, rarely used, assesses validation by collecting reference measures by direct observation or feeding studies for a time period that is exactly consistent with when each exposure measurement is taken. ○ A second type collects reference measures such as recovery biomarkers or less biased assessment instruments for a time period that is not exactly the same as the measured exposure assessment. Bias is the difference between the average reported intake and average true intake at the group level. ○ Correlation coefficients between measured and true usual intakes are related to the loss of power to detect exposure–outcome relations. ○ Unbiased reference measures are limited, but include some recovery biomarkers and accurately collected data from feeding studies or direct observation. ○ Validation of one method using a comparison with a second method that has a known bias or imprecision (i.e., a weak or flawed reference instrument) can propagate the bias. A common example of this is when one FFQ with a known bias is used as the comparator for another new FFQ. • Sources of error and lack of validation for exposure assessments can result in misleading results for assessing diet–health relations, intake–response curves, and population prevalence estimates. • Differences in research objectives can affect the appropriateness of different dietary assessment approaches. • Failure to consider measurement error and validation can lead to erroneous and misleading study conclusions.
Baseline exposure	<p>This issue refers to the need for consideration of baseline exposure in assessing exposure–outcome relations:</p> <ul style="list-style-type: none"> • Allows for correct description/identification of study groups. • Allows for comparison of nutritional and dietary similarities and differences between study groups. • Facilitates the evaluation of generalizability of study populations to user target populations. • Facilitates the identification of potential nutrition-based confounders and nutrient–nutrient interactions that can influence the relations between the exposure of interest and the outcome, and which therefore should be accounted for in the analysis. • Observational studies: allows for appropriate assignment of participants to study groups. • Intervention trials: baseline exposures can inform on total exposure (from both diet/supplements and intervention, for example), allowing for the determination of accurate intake–response curves.
Exposure throughout study	<p>This issue refers to the need to document changes to exposure throughout the study that could affect exposure–outcome relations:</p> <ul style="list-style-type: none"> • Identifies exposure changes during the study that can occur owing to self-initiated changes in participant dietary patterns and supplement use or compositional changes in marketed food products used by study participants that occur after baseline assessment. These changes can affect exposure–outcome relations in unpredictable ways and lead to erroneous conclusions. Accounting for these changes in analyses is necessary to prevent misleading and erroneous results. • Intervention studies: adherence to intervention products should be monitored because it can affect the actual exposure and thus the exposure–outcome relation.

(Continued)

TABLE 1 (Continued)

Nutrition-specific issue	Rationale for consideration of the issue in the NUQUEST checklists
Dietary context	<p>This issue refers to the need to consider the complexity of the dietary exposure to determine its impact, if any, on the exposure–outcome relation:</p> <ul style="list-style-type: none"> • Food substances may correlate with other food substances in the diet, thereby confounding the attribution of an exposure–outcome relation to a single specific food substance. • Bioavailability differences among food substance sources (e.g., naturally occurring or synthetic) and/or the varying effects of different dietary patterns on the bioavailability of the food substance of interest can affect the exposure–outcome relation. • Direct or indirect interactions between the food substance and other dietary components can alter its effect on the outcome. • Dietary context provides information on total intakes or exposure of the food substance (e.g., background diet plus supplements, bioavailable vs. less bioavailable formats), thus more accurately capturing the effect of the food substance on the outcome. • Dietary context can provide information on energy intakes. This information is important in adjusting for potential biases in nutrient exposures related to energy intakes, when evaluating equivalencies of food or dietary pattern substitutions, or in identifying the potential for under- or overreporting of food substances of interest. • Dietary context provides information to calculate ranges or distributions of the food substance, thus enhancing the ability to create intake–response curves. • Dosing or eating conditions (e.g., single bolus vs. multiple exposures per day; with meals vs. between meals; chemical forms of the supplement) can affect the intake–response relation. • Intervention studies: the ability to “blind” a study may be affected when foods or dietary patterns are the intervention; in studies where there is an addition or deletion of a food, or a food substitution is made, energy equivalency between comparison groups must be considered.
Duration of exposure relative to outcome	<p>This issue refers to the need for the exposure to the food substance to be of sufficient duration to see the full effect on the outcome:</p> <ul style="list-style-type: none"> • The time needed to show a significant and nontransient change in the outcome may vary by food substance and outcome assessed. • Baseline exposure of study participants can affect the length of time of exposure needed to see an effect of the food substance on the outcome. • The dose of the exposure may affect the duration required to observe a change in the outcome. For example, low doses consumed for a longer duration can have the same effect on the outcome as high doses consumed for a shorter time. • Similarly, the format of the exposure may affect the duration required to observe a change in the outcome. For example, a food substance in supplement form may be more bioavailable than as part of a dietary intervention. • Outcomes based on a surrogate marker might reasonably be expected to occur in a shorter time period than the occurrence of a disease outcome.

¹For the purpose of this work, in the context of a nutrition study, “exposure” refers to intake or nutritional status of the food substance or dietary pattern (e.g., folate, Mediterranean diet) of interest. The food substance definition in the table can be found in the checklists. We considered “dietary exposure” broadly to encompass multiple components of exposure assessment including accuracy and completeness of exposure assessment methodologies, assurances of adherence to interventions, and useful context information for interpreting dietary exposure data of interest. The National Cancer Institute’s Dietary Assessment Primer provides comprehensive and detailed information on how to assess and use exposure assessment data (17, 48). NUQUEST, Nutrition QUality Evaluation Strengthening Tools.

selected from SRs in the nutrition literature in 2 separate rounds (10 of each study design in round 1 without a worksheet; 5 in a second round with a worksheet; **Supplemental File 6**). The working group developed the NUQUEST worksheet based on rater feedback after the first round of testing. This worksheet is intended to provide, as needed, a framework for raters to a priori learn how to assess each item as it relates to the research question; the Population, Intervention (or exposure), Comparison, and Outcome (PICO); and methods of assessing intake/nutritional status, outcomes, and confounders. Each of the NUQUEST sections, based on items grouped according to selection, comparability, ascertainment of outcomes (or exposure), and the overall assessment, was rated independently by each rater as “good,” “neutral,” or “poor.”

Reliability.

Overall interrater reliability for NUQUEST across all study designs ($n = 45$ studies) was substantial (0.73; 95% CI: 0.52, 0.93) and moderate to almost perfect for the individual instruments for RCTs (0.59; 95% CI: 0.19, 1.00; $n = 15$ studies), cohort studies (0.57; 95% CI: 0.15, 1.00; $n = 15$ studies), and case-control studies (1.00; 95% CI: 1.00, 1.00; $n = 15$ studies). κ Values were not robust when considering only the first round of reliability testing owing to the limited sample size ($n = 30$ studies) and the corresponding number of nominal categories, therefore only the combined κ values are presented ($n = 45$ studies).

The proportion of agreement of the 2 raters was generally satisfactory for the individual sections and NUQUEST overall rating (Table 5). The proportion of “near agreement” (i.e., the raters’ assessments of good, neutral, or poor were either exactly

TABLE 2 Nutrition-specific items in Nutrition Quality Evaluation Strengthening Tools (NUQUEST)

Study design	Nutrition-specific items
Randomized controlled trial	<ul style="list-style-type: none"> • The frequency and quantity of the intervention exposure under study are accurately and reliably measured. • The relevant baseline exposure is measured and taken into account. • The baseline exposures of all groups have been maintained over the course of the study. • Adherence to the intervention and control intakes is monitored in a reliable and accurate manner. • The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.
Cohort study	<ul style="list-style-type: none"> • The frequency and quantity of the exposure under study are accurately and reliably measured. • The relevant exposure at baseline is measured and taken into account. • The baseline exposure differences between the groups have been maintained over the course of the study. • The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.
Case-control study ¹	<ul style="list-style-type: none"> • The exposure under study is accurately and reliably measured. • The baseline exposure differences between the groups have been maintained over the course of the study. • The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.

¹All items except for the exposure assessment in NUQUEST for case-control studies are found in the nutrition-specific section of NUQUEST. An exception was made for NUQUEST for case-control studies because the generic assessment items for this study design included an exposure ascertainment item. Nutrition-specific guidance was developed for this generic item to highlight the specific issues in assessing nutrition exposures, and to highlight the fact that retrospective case-control study designs are particularly vulnerable to recall bias.

the same or differed by only 1 category of good, neutral, or poor) on the selection, comparability, ascertainment, and nutrition sections of NUQUEST, as well as the overall rating, ranged from 80% to 100% for RCTs, from 80% to 100% for cohort studies, and from 87% to 100% for case-control studies. The proportion for “agreement” (i.e., raters’ assessments of good, neutral, or poor were exactly the same) was lower. In particular, for the overall rating, agreement was only 40% based on the

first 10 studies assessed for each of the 3 study designs. For the assessment of the remaining 5 studies for each study design, the agreement improved to 100% for RCTs, 80% for cohort studies, and 100% for case-control studies when raters used the completed worksheets.

Rater feedback indicated that it would have been helpful to meet to review their consensus on NUQUEST items after assessment of the first 2–3 studies before completing the

TABLE 3 Examples of nutrition-specific guidance developed for generic items of Nutrition Quality Evaluation Strengthening Tools (NUQUEST) for cohort studies

Concept	Item	Sample guidance
Selection of participants/creation of study groups	1.1 The groups being studied are selected from source populations that are comparable in all respects other than the exposure under investigation.	...Groups that are formed from 2 different source populations or are selected from the same source population using different approaches are at higher risk of selection bias. For example, comparing a group of women of childbearing age from an urban center who took folic acid during pregnancy with those from a rural area who did not would be problematic...
	1.2 The study indicates how many of the people asked to take part did so in each of the groups being studied.	...Groups selected from different source populations, or from the same source population using different approaches, may have differential participation rates and should be evaluated more closely. For example, one would be less concerned about differential participation rates in a cohort of women of childbearing age who were selected from the same urban area and among whom groups were formed based on whether or not they took folic acid during their pregnancy. In contrast, if populations from rural and urban areas are compared, we would have more concern that participation rates might differ because, for instance, access to the study centers may be more difficult for the rural participants.
Comparability of groups	1.8 The main potential confounders are identified and/or taken into account in the design and analysis.	...There may be fundamental differences between the groups that may affect nutrient intake or exposure and these should be considered and accounted for in a study (e.g., health-related behavior or personal/lifestyle variables). These factors may be known or unknown. For example, in a cohort of women of childbearing age where folic acid intake (compared with no intake) is being studied, we may want to consider potential confounders that may be related to the nutrient intake such as education, physical activity, and race or ethnicity. In addition, and particularly when we have different source populations (e.g., different geographies), a variety of known and unknown differences between the groups may confound the nutrient intake or exposure, including but not limited to population structure, culture, genetic diversity, and health or personal behaviors.

TABLE 4 Examples of nutrition-specific items and guidance developed for the nutrition-specific items of Nutrition Quality Evaluation Strengthening Tools (NUQUEST) for cohort studies¹

Concept	Item	Sample guidance
Exposure of interest/intervention intake	2.1 The frequency and quantity of the exposure under study are accurately and reliably measured.	<p>...<i>The study should report the method used to assess exposures or status and the validity of such methods.</i> Validity refers to the degree to which the exposure assessment methodology accurately measures the aspect of exposure it was intended to measure. Exposure assessment methods are validated in different ways. Comparing an exposure method to a “gold-standard” or reference method thought to capture all the food and supplement sources containing the exposure of interest provides the best validation approach. Comparing a self-reported method to another self-reported method requires caution because the 2 methods may be affected by the same types of errors and biases. An exposure assessment method could be modified to reflect the population under study if the food/supplement supply differs for the new study population. Validation of a new questionnaire against a previous questionnaire targeting the exposure under study allows for comparability among similarly validated studies...</p> <p>Currently available assessment methods are often subject to random errors and systematic biases that can result in misleading and erroneous conclusions. Optimal methodologies as well as the nature and severity of potential errors and biases vary among food substances and, therefore, must be considered on a case-by-case basis. Methods used in the study under evaluation need to be evaluated for their potential for RoB that includes both general and topic-specific criteria. Topic-specific RoB criteria are issues related specifically to the exposure of interest. For example, issues related to assessing sodium intake will be different from those of other nutrients or total energy...</p> <p><i>Food substance intakes from self-reported diets</i> (e.g., FFQ, food record, or 24-h recall). Methods have different strengths and weaknesses (known errors and biases)... The frequency of evaluation of the exposure will often depend on the food substance, the dosage conditions, and the outcome.</p> <p><i>Food substance intakes from supplements or specific food products</i> (e.g., bars, liquid drinks, fortified foods). The study should report product composition details and the methods used to confirm the composition of the product (i.e., assay procedure), and how adherence to the use of these products was monitored (e.g., supplement count, assessment of changes in nutritional status).</p> <p><i>For dietary pattern or food intakes</i>, the study should include ranges or distributions of any prespecified food substance or status exposures of interest. The study should also include exposure measurement characteristics and errors, of which some can be controlled for statistically (e.g., “usual” intakes). This can be accomplished by intake methodologies that represent averages over a specified time period or by statistically adjusting for within-person variabilities in 24-h recalls. Some intake methodologies cannot adjust for usual intakes (e.g., if they are unknown or unmeasured; or subject to a systematic bias such as underreporting) and this can introduce bias.</p> <p><i>Biomarkers of intake and exposure</i> are usually metabolites recovered from the blood or urine that objectively and accurately assess intake over a period of time. However, biomarker assays can vary in accuracy and reliability. If a biomarker has been accurately measured and appropriately qualified for its intended purpose, it is more reliable than exposure assessed by dietary intake estimation methods.</p> <p>Note that use of nonstandardized methodologies could affect status measures, even when the method itself includes standards and controls. For example, 25-hydroxyvitamin D values differ depending on the assay platform and the laboratory performing the measure. A measure from a laboratory that has been certified for the method will be more reliable than one from an uncertified laboratory and standardized procedures are more reliable than nonstandardized procedures.</p> <p><i>Other important considerations for assessing exposure include the composition, the form, any potential interactions with other exposure components, supplement brand name, and food types or recipes. The importance and weight of these issues (and other relevant characteristics of the exposure) must be determined a priori by the reviewer.</i> Knowledge about the exposure form (e.g., naturally occurring vs. synthetic food substance) and intake conditions (e.g., taken with or without meals; bolus vs. multiple daily exposures; preparation practices for meal components) can be important, especially when bioavailability and/or bioactivity are issues. Ideally, product composition would be confirmed by analysis because food/supplement label values and/or composition databases do not always accurately reflect actual composition. <i>If the exposure is not accurately measured or adequately characterized, confidence in the degree of exposure is reduced. A value on a product label could be acceptable but might not accurately reflect composition and should be interpreted with caution.</i></p> <p>Energy intakes are another consideration. They are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, and for identifying potential errors in dietary intake reporting (e.g., under- or overreporting of dietary intake).</p>

(Continued)

TABLE 4 (Continued)

Concept	Item	Sample guidance
		<i>Rating:</i> Documentation of exposure should be sufficient to determine its validity. For example, intake/nutritional status based on a comparison to an accurate external method would have the highest rating. Self-reported intake or use of a biomarker assay procedure with accuracy or bias concerns that have been described/evaluated would have higher uncertainty. Self-reported intake with no description of accuracy or bias would have the lowest rating. In addition, because of large day-to-day variabilities in intakes by a given individual, higher rating should be given to results expressed as “usual” or long-term intakes. If adherence to the exposure is assessed by a supplement count then there is greater confidence in the measurement than if the assessment is achieved by recall. The tolerance for uncertainty will vary depending on the purpose of your evaluation and will be reflected in the study rating. The rationale for the rating system used, including a description of tolerance for uncertainty, should be defined a priori.
Baseline nutrient intake or exposure	2.2 The relevant exposure at baseline is measured and taken into account.	Accurate baseline intake/status data are necessary to ensure appropriate group assignments; to determine whether changes in intake occurred during the study period (which could confound study results); and to compare results across studies in different populations. If supplement use is the exposure, baseline intakes can be used to ascertain the nutritional similarity of exposed and unexposed groups at baseline and estimate total intake. If expressed as distributions, baseline intakes of study groups can aid in the interpretation of results. For example, if a wide range of baseline statuses is present among groups of a supplement study, the overall effect of the exposure may be null; this could happen if nutrient-depleted participants experience benefit but replete participants experience no effect or even harm...
Intake adherence and maintenance	2.3 The baseline exposure differences between the groups have been maintained over the course of the study.	...All participants are exposed to diets that include food substances, foods, or dietary patterns of interest. Changes in intake over the duration of the study for any or all of the groups could make it more difficult to detect differences in outcome due to the effect of the baseline exposures. Has the potential for this been monitored? If a change in intake has occurred, does the study have appropriate methods to deal with it?...
Intake/outcome interval	2.4 The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.	In order to determine whether a “true” relation exists, the duration of the interval between the exposure and the outcome must be sufficient to allow the outcome to occur. False positives can occur in studies of inadequate duration. For example, transient decreases in LDL cholesterol may be observed in short-term lipid studies but may not persist in long-term studies. Confounding factors such as seasonal changes can transiently affect status or intake. Even study enrolment can transiently affect intake or status in the short term...

¹RoB, risk of bias.

assessment for all of the studies. They felt that this would have increased their agreement for items for which ratings were inconsistent. The multidisciplinary backgrounds of the raters underscored the need to understand the nature and sources of discrepancies between their ratings, which were generally dependent on the expertise of each individual rater (study appraisal methodology or nutrition).

Validity.

As noted, the SIGN instruments on which NUQUEST was built demonstrate high content validity (26) by virtue of their original development process and extensive history of use. When the working group compared the overall study ratings from NUQUEST (based on consensus from 2 independent raters) to the overall study ratings from the original SR (Figure 2), there was agreement for 21 of the 45 studies (46.7%; 10 RCTs, 9 cohort studies, 2 case-control studies) and at least near agreement for 42 studies (93.3%; 14 RCTs, 15 cohort studies, 13 case-control studies). When the overall ratings differed between NUQUEST and the original SR, the rating was lower 79% of the time (19 of 24 studies) and higher 21% of the time (5 of 24 studies). For the 3 studies for which there was not at least near agreement (study numbers 1, 32, and 39; see Figure 2), the nutrition bolt-on

section affected the NUQUEST overall study rating negatively and led to greater disagreement between the NUQUEST overall rating and SR overall rating. In summary, the overall study ratings generally demonstrated agreement between NUQUEST and the original SR ratings and, in instances of major disagreement, the nutrition-specific section was a key contributor to the difference.

Assess formatting and usability

Based on feedback from raters and consensus discussions among the working group members, we modified the format and organization of NUQUEST to improve the logical structure and usability while maintaining the scientific concepts outlined in the original SIGN instruments (Supplemental Files 1–3). We edited the original SIGN items and guidance slightly to improve clarity and, where appropriate, to highlight their relevance to nutrition studies. NUQUEST was reorganized into 5 sections that varied in title depending on the study design: 1) *Selection*—Selection of participants (RCT)/Selection of cohorts (cohort)/Creation of study groups (case-control); 2) *Comparability*—Comparability of study groups (RCT, case-control)/Comparability of cohorts (cohort); 3) *Ascertainment*—Ascertainment of outcomes (RCT, cohort)/Exposure ascertainment (case-control); 4) *Nutrition*—Nutrition-specific issues (RCT, cohort, and case-control); and 5)

TABLE 5 Rater agreement on individual sections and overall rating using Nutrition Quality Evaluation Strengthening Tools (NUQUEST)¹

Study design	Studies, <i>n</i>	Degree of agreement	NUQUEST assessment in agreement, %				
			Selection ²	Comparability ³	Ascertainment ⁴	Nutrition ⁵	Overall ⁶
Randomized controlled trial	15	Near agreement	87	80	93	100	87
		Agreement	40	47	93	80	60
	10	Agreement without worksheet	40	40	80	100	40
	5	Agreement with worksheet	40	60	80	80	100
Cohort	15	Near agreement	100	80	80	93	93
		Agreement	67	47	33	87	53
	10	Agreement without worksheet	60	30	30	80	40
	5	Agreement with worksheet	80	80	40	100	80
Case-control	15	Near agreement	93	87	93	100	100
		Agreement	60	47	40	60	60
	10	Agreement without worksheet	70	20	40	40	40
	5	Agreement with worksheet	40	100	40	100	100

¹Agreement/near agreement was determined section by section and overall. Percentage in agreement was calculated for each section/overall using the number of studies with rater agreement or near agreement out of the total number of studies assessed. Near agreement: raters' assessments of good, neutral, or poor were either exactly the same or within 1 category of good, neutral, or poor. Agreement: raters' assessments of good, neutral, or poor were exactly the same. RCT, randomized controlled trial.

²Selection = selection of participants (RCT)/selection of cohorts (cohort)/creation of study groups (case-control).

³Comparability = comparability of study groups (RCT, case-control)/comparability of cohorts (cohort).

⁴Ascertainment = ascertainment of outcomes (RCT, cohort)/exposure ascertainment (case-control).

⁵Nutrition = nutrition-specific (RCT, cohort, and case-control).

⁶Overall = NUQUEST overall rating.

Overall study rating. A table in the guidance section of each checklist describes the individual sources of bias that NUQUEST is designed to address (i.e., performance, attrition, detection, selection, misclassification, and dietary exposure assessment biases). The revised structure and improved guidance brought focus to the generic concepts of selection, comparability, and ascertainment and to the de novo developed bolt-on nutrition concepts, and facilitated the determination of the overall rating for a specific concept (e.g., selection), which in turn assisted the decision-making required for the overall study rating.

The original SIGN rating scheme for each item included the following response options: yes/no/can't say/does not apply. In the first round of validation (30 studies), the raters frequently gave "fence-sitting" responses where they used the response option of "can't say"; however, the working group determined that this rating was not always justified. In response, we modified the rating system for individual items to include a broader range of response options (yes/probably yes/probably no/no) and removed the "can't say" option. The renamed response option "not applicable" was included only for items where a response would not be pertinent.

The overall rating options for each section are: good (+), where almost all criteria are met, there is little or no concern, and low RoB; neutral (0), where most criteria are met but there are some flaws with an associated concern, and there is a moderate RoB; and poor (−), where either most or all criteria are not met, there are significant flaws, and there is a high RoB. The overall ratings for each section are carried forward to the Overall Study Rating section and used when considering the final assessment of the study (response options: poor, neutral, or good). Guidance for determining overall ratings, including the consideration of the relative importance of specific items within the context of the study design, was edited to reflect the updated format and response options.

Based on qualitative feedback, the raters felt that NUQUEST was easy to apply and feasible to complete within a reasonable time by raters with general knowledge of epidemiology or nutrition science (average: 16.5 min/study, range: 10–33 min). NUQUEST for cohort studies generally took more time for raters to complete than the others, but time to completion improved with each completed study assessment. To aid the future use of NUQUEST, we identified points to consider when developing a review (**Box 1**) and when training raters to use the worksheet (**Box 2**).

The raters indicated that they felt more comfortable rating items that were within their area of expertise than rating items that were not. Specifically, raters with expertise in epidemiology/study design methodology felt confident to assign stricter ratings to the generic items, whereas raters with expertise in nutrition felt confident to assign stricter ratings for the nutrition-specific items. In addition, raters were more likely to use the "no information" option for items outside of their expertise. Often other information in the article could have permitted the rater to assess whether something likely occurred or not. With the addition of a completed worksheet in the second validation round (completed example in **Supplemental File 7**), the raters indicated that they were more confident in applying a rating for items for which they had less expertise and overall more likely to be stricter when applying ratings to all items. This resulted in a tendency to downgrade individual items and overall study ratings. The addition of the worksheet also decreased the use of the "no information" rating; however, raters still tended to use this option in areas for which they had less expertise. As such, the working group decided to remove this option in the final checklists to avoid uninformative answers. Users are now forced to make a judgment on each item, even in cases where details are reported poorly, while taking into account stated accounts of study design, conduct, and analysis and items that the rater judges are likely to have (or have not) occurred in the study.

STUDY DESIGN	STUDY ID	NUQUEST ASSESSMENT ¹ :					Overall ⁶	SR:
		Selection ²	Comparability ³	Ascertainment ⁴	Nutrition ⁵	Overall ⁶		Overall
RCT	1	●	○	○	○	○ ⁷	● ⁸	
	2	●	●	●	●	●	○ ⁸	
	3	●	●	●	●	●	● ⁹	
	4	●	●	●	●	● ⁷	● ⁸	
	5	○	○	●	●	●	● ⁹	
	6	○	●	○	○	○ ⁷	○ ⁸	
	7	●	○	●	●	●	● ⁹	
	8	●	●	●	●	●	● ⁸	
	9	●	●	●	●	●	● ⁹	
	10	●	●	●	●	●	● ⁸	
	11	●	●	○	○	● ⁷	● ⁸	
	12	○	○	●	○	○	● ⁹	
	13	●	●	○	●	●	● ⁸	
	14	○	●	○	○	○ ⁷	○ ⁸	
	15	○	○	●	○	○	○ ⁹	
Cohort study	16	●	●	●	●	●	● ¹⁰	
	17	●	○	●	●	●	○ ¹⁰	
	18	●	○	●	●	○	○ ¹⁰	
	19	●	○	○	●	○	○ ¹⁰	
	20	○	○	●	○	○ ⁷	○ ¹⁰	
	21	●	○	●	●	●	● ¹⁰	
	22	●	○	●	●	●	● ¹⁰	
	23	●	●	●	●	●	● ¹⁰	
	24	●	●	●	●	●	● ¹⁰	
	25	●	○	●	●	● ⁷	● ¹⁰	
	26	●	○	●	●	●	○ ¹⁰	
	27	○	○	○	○	○ ⁷	○ ¹⁰	
	28	○	○	●	○	○ ⁷	○ ¹⁰	
	29	●	○	○	●	●	○ ¹⁰	
	30	○	○	○	○	○ ⁷	● ¹⁰	
Case-control study	31	●	○	○	○	○ ⁷	● ¹¹	
	32	●	●	○	○	○	● ¹¹	
	33	●	○	○	○	○ ⁷	● ¹¹	
	34	●	●	●	●	●	● ¹¹	
	35	●	●	●	●	●	● ¹¹	
	36	●	●	●	●	●	● ¹¹	
	37	●	●	○	○	○	● ¹²	
	38	●	○	○	○	○ ⁷	● ¹¹	
	39	●	○	●	○	○	● ¹²	
	40	●	○	○	○	○ ⁷	● ¹¹	
	41	●	○	●	○	○	○ ¹³	
	42	○	○	●	○	○ ⁷	● ¹¹	
	43	●	●	●	●	●	● ¹¹	
	44	●	○	●	●	●	● ¹¹	
	45	●	●	○	●	●	● ¹¹	

FIGURE 2 Nutrition Quality Evaluation Strengthening Tools (NUQUEST) ratings for individual sections and overall and the overall rating from the original SRs. ¹NUQUEST assessments are by section and overall by study: ● *good*, where almost all criteria are met, little or no concern, and low RoB; ● *neutral*, where most criteria are met, there are some flaws, and moderate RoB; and ○ *poor*, where either most or all criteria are not met, there are significant flaws, and high RoB. ²Selection = selection of participants (RCT)/selection of cohorts (cohort)/creation of study groups (case-control). ³Comparability = comparability of study groups (RCT, case-control)/comparability of cohorts (cohort). ⁴Ascertainment = ascertainment of outcomes (RCT, cohort)/exposure ascertainment (case-control). ⁵Nutrition = nutrition-specific (RCT, cohort, case-control). ⁶Overall = overall NUQUEST rating. ⁷Assessed using the NUQUEST worksheet. ⁸Evaluated by the Cochrane RoB Tool modified for nutrition studies. ⁹Evaluated by the Cochrane RoB Tool. ¹⁰Evaluated by the NOS modified for nutrition cohort studies. ¹¹Evaluated by the NOS. ¹²Evaluated by the JBI critical appraisal tool for case-control studies. ¹³Evaluated by the JBI critical appraisal tool for cross-sectional studies. Because the original SRs used a variety of tools to assess RoB, the original scores were recoded to “poor, neutral, and good” to allow for comparison to NUQUEST scores. Recoding for the Cochrane RoB Tool or the NOS modified for nutrition cohort studies was as follows: low RoB, it was assigned good ●; moderate RoB, it was assigned neutral ●; and high RoB, it was assigned poor ○. Recoding for the NOS was as follows: 1–3, it was assigned ○; 4–6, it was assigned ●; and 7–9, it was assigned ●. Recoding for the JBI cross-sectional tool was as follows: 1–4, it was assigned ○; and 5–8, it was assigned ●. Recoding for the JBI case-control tool was as follows: 1–3, it was assigned ○; 4–7, it was assigned ●; and 8–10, it was assigned ●. Where the authors of the original SR did not make an overall RoB judgment, section-specific judgments and related text within the report were examined to assign an overall RoB assessment by 2 members of the working group. ID, identification; JBI, Joanna Briggs Institute; NOS, Newcastle-Ottawa Scale; RCT, randomized controlled trial; RoB, risk of bias; SR, systematic review.

BOX 1**If Initiating a Review, Points to Consider When Developing the Worksheet¹**

- Is there a clearly stated key research question for the review?
 - In cases where the study's purpose differs from the key research question, does the study design provide information that can still be used to evaluate the key research question?
- Has a PICO been prepared before the identification of articles to assist in study selection/exclusion decisions?
 - Are the selection/exclusion criteria sufficiently broad to include a range of acceptable study methodologies that can subsequently be subjected to assessment with application of the NUQUEST criteria?
- Have the rating criteria to assist reviewers in using NUQUEST been prepared before initiation of the reviews?
 - Does the worksheet include criteria for evaluating a range of exposure methodologies that are specific for the food substance and for the analysis and expression of results?
 - Does the worksheet differentiate between key confounders and other potential confounders of lesser importance?

¹NUQUEST, NUtrition QUality Evaluation Strengthening Tools; PICO, population, intervention/exposure, comparator, outcome.

Finalize NUQUEST

NUQUEST for RCTs, cohort studies, and case-control studies, and guidance for each instrument are available in the Supplementary data, where we have also included an example of a completed worksheet (Supplemental File 7). The wording of the guidance is critical for the application of NUQUEST and refinement is expected to continue with future use.

Discussion

With an emphasis on using the best available evidence to inform public health and nutrition initiatives (52), several evidence appraisal instruments have been developed; however, current instruments do not adequately address methodological challenges inherent to nutrition studies. These challenges contribute to uncertainty and lower our ability to establish nutrition–health outcome associations with confidence. Here we present a series of developed and tested RoB tools that retain generic study appraisal components but with the addition of nutrition-specific appraisal items to address challenges common to nutrition studies. NUQUEST presents an integrated approach for combining generic and nutrition-specific issues in a single tool for assessing the RoB of individual nutrition studies. Other complementary tools include guidelines for designing and conducting nutrition studies (46), guidelines for reporting

nutrition studies (2, 34, 35), procedures for conducting nutrition SRs and meta-analyses (53), and guidelines for grading the evidence (32, 54–57). The rigorous evidentiary evaluation and transparency resulting from the use of RoB tools for nutritional epidemiology can identify problematic areas for which more research or creative statistical approaches for dealing with limitations of dietary exposure data are needed (17).

BOX 2**For the Initiators of the Review, Points to Consider When Preparing Raters to Use the Worksheet in Study Assessments¹**

- Do raters understand the worksheet's purpose for the review, the PICO, and the rating criteria?
- If possible, do raters represent multidisciplinary expertise (e.g., epidemiology, nutrition)?
- If there is >1 rater, are differences adjudicated?
- Do raters recognize the types of situations that may require judgment?

For example:

- If detailed study design and methods information is primarily available in companion articles, have these articles been accessed?
- Where double blinding may not be possible (e.g., 1 of the groups receives an educational program or a whole food substitution), have the study investigators taken steps to reduce the potential biases (e.g., detection bias)?
- If randomization failed to some degree, have the study investigators taken steps to reduce potential biases by treating important variables as covariates in analyses?
- If comparability of groups is based on a broad range of potential confounders or limited to commonly used demographic factors, how well have study investigators dealt with the key critical confounders?
- If intention-to-treat analysis was not specifically mentioned, was there other information (e.g., analysis descriptions) to suggest that it was or was not the approach used?
- When evaluating nutrient exposures, have the complexities and nuances of exposure assessments for different nutrients, different purposes, different study designs, and different statistical analyses been taken into account?
- If nutrient exposure is an intermediate factor (e.g., when education is the intervention), is it still possible to evaluate the exposure–outcome relation of interest?
- Has the rater maintained focus on the evaluation of the scientific quality of reviewed studies and not been distracted by the potentially trivial nature of a study topic? Trivial or not, does the study provide information that can be used to evaluate the key research question?

¹In many cases, ≥ 2 raters will perform the assessment. However, we recognize that in some cases, a single rater will conduct the review. In this latter case, adaptation of the “Points to Consider” may be appropriate. PICO, population, intervention/exposure, comparator, outcome.

We built on the existing properties of the SIGN instruments, which are valid, reliable, and have extensive use experience (26). The SIGN instruments have a long history of application and have performed well identifying studies of different quality over the years to the satisfaction of authors and users. The SIGN instruments provided a complementary suite of tools for evaluating 3 different study designs commonly used in nutrition studies on which a nutrition-specific instrument could be developed. We enhanced the clarity of the SIGN items by reorganizing them into sections focused on selection, comparability, and ascertainment, and supplemented the items by adding nutrition-related guidance. Nutrition-specific RoB items were added to recognize the nature and impact of the challenges when considering nutrition exposures. We offer suggestions in Boxes 1 and 2 and in the guidance and a worksheet template to elucidate topic-specific issues important for study assessment (see Supplemental File 6). We provide an example of a completed worksheet (Supplemental File 7) to illustrate its use in the context of a specific research question. We also include reminders in the guidance to consider both general and topic-specific criteria when rating studies. NUQUEST focuses on the extent to which nutrition studies were designed, conducted, analyzed, and reported to the highest possible methodological standards, but is not intended to quantify the magnitude of bias that may be attributable to methodological flaws.

Our nutrition-specific RoB tools can be used for assessing single studies or multiple studies included in an SR, the latter being similar to the adaptation of ROBINS-I for evaluation of nonrandomized studies of exposures (58). Realizing that some existing assessment instruments are challenging to use (59, 60), we aimed to make NUQUEST user-friendly for both epidemiology and nutrition users via an improved rating system, improved organization, and detailed nutrition-specific guidance. Our multidisciplinary raters felt that NUQUEST was feasible: assessments could be completed in a reasonable amount of time, time for completion improved with repeated use, and the inclusion of a completed worksheet facilitated study evaluation. Importantly, we wanted NUQUEST to be educational for users and to promote improvement in nutrition study design, conduct, and reporting.

Although SRs and meta-analyses are important for informing evidence-based nutrition policy, some nutrition scientists have expressed concern that there is potential for misuse if they are poorly conducted (61). Indeed, a recent SR and meta-analysis revealed that the vast majority of a large sample of nutrition SRs had serious methodological issues, including no RoB assessment of included studies and a lack of formal evaluation of the certainty of evidence (62). Given that most nutrition evidence comes from observational studies where exposure is based on self-reported intakes with systematic biases and random errors that can produce inaccurate and misleading results (4), it is imperative that nutrition-focused SRs are rigorous and that RoB is assessed in a consistent and transparent fashion. Our results showed that the effect of nutrition-specific criteria on overall ratings varied, with some ratings increasing and others decreasing when comparing NUQUEST to other instruments. The use of NUQUEST has several advantages for supporting SR development and nutrition guidance decisions, including clarity of the key research question, flexibility and transparency in developing the PICO and worksheet, and transparency of

the strengths and weaknesses of the relied-upon evidence. In addition, policy and guidance decisions are not solely dependent on the strength of the evidence but also consider other factors (e.g., public health impact, economics) (52). We believe that NUQUEST, with its emphasis on assessment of exposure data, can support future nutrition guidance initiatives.

A common challenge in assessing the RoB in nutrition studies, particularly for observational studies, is the failure of study authors to adequately describe the procedure used to establish the accuracy and appropriateness of their dietary exposure assessment. Researchers completing the NUQUEST worksheet will be responsible for establishing rating criteria for assessing the accuracy, reliability, and appropriateness of dietary intake in selected studies, as well as the appropriateness of the dietary methodology for meeting research objectives. This need can be partially addressed if journal editors require that authors follow reporting guidelines for dietary studies (2, 34, 35). Resources exist to aid in these decisions, such as the National Cancer Institute's Dietary Assessment Primer (17), which provides a comprehensive description of the available dietary assessment instruments, the key concerns in dietary assessment (i.e., measurement error and validation), and recommendations for choosing a dietary assessment approach for different research objectives. In addition, the evolving use of statistical procedures to improve the ability to estimate "true" intakes in nutritional epidemiology research (47) can also provide useful information for establishing NUQUEST worksheet criteria. The failure to document validation procedures or the use of validation procedures generally considered to increase the potential for RoB (e.g., repeatability properties of an exposure methodology) would likely result in lower nutrition-specific ratings; documentation of validation against an external "gold standard" (e.g., use of a recovery biomarker or controlled feeding trials), particularly when also combined with statistical adjustments for other dietary components and personal characteristics, would warrant higher ratings.

Although NUQUEST strengthens the ability to assess nutrition studies, these instruments have limitations. As is common practice, we recommend training on a set of studies to test topic-specific guidance and use of the worksheet, and discussing consensus on items among raters. The working group did not address issues of outcome specificity. However, in practice, any tool can be made outcome-specific, as can be done with the use of specific criteria in the outcome section of the worksheet, keeping in mind potential biases related to the ascertainment of each outcome of interest (63). Ultimately, even with the guidance and assessment criteria defined, user judgment will be required. We found that the inclusion of a worksheet improved interrater reliability and increased discussion among raters to achieve consensus. These limitations underscore the need for an interdisciplinary team, including experts with knowledge of study rating systems, study design, and nutrition. The effective application of NUQUEST will require the team to develop a priori rating criteria and estimate their relative impact on study RoB. Because there is no current gold-standard approach for the assessment of RoB in nutrition studies, the working group could not assess the criterion validity of NUQUEST. Instead, the exercise here represents an initial assessment of content, construct, and concurrent validity. Our results suggest NUQUEST performed well in comparison to other RoB tools

and that, where there was disagreement, the nutrition-specific items influenced the overall rating. Finally, the absence of a SIGN checklist for cross-sectional studies precluded the development of a NUQUEST for cross-sectional studies, at least for the moment. We recognize the need for such a tool given the plethora of cross-sectional nutrition studies. As such, NUQUEST for cross-sectional studies is being developed de novo and will be reported separately.

In conclusion, we developed a set of validated nutrition-specific RoB tools to assess nutrition studies of commonly used study designs. NUQUEST is grounded in the scientific concepts reported in a broadly used generic assessment instrument, while improving the usability for application to nutrition studies. These nutrition-specific RoB tools for RCTs, cohort studies, and case-control studies are intended for the evaluation of individual nutrition studies or studies included in SRs. In addition, NUQUEST will help researchers improve the design, conduct, and reporting of future nutrition research studies through its focus on nutrition-specific appraisal items. Finally, the availability of these new tools will foster consistency among future reviews of nutrition studies while enabling future systematic refinement, as warranted, by experience and as science evolves.

We gratefully acknowledge the SIGN for their permission to adapt the SIGN checklists in the development of the NUQUEST checklists. We thank Joan Peterson, Dr. Zemin Bai, Amy Johnston, Dr. Shu-Ching Hsieh, Dr. Matthew Parrot, Alexandre Martel, Dr. Jesse Elliott, and Cynthia Boudrault for their contributions as raters for the validation of the QAIs. We greatly appreciate Drs. Stephanie Atkinson, Alice Lichtenstein, and Maria Makrides for their time and insight on the nutrition-specific items and guidance, and/or individual study RoB.

The authors' responsibilities were as follows—AJM, EAY, GAW, and LSG-F: conceptualized the project; SEK: drafted the manuscript; AJM and GAW: had primary responsibility for the final content; and all authors: developed the research plan, conducted the research, analyzed the data, and contributed to the final manuscript. All authors read and approved the final manuscript. The authors report no conflicts of interest.

Data Availability

Data described in the article are publicly and freely available without restriction (available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at <https://academic.oup.com/ajcn/>).

References

1. Yetley EA, MacFarlane AJ, Greene-Finestone LS, Garza C, Ard JD, Atkinson SA, Bier DM, Carriquiry AL, Harlan WR, Hattis D, et al. Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. *Am J Clin Nutr* 2017;105(1):249S–85S.
2. Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, Lichtenstein AH, Yetley EA, Lau J. Reporting of systematic reviews of micronutrients and health: a critical appraisal. *Am J Clin Nutr* 2009;89(4):1099–113.
3. Lichtenstein AH, Yetley EA, Lau J. Application of systematic review methodology to the field of nutrition. *J Nutr* 2008;138(12):2297–306.
4. National Academies of Sciences, Engineering, and Medicine. Guiding principles for developing Dietary Reference Intakes based on chronic disease. Washington (DC): The National Academies Press; 2017.
5. National Academies of Sciences, Engineering, and Medicine. Dietary Reference Intakes for sodium and potassium. Washington (DC): The National Academies Press; 2019.
6. Trepanowski JF, Ioannidis JPA. Perspective: limiting dependence on nonrandomized studies and improving randomized trials in human nutrition research: why and how. *Adv Nutr* 2018;9(4):367–77.
7. Zeraatkar D, Johnston BC, Guyatt G. Evidence collection and evaluation for the development of dietary guidelines and public policy on nutrition. *Annu Rev Nutr* 2019;39:227–47.
8. Suh JR, Herbig AK, Stover PJ. New perspectives on folate catabolism. *Annu Rev Nutr* 2001;21:255–82.
9. Academy of Nutrition and Dietetics. Evidence analysis manual: steps in the Academy Evidence Analysis Process. Chicago (IL): Academy of Nutrition and Dietetics; 2016.
10. Newberry SJ, Chung M, Anderson CAM, Chen C, Fu Z, Tang A, Zhao N, Booth M, Marks J, Hollands S, et al. Sodium and potassium intake: effects on chronic disease outcomes and risks. Comparative Effectiveness Review, No. 206. Rockville (MD): Agency for Healthcare Research and Quality (US); 2018.
11. Guyatt G, Busse J. Methods commentary: risk of bias in randomized trials 1. [Internet]. Ottawa (Canada): Evidence Partners; 2021. [Accessed 2021 Oct 8]. Available from: <https://www.evidencepartners.com/resources/methodological-resources/risk-of-bias-commentary>.
12. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, Klassen TP. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339(7728):b4012.
13. Handel MN, Rohde JF, Jacobsen R, Nielsen SM, Christensen R, Alexander DD, Frederiksen P, Heitmann BL. Processed meat intake and incidence of colorectal cancer: a systematic review and meta-analysis of prospective observational studies. *Eur J Clin Nutr* 2020;74(8):1132–48.
14. Balk EM, Adams GP, Langberg V, Halladay C, Chung M, Lin L, Robertson S, Yip A, Steele D, Smith BT, et al. Omega-3 fatty acids and cardiovascular disease: an updated systematic review. *Evid Rep Technol Assess (Full Rep)* 2016;(223):1–1252.
15. Newberry SJ, Chung M, Shekelle PG, Booth MS, Liu JL, Maher AR, Motala A, Cui M, Perry T, Shanman R, et al. Vitamin D and calcium: a systematic review of health outcomes (update). *Evid Rep Technol Assess (Full Rep)* 2014;(217):1–929.
16. Newberry SJ, Chung M, Booth M, Maglione MA, Tang AM, O'Hanlon CE, Wang DD, Okunogbe A, Huang C, Motala A, et al. Omega-3 fatty acids and maternal and child health: an updated systematic review. *Evid Rep Technol Assess (Full Rep)* 2016;(224):1–826.
17. Thompson FE, Kirkpatrick SI, Subar AF, Reedy J, Schap TE, Wilson MM, Krebs-Smith SM. The National Cancer Institute's Dietary Assessment Primer: a resource for diet research. *J Acad Nutr Diet* 2015;115(12):1986–95.
18. Del Valle HB, Yaktine AL, Taylor CL, Ross AC. Dietary Reference Intakes for calcium and vitamin D. Washington (DC): National Academies Press; 2011.
19. European Food Safety Authority (EFSA). Outcome of public consultations on the Scientific Opinions of the EFSA Panel on Nutrition, Novel Foods and Food Allergens (NDA) on Dietary Reference Values for sodium and chloride. *EFSA Support Publ* 2019;16(9):1679E.
20. National Heart, Lung, and Blood Institute (NHLBI). Lifestyle interventions to reduce cardiovascular risk: systematic evidence review from the Lifestyle Work Group, 2013. Bethesda (MD): NHLBI; 2015.
21. World Health Organization (WHO). Evidence-informed guideline development for World Health Organization nutrition-related normative work: continuous quality improvement for efficiency and impact. Geneva (Switzerland): WHO; 2018.
22. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20(1):37–46.
23. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70(4):213–20.
24. Garrity C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, Affengruber L, Stevens A. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. *J Clin Epidemiol* 2021;130:13–22.
25. Tricco AC, Antony J, Zarin W, Striffler L, Ghassemi M, Ivory J, Perrier L, Hutton B, Moher D, Straus SE. A scoping review of rapid review methods. *BMC Med* 2015;13(1):224.
26. Bai A, Shukla VK, Bak G, Wells G. Quality Assessment Tools project report. Ottawa (Canada): Canadian Agency for Drugs and Technologies in Health; 2012.

27. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VS, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Med Res Method* 2004;4(1):22.
28. Lohr KN. Rating the strength of scientific evidence: relevance for quality improvement programs. *Int J Qual Health Care* 2004;16(1):9–18.
29. Sanderson S, Tatt ID, Higgins J. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36(3):666–76.
30. Whiting P, Davies PA, Savović S, Caldwell DM, Churchill R. Evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews. Bristol (UK): School of Social and Community Medicine, University of Bristol; 2013.
31. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, Petticrew M, Altman DG. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7(27):iii–x, 1–173.
32. Schwingshackl L, Knuppel 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(6):994–1004.
33. Yang C, Pinart M, Kolsteren P, Van Camp J, De Cock N, Nimptsch K, Pischon T, Laird E, Perozzi G, Canali R, et al. Perspective: essential study quality descriptors for data from nutritional epidemiologic research. *Adv Nutr* 2017;8(5):639–51.
34. Hörnell A, Berg C, Forsum E, Larsson C, Sonestedt E, Åkesson A, Lachat C, Hawwash D, Kolsteren P, Byrnes G, et al. Perspective: an extension of the STROBE statement for observational studies in nutritional epidemiology (STROBE-nut): explanation and elaboration. *Adv Nutr* 2017;8(5):652–78.
35. Lachat C, Hawwash D, Ocké MC, Berg C, Forsum E, Hörnell A, Larsson C, Sonestedt E, Wirfält E, Åkesson A, et al. Strengthening the Reporting of Observational Studies in Epidemiology—Nutritional Epidemiology (STROBE-nut): an extension of the STROBE statement. *PLoS Med* 2016;13(6):e1002036.
36. Farrah K, Young K, Tunis MC, Zhao L. Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols. *Syst Rev* 2019;8(1):280.
37. Charlton P, Doucet S, Azar R, Nagel DA, Boulos L, Luke A, Mears K, Kelly KJ, Montelpare WJ. The use of the environmental scan in health services delivery research: a scoping review protocol. *BMJ Open* 2019;9(9):e029805.
38. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L. Systems to rate the strength of scientific evidence. *Evid Rep Technol Assess (Summ)* 2002;(47):1–11.
39. Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50 methodology. Critical appraisal: checklists. Checklist 2: randomised controlled trials, methodology. Checklist 3: cohort studies, methodology. Checklist 4: case-control studies. Edinburgh (United Kingdom): SIGN; 2014 [Internet]. Available from: <https://www.sign.ac.uk/what-we-do/methodology/checklists/>.
40. Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50 methodology. Critical appraisal: notes. Edinburgh (United Kingdom): SIGN; 2014 [Internet]. Available from: <https://www.sign.ac.uk/what-we-do/methodology/checklists/>.
41. Joanna Briggs Institute (JBI). Critical appraisal checklist for randomized controlled trials. Adelaide (Australia): JBI; 2020.
42. Joanna Briggs Institute (JBI). Critical appraisal checklist for cohort studies. Adelaide (Australia): JBI; 2020.
43. Joanna Briggs Institute (JBI). Critical appraisal checklist for case control studies. Adelaide (Australia): JBI; 2020.
44. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, Midthune D, Moshfegh AJ, Neuhaus ML, Prentice RL, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol* 2014;180(2):172–88.
45. Freedman LS, Commins JM, Moler JE, Willett W, Tinker LF, Subar AF, Spiegelman D, Rhodes D, Potischman N, Neuhaus ML, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. *Am J Epidemiol* 2015;181(7):473–87.
46. Lichtenstein AH, Petersen K, Barger K, Hansen KE, Anderson CAM, Baer DJ, Lampe JW, Rasmussen H, Matthan NR. Perspective: design and conduct of human nutrition randomized controlled trials. *Adv Nutr* 2021;12(1):4–20.
47. Zheng C, Beresford SA, Van Horn L, Tinker LF, Thomson CA, Neuhaus ML, Di C, Manson JE, Mossavar-Rahmani Y, Seguin R, et al. Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women. *Am J Epidemiol* 2014;180(5):526–35.
48. NIH National Cancer Institute (NCI). Dietary assessment primer. [2021 Internet]. Bethesda (MD): NCI. [Accessed 2021 Oct 28]. Available from: <https://dietassessmentprimer.cancer.gov/>.
49. Houghton LA, Gray AR, Rose MC, Miller JC, Hurthouse NA, Gregory JF 3rd. Long-term effect of low-dose folic acid intake: potential effect of mandatory fortification on the prevention of neural tube defects. *Am J Clin Nutr* 2011;94(1):136–41.
50. Nguyen P, Tam C, O'Connor DL, Kapur B, Koren G. Steady state folate concentrations achieved with 5 compared with 1.1 mg folic acid supplementation among women of childbearing age. *Am J Clin Nutr* 2009;89(3):844–52.
51. Tighe P, Ward M, McNulty H, Finnegan O, Dunne A, Strain J, Molloy AM, Duffy M, Pentieva K, Scott JM. A dose-finding trial of the effect of long-term folic acid intervention: implications for food fortification policy. *Am J Clin Nutr* 2011;93(1):11–8.
52. Johnston BC, Seivenpiper JL, Vernooij RWM, de Souza RJ, Jenkins DJA, Zeraatkar D, Bier DM, Guyatt GH. The philosophy of evidence-based principles and practice in nutrition. *Mayo Clin Proc Innov Qual Outcomes* 2019;3(2):189–99.
53. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JPT, Thomas J. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Database Syst Rev* 2019;10:ED000142.
54. Johnston BC, Alonso-Coello P, Bala MM, Zeraatkar D, Rabassa M, Valli C, Marshall C, El Dib R, Vernooij RWM, Vandvik PO, et al. Methods for trustworthy nutritional recommendations NutriRECS (Nutritional Recommendations and accessible Evidence summaries Composed of Systematic reviews): a protocol. *BMC Med Res Method* 2018;18(1):162.
55. EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, et al. Guidance on the use of the weight of evidence approach in scientific assessments. *EFSA J* 2017;15(8):e04971.
56. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924–6.
57. 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 Method* 2019;19(1):178.
58. Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, Goldstone AE, Ross P, Ansari M, Akl EA, et al. 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;122:168–84.
59. Minozzi S, Cinquini M, Gianola S, Castellini G, Gerardi C, Banzi R. Risk of bias in nonrandomized studies of interventions showed low inter-rater reliability and challenges in its application. *J Clin Epidemiol* 2019;112:28–35.
60. Thomson H, Craig P, Hilton-Boon M, Campbell M, Katikireddi SV. Applying the ROBINS-I tool to natural experiments: an example from public health. *Syst Rev* 2018;7(1):15.
61. Tobias DK, Wittenbecher C, Hu FB. Grading nutrition evidence: where to go from here? *Am J Clin Nutr* 2021;113(6):1385–7.
62. Zeraatkar D, Bhasin A, Morassut RE, Churchill I, Gupta A, Lawson DO, Miroshnychenko A, Siroitch E, Aryal K, Mikhail D, et al. 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.
63. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, Murad MH, Treadwell JR, Kane RL. Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. *J Clin Epidemiol* 2018;97:26–34.