Nutrition users' guides: systematic reviews part 1 –structured guide for methodological assessment, interpretation and application of systematic reviews and meta-analyses of non-randomised nutritional epidemiology studies

Dena Zeraatkar,[1,2] Russell J de Souza,[1] Gordon H Guyatt,[1,3] Malgorzata M Bala,[4] Pablo Alonso-Coello,[5,6,7] Bradley C Johnston [ID] [8,9,10]

ABSTRACT

Due to the challenges of conducting randomised controlled trials (randomised trials) of dietary interventions, evidence in nutrition often comes from non-randomised (observational) studies of nutritional exposures—called nutritional epidemiology studies. When using systematic reviews of such studies to advise patients or populations on optimal dietary habits, users of the evidence (eg, healthcare professionals such as clinicians, health service and policy workers) should first evaluate the rigour (validity) and utility (applicability) of the systematic review. Issues in making this judgement include whether the review addressed a sensible question; included an exhaustive literature search; was scrupulous in the selection of studies and the collection of data; and presented results in a useful manner. For sufficiently rigorous and useful reviews, evidence users must subsequently evaluate the certainty of the findings, which depends on assessments of risk of bias, inconsistency, imprecision, indirectness, effect size, dose-response and the likelihood of publication bias. Given the challenges of nutritional epidemiology, evidence users need to be diligent in assessing whether studies provide evidence of sufficient certainty to allow confident recommendations for patients regarding nutrition and dietary interventions.

WHAT IS A SYSTEMATIC REVIEW AND META-ANALYSIS?

A systematic review identifies, evaluates and summarises the findings of all relevant primary studies addressing a particular question. It is often, but not always, accompanied by a meta-analysis, which is the statistical pooling of data across studies to produce a single effect estimate.

Systematic reviews and meta-analyses have become foundational to evidence-based clinical practice. For example, standards for producing trustworthy guidelines require all recommendations be based on rigorous and comprehensive systematic reviews of the evidence.[1] Failing to base clinical care (box 1) and guideline recommendations on systematic reviews risks overlooking important evidence, selectively choosing favourable evidence or neglecting important limitations of the evidence.

There are two aspects to evaluating a systematic review and meta-analysis (SRMA), both within and outside of nutritional epidemiology (box 2). The first has to do with the methodological rigour (validity) and utility (applicability) of SRMA. These issues are discussed in the first half of this manuscript. The second set of issues addresses the certainty (quality) of evidence summarised in the SRMA. These issues are discussed in the second half of the manuscript. While a poorly conducted SRMA is of little use and may be misleading if the body of evidence has important limitations, a well-conducted SRMA may only provide low or very low certainty evidence, evidence that is still important for transparent and informed clinical decision-making.

This article describes considerations in evaluating and applying SRMAs in nutrition and draws from the JAMA Users' Guides series.[2] Given that SRMAs of non-randomised studies (eg, cohort, case–control studies) addressing nutritional exposures—often referred to as nutritional epidemiology studies—require very unique considerations,[3] the primary

Box 1   Clinical scenario

You are a family doctor (general practitioner) following a healthy 45-year-old female patient. She reports having recently come across a series of news articles on the alleged adverse health effects of red and processed meat. She is concerned about her cardiovascular health and her risk for cancer and inquires whether reducing her intake of red and processed meat (currently at four and two servings/week, respectively) is important. She also indicates that her meat consumption is mostly from local sources using sustainable (regenerative), ethical and humane farming practices. You recall having read a series of systematic review articles on this topic of meat and health outcomes. You ask the patient to return in 2 weeks for further advice.

You formulate the relevant clinical question: in healthy middle-aged adults, does the reduction of red and processed meat intake reduce the risk of adverse cardiovascular and cancer health outcomes? You refer to the series of articles you recall having read on the topic—a guideline and a series of supporting systematic reviews. The guideline, published by an international consortium of methodologists, nutrition researchers and clinicians, called NutriRECS, provides a weak recommendation for adults to continue their current levels of red and processed meat consumption.[46] The systematic reviews summarise the evidence from randomised trials,[42] cohort studies reporting on cardiovascular outcomes,[43] cohort studies reporting on cancer outcomes[44] and cohort studies reporting on the effects of dietary patterns low or high in red and processed meat on cardiovascular and cancer outcomes.[45]

focus of the first of two articles on nutrition systematic reviews will focus on synthesising non-randomised studies. Part 2 will focus on systematic reviews of randomised clinical trials.

EVALUATE THE VALIDITY AND APPLICABILITY OF THE SYSTEMATIC REVIEW

In using the results of an SRMA to guide clinical or public health decisions, evidence users will need to judge the validity and applicability of the SRMA. The validity of

Box 2   Guide for evaluating validity, applying the results and assessing the certainty of evidence of a systematic review and meta-analysis

Evaluate validity and applicability of the systematic review
1. Did the review explicitly address a relevant question?
2. Were methods for identifying and selecting studies and collecting data sufficiently rigorous?
3. Did the review appropriately synthesise data and report results that are ready for application?

Assess the certainty (quality) of the evidence
4. How serious is the risk of bias in the body of evidence?
5. Are the results consistent across studies?
6. How precise are the results?
7. Do the results directly apply to my patient?
8. Is there concern about publication bias?
9. Are there reasons to be more certain of findings based on effect size, credible dose-response gradient and/or direction of plausible confounders?

Box 3   Substitution models

A review addressing sugar-sweetened beverages, for example, summarised results from substitution models to estimate the effects of replacing sugar-sweetened beverages with other beverage alternatives (eg, water, milk, juice, coffee).[52] The review found the substitution of sugar-sweetened beverages with other beverage alternatives may have a positive effect on body weight and composition, but this positive effect is likely larger for the replacement of sugar-sweetened beverages with water.

an SRMA may be undermined by limitations such as an incomprehensive search for eligible studies or inappropriate eligibility criteria[2] and its applicability by issues such as addressing an irrelevant question or failing to synthesise data and report results that are ready for application.[2]

1. Did the review address a relevant question?

The first step in addressing the applicability of an SRMA of nutritional epidemiology studies is assessing whether authors have stated explicit eligibility criteria that specify the population, exposure, comparator and outcome(s) of interest (frequently referred to as PECO—a variation of PICO used for therapeutic interventional studies). In nutritional epidemiology, the exposure may be a food, food compound (eg, micronutrient, macronutrient, bioactive compound) or dietary pattern. Unlike SRMAs of therapeutic interventions, where an intervention is compared with an alternative intervention (or standard care), SRMAs of nutritional exposures typically compare individuals with a higher intake of exposure (or higher adherence to a dietary pattern) to those with lower intake (or lower adherence to a dietary pattern).

Box 4   Dose-response meta-analyses from NutriRECS reviews of cohort studies

The NutriRECS reviews of cohort studies[43 44] present results from dose-response meta-analyses corresponding to a reduction of red and processed meat intake by three servings/week. For analyses for which a non-linear relationship is observed, the reviews present graphs relating the quantity of intake with the relative risk for the outcome (figure 1). Evidence users can convert effect estimates from linear dose-response meta-analyses to correspond to different quantities of intake by multiplying the natural logarithm of the point estimate and the CIs by a conversion factor (eg, the conversion factor for converting three servings/week to five servings/week is 5/3) and subsequently exponentiating the results (table 1). For example, to convert the relative risk (RR) of cardiovascular mortality (0.90) corresponding to a reduction of three servings/week of unprocessed red meat to five servings/week, evidence users could use the following procedure: $RR = e^{lnln\ (0.90) \times 5/3}$

Below, figure 1 presents the relative risk of type 2 diabetes corresponding to various servings of processed meat intake per week compared with zero servings/week, derived from a non-linear model. Since these results are derived from a non-linear model, the magnitude of change in the risk of type 2 diabetes will depend on the baseline intake of processed meat.
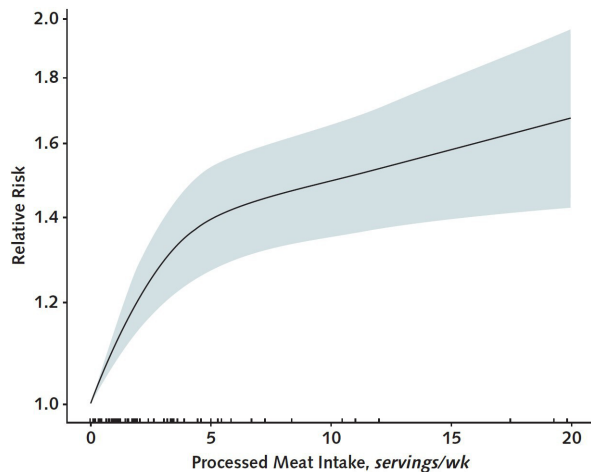
Figure 1 Non-linear effect of processed meat on type 2 diabetes. The solid black line represents the point estimate, the shaded region represents the 95% CIs and the tick marks represent the positions of the study-specific estimates.

For an SRMA to be optimally useful, it should account for the foods or food compounds that are consumed instead of the exposure of interest, which may also impact the risk for the outcome under study. Investigators may address this issue by summarising results from substitution models or joint analyses, analytical approaches to estimating the effects of the substitution of one food or food compound for another or the joint effects of two or more exposures (box 3).[4] Despite addressing a more specific causal question, substitution models also come with limitations. They are statistical exercises and do not actually account for how people in the real world adapt their diet to accommodate a dietary modification. Foods and diets are also complex mixtures and substitution models may oversimplify their interactions. These models are also subject to the same limitations as other nutrition models, which are described later in this manuscript.

SRMAs of observational studies addressing the impact of alternative diets are almost always concerned with the effect of those diets—that is, they address their causal impact on patient-important outcomes. Given that causation is the issue at hand, authors should use causal language—though they should, they may not and this is potentially confusing.[5] While many SRMAs of diet exposures or interventions do not conduct or report on the certainty of evidence,[6] it will be up to evidence users to discern the extent to which SRMA results indeed support causal inference (ie, based on Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidance provide moderate, low or very low certainty evidence—they will very seldom provide high certainty evidence) on an outcome by outcome basis.

In this manuscript, to align with their causal intent we use causal language to refer to estimates from SRMAs of observational studies. However, we qualify causal statements with recommended GRADE language to express uncertainty (eg, 'likely' or 'probably' for moderate certainty evidence; 'possibly' or 'may' for low certainty; and 'very uncertain' for very low certainty evidence).[7]

2. Were methods for identifying and selecting studies and collecting data sufficiently rigorous?

SRMAs are at risk of presenting misleading results if they fail to include all eligible studies. For most questions, SRMAs that search MEDLINE and EMBASE, or databases with similar coverage, likely include all or nearly all relevant published studies. For some questions, however, MEDLINE and EMBASE may not be sufficient and it is difficult to know in advance whether a more extensive search is necessary.[8 9] Hence, reviews should always conduct an exhaustive search of the literature that includes the following: study registries, bibliographies of included studies, doctoral theses (eg, EBSCO Open Dissertations, ProQuest, BIOSIS), abstracts from relevant scientific meetings and soliciting experts for relevant published and unpublished studies.[8 10] Further, SRMAs should reduce the opportunity for biases and errors in the selection of studies and the collection of data by ensuring that these processes are conducted independently and in duplicate.

3. Did the review appropriately synthesise and report results that are ready for application?

When using an SRMA to inform inferences about the effect of the intake of a food or food compound on the risk for a health outcome, evidence users should look for results from a dose-response meta-analysis that summarises the relationship between the quantity of the exposure and the risk for the outcome.[11 12] Evidence users should also pay close attention to the quantity of the exposure corresponding to which results are presented and judge whether the quantity is reasonable (box 4). A review may, for example, present the risk difference for adverse cardiovascular outcomes associated with 5 servings/day or 20 servings/day of fruits and vegetables; the latter may be a larger and more impressive effect but may

Table 1 Relative effect estimates for cardiovascular mortality corresponding to different quantities of intake of unprocessed red and processed meat from a dose-response meta-analysis

Exposure	Reduction of three servings/week RR (95%CI)	Reduction of five servings/week RR (95%CI)	Reduction of seven servings/week RR (95%CI)
Unprocessed red meat	0.90 (0.88 to 0.91)	0.84 (0.81 to 0.85)	0.78 (0.74 to 0.80)
Processed meat	0.90 (0.84 to 0.97)	0.84 (0.75 to 0.95)	0.78 (0.67 to 0.93)

RR, relative risk.

Box 5   Absolute effects from NutriRECS reviews of cohort studies

The NutriRECS systematic review and meta-analyses of cohort studies present absolute effect estimates corresponding to a reduction of red and processed meat of three servings/week,[43 44] a reduction the authors thought might be feasible for most people. Population risks for cardiometabolic outcomes were sourced from the Emerging Risk Factors Collaboration, a consortium of over 100 cohorts, primarily from North America and Western Europe, that includes mostly middle-aged to older adults who were omnivores, while for cancer outcomes NutriRECS used GLOBOCAN, a repository of risk estimates from national cancer registries. Table 2 presents the absolute effects of cardiovascular mortality, type 2 diabetes and cancer mortality.[46]

---

not be reasonable because few studies, if any, reported on the effects of 20 servings/day and the review likely extrapolated the effect from lower quantities of intake. Further, consuming 20 servings/day of fruits and vegetables is not practical for most people.

The dose-response effect of nutritional exposure on a health outcome may be linear or non-linear, both of which may be modelled using dose-response meta-analysis. Evidence users can convert effect estimates from linear dose-response meta-analysis to correspond to different quantities of the exposure by simple algebraic formulae. For non-linear dose-response meta-analysis, SRMAs typically present a graph that depicts the relationship between the exposure and outcome, which evidence users can use to approximate effects corresponding to different quantities of the exposure.

SRMAs may also present results from meta-analyses comparing extreme categories of exposure (eg, a meta-analysis comparing the incidence of an outcome in the highest quantile of intake of a food or food compound to the incidence in the lowest quantile). Despite

challenges with their interpretation, such analyses remain common.[6 13] For exposures for which there is no comparable measure of consumption or adherence across studies, such as a posteriori derived dietary patterns (ie, dietary patterns that are derived by methods like factor analysis and principal component analysis), meta-analyses comparing extreme quantiles may be the only feasible approach to pool data. Because such analyses are difficult to interpret, evidence users should be cautious when SRMAs present only comparisons of extreme categories of exposure.[14 15]

When reporting SRMA results, ideally, authors should present absolute effects—the effect of an exposure expressed as the rate of the outcome (eg, risk differences, number needed to treat or harm) rather than the relative effects (ie, relative risk, OR, HR) alone. SRMAs most often meta-analyse relative effects because they tend to be similar across populations (while also providing more compelling [exaggerated] results).[16] Absolute effects, however, are essential for realistic and intuitive decision-making and without them evidence users cannot help their patients make rational decisions.[17 18] When absolute effects are presented, evidence users should judge whether the population risk estimates used for their calculation are sufficiently similar to the baseline risk of the outcome in their patient.[13] Unfortunately, SRMAs of nutritional epidemiology studies seldom present absolute effects.[6] When absolute effects are not presented, evidence users can calculate them using population risk estimates (box 5).[13 19]

To calculate absolute effects, evidence users can multiply the relative risk (including the lower and upper CIs) by the population risk. For reviews that present results as ORs or HRs, users can calculate absolute effects using alternative formulae:[19 20]

Table 2   Relative and absolute effect estimates for cardiovascular mortality, type 2 diabetes and cancer mortality

Exposure/Outcome	Relative risk from dose-response meta-analysis corresponding to a reduction of three servings/week (95%CI)	Population risk	Risk difference corresponding to a reduction of three servings/week (95%CI)
Unprocessed red meat			
Cardiovascular mortality	0.90 (0.88 to 0.91)	41/1000 over 10.8 years	4 fewer per 1000 persons (5 fewer to 4 fewer) over 10.8 years
Type 2 diabetes	0.90 (0.88 to 0.92)	56/1000 over 10.8 years	6 fewer per 1000 persons (7 fewer to 4 fewer) over 10.8 years
Cancer mortality	0.93 (0.91 to 0.94)	105/1000 over lifetime	7 fewer per 1000 persons (9 fewer to 6 fewer) over lifetime
Processed meat			
Cardiovascular mortality	0.90 (0.84 to 0.97)	41/1000 over 10.8 years	4 fewer per 1000 persons (7 fewer to 1 fewer) over 10.8 years
Type 2 diabetes	0.78 (0.72 to 0.84)	56/1000 over 10.8 years	12 fewer per 1000 persons (16 fewer to 9 fewer) over 10.8 years
Cancer mortality	0.92 (0.89 to 0.94)	105/1000 over lifetime	8 fewer per 1000 persons (12 fewer to 6 fewer) over lifetime

---

Box 6 Risk of bias in NutriRECS reviews of cohort studies

In the NutriRECS systematic reviews with meta-analyses (SRMAs) of cohort studies, the risk of bias is assessed using ad hoc criteria instead of a known risk of bias tools (eg, Ottawa-Newcastle instrument) due to the limitations of such tools at the time of NutriRECS publications. The ad hoc criteria addressed confounding, the selection of participants, measurement of the exposure, errors in the measurement of the outcome and missing outcome data.[43–45 53]

Reviews that use ad hoc criteria to assess risk of bias rather than established risk of bias tools may neglect important considerations of risk of bias. The ad hoc risk of bias criteria used in the NutriRECS SRMAs of cohort studies, for example, do not include selective reporting. The discussion section of the SRMA report, however, acknowledges that all studies are at risk of selective reporting bias due to the lack of standard practices for the registration of protocols and statistical analysis plans of nutritional epidemiology studies.

In the application of GRADE, the NutriRECS SRMAs rated down the certainty of evidence when the evidence came primarily from studies at high risk of bias. For example, the dose-response evidence on unprocessed red meat and cardiovascular mortality came from seven cohort studies, four of which were at high risk of bias due to the lack of periodically repeated measurement of diet and inadequate adjustment for confounders.[43] Given these issues, the authors rated the evidence as very low certainty.

$$Absolute\ risk\ from\ odds\ ratio\ \left(OR\right) = \frac{OR \times population\ risk}{1 - population\ risk + OR \times population\ risk}$$

$$Absolute\ risk\ from\ hazard\ ratio\ \left(HR\right) = 1 - \left(1 - population\ risk\right)^{HR}$$

RATE THE CERTAINTY (QUALITY) OF THE EVIDENCE

While results from a body of evidence may appear highly effective, our certainty in the body of evidence may be undermined by issues such as limitations in study designs or differences between the questions addressed in studies and the clinical or public health question of interest. Hence, optimal decision-making also requires consideration of the certainty (quality) of evidence. Several approaches to evaluate the certainty of evidence are available.[21] One such system is the GRADE approach.[22] The GRADE approach is by far the most commonly used system to evaluate the certainty of evidence—is it used by over 100 organisations worldwide, including the WHO, the Cochrane Collaboration and the American Academy of Nutrition and Dietetics and has evolved into a benchmark for the validity of reviews and guidelines. The GRADE approach intends to improve the interpretation of evidence for healthcare decision-making based on a common and transparent standard.[22]

Original GRADE guidance proposed that for questions of causal inference, a body of evidence comprised of randomised trials starts at high certainty and non-randomised studies start at low certainty.[23] This is because, randomised trials, by virtue of randomisation, may achieve balance (more or less) in both known and unknown prognostic factors such that any observed

---

Box 7 Example of inconsistency

A systematic review and meta-analysis addressing the association between vegetarian diets and metabolic syndrome reports on five cross-sectional studies and calculates a pooled OR of 0.96 (95% CI 0.50 to 1.85).[54] Figure 2 shows the forest plot for the meta-analysis of these studies. The results of each study are represented by squares with horizontal lines depicting 95% CIs. Studies with results to the left of the vertical line suggest that vegetarian diets reduce the risk of metabolic syndrome. Conversely, studies to the right favour an omnivorous diet. Three studies show a moderate to large effect favouring a vegetarian diet whereas two studies show a moderate to large effect favouring an omnivorous diet and there is little overlap in the CIs of these two groups of studies. The I^2 statistic indicates substantial heterogeneity (I^2=85%). In such a situation, we are less certain of the pooled effect estimate due to the inconsistency of the results of primary studies.

differences in outcomes between randomised arms can be more confidently attributed to the intervention under investigation.[24] In non-randomised studies, however, participants with and without exposure may differ with regard to prognostic factors such that any differences in outcomes between participants may be an artefact of differences in other prognostic factors.[24] Even if investigators use sophisticated design and analytical methods to adjust for a comprehensive list of prognostic factors, important factors that are unknown or unmeasured may still influence results. This phenomenon is called residual confounding and is the reason why non-randomised studies are initially rated at low certainty.

Newer GRADE guidance now suggests that a body of evidence comprised of non-randomised studies can also start at high certainty and the certainty of evidence may be downgraded by considering limitations of the evidence in comparison to a 'target trial'—a hypothetical trial, without any limitations, that may or may not be feasible, addressing the question of interest.[25] Using this approach, however, a body of evidence comprised of non-randomised studies will almost always still land at low or very low certainty due to concerns with residual confounding.

The certainty of a body of evidence may be rated down by one or more levels due to concerns related to five factors: risk of bias (ie, study limitations that may lead to systematic underestimation or overestimation), inconsistency (ie, unexplained heterogeneity in results across studies), indirectness (ie, differences between the questions addressed in studies and the clinical or public health question of interest), imprecision (ie, number of events or participants and the magnitude of CIs around an estimate in relation to the minimum difference in the outcome that patients or the target population finds important) and publication bias (ie, the tendency for studies with statistically significant results or positive results to be published, published faster or published in journals with higher visibility). The certainty of a body of non-randomised studies may also be rated up in select scenarios: when there is a valid dose-response
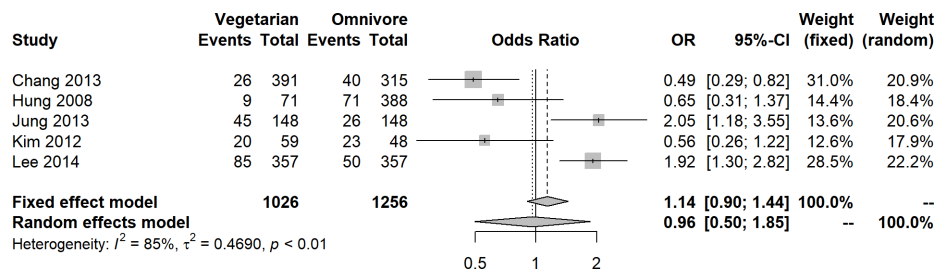
Study	Vegetarian Events	Total	Omnivore Events	Total	Odds Ratio	OR	95%-CI	Weight (fixed)	Weight (random)
Chang 2013	26	391	40	315		0.49	[0.29; 0.82]	31.0%	20.9%
Hung 2008	9	71	71	388		0.65	[0.31; 1.37]	14.4%	18.4%
Jung 2013	45	148	26	148		2.05	[1.18; 3.55]	13.6%	20.6%
Kim 2012	20	59	23	48		0.56	[0.26; 1.22]	12.6%	17.9%
Lee 2014	85	357	50	357		1.92	[1.30; 2.82]	28.5%	22.2%
Fixed effect model		1026		1256		1.14	[0.90; 1.44]	100.0%	--
Random effects model						0.96	[0.50; 1.85]	--	100.0%

Heterogeneity: $I^2 = 85\%$, $\tau^2 = 0.4690$, $p < 0.01$

Figure 2 Forest plot showing the pooled results of five cross-sectional studies on the association between.

relationship, a large effect (eg, relative risk <0.5or >2.0), or when all plausible confounders act in the opposite direction than the observed effect.[26] A description of these issues follows.

GRADE, initially critiqued for its limited applicability to environmental epidemiology, has broadened its scope. For example, it has established the Environmental and Occupational Health Project Group, which has published guidance on the application of GRADE to environmental and occupational hazards, which also apply to nutritional epidemiology.[3 27–30]

Below, we describe considerations in assessing the certainty of evidence presented by SRMAs. If an SRMA assesses the certainty of evidence using these considerations, evidence users will need to consider whether they agree with the judgements presented by the authors. Conversely, if an SRMA does not present an assessment of the certainty of evidence, evidence users can make judgements about the certainty considering the aforementioned criteria, while keeping in mind that study results that firmly support a causal inference statement should typically be accompanied by high certainty evidence based on the GRADE approach.

4. How serious is the risk of bias in the body of evidence?

Users of the evidence should be cautious about applying evidence from SRMAs when most of the evidence comes from studies that are at high risk of bias. A well-conducted review will assess the risk of bias of primary studies—though based on one systematic survey, most SRMAs of nutritional epidemiology studies do not use appropriate

and comprehensive criteria and so evidence users should be cautious about accepting the authors' interpretation of the degree of risk of bias without further considerations.[31]

Bias in non-randomised studies may arise due to confounding bias, inappropriate criteria for the selection of participants, errors in the measurement of the exposure, missing data, errors in the measurement of the outcome and selective reporting of results.[32 33] To our knowledge, until recently, there was no risk of bias tools that comprehensively accounted for these biases. The ROBINS-E (Risk Of Bias In Non-Randomized Studies - Exposure) tool, a new risk-of-bias tool for non-randomised studies of exposures, improves on previous risk-of-bias tools by addressing these sources of bias (box 6).[32]

In nutritional epidemiology studies, confounding and bias related to measurement of the exposure represent major areas of concern. Confounders are variables that are correlated with the exposure of interest and have a causal effect on the outcome of interest. Confounding occurs when there are differences in confounding variables between exposure groups. If those who consume more red meat, for example, are more likely to be obese and exercise less, results may appear to show an increase in cardiovascular disease with increased consumption of red meat, which may actually be due to these other effect-modifying variables. Investigators deal with confounding by restricting the study sample to ≥1 levels of confounding factors, matching or statistical methods, such as stratified analyses or adjustment through regression models.[34] At minimum, evidence users should ensure that primary studies control for age, sex, smoking and socioeconomic status. When studies included in the review do not control for these variables, confounding is highly probable. Further, it must be noted that adjusting for factors like smoking, which vary extensively by person-year or decade, is a crude adjustment at best.

Depending on the question of interest, adjustment for other confounding variables will also be necessary. The confounders that are adjusted for in studies are typically presented in a table of study characteristics (this is usually the first table of the SRMA). Users of the evidence should carefully review the confounders for which primary studies adjust and ensure that studies adjust for all suspected confounders.

Even when studies control for all suspected confounders, however, the problem that remains is that studies will yield unbiased results only if all confounders

Box 8   Judgements of Imprecision in NutriRECS reviews of cohort studies

The NutriRECS systematic review and meta-analysis (SRMA) of cohort studies addressing cancer reports a relative risk of 0.99 (95% CI 0.89 to 1.09) among people who consume three fewer servings/week of processed meat and overall cancer incidence.[44] Assuming a lifetime population risk of 185 per 1000 people, the reduction of processed meat intake is associated with 2 fewer cases of cancer in 1000, with CIs ranging from 20 fewer to 17 more. The authors of the SRMA considered the results imprecise because they anticipated that people may consider reducing their intake of processed meat if it would prevent 20 cancers in 1000 people followed over a lifetime (the a priori minimal threshold identified by the guideline panel), but would not reduce their intake if it would result in an additional 17 cases of cancer.

are known—which they seldom are—and if known confounders are measured accurately and can easily be adjusted for (eg, biological sex). Because of this inevitable uncertainty, in most cases, SRMAs of nutritional epidemiology studies will yield only low certainty evidence.

A second issue is the measurement of dietary exposures. Users of the evidence should be cautious of reviews that primarily include studies that measure diet using memory recall-based instruments as these measures are subject to serious limitations.[35] Evidence users can typically express more certainty in the results of SRMAs of studies that measure diet using weighted dietary records or established biomarkers (eg, urine sodium excretion, adipose tissue fatty acids). While we are not yet aware of other reliable and valid dietary measures, future advancements and technologies may eventually facilitate improved dietary measurement.[35]

5. Are the results consistent across studies?

When results across primary studies in an SRMA are inconsistent, we are less certain of the findings. Evidence users can look for inconsistency by visually inspecting a forest plot for large differences in point estimates or CIs that do not overlap. Users may also gauge inconsistency by looking for statistical indicators of heterogeneity, such as the I^2 statistic (a summary of the magnitude of heterogeneity that ranges from 0% to 100%). A high I^2 statistic (usually $I^2 > 50\%$ or $> 60\%$) suggests that there is substantial heterogeneity
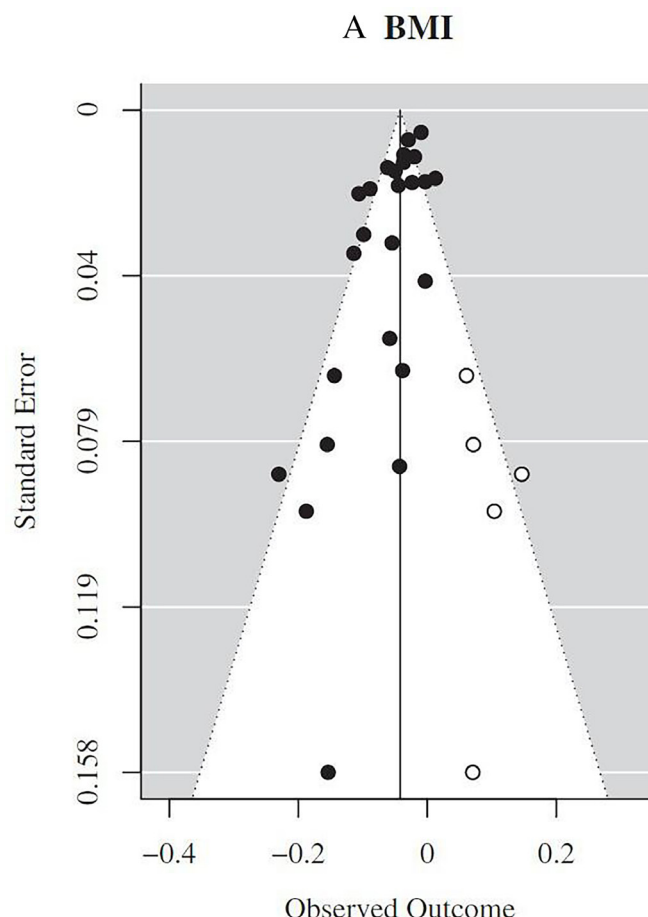
Figure 3   A funnel plot to help assess publication bias from a systematic review and meta-analysis on the relationship between the frequency of family meals and children's BMI. White circles represent imputed studies. BMI, body mass index.

(box 7). Statistical methods for detecting heterogeneity, however, have important limitations. For example, the I^2 value is prone to misinterpretation since even small degrees of unimportant inconsistency may translate to high I^2 values if estimates from studies are highly precise.[36]

When deciding whether results are consistent, evidence users should not rely completely on statistical indicators of heterogeneity and should instead judge whether differences in the results of studies, if any, would be important to patients or target populations. Evidence users may be concerned that SRMAs include studies too variable in participants to yield useful insights. They need not be too concerned, however. If study participants differ widely and results are nevertheless similar across studies, it tells us that individuals with different characteristics are likely to respond similarly to the intervention, allowing the application of the results to a wide range of patients or populations.[37] Evidence users might also encounter heterogeneity that is not clinically important. For example, if all studies suggest a dietary exposure is protective against an adverse health outcome but

Box 11 Clinical scenario resolution

We now return to our opening clinical scenario. Although the reviews do not address the effects associated with replacing red or processed meat with particular alternative foods, the reviews did have applicability, addressing a relevant question: the effects of reducing red and processed meat intake on adverse cardiovascular and cancer health outcomes. The systematic review and meta-analyses are methodologically rigorous (valid): they followed a priori specified methods; included an exhaustive search for all relevant studies; performed screening and extraction of data in duplicate; and used appropriate methods for the quantitative synthesis of results across studies, including dose-response meta-analyses.

The authors apply the Grading of Recommendations, Development and Evaluation approach to assess the certainty of evidence. Randomised trials provided low to very low certainty evidence that diets lower in red meat may result in some small reductions in adverse cardiovascular and cancer health outcomes.[42] Cohort studies sometimes provided low certainty evidence (if adequately adjusted, with consistent results) and sometimes very low certainty (certainty rated down for risk of bias concerns of inadequate adjustment for confounding and lack of periodic repeated measurement of diet, inconsistency and imprecision).[42 44 45]

You present the results to your patient (table 3 shows results of select outcomes from the NutriRECS systematic reviews of cohort studies[43–45]). Given the uncertainty (low to very certainty evidence) and small to very small magnitude of any benefit of reducing consumption (by three servings/week) that might exist, the patient considers the inconvenience and reduction in the pleasure of eating not worth the possible benefits and chooses to continue her current levels of red and processed meat consumption, meats that come from local, regenerative and ethical farming practices.

in subgroups, whether the studies involve within or between study comparisons (ie, if the hypothesis is the effect differs in men and women, whether men and women were included in the same study or studied separately) and whether authors investigated only a small number of a priori specified hypotheses, to differentiate between spurious and trustworthy subgroup findings.[38]

6. How precise are the results?

The width of the 95% CI represents the range in which the true effect plausibly lies (for a more detailed explanation of CIs, see a study by O'Brien and Yi[39]) and is the primary indicator of the precision of SRMA results. We are less certain of the pooled point estimate if the lower and upper boundaries of the CI, were they to represent the true effect, would lead to different dietary advice or actions.[40] We are more certain of the pooled estimate if the lower and upper boundaries of the CI both suggest that there is no important effect, or that there is an important effect.[40] Deciding whether results are sufficiently precise is subjective and evidence users should consider minimal important differences or decision thresholds based on the values and preferences of their patients or target population when making this judgement.[40 41] For example, for the red and processed meat systematic reviews,[42–45] based on the values and preferences of a guideline panel that included members of the public,[46] for fatal outcomes, <10 events per 1000 was considered as 'little to no benefit'. For non-fatal outcomes (eg, cancer incidence), <20 per 1000 were considered 'little to no benefit' over a lifetime (box 8).

7. Do the results directly apply to my patient?

If the populations, exposures or outcomes investigated in primary studies contained in an SRMA differ from the patient or target population of interest, the evidence may not be applicable (called indirectness). For example, SRMAs of nutritional epidemiology studies may investigate the effects of a nutritional exposure on cardiovascular health in individuals who have already developed symptoms of cardiovascular disease (secondary prevention)

studies suggest that the exposure is protective to different degrees, heterogeneity is less important.[37]

vegetarianism and metabolic syndrome. Ideally, reviewers will anticipate inconsistency, generate a priori hypotheses to explain inconsistency and test their hypotheses using subgroup analyses or meta-regression. Even when reviewers do find a possible subgroup effect, it may still be spurious. Evidence users should use established criteria, considering how likely it is that chance explains the difference

Table 3 Summary of select findings of the NutriRECS systematic reviews

Outcome	Unprocessed red meat (reduction of three servings/week)		Processed meat (reduction of three servings/week)		Dietary patterns (lower vs higher in red meat)	
	Risk difference	Certainty of evidence	Risk difference	Certainty of evidence	Risk difference	Certainty of evidence
Cardiovascular mortality	4 fewer per 1000 persons (from 5 fewer to 4 fewer) over 10.8 years	⊕◯◯◯ VERY LOW	4 fewer per 1000 persons (from 7 fewer to 1 fewer) over 10.8 years	⊕◯◯◯ VERY LOW	6 fewer per 1000 persons (from 9 fewer to 2 fewer) over 10.8 years	⊕◯◯◯ VERY LOW
Type 2 diabetes	6 fewer per 1000 persons (from 7 fewer to 4 fewer) over 10.8 years	⊕⊕◯◯ LOW	12 fewer per 1000 persons (from 16 fewer to 9 fewer) over 10.8 years	⊕◯◯◯ VERY LOW	14 fewer per 1000 persons (from 18 fewer to 8 fewer) over 10.8 years	⊕◯◯◯ VERY LOW
Overall cancer mortality	7 fewer per 1000 persons (from 9 fewer to 6 fewer) over lifetime	⊕⊕◯◯ LOW	8 fewer per 1000 persons (from 12 fewer to 6 fewer) over lifetime	⊕⊕◯◯ LOW	12 fewer per 1000 persons (from 18 fewer to 4 fewer) over lifetime	⊕◯◯◯ VERY LOW

but may extrapolate these results to individuals without any signs of cardiovascular disease (primary prevention). Similarly, SRMAs may investigate the effects of nutritional exposures on surrogate outcomes (ie, outcomes that are only important to patients or populations due to their correlation, or assumed link, with other outcomes such as blood pressure) as an indirect measure of patient-important outcomes such a stroke or myocardial infarction.[47] Judgements regarding indirectness depend on whether any differences between the characteristics of studies and the clinical or public health question of interest would lead to an appreciable change in the direction or magnitude of the effect (box 9).

8. Is there concern about publication bias?

Findings from an SRMA will be biased when results from studies that are published and available for inclusion in the SRMA differ from those that are unpublished or unavailable. Sometimes, studies that are published and available show results that are more interesting or provocative than those that are unpublished and unavailable. Bias due to missing studies, called publication bias, is likely frequent in nutritional epidemiology due to the lack of standard registration practices for protocols, which allows investigators to explore a large number of exposures and outcomes and to report only results that yield interesting findings.

One clue that publication bias exists is if results from smaller studies—that are typically more likely to be published and are thus at higher risk for publication bias—are different from larger studies—studies that are at lower risk of publication bias. This is often referred to as 'small study effects'. To assess for publication bias, evidence users should start by evaluating how comprehensive (ie, systematic) the search was (see above). SRMA authors should then look for publication bias using visual inspection of funnel plots, and look for statistical tests, such as Egger's test and Begg's test, that relate the precision of studies to their effect estimates (box 10).[48 49] These methods, however, have important limitations. Statistical tests for publication bias, for example, are almost always underpowered. Further, even when there is statistical evidence that effects from smaller studies are different from larger studies (ie, 'small study effects'), publication bias is only one plausible explanation (another is that smaller and larger studies differ in other important ways, such as risk of bias).

9. Are there reasons to be more certain of findings based on effect size, credible dose-response gradient and/or direction of plausible confounders?

Three uncommon situations can sometimes make us more certain of findings of non-randomised studies. First, when the observed effect is large (typically a relative risk (RR)>2.0 or RR<0.5), biases, such as confounding, are less likely to fully explain the observed effect. Large effects will very rarely, if ever, occur in nutritional epidemiology studies since the foods and nutrients to which participants are typically exposed typically have small effects on health, though these effects may accumulate over long durations of exposure.[50]

Second, we may be more certain of results when we observe a dose-response gradient, particularly if biases about which we are concerned are unlikely to produce spurious dose-response associations. It must be noted, however, that nutritional exposures are highly correlated with one another,[29 51] which makes spurious dose-response associations highly plausible.[29]

The third situation—when all plausible confounders would act in the opposite direction than the direction of the observed effect—is also unlikely to occur in nutrition because we seldom are aware of all plausible confounders. Overall, because situations that make us more certain of findings of non-randomised studies seldom occur in nutrition, SRMAs of nutritional epidemiology studies usually provide only low to very low certainty evidence (box 11).

Author affiliations
[1]Department of Health Research Methods, Evidence & Impact, McMaster University, Hamilton, Ontario, Canada
[2]Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada
[3]Department of Medicine, McMaster University, Hamilton, Ontario, Canada
[4]Chair of Epidemiology and Preventive Medicine, Department of Hygiene and Dietetics, Jagiellonian University Medical College, Krakow, Poland
[5]Iberomerican Cochrane Centre, Clinical Epidemiology and Public Health Department, Institute of Biomedical Research of Barcelona, Barcelona, Spain
[6]Institut de Recerca Sant Pau (IR SANT PAU), Sant Quintí, Barcelona, Spain
[7]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[8]Department of Nutrition, College of Agriculture and Life Sciences, Texas A&M University, College Station, Texas, USA
[9]Department of Epidemiology & Biostatistics, School of Public Health, Texas A&M University, College Station, Texas, USA
[10]Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

X Dena Zeraatkar @DenaZera, Pablo Alonso-Coello @PabloACoello and Bradley C Johnston @methodsnerd

ORCID iD
Bradley C Johnston http://orcid.org/0000-0001-8872-8626

REFERENCES

1 World Health Organization. WHO Handbook for Guideline Development. 2nd Ed. Geneva: World Health Organization, 2014.
2 Murad MH, Montori VM, Ioannidis JPA, *et al*. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA* 2014;312:171–9.
3 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.
4 Song M, Giovannucci E. Substitution analysis in nutritional epidemiology: proceed with caution. *Eur J Epidemiol* 2018;33:137–40.
5 Han MA, Leung G, Storman D, *et al*. Causal language use in systematic reviews of observational studies is often inconsistent with intent: a systematic survey. *J Clin Epidemiol* 2022;148:65–73.
6 Zeraatkar D, Bhasin A, Morassut RE, *et al*. Characteristics and quality of systematic reviews and meta-analyses of observational nutritional epidemiology: a cross-sectional study. *Am J Clin Nutr* 2021;113:1578–92.
7 Santesso N, Glenton C, Dahm P, *et al*. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol* 2020;119:126–35.
8 Higgins JPT, Thomas J, Chandler J. *Cochrane handbook for systematic reviews of interventions version 6.0*. Cochrane, 2019:Ed000142.
9 Sampson M, Barrowman NJ, Moher D, *et al*. Should meta-analysts search embase in addition to medline? *J Clin Epidemiol* 2003;56:943–55.
10 Mahood Q, Van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. *Res Synth Methods* 2014;5:221–34.
11 Orsini N, Li R, Wolk A, *et al*. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. *Am J Epidemiol* 2012;175:66–73.
12 Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* 1992;135:1301–9.
13 Talukdar JR, Steen JP, Goldenberg JZ, *et al*. Saturated fat, the estimated absolute risk and certainty of risk for mortality and major cancer and cardiometabolic outcomes: an overview of systematic reviews. *Syst Rev* 2023;12:179.
14 Yu WW, Schmid CH, Lichtenstein AH, *et al*. Empirical evaluation of meta-analytic approaches for nutrient and health outcome dose-response data. *Res Synth Methods* 2013;4:256–68.
15 Khan TA, Chiavaroli L, Zurbau A, *et al*. A lack of consideration of a dose–response relationship can lead to erroneous conclusions regarding 100% fruit juice and the risk of cardiometabolic disease. *Eur J Clin Nutr* 2019;73:1556–60.
16 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;21:1575–600.
17 Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med* 1992;92:121–4.
18 Alonso-Coello P, Carrasco-Labra A, Brignardello-Petersen R, *et al*. Systematic reviews experience major limitations in reporting absolute effects. *J Clin Epidemiol* 2016;72:16–26.
19 Guyatt GH, Oxman AD, Santesso N, *et al*. GRADE guidelines: 12. Preparing summary of findings tables—binary outcomes. *J Clin Epidemiol* 2013;66:158–72.
20 Tierney JF, Stewart LA, Ghersi D, *et al*. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;8:16.
21 Atkins D, Eccles M, Flottorp S, *et al*. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches the GRADE working group. *BMC Health Serv Res* 2004;4:38.
22 Guyatt GH, Oxman AD, Vist GE, *et al*. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
23 Balshem H, Helfand M, Schünemann HJ, *et al*. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6.
24 Vernooij R, Guyatt GH, Zeraatkar D, *et al*. Reconciling contrasting guideline recommendations on red and processed meat for health outcomes. *J Clin Epidemiol* 2021;138:215–8.
25 Schünemann HJ, Cuello C, Akl EA, *et al*. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol* 2019;111:105–14.
26 Guyatt GH, Oxman AD, Sultan S, *et al*. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011;64:1311–6.
27 Verbeek JH, Whaley P, Morgan RL, *et al*. An approach to quantifying the potential importance of residual confounding in systematic reviews of observational studies: a GRADE concept paper. *Environ Int* 2021;157:106868.
28 Hilton Boon M, Thomson H, Shaw B, *et al*. Challenges in applying the GRADE approach in public health guidelines and systematic reviews: a concept article from the GRADE public health group. *J Clin Epidemiol* 2021;135:42–53.
29 Murad MH, Verbeek J, Schwingshackl L, *et al*. GRADE guidance 38: updated guidance for rating up certainty of evidence due to a dose-response gradient. *J Clin Epidemiol* 2023;164:45–53.
30 Morgan RL, Beverly B, Ghersi D, *et al*. GRADE guidelines for environmental and occupational health: a new series of articles in environment international. *Environ Int* 2019;128:11–2.
31 Zeraatkar D, Kohut A, Bhasin A, *et al*. Assessments of risk of bias in systematic reviews of observational nutritional epidemiologic studies are often not appropriate or comprehensive: a methodological study. *BMJ Nutr Prev Health* 2021;4:487–500.
32 Morgan R, Sterne J, Higgins J, *et al*. *A new instrument to assess risk of bias in non-randomised studies of exposures (ROBINS-E): application to studies of environmental exposure*. Cape Town, South Africa: Global Evidence Summit, 2017.
33 Sterne JA, Hernán MA, Reeves BC, *et al*. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
34 Agoritsas T, Merglen A, Shah ND, *et al*. Adjusted analyses in studies addressing therapy and harm: users' guides to the medical literature. *JAMA* 2017;317:748–59.
35 Kirkpatrick SI, Baranowski T, Subar AF, *et al*. Best practices for conducting and interpreting studies to validate self-report dietary assessment methods. *J Acad Nutr Diet* 2019;119:1801–16.
36 Rücker G, Schwarzer G, Carpenter JR, *et al*. Undue reliance on I2 in assessing heterogeneity may mislead. *BMC Med Res Methodol* 2008;8:79.
37 Guyatt G, Zhao Y, Mayer M, *et al*. GRADE guidance 36: updates to GRADE's approach to addressing inconsistency. *J Clin Epidemiol* 2023;158:70–83.
38 Schandelmaier S, Briel M, Varadhan R, *et al*. Development of the instrument to assess the credibility of effect modification analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ* 2020;192:E901–6.
39 O'Brien SF, Yi QL. How do I interpret a confidence interval? *Transfusion* 2016;56:1680–3.
40 Zeng L, Brignardello-Petersen R, Hultcrantz M, *et al*. GRADE guidance 34: update on rating imprecision using a minimally contextualized approach. *J Clin Epidemiol* 2022;150:216–24.
41 Valli C, Rabassa M, Johnston BC, *et al*. Health-related values and preferences regarding meat consumption: a mixed-methods systematic review. *Ann Intern Med* 2019;171:742–55.
42 Zeraatkar D, Johnston BC, Bartoszko J, *et al*. Effect of lower versus higher red meat intake on cardiometabolic and cancer outcomes: a systematic review of randomized trials. *Ann Intern Med* 2019;171:721–31.
43 Zeraatkar D, Han MA, Guyatt GH, *et al*. Red and processed meat consumption and risk for all-cause mortality and cardiometabolic outcomes: a systematic review and meta-analysis of cohort studies. *Ann Intern Med* 2019;171:703–10.
44 Han MA, Zeraatkar D, Guyatt GH, *et al*. Reduction of red and processed meat intake and cancer mortality and incidence: a systematic review and meta-analysis of cohort studies. *Ann Intern Med* 2019;171:711–20.
45 Vernooij RWM, Zeraatkar D, Han MA, *et al*. Patterns of red and processed meat consumption and risk for cardiometabolic and cancer outcomes: a systematic review and meta-analysis of cohort studies. *Ann Intern Med* 2019;171:732–41.
46 Johnston BC, Zeraatkar D, Han MA, *et al*. Unprocessed red meat and processed meat consumption: dietary guideline recommendations

from the nutritional recommendations (NutriRECS) consortium. *Ann Intern Med* 2019;171:756–64.

47 Svensson S, Menkes DB, Lexchin J. Surrogate outcomes in clinical trials: a cautionary tale. *JAMA Intern Med* 2013;173:611–2.

48 Egger M, Davey Smith G, Schneider M, *et al*. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34.

49 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088–101.

50 Siontis GCM, Ioannidis JPA. Risk factors and interventions with statistically significant tiny effects. *Int J Epidemiol* 2011;40:1292–307.

51 Patel CJ, Ioannidis JPA. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *J Epidemiol Community Health* 2014;68:1096–100.

52 Zheng M, Allman-Farinelli M, Heitmann BL, *et al*. Substitution of sugar-sweetened beverages with other beverage alternatives: a review of long-term health outcomes. *J Acad Nutr Diet* 2015;115:767–79.

53 Evidence Partners. Tool to assess risk of bias in cohort studies. Available: www.evidencepartners.com/wp-content/uploads/2014/02/Tool-to-Assess-Risk-of-Bias-in-Cohort-Studies [Accessed 27 Feb 2019].

54 Picasso MC, Lo-Tayraco JA, Ramos-Villanueva JM, *et al*. Effect of vegetarian diets on the presentation of metabolic syndrome or its components: a systematic review and meta-analysis. *Clin Nutr* 2019;38:1117–32.

55 Ierodiakonou D, Garcia-Larsen V, Logan A, *et al*. Timing of allergenic food introduction to the infant diet and risk of allergic or autoimmune disease: a systematic review and meta-analysis. *JAMA* 2016;316:1181–92.

56 Moynihan PJ, Kelly SAM. Effect on caries of restricting sugars intake: systematic review to inform WHO guidelines. *J Dent Res* 2014;93:8–18.

57 Dallacker M, Hertwig R, Mata J. The frequency of family meals and nutritional health in children: a meta-analysis. *Obes Rev* 2018;19:638–53.