



## Systematic reviews and meta-analysis in nutrition research

George A. Kelley\* and Kristi S. Kelley

*Meta-Analytic Research Group, School of Public Health, Department of Biostatistics, Robert C. Byrd Health Sciences Center, West Virginia University, P.O. Box 9190, Morgantown, WV 26506-9190, USA*

(Submitted 28 June 2019 – Final revision received 10 August 2019 – Accepted 14 August 2019)

### Abstract

There exists an ever-increasing number of systematic reviews, with or without meta-analysis, in the field of nutrition. Concomitant with this increase is the increased use of such to guide future research as well as both practice and policy-based decisions. Given this increased production and consumption, a need exists to educate both producers and consumers of systematic reviews, with or without meta-analysis, on how to conduct and evaluate high-quality reviews of this nature in nutrition. The purpose of this paper is to try and address this gap. In the present manuscript, the different types of systematic reviews, with or without meta-analyses, are described as well as the description of the major elements, including methodology and interpretation, with a focus on nutrition. It is hoped that this non-technical information will be helpful to producers, reviewers and consumers of systematic reviews, with or without meta-analysis, in the field of nutrition.

**Key words:** Systematic reviews: Meta-analyses: Guidelines: Nutrition

Systematic reviews with meta-analyses have the potential to play an important role in quantitatively synthesising evidence when numerous studies on a similar topic exist, especially when disagreement persists among those studies. The potential strengths of meta-analysis include (1) increased statistical power for primary outcomes, (2) ability to reach agreement when original studies yield conflicting findings, (3) improving effect size estimates and (4) answering questions not addressed in original trials<sup>(1)</sup>. In addition, meta-analyses provide the opportunity to generate hypotheses that can be tested in subsequent original trials. Furthermore, systematic reviews, with or without meta-analysis, often play a major role in guideline development<sup>(2)</sup>. In a recent special issue devoted entirely to *P* values in the *American Statistician*, Wasserstein *et al.* suggested that since one study is usually not definitive, meta-analysis is critical to determining the uncertainty in the evidence<sup>(3)</sup>. Recognising their potential value, the number of systematic reviews, with or without meta-analysis, has increased dramatically over approximately the last 40 years. For example, a simple PubMed search conducted by the authors on 10 May 2019, using the search phrase “systematic review” OR meta-analy\* yielded four citations in 1978 *v.* 31 295 in 2018, the most recent complete year for which data were available. The number of systematic reviews with meta-analyses in the area

of nutrition has also increased dramatically over the same time period. A simple PubMed search conducted by the authors on 10 May 2019, using the search phrase (“systematic review” OR meta-analy\*) AND (food OR beverages OR diet OR nutrition) yielded one citation in 1978 *v.* 2743 in 2018, the most recent complete year in which data were available.

### Types of systematic reviews

Table 1 lists the different types of systematic reviews with a description provided hereafter.

### Scoping reviews

While no one universal definition exists, a scoping review may be best defined as a type of research synthesis that aims to ‘map the literature on a particular topic or research area and provide an opportunity to identify key concepts; gaps in the research; and types and sources of evidence to inform practice, policy-making, and research’<sup>(4)</sup>. Thus, scoping reviews can be beneficial from both a research and practice perspective. To illustrate its use in the field of nutrition, Amouzandeh *et al.* recently conducted a scoping review of the validity, reliability and conceptual alignment of food literacy measures for adults<sup>(5)</sup>. The authors concluded that most tools provided a theoretical

**Abbreviations:** AD, aggregate data; IPD, individual participant/patient data; IVhet, inverse heterogeneity; LFK, Luis Furuya-Kanamori; NI, non-inferiority; NMA, network meta-analysis; PI, prediction intervals; PRISMA-ScR, Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews; QE, quality effects; SRPSR, systematic reviews of previous systematic reviews; TC, total cholesterol.

\* **Corresponding author:** George A. Kelley, email [gkelley@hsc.wvu.edu](mailto:gkelley@hsc.wvu.edu)

**Table 1.** Types of systematic reviews

Type	Description
Scoping review	Type of research synthesis that aims to ‘map the literature on a particular topic or research area and provide an opportunity to identify key concepts, gaps in the research, and types and sources of evidence to inform practice, policymaking, and research’ <sup>(4)</sup> . Also known as ‘systematic scoping review’, ‘scoping report’, ‘scope of the evidence’, ‘rapid scoping review’, ‘structured literature review’, ‘scoping project’, ‘scoping meta review’
Systematic review of previous systematic reviews	A systematic review of previous systematic reviews on the same topic. Also known as ‘umbrella reviews’, ‘overviews of reviews’, ‘reviews of reviews’, ‘summary of systematic reviews’, ‘synthesis of reviews’, or ‘meta-reviews’ <sup>(6)</sup>
Systematic review without meta-analysis	‘A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review’ <sup>(6)</sup> . Data are synthesised qualitatively
Systematic review with meta-analysis	Same as a systematic review without meta-analysis (see above) except that data are synthesised quantitatively, that is, meta-analysis
– AD meta-analysis	A systematic review that includes a meta-analysis based on summary data (sample sizes, means, standard deviations) extracted from eligible studies
– IPD meta-analysis	A systematic review that includes a meta-analysis based on individual participant/patient data v. summary data. This de-identified data almost always has to be requested from the author(s) of the original studies
– Network meta-analysis (AD or IPD)	A systematic review with a meta-analysis comparing at least three treatments that includes both direct (head to head) and indirect (comparing two treatments via a comparative control group) evidence. This can be based on either aggregate or individual participant data
– Non-inferiority meta-analysis (AD or IPD)	Meta-analysis that attempts to assess whether a new intervention is no worse than a reference intervention

AD, aggregate data; IPD, individual participant/patient data.

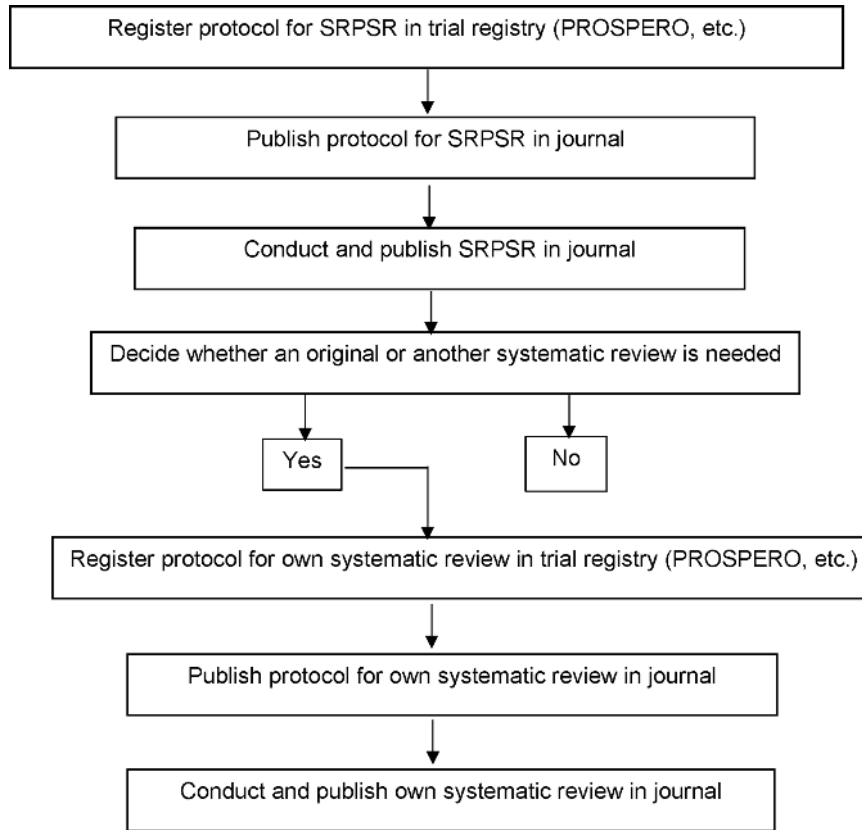
framework, which is valid and reliable<sup>(5)</sup>. In addition, they believed that their results will assist practitioners in selecting and developing tools for the measurement of food literacy<sup>(5)</sup>. Congruent with other types of reviews, the number of scoping reviews in the field of nutrition is increasing. As an example, a PubMed search conducted on 11 May 2019, using the search phrase (“scoping review” OR “systematic scoping review” OR “scoping report” OR “scope of the evidence” OR “rapid scoping review” OR “structured literature review” OR “scoping project” OR “scoping meta review”) AND (food OR beverages OR diet OR nutrition) demonstrated that the number of citations has increased from one in 1981 to 161 in 2018, the most recent complete year for which data were available. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) provides an excellent guide, including a checklist, for conducting and reporting a scoping review<sup>(7)</sup>. Checklists such as the PRISMA series provide very helpful information to producers, reviewers and consumers (clinicians, guideline developers, etc.) for ensuring that high-quality reviews are conducted. Therefore, the authors advocate that journals require the appropriate checklist when authors submit their manuscript for publication consideration.

### Systematic reviews of previous systematic reviews

Given the proliferation of systematic reviews, with or without meta-analysis, on the same topic, there is now a need to assess these previous reviews. As an example of a systematic review of previous systematic reviews (SRPSR) in nutrition, Agostoni *et al.* recently conducted a SRPSR on the long-term effects of dietary nutrient intake during the first 2 years of life in healthy infants from developed countries<sup>(8)</sup>. The overall conclusion of the authors was that a large degree of uncertainty currently exists

on the health effects of differences in early nutrition among healthy full-term infants<sup>(8)</sup>.

There are at least two important reasons for conducting a SRPSR. First, for those desiring to conduct their own systematic review, with or without meta-analysis, such a review can help justify the conduct of a new or updated review. If an updated or new review is deemed warranted, then this information should be included in the introduction section of the new or updated review. Ideally, this should include reference to a previously published SRPSR. If after searching the literature the authors believe that no previous reviews exist, then this should be stated. The inclusion of this information may be especially important given the recent criticism regarding the publication of redundant reviews on the same topic<sup>(9)</sup>. Fig. 1 depicts a stepwise process suggested by the authors for moving from a SRPSR to one’s own review, details of which can be found elsewhere<sup>(10)</sup>. Briefly, a major decision that needs to be made is whether a new systematic review, with or without meta-analysis, is needed. The Cochrane Collaboration recommends that another systematic review be based on needs and priorities, with consideration of strategic importance, practical aspects as it pertains to organising the review, and impact of another review<sup>(11)</sup>. The Agency for Healthcare Research and Quality in the United States approaches this from a needs-based perspective in which the focus is on stakeholder impact as well as currency and necessity<sup>(12)</sup>. A determination is then made to create, archive or continue surveillance<sup>(12)</sup>. The Panel for Updating Guidance for Systematic Reviews (PUGS) created a consensus and checklist for when and how to perform another systematic review<sup>(13)</sup>. This process includes assessing the currency as well as previous review(s), if any exist, identifying relevant new methods, studies or other information that may justify another review, and assessing the potential impact of



**Fig. 1.** Suggested stepwise approach for deciding whether a new or updated systematic review, with or without meta-analysis, should be conducted. Adapted from Kelley & Kelley<sup>(10)</sup>. SRPSR, systematic reviews of previous systematic reviews.

another review<sup>(13)</sup>. The PUGS guidelines and checklist may be the most suitable method for researchers interested in conducting another systematic review, with or without meta-analysis. Any new reviews should also address an important research question, something that should be explained in the introduction section of the manuscript.

A second reason for conducting a SRPSR is that given the large number of reviews of this type on many of the same topics, a need exists to evaluate these in order to provide decision makers (clinicians, guideline developers, policymakers, etc.) with the information they need to make informed choices on the topic of interest. A simple PubMed search conducted by the authors on 10 May 2019, using the search criteria “systematic review of previous systematic reviews” OR “umbrella review” OR “overview of reviews” OR “review of reviews” OR “summary of systematic reviews” OR “meta-reviews” AND (food OR beverages OR diet OR nutrition) yielded 173 citations associated with nutrition-related SRPSR in 2018, the most recent complete year for which data were available. As part of the conduct of a SRPSR, an evaluation regarding the quality and/or risk of bias of each included systematic review, with or without meta-analysis, should be included. Instruments for assessing such include, but are not limited to, (1) a Measurement Tool to Assess systematic Reviews 2<sup>(14)</sup>, (2) Risk of Bias in Systematic Reviews<sup>(15)</sup> (3) Grading of Recommendations, Assessment, Development and Evaluations (GRADE)<sup>(16)</sup> and (4) Quality

Assessment of Diagnostic Accuracy Studies 2<sup>(17)</sup>. The importance of SRPSR is supported by a recent thematic series devoted to this topic<sup>(18–20)</sup>. In addition, Ballard & Montgomery also provide methodological guidance, including a four-item checklist, for evaluating a SRPSR<sup>(21)</sup>. Finally, for the reasons previously given as well as to improve efficiencies and avoid research waste<sup>(18)</sup>, the authors believe that funding agencies should support high-quality SRPSR. Detailed information regarding SRPSR can be found elsewhere<sup>(18–28)</sup>.

### Systematic review without meta-analysis

The Cochrane Collaboration defines a systematic review as a ‘review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review<sup>(6)</sup>. The key characteristics of a systematic review include (1) clearly stated objectives with predefined eligibility criteria for studies, (2) an explicit, reproducible methodology, (3) a systematic search that attempts to identify all studies that meet the eligibility criteria, (4) an assessment of the validity of the findings of the included studies (risk of bias, etc.) and (5) a systematic presentation and synthesis of the characteristics and findings of the included studies<sup>(6)</sup>. A systematic review without a meta-analysis is often conducted because the authors feel that the studies are not combinable

quantitatively given that they are too different and/or cannot be combined into some type of common metric. This is usually not an easy task since no one study is exactly alike, nor should they be. For example, some people may decide *a priori* that the studies will be too different to combine quantitatively (apples and oranges) while others may decide that the eligible studies can be combined (fruit salad). If a meta-analysis is not included, then the reason for not doing so should be stated in the research synthesis sub-section of the Methods section of the manuscript. When a meta-analysis is not included, the results are synthesised qualitatively. As an example, Calder *et al.* conducted a systematic review without meta-analysis with respect to increasing arachidonic acid intake and PUFA status, metabolism and health-related outcomes in humans<sup>(29)</sup>. Based on twenty-two articles from fourteen randomised controlled trials, the authors concluded that insufficient evidence currently exists to support any recommendation regarding the specific health effects of arachidonic acid intake<sup>(29)</sup>. The original PRISMA statement provides guidance, including a checklist, for conducting and reporting a systematic review, with or without meta-analysis<sup>(30)</sup>.

### Systematic review with meta-analysis

A systematic review with meta-analysis is similar to a systematic review without a meta-analysis with the exception that the former includes a quantitative synthesis, that is, meta-analysis of the data. Generally, systematic reviews with a meta-analysis consist of the following types: (1) aggregate data (AD) meta-analysis, (2) individual participant/patient data (IPD) meta-analysis, (3) network meta-analysis (NMA), which can be based on either AD or IPD and (4) non-inferiority (NI) meta-analysis (AD or IPD).

**Aggregate data meta-analysis.** An AD meta-analysis is a quantitative approach in which summary data, for example, sample sizes, means and standard deviations are abstracted for outcomes of interest (kJ consumed, cholesterol intake, etc.) from previously published studies and then pooled for analysis. These are by far the most common types of meta-analyses conducted today and often focus on pairwise comparisons, for example, changes in an intervention *v.* control group. A simple PubMed search conducted by the authors on 13 May 2019, using the search string (“systematic review” OR meta-analy\*) AND (food OR beverages OR diet OR nutrition) NOT (“individual participant data” OR “individual patient data” OR “IPD” OR “systematic review of previous systematic reviews” OR “umbrella review” OR “overview of reviews” OR “review of reviews” OR “summary of systematic reviews” OR “meta-reviews”) yielded a total of one citation in 1978 *v.* 2557 in 2018, the most recent and complete year in which data were available. As an example of an AD meta-analysis in nutrition, Zhang *et al.*, conducted a systematic review with meta-analysis on the efficacy and safety of iron supplementation in patients with heart failure and iron deficiency<sup>(31)</sup>. Based on nine randomised controlled trials representing 789 patients who received iron therapy, significant improvements were observed for the 6-min walk test and peak maximum oxygen consumption as well as fewer patients being hospitalised for heart failure<sup>(31)</sup>. No associations were found for total re-hospitalisation or mortality<sup>(31)</sup>.

As previously mentioned, the original PRISMA statement provides guidance, including a checklist, for conducting and reporting a systematic review with AD meta-analysis<sup>(30)</sup>. In addition, recent guidance for conducting systematic reviews and meta-analyses of observational studies in aetiology is also available<sup>(32)</sup> and the Cochrane Handbook provides extensive information on the conduct of systematic reviews with AD meta-analysis<sup>(6)</sup>.

**Individual participant/patient data meta-analysis.** An IPD meta-analysis is a systematic review that includes a meta-analysis based on IPD and often comprises a consortium made up of a large number of investigators such as the European Consortium that recently conducted an IPD meta-analysis on vitamin D and mortality<sup>(33)</sup>. Since de-identified IPD is usually not available in the original studies, it needs to be requested from the author(s). Considered the ‘gold standard’ of meta-analyses, the potential advantages of an IPD meta-analysis, described in detail elsewhere<sup>(34)</sup>, include, but are not limited to, ‘standardizing statistical analyses in each study; deriving desired summary results directly, independent of study reporting; checking modelling assumptions; and assessing participant-level effects, interactions and non-linear trends’<sup>(35)</sup>. However, one of the major disadvantages of an IPD meta-analysis is the ability to retrieve original data from study authors, with ranges of 25–100% reported across different subject areas<sup>(36–39)</sup>. As a result, this can lead to an increased risk of bias. While at least one approach has been recommended for integrating both IPD and AD<sup>(40)</sup>, one is still left with AD from those studies in which IPD cannot be retrieved. A second disadvantage of an IPD *v.* AD meta-analysis is the increased time and resources associated with such analysis. For example, one study estimated the costs of a previous IPD meta-analysis<sup>(41)</sup> to be eight times greater than an AD meta-analysis<sup>(42)</sup>. Finally, several studies have shown a lack of statistically and practically important differences between AD and IPD meta-analyses when an indistinguishable, or nearly indistinguishable, number of studies are included<sup>(41,43–45)</sup>. Despite these disadvantages, the number of IPD meta-analyses is increasing, including the field of nutrition. A simple PubMed search conducted by the authors on 13 May 2019, using the search string (“systematic review” OR meta-analy\*) AND (food OR beverages OR diet OR nutrition) AND (“individual participant data” OR “individual patient data” OR “IPD”) NOT (“systematic review of previous systematic reviews” OR “umbrella review” OR “overview of reviews” OR “review of reviews” OR “summary of systematic reviews” OR “meta-reviews”) yielded one citation in the year 2002 *v.* twenty-six in 2018, the most recent year in which complete data were available. As an example in the field of nutrition, Smelt *et al.* recently conducted an IPD meta-analysis of randomised controlled trials on the effects of vitamin B<sub>12</sub> and folic supplementation on routine haematological parameters in adults 60 years of age and older<sup>(46)</sup>. The authors concluded that there is currently a lack of evidence to support the effects of supplementation of low concentrations of vitamin B<sub>12</sub> and folate on haematological parameters in community-dwelling adults 60 years of age and older<sup>(46)</sup>. A set of PRISMA guidelines, including a checklist, for conducting and reporting an IPD meta-analysis (PRISMA-IPD)



are available<sup>(47)</sup>. Additional details regarding the conduct of an IPD have been reported elsewhere<sup>(6,34,48)</sup>.

**Network meta-analysis.** A more recent and increasingly used approach, including the field of nutrition<sup>(49)</sup>, is the conduct of a systematic review with NMA, usually in the form of an AD NMA *v.* IPD NMA. NMA, also known as ‘multiple treatments meta-analysis’ or ‘mixed treatment comparisons meta-analysis’, is a type of meta-analysis that compares at least three treatments and includes both direct (comparing two treatments head to head) and indirect (comparing two treatments via a comparative control group) evidence. One of the major reasons for its increased use is the ability to include multiple treatments in the same analysis, thereby facilitating treatment recommendations. For example, Galaviz *et al.* recently conducted an NMA on the real-world impact of global diabetes prevention interventions on diabetes incidence, body weight and glucose<sup>(50)</sup>. The overall conclusion of the authors’ NMA of sixty-three studies was that real-world lifestyle modification strategies can reduce diabetes risk<sup>(50)</sup>. A simple PubMed search conducted by the authors on 14 May 2019, using the search string (“network meta-analysis” OR “multiple treatments meta-analysis” OR “mixed treatment comparisons meta-analysis”) AND (food OR beverages OR diet OR nutrition) NOT (“systematic review of previous systematic reviews” OR “umbrella review” OR “overview of reviews” OR “review of reviews” OR “summary of systematic reviews” OR “meta-reviews”) yielded one initial citation in the year 2007 *v.* thirty-three in 2018, the most recent year in which complete data were available. Not surprisingly, NMA is more time and resource intensive than a traditional AD meta-analysis given the large number of treatments that are usually included as well as the inclusion of both direct and indirect evidence. PRISMA guidelines, including a checklist, for conducting and reporting a NMA (PRISMA-NMA) are available<sup>(51)</sup>. Additional details regarding this emerging and important approach have been described elsewhere<sup>(52–55)</sup>.

**Non-inferiority meta-analysis.** The most recent, but still infrequent type of meta-analysis to emerge is a NI meta-analysis. A NI meta-analysis attempts to assess whether a new intervention is no worse than a reference intervention<sup>(56)</sup>. A major challenge of a NI meta-analysis is the NI margin used<sup>(56)</sup>. These types of meta-analyses could be based on either AD or IPD and could also take the form of a NMA (AD or IPD)<sup>(57)</sup>. While the authors are not aware of any NI meta-analyses in the field of nutrition, Acuna *et al.* recently conducted a NI meta-analysis that examined the quality of surgical outcomes using laparoscopic *v.* open resection for rectal cancer<sup>(58)</sup>. Based on their analysis of fourteen randomised controlled trials, the authors concluded that laparoscopy was non-inferior to open surgery for rectal cancer<sup>(58,59)</sup>. More detailed information regarding NI meta-analyses can be found elsewhere<sup>(56,57,60)</sup>.

### Primary components of systematic reviews with meta-analysis

Given that traditional AD meta-analyses still dominate the literature, the emphasis of the rest of this manuscript will centre

on this type of quantitative review but while noting that much of this information can be applied to many of the other types of systematic reviews with meta-analyses that have been previously described. For more detailed information, readers are referred to the PRISMA Guidelines, including a twenty-seven-item checklist, for the conduct and reporting of systematic reviews with AD meta-analysis<sup>(30)</sup>.

### Overview

Similar to most research studies, a systematic review with meta-analysis manuscript (broadly) should consist of an abstract, introduction, methods, results, discussion and conclusion(s) section.

### Abstract

The structure of the abstract of a systematic review with meta-analysis generally mirrors that of an original study. The PRISMA guidelines provide specific information, including a twelve-item checklist, regarding information to report in the abstract of a systematic review, with or without meta-analysis<sup>(61)</sup>. However, adherence to all items in the checklist may be difficult given the word limitations on abstracts imposed by journals and conference abstracts. Thus, one may have to prioritise the most important information to be included, especially since many readers may not read beyond the abstract. For example, Saint *et al.* reported that almost two thirds (63%) of internists only read the abstracts of medical journal articles<sup>(62)</sup>. Given the former, a clear and concise abstract would seem to be important.

### Introduction

In the introduction section of the manuscript, the authors should provide a strong rationale for why the present study is needed. This should include the importance of the issue to be addressed as well as a review of prior research on the topic. Based on the authors’ experiences, producers of systematic reviews with meta-analysis usually provide an adequate description of the importance of the topic to be addressed but often lack information regarding previous original studies on the topic as well as previous systematic reviews with meta-analysis, if any, to justify their own systematic review with meta-analysis. The former is important because the conflicting findings of previous original studies are often one of the very reasons for conducting reviews of this nature. The latter is equally important because of the increasing concern about redundant systematic reviews, with or without meta-analysis, that is, value added<sup>(9)</sup>. If the authors are not aware of any previous systematic reviews with meta-analysis on the topic, then it should be stated. For example, in a systematic review with AD meta-analysis of randomised controlled trials examining the impact of modified dietary interventions on maternal glucose control and neonatal birth weight, Yamamoto *et al.* cited three previous systematic reviews and meta-analyses related to the topic but none specific to their proposed work regarding the impact of modified dietary interventions on detailed maternal glycaemic parameters, including changes in glucose-related variables<sup>(63)</sup>. As previously mentioned, one approach to help justify one’s own work, though

more time-consuming and resource intensive, is to conduct and publish a systematic review of previous systematic reviews with meta-analysis on the topic and describe this in the introduction section of the manuscript<sup>(40)</sup>. Finally, the end of the introduction should clearly delineate the purpose/objective(s)/research question(s) of the intended systematic review with AD meta-analysis.

### Methods and results

Any systematic review, with or without meta-analysis, should include an *a priori* research plan and at a minimum, register the protocol in a systematic review trials registry such as PROSPERO<sup>(64)</sup>. At the beginning of the methods section of the paper, the registration number should be reported. Registering a systematic review with meta-analysis is important for (1) promoting transparency, (2) helping to reduce potential bias and (3) helping to avoid unintended duplication of effort<sup>(65)</sup>. Registration is beneficial for researchers, commissioning and funding organisations, journal editors and peer reviewers<sup>(65)</sup>. Based on these benefits, the authors would advocate that journals require all manuscript submissions to include a registration number before being considered for peer review. In addition to the protocol being registered in PROSPERO, it is suggested that authors consider publishing their protocol in a peer-reviewed journal, thereby enhancing reach and possibly improving their study design. As an example, Asghari *et al.* recently published a protocol for a systematic review with AD meta-analysis in which they plan to examine the effects of vitamin D supplementation on serum 25-hydroxyvitamin D concentration in children and adolescents<sup>(66)</sup>. The PRISMA group provides detailed guidelines, including a seventeen-item checklist, for developing and reporting the protocol for a systematic review, with or without meta-analysis (PRISMA-P)<sup>(67)</sup>. To enhance the field of research, the authors would also advocate that peer-reviewed journals consider publishing high-quality protocols, including requiring a completed PRISMA-P checklist upon submission.

Congruent with PRISMA guidelines,<sup>(30)</sup> the methods section of a systematic review with AD meta-analysis should usually be partitioned into the following sections: (1) study eligibility, (2) data sources, (3) study selection, (4) data abstraction, (5) risk of bias assessment and (6) data synthesis.

**Study eligibility.** This section should describe the studies that should be included in a systematic review with AD meta-analysis. To aid in determining eligible studies as well as searching the literature, one may consider using the PICO or PICOS framework<sup>(30)</sup>. Where applicable, the PICO/PICOS structure includes participants/population (P), interventions (I), comparisons (C), outcomes (O) and study design/setting (S)<sup>(30)</sup>. For example, in a recent systematic review with AD meta-analysis on dietary patterns, bone mineral density and fracture risk, the PICO framework included an open population (P), dietary patterns as the intervention (I), other dietary patterns as the comparison (C), bone mineral density, bone mineral content or fracture as the outcomes (O) and observational study designs (S)<sup>(68)</sup>. For observational studies dealing with aetiology, the population, exposure, control and outcomes framework has recently been suggested<sup>(32)</sup>. In addition, the type of study designs

included should also be reported. For example, in a meta-analysis that examined the effects of Ca intake on breast cancer risk, the population consisted of females, the exposure was Ca intake (dietary and/or supplemental), the control/comparator was no dietary or supplemental Ca intake, the outcome was breast cancer risk and the study designs included were prospective cohort, case-control or case-cohort studies<sup>(69)</sup>.

In addition to providing a description of potential eligible studies, reasons for excluding studies may also be provided, though it is perfectly reasonable to assume that any study not meeting one's eligibility criteria would be excluded. However, this does not exclude one from including a supplementary file of excluded citations, including the reasons for exclusion after each reference. A systematic review may include studies in any language, especially given the free online language translators that are currently available. However, there is no clear consensus regarding increased bias whether a systematic review is limited to English-language articles published in peer-reviewed journals<sup>(6)</sup>. In addition, studies may be derived from both published and unpublished sources (master's theses, dissertations, abstracts from conference proceedings, clinical trials registries, etc.). However, van Driel *et al.* concluded that (1) the difficulty in retrieving unpublished work could lead to selection bias, (2) many unpublished trials are eventually published, (3) the methodological quality of such studies are poorer than those that are published and (4) the effort and resources required to obtain unpublished work may not be warranted<sup>(70)</sup>.

**Data sources.** The data sources subsection of the methods describes the sources that are to be used to try and locate potential eligible studies. While there will always be a margin of search error, the goal is to try and obtain as many studies as possible that meets one's eligibility criteria. To achieve this goal, a list of electronic databases that were searched should be provided (PubMed, Embase, etc.) as well as the search criteria for the databases. While there is no clear consensus, it has been suggested that at least two electronic databases be searched<sup>(6)</sup> because no one database indexes all journals. While a minimum of two databases is one suggestion<sup>(6)</sup>, Bramer *et al.* recently suggested that at least Embase, MEDLINE, Web of Science and Google Scholar be searched to ensure adequate coverage<sup>(71)</sup>. However, Google Scholar may not be worth the time and effort, given its lack of sensitivity and specificity<sup>(72)</sup>. For those researchers who do not have easy access to Embase but can access Scopus, searching the latter may be acceptable since Scopus has been reported to provide 100% coverage of both MEDLINE and Embase<sup>(73)</sup>. It is also relevant to point out that MEDLINE is nested within the PubMed database. If grey literature is included, sources such as ProQuest master's theses and dissertations and the System for Information on Grey Literature in Europe databases could be searched. When searching electronic databases, the detailed search strategy for at least one of them, for example, PubMed, should be included. This may be embedded in the text or included as a supplementary file. To ensure adequate coverage, it is recommended that nutritionists search a minimum of three databases, inclusive of the following: (1) PubMed, (2) Embase or Scopus and (3) Web of Science.

In addition to searching electronic databases, other methods should be used. These include such things as cross referencing from retrieved studies, searching clinical trials databases,

hand-searching selected journals and expert review. The start and end dates for all searches should be provided, including the reason(s) for the chosen start date. Finally, the name(s) of the individual(s) who conducted the searches should also be provided<sup>(30)</sup>.

**Study selection.** The study selection section describes the process that was used to select studies. To avoid study selection bias, studies should be reviewed by at least two people, independent of each other. Those individuals should then meet and review their selections for agreement. However, prior to doing so, one may provide data on the level of agreement before addressing discrepancies. One common statistic used to address this is the kappa statistic ( $\kappa$ )<sup>(74)</sup>. If agreement cannot be reached for one or more studies when the selectors meet, at least one other person should make a recommendation. For all excluded studies, the reason(s) for exclusion should be recorded. One broad way to address exclusions is to follow the PICOS structure: (1) participants/population, (2) intervention, (3) comparison, (4) outcomes, (5) study design/setting and (6) other. The names of all individuals involved in the study selection process, including their role, should also be provided.

**Data abstraction.** The data abstraction/extraction section describes the process used to code the eligible studies. A first step is to provide a brief description of how the codebooks were developed to abstract data, including a list and description of the information that was coded. Generally, this may include (1) study characteristics (authors, year of publication, journal, study design, etc.), (2) participant characteristics (age, gender, race/ethnicity, morbidities, etc.), (3) intervention characteristics (length of study, etc.) and (4) outcome characteristics (sample sizes, means, standard deviations, etc.). Additional information for abstracting data, including for complex meta-analyses, is provided elsewhere<sup>(75)</sup>. The same process for selecting studies should be used for abstracting data. In addition, the authors should provide information on the process used for obtaining missing data. If no attempt was made to obtain missing data, then this should be stated.

**Risk of bias assessment.** A systematic review, with or without meta-analysis, should usually include some type of risk of bias assessment for each included study. It is important here to distinguish between the risk of bias and study quality, something that appears to often be overlooked given the authors' more than 25 years of experience in reviewing manuscripts and grant proposals. The Cochrane Collaboration recommends that the focus be on the risk of bias, amongst other factors, given that the ultimate goal should be the degree to which the results of the concluded studies are to be believed<sup>(6)</sup>. It also overcomes the uncertainty in differentiating between the quality in the conduct of a study *v.* the conduct in the reporting of a study<sup>(6)</sup>. While this does not negate the use of study quality scales, the potential limitations should be clearly delineated in the manuscript. However, the use of quality scales to decide what studies should be included or excluded is strongly discouraged, as previously mentioned, given the difficulty in distinguishing between the quality of the reporting of a study and the quality in the conduct of a study<sup>(6)</sup>. There are at least eighty-six risk of bias/study quality assessment instruments<sup>(76)</sup>.

Seehra *et al.* reported that the Cochrane risk of bias was the most common tool used for assessing randomised controlled trials (26.1%), while the Newcastle–Ottawa scale, a study-quality instrument, was used most commonly for assessing non-randomised studies (15.3%), including case–control and cohort studies<sup>(77)</sup>. However, since the time of this publication, the Cochrane Collaboration has updated their risk of bias tool for randomised controlled trials<sup>(78)</sup> and also created an instrument for assessing the risk of bias in non-randomised studies in which the health effects of two or more interventions are compared<sup>(79)</sup>. For authors, the important point here is to carefully consider the instrument(s) to be used and provide a rationale for the choice(s). For example, the authors may choose to use some type of risk of bias assessment instrument as well as some type of study quality tool. Finally, the processes for evaluating the risk of bias and/or the study quality are the same as those for selecting studies and extracting data. While not without limitations, the risk of bias and/or study quality results can help consumers of meta-analyses with decisions regarding the strengths and potential limitations of included studies.

**Data synthesis (effect size calculation).** The data synthesis piece of a systematic review can be either qualitative or quantitative (meta-analysis). The focus here will be on the meta-analytic approach. The initial step in conducting a meta-analysis is deciding on the method that will be used to calculate a common effect size for each outcome from each study so that the findings might be pooled into an overall result. The calculation of an effect size traditionally comprises sample sizes as well as measures of central tendency (e.g. means) and dispersion (e.g. standard deviations). If feasible, the focus should be on calculating and reporting effect sizes using the original metric, for example, kJ/d. The primary reason for this approach is based on the belief that it will be easier for consumers (nutritionists, clinicians, policymakers, etc.) to understand. However, in many situations, the calculation of something like a standardised mean difference effect size (Hedge's *g*, Cohen's *d*, etc.) may be necessary if the outcome of interest is assessed using different scales, for example, the effects of dietary improvement on symptoms of depression and anxiety, given that depression and anxiety outcomes were assessed using different scales<sup>(80)</sup>. Another strength of the standardized mean difference effect size is the ability to calculate this statistic from a number of different tests (*t* tests, *F* ratios, correlations, etc.)<sup>(6,81)</sup>. Alternatively, one potential weakness of the standardized mean difference effect size is the inability of consumers to understand this metric. For example, it is usually much easier for consumers to understand and interpret a decrease in resting systolic blood pressure of 8 mmHg *v.* a mean reduction of 0.50 standardised deviation units. Given the former, it is recommended that the original metric be used if all of the studies for the outcome of interest report the results for that outcome using the same metric or if the results can be converted into a metric that is easier for the reader to interpret, for example, converting total cholesterol (TC) from mg/dl to mmol/l by multiplying TC in mg/dl by 0.02586. If the outcome of interest is assessed using different instruments with various scales that cannot be converted into a more easily understood metric, then the standardised mean difference effect size is recommended. If the standardised mean difference

effect size is used, we recommend that results based on the original scale, including variance statistics, also be reported in a table or figure.

**Data synthesis (effect size pooling).** After deciding on the metric used to pool results, a decision needs to be made on the type of model that will be used to pool results. However, prior to that decision, the investigators need to decide which study designs to include. For intervention studies, we recommend that only randomised controlled trials be included because they are the only way to control for confounders that are not known or measured as well as the observation that non-randomised controlled trials and single group trials tend to overestimate the effects of healthcare interventions<sup>(82,83)</sup>. For observational studies, we recommend that case-control, cross-sectional as well as retrospective and prospective study designs be analysed separately. These separate results can easily be displayed in a table and/or forest plot.

For pooling, there is currently no clear consensus on the one best model for combining results, necessitating a clear need for a large simulation study that tests all the different models under various conditions. With a focus on frequentist meta-analysis, historically two basic types of models are used, the traditional fixed-effect model and the random-effects model. In a traditional fixed-effect model, the assumption is that all the included studies share the same common effect size. Thus, any differences in the observed effects are considered to be the result of within-study sampling error while between-study variance is not accounted for. In contrast, random-effects models assume that the true effect size may differ both within (within-study sampling error) and between (between-study variance) studies. Thus, random-effects models attempt to account for both within- and between-study variance. Multiple random-effects models exist, all of which use different statistical approaches to estimate the between-study variance<sup>(84-89)</sup>. Therefore, if a random-effects model is used, it is important for authors to report and cite that random-effects model since they can lead to different results<sup>(90)</sup>. The most commonly used, but not necessarily the best model, is the original random-effects, method-of-moments approach of Dersimonian & Laird<sup>(85)</sup>. Its common use is most likely the consequence of its longevity as well as presence in numerous statistical packages for meta-analysis. The former notwithstanding, caution may be warranted in the *a priori* use of the traditional fixed-effect model and various random-effects models that are currently available<sup>(84-89)</sup>. For the traditional fixed-effect model, the issue has to do with not accounting for potential between-study variance that may exist. For random-effects models, an attempt is made to account for between-study variance that usually results in wider CI but also results in an increased mean squared error, which is a problem. In addition, the pooled mean effect for random-effects models is not always more conservative than the traditional fixed-effect model<sup>(91)</sup>. Alternatively, fixed-effect models with robust error estimation may currently be the best choice<sup>(92-94)</sup>. In the presence of statistical homogeneity, these models will collapse into the traditional fixed-effect model. Both the inverse heterogeneity (IVhet) and quality effects (QE) models are examples of fixed-effect models with robust error estimation<sup>(92,93)</sup>. Both have been shown to be more robust than the traditional Dersimonian and Laird

approach, with regard to coverage probabilities<sup>(92,93)</sup>. The IVhet model uses an estimator under the fixed-effect model assumption but importantly has a quasi-likelihood-based variance structure<sup>(92)</sup>, while the QE model weights studies by including a quality score for each study, derived from a pre-existing or self-developed scale<sup>(93)</sup>. The relationship between the two models is that the IVhet model is the QE model with quality set to equal. Thus, no quality scores need to be imputed when using the IVhet model<sup>(93)</sup>.

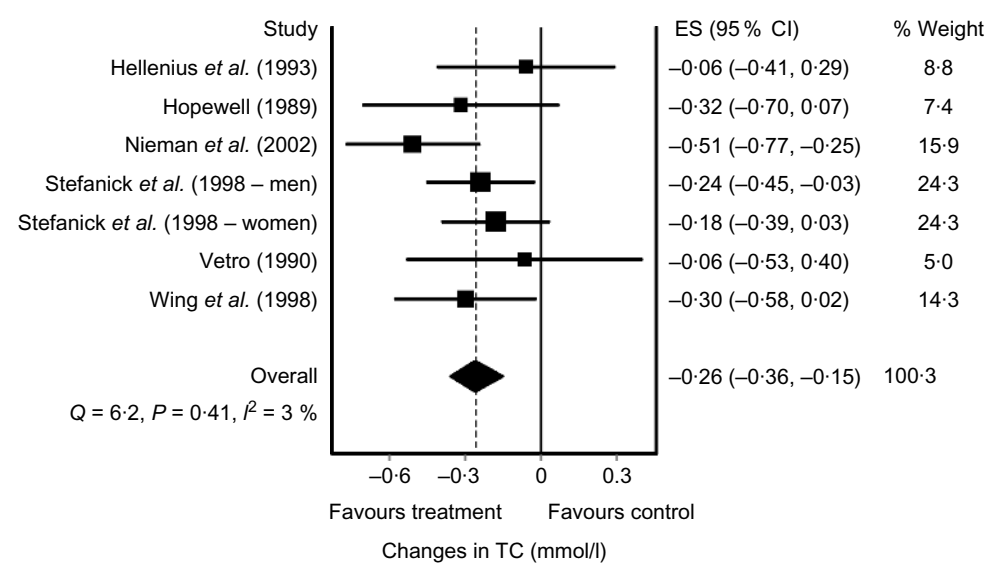
While acknowledging the current and ever-changing state of the evidence as well as the prioritisation of coverage probabilities over point estimates, we recommend that the IVhet and QE models be used when conducting an AD meta-analysis<sup>(92-94)</sup>. However, it's also important to understand that no statistical model is perfect. In addition, the choice of which model to use will often depend on how a meta-analyst poses the question and what modelling assumptions they make *a priori*, including what the parameter of interest is. Both the IVhet and QE models are currently available in a free, easy-to-use Excel meta-analysis add-in program (Meta XL)<sup>(95)</sup>. A Stata module (admetan) is also available to execute the IVhet and QE models.

Irrespective of model choice, and assuming a frequentist approach is used, pooled results should typically be reported using point estimates and 95 % CI as well as *z*- or *t*-based  $\alpha$  values. While not germane to meta-analysis, one should consider when reporting and interpreting results the recent recommendations in an editorial by Wasserstein *et al.*<sup>(3)</sup> as well as the rest of an entire issue of *The American Statistician* devoted to the use and over-reliance on 'statistical significance'. Similar recommendations were made in a recent commentary by Amrhein *et al.*<sup>(96)</sup>.

In addition to 95 % CI<sup>(96)</sup>, 95 % prediction intervals (PI) may also be reported when findings are pooled from those based on models such as random-effects<sup>(97)</sup>. The concept behind PI is that they tell one how effects are distributed around a summary effect<sup>(97)</sup>. This is in contrast to point estimates and CI, which provide an estimate of the overall effect and precision, respectively<sup>(97)</sup>. From an applied perspective, PI may make more sense because they help to determine uncertainty about whether an intervention works or not<sup>(97)</sup>. However, it has been recommended that caution be derived in drawing strong conclusions from 95 % PI because of coverage problems<sup>(98)</sup>. In addition, it has been suggested that because PI are calculated based on trials that are generally homogeneous, that is, patient populations and comparator treatments are interchangeable, the overall effect estimates may not be accurate if they do not meet this criterion<sup>(99)</sup>. As an example of PI use in nutrition, Cariolou *et al.* recently conducted an AD meta-analysis on the association between 25-hydroxyvitamin D deficiency and mortality in children with acute or critical conditions<sup>(100)</sup>. Based on a random-effects model, the pooled OR and 95 % CI of the risk of mortality in vitamin D deficient *v.* vitamin D non-deficient acute and critically ill children was 1.81 (95 % CI 1.24, 2.64). However, based on 95 % PI (0.71, 4.20), there was much less certainty, that is, wider intervals that also included 1, regarding this association<sup>(100)</sup>.

Similar to original studies, it is important to examine and report data on heterogeneity and inconsistency in meta-analysis. In meta-analysis, heterogeneity refers to any type of variability





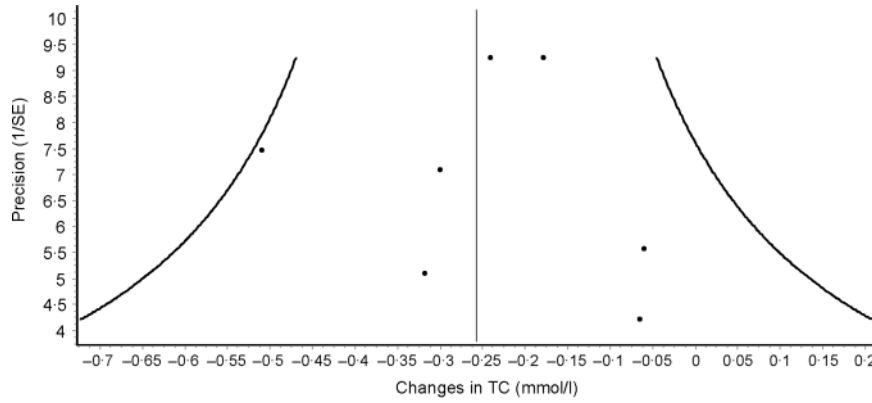
**Fig. 2.** Forest plot example of diet-induced changes in total cholesterol (TC) in adults based on the inverse variance heterogeneity (IVhet) model. The black squares represent mean changes in TC from each study while the left and right extremes of the squares represent the corresponding 95% CI, that is, compatibility intervals for the mean changes. The middle of the black diamond represents the pooled mean change in TC, while the left and right extremes of the diamond represent the corresponding 95% CI of the pooled mean change. The vertical dashed line represents the pooled mean change in TC while the solid vertical line represents zero (0) effect. As can be seen, the pooled 95% CI did not include zero (0), suggesting compatibility regarding the association between diet and reductions in TC. The results for Cochran's Q statistic, P value for Q and I<sup>2</sup> suggest a lack heterogeneity and inconsistency. The ES represents effect size changes in TC in mmol/l, while % weight represents the percentage weight attributed by each study to the overall pooled mean effect. Results were similar when the two results by Stefanick *et al.* were pooled into one overall ES. Data adapted from Kelley *et al.*<sup>(104)</sup>.

between studies and may be categorised broadly as clinical (patient characteristics, etc.), methodological (blinding, allocation concealment, etc.) and statistical (differences in outcome assessments, etc.)<sup>(6)</sup>. The Cochran Q statistic is typically used to examine heterogeneity<sup>(101)</sup>, while the I<sup>2</sup> statistic, an extension of Q, is used to examine inconsistency<sup>(102)</sup>. The Q statistic is a measure of statistical significance and given power problems, is typically reported as significant if the alpha (α) value is < 0.10 as opposed to < 0.05<sup>(102)</sup>. I<sup>2</sup> is a relative measure that ranges from 0 to 100%, with higher values representative of greater inconsistency<sup>(102)</sup>, while τ<sup>2</sup> is an absolute measure of between-study heterogeneity. However, like any statistic, Q, I<sup>2</sup> or τ<sup>2</sup> are not perfect with respect to explaining all the potential sources of heterogeneity<sup>(103)</sup>.

A standard graphical method of reporting results from each study as well as the overall pooled effect is through the use of a forest plot. An example of a forest plot using the IVhet model<sup>(92)</sup> is shown in Fig. 2<sup>(104)</sup>. While not common given the different ways in which data are reported, sample sizes as well as change outcome means and standard deviations from each intervention group may also be displayed in a forest plot. However, to reduce bias, including studies that only report data in exactly the same way is strongly discouraged if the overall treatment effect and variance from each study can be calculated from other reported statistics.

**Data synthesis (small-study effects).** An assessment for potential small-study effects (publication bias, etc.) is usually important in meta-analysis. Historically, this has most often been assessed qualitatively using some type of funnel plot and quantitatively using Egger's test<sup>(105)</sup>, though other methods exist for the assessment of both<sup>(106,107)</sup>. Briefly, a funnel plot is

a scatterplot in which the precision of each included study (standard error, inverse of the standard error, etc.) is plotted on the vertical (y) axis and the effect size for each included study (mean difference, standardised mean difference, OR, etc.) is plotted on the horizontal (x) axis. In the absence of small-study effects, the values should appear as an inverted funnel, with smaller sample size studies showing greater dispersion, that is, larger standard errors, at the bottom of the plot, while studies with larger sample sizes showing less dispersion towards the top. Smaller missing studies without statistically significant effects will lead to an asymmetrical appearance of the funnel plot with a gap in the bottom corner of the plot. However, the funnel plot can be difficult to interpret<sup>(108)</sup>. An example of a funnel plot using the same data as for the forest plot<sup>(104)</sup> is shown in Fig. 3. Egger's regression–intercept test is used for the Y intercept = 0 from a linear regression of a normalised effect estimate, that is, estimate divided by its standard error, against precision, that is, the reciprocal of the standard error of the estimate<sup>(105)</sup>. Unfortunately, the power to detect asymmetry with Egger's test is low when the number of studies is small<sup>(109)</sup>. Present recommendations suggest that if there are at least ten studies, a funnel plot and Egger's test may be used to examine for the small-study effects if the outcome of interest is continuous in nature, for example, changes in TC. However, since the time of the publication of these recommendations, an alternative qualitative (Doi plot) and quantitative (Luis Furuya-Kanamori (LFK) index) approach have been suggested to be more robust with respect to ease in visualising asymmetry (Doi plot) as well as greater diagnostic accuracy in differentiating between asymmetry and no asymmetry (LFK index)<sup>(107)</sup>. Rather than use a scatterplot, the Doi plot uses a normal quantile plot v. effect rather than precision



**Fig. 3.** Example of funnel plot based on diet-induced changes in total cholesterol (TC) following a dietary intervention. The solid vertical line represents the overall pooled mean change in TC in mmol/l after a dietary intervention. The x-axis represents changes in TC in mmol/l from each study while the y-axis represents the inverse of the standard error for changes in TC from each study. Each dot represents changes in TC plotted against its precision. In the absence of small-study effects, the plot should resemble a pyramid or inverted funnel, with scatter due to sampling variation. In the presence of potential small-study effects, the results from smaller studies with smaller/null findings will be missing in that region of the plot. While difficult to interpret, especially given the small number of effect estimates, there do not appear to be any small-study effects. Results were similar when the two results by Stefanick *et al.* were pooled into one overall effect size. Data adapted from Kelley *et al.*<sup>(104)</sup>.

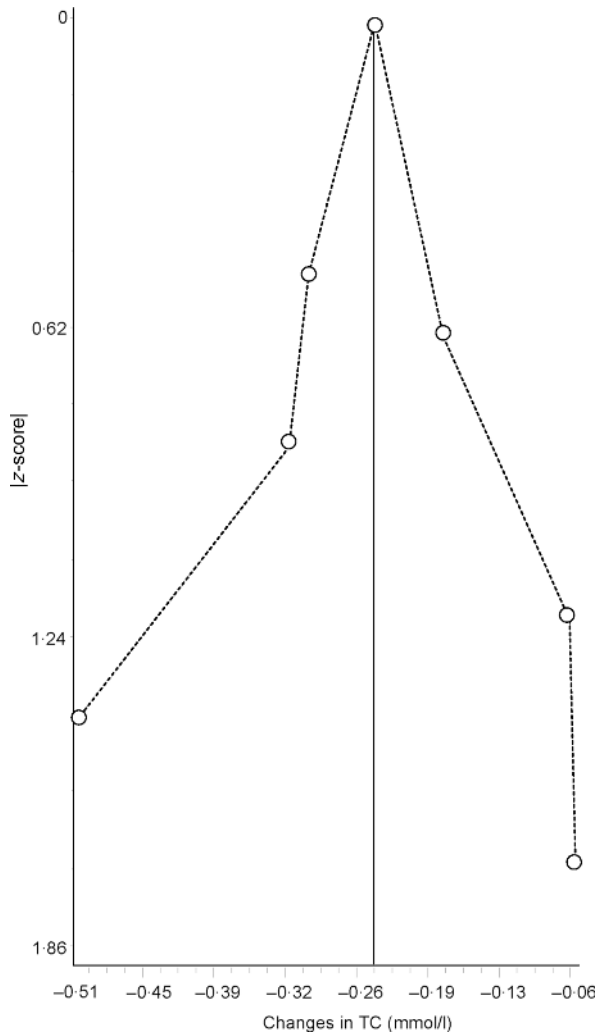
*v.* effect, providing better visualisation than a dot plot<sup>(107)</sup>. The LFK index, an index based on the Doi plot, assesses asymmetry quantitatively, with a value of zero (0) representing perfect symmetry, and thus, no apparent small-study effects<sup>(107)</sup>. It is based on the concept in which symmetry would be considered with respect to a vertical line on the horizontal (*x*) axis from the effect size with the lowest absolute *z* score on the Doi plot, dividing the plot into two regions with the same areas. The LFK index then quantifies the difference between these two regions in terms of the areas below the plot and the difference in the number of studies included in each arm of the plot<sup>(107)</sup>. Values  $\pm 1$ , greater than  $\pm 1$  and within  $\pm 2$  and greater than  $\pm 2$  are considered to represent no, minor and major asymmetry, respectively<sup>(107)</sup>. An example of the Doi plot and LFK index using the same data as for our previous examples is shown in Fig. 4.

**Data synthesis (influence and cumulative meta-analysis).** Many meta-analyses include a small number of trials. For example, it has been reported that the typical number of studies included in a Cochrane systematic review is six<sup>(110)</sup>. Given the former, it is usually relevant to conduct influence analysis with each study deleted from the model once in order to examine the effect that each study has on the overall results. Fig. 5 provides an example of influence analysis using the same data as for our other examples<sup>(104)</sup>.

In addition to influence analysis, it is often relevant to conduct cumulative meta-analysis, traditionally ranked by year of publication, to examine the accumulation of results over time<sup>(111)</sup>. The inclusion of findings from a cumulative meta-analysis can aid in making more educated choices based on past years of research as well as leading to more timely and increased use of successful interventions in practice<sup>(111)</sup>. Using this method, findings are pooled as each additional study is added to the model. An example of cumulative meta-analysis using the same data as for our previous examples is shown in Fig. 6.

**Data synthesis (subgroup and/or meta-regression analysis).** Given an adequate number of studies, subgroup and/or meta-regression may be conducted to explore the effect of selected

covariates, for example, age, on the outcome(s) of interest, for example, changes in fat mass as a result of a weight-loss intervention. Traditionally, these are based on weights derived from fixed and random-effects models, and more recently, approaches such as the IVhet and QE models, details for all of which have been described elsewhere<sup>(6,81,92,93,112,113)</sup>. While there may be a propensity for investigators to only conduct analyses when statistically significant and/or a large amount of inconsistency is found, this is generally not advised, given the current limitations of measures for heterogeneity and inconsistency<sup>(114)</sup>. With respect to the number of studies needed to conduct analyses such as meta-regression, currently no firm consensus exists regarding this. However, as a broad recommendation, and while understanding the potential arbitrariness of any definitive number given the numerous factors to consider, we support the recommendation of Fu *et al.*, in which there should be at least six studies per covariate for a continuous variable, for example, age, and at least four studies per group for a categorical variable, for example, sex (female, male)<sup>(115)</sup>. Exclusive of dose-response analyses, the four studies per group for a categorical variable is also recommended for any subgroup analyses conducted. If multiple meta-regression analysis is conducted, one should also consider conducting and reporting results for all simple meta-regression analyses performed. This may be especially relevant, given that such analyses in meta-analysis are considered to be exploratory. As a result, such findings would need to be tested in original studies because studies are not randomly allocated to covariates in meta-analysis. Consequently, they are regarded as observational. For categorical variables such as sex, there may be a lack of studies in one or more categories to conduct any type of meta-regression or subgroup comparisons. If this is the case, there are more than two categories, and it is scientifically plausible, one may collapse one or more categories, so that at least two exist. One can then conduct their meta-regression and/or subgroup analyses. If this is not possible, one may then consider additional forms of sensitivity analyses by omitting the results



**Fig. 4.** Example of Doi plot based on diet-induced changes in total cholesterol (TC) following a dietary intervention. The vertical line on the horizontal (x) axis represents the effect size (ES) with the lowest absolute z score, dividing the plot into two regions with the same areas. Visualisation of the plot suggests no asymmetry and thus no small-study effects such as publication bias. The obtained Luis Furuya-Kanamori index of 0.30 also suggests no asymmetry. Results were similar when the two results by Stefanick *et al.* were pooled into one overall ES. Data adapted from Kelley *et al.*<sup>(104)</sup>.

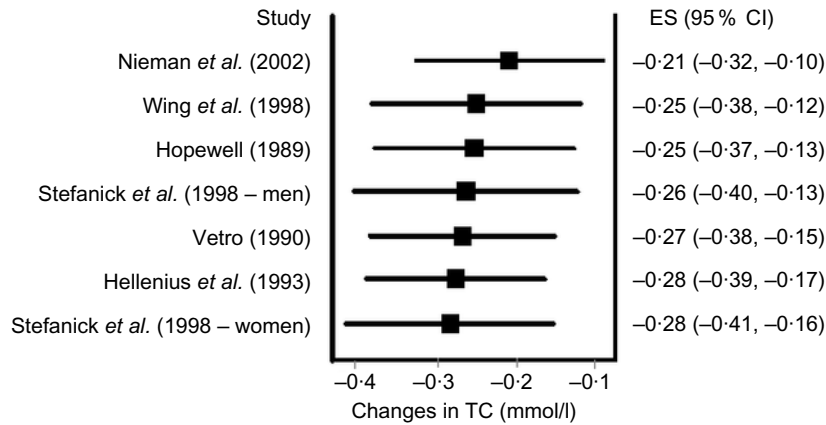
from the category with the smaller number of studies to see how it effects one's overall results. As an example, if there are results from ten studies, eight in males and two in females, one may choose to run their analyses with only the results from the males to see how it compares with the overall pooled results.

One aspect of meta-analysis in nutrition as well as other fields is that some studies conduct and report on highest *v.* lowest tertile comparisons. However, these are almost always difficult to interpret in terms of what nutritionists should recommend, given that there is overlap between studies with respect to what is considered high and low. Indeed, some low categories could be minimal and well below current recommended daily allowances while others could be considered close to pharmaceutical. Since nutritionists tend to prefer a recommended intake that can be applied to various populations and groups with confidence, it is recommended that any such comparisons

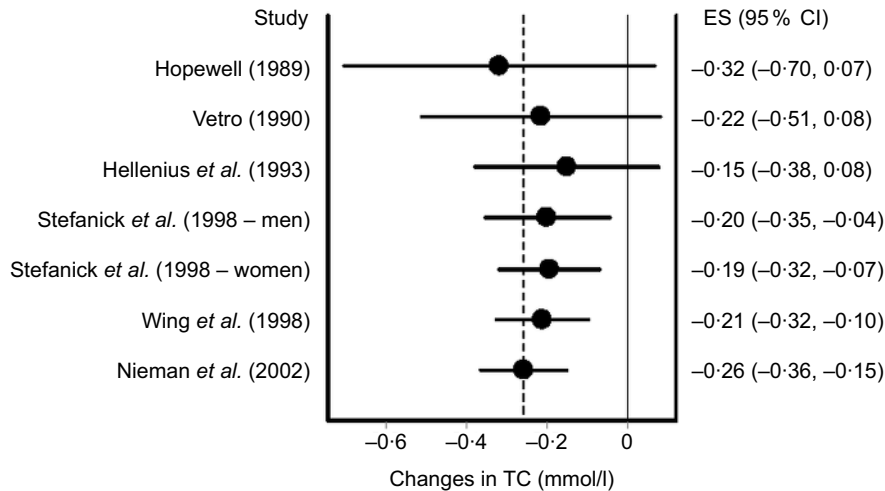
be conducted using a dose–response approach. This consists of modelling the association between the exposure and outcome to estimate the increase or decrease associated with one unit, or some other appropriate unit change, in exposure<sup>(32)</sup>. For example, using linear dose–response meta-analysis, Morze *et al.* found no significant associations between a 10-g/d increase in chocolate intake and heart failure (relative risk = 0.99, 95 % CI 0.94, 1.04) as well as type 2 diabetes (relative risk = 0.94, 95 % CI 0.88, 1.01)<sup>(116)</sup>. However, a small inverse association was observed for CHD (relative risk = 0.96, 95 % CI 0.93, 0.99), and stroke (relative risk = 0.90, 95 % CI 0.82, 0.98)<sup>(116)</sup>. Greenland & Longnecker<sup>(117)</sup>, Hartemink *et al.*<sup>(118)</sup> and Xu *et al.*<sup>(112)</sup> provide detailed information regarding dose–response methods for meta-analysis.

*Data synthesis (practically relevant information).* An aspect that is sometimes overlooked when conducting a meta-analysis is the need to provide practically relevant information to readers. In addition to reporting both absolute and relative results whenever possible, the use of metrics such as the number needed to treat (NNT)<sup>(6,119)</sup> and percentile improvement based on values such as Cohen's  $U_3$  index<sup>(120)</sup>, when appropriate, could be considered. For example, using the diet and TC data from our previous examples<sup>(104)</sup>, the method of Hasselblad and Hedges for estimating the NNT from continuous data<sup>(121)</sup>, and a control group risk of 30 %, the NNT for diet-associated reductions in TC was 5, meaning that one in five (20 %) people would reduce their TC if they dieted. Using the same data, Cohen's  $U_3$  index for percentile improvement was 16.9, meaning an improvement from the 50th to 66.9th percentile. In addition, one should also consider both the clinical and population health importance of any findings from a meta-analysis. For example, a 2-mmHg reduction in resting systolic blood pressure as a result of lower sodium intake may not be very important at the patient level but may have significant implications at the population level, given that lower sodium intake has been associated with a 4 % reduction in CHD and a 6 % reduction in stroke<sup>(122)</sup>.

*Data synthesis (strength of evidence).* An assessment for the strength of the evidence for the outcome(s) of interest should usually be conducted and reported. One of the most common instruments used is the GRADE instrument, details of which are provided elsewhere<sup>(123)</sup>. In brief, GRADE is a subjective tool that assesses the strength of evidence for a specific outcome across five areas: (1) risk of bias, (2) imprecision, (3) inconsistency, (4) indirectness and (5) publication bias<sup>(123)</sup>. For each of these items, the evidence can be rated down by one to two levels. There can also be an increase of one or two levels if there is a large effect and/or an increase of one level if either a dose–response relationship is observed or all plausible confounding would reduce the effect or increase the effect if no effect was identified<sup>(123)</sup>. For the GRADE instrument, risk of bias focuses on study limitations that include lack of allocation concealment and blinding, incomplete accounting of participants and outcome events, selective outcome reporting as well as any other limitations that reviewers believe may impact the outcome<sup>(123)</sup>. Imprecision is the degree of uncertainty about the findings and includes such things as a wide CI around the estimate of effect, while inconsistency signifies unexplained heterogeneity in results<sup>(123)</sup>. Indirectness is the evaluation of findings based



**Fig. 5.** Influence analysis based on the inverse variance heterogeneity model with each result deleted from the overall analysis once. The black squares represent mean changes in total cholesterol (TC) with the corresponding study deleted from the model, while the left and right extremes of the squares represent the corresponding 95% CI for the mean changes. As can be seen, changes ranged from  $-0.21$  to  $-0.28$  mmol/l with non-overlapping 95% CI for all. These findings suggest that no one result had a significant impact on the overall findings. Results were similar when the two results by Stefanick *et al.* were pooled into one overall effect size (ES). Data adapted from Kelley *et al.*<sup>(104)</sup>.



**Fig. 6.** Cumulative meta-analysis ranked by year and based on the inverse variance heterogeneity model. The black circles represent mean changes in total cholesterol (TC) with the corresponding study, and all earlier studies pooled while the left and right extremes of the circles represent the corresponding 95% CI for the mean pooled changes. As can be seen, non-overlapping 95% CI have been observed since 1998. Results were similar when the two results by Stefanick *et al.* were pooled into one overall effect size (ES). Data adapted from Kelley *et al.*<sup>(104)</sup>.

on whether the included studies directly compare the interventions and populations in which one is interested in as well as measuring outcomes believed to be important by participants, for example, self-reported health-related quality of life as a result of weight loss in obese participants. Lastly, publication bias is the selective publication of studies in which improvements are embellished and harms are underestimated<sup>(123)</sup>. The overall certainty of the evidence is then rated by the authors as either (1) very low, (2) low, (3) moderate or (4) high<sup>(123)</sup>. As an example of the use of the GRADE instrument in nutrition, Baranski *et al.* rated the overall strength of evidence as moderate or high for the majority of parameters for which significant differences were detected in a systematic review with meta-analysis on differences in composition between organic and non-organic crops and crop-based foods<sup>(124)</sup>.

### Discussion and conclusions

Where appropriate, the discussion and conclusions sections of a systematic review with meta-analysis should include (1) a summary of the overall findings, (2) a discussion of how the findings compare with previous research on the topic, (3) the potential clinical, public health and policy implications of the findings, (4) directions for future research with respect to both the reporting of future studies on the topic and additional studies that might be needed, for example, the dose-response effects of vitamin D on bone mineral density and (5) the strengths and potential limitations of one's systematic review with meta-analysis. With respect to the latter, one of the inherent limitations of any AD systematic review with meta-analysis is the potential for ecological fallacy<sup>(125)</sup>. The PRISMA guidelines provide



greater details regarding items to include in the discussion and conclusion sections of a systematic review with meta-analysis<sup>(30)</sup>.

With respect to interpretation on the part of the consumer, the results of a systematic review with meta-analysis should be considered, broadly, with respect to several potential factors. First and foremost, were any significant findings also found practically important? Second, were the included studies representative of the population, exposures and outcomes that one is interested in and deemed to be important? Third, do any potential benefits outweigh the risks involved? Fourth, is the evidence considered to be strong?

Finally, meta-analysis, like many fields today, is progressing at a rapid pace. As a result, it is very difficult for generic statisticians, biostatisticians and other relevant professionals to stay current unless they have a specific and current focus in this burgeoning field. Given the former, we strongly recommend that not only a content expert but also a meta-analytic expert be included in any meta-analysis that is conducted.

### Conclusion

The number of systematic reviews, with or without meta-analysis, is increasing in the field of nutrition. The purpose of this article was to provide a non-technical introduction to producers, reviewers and consumers of these important reviews, with a focus on nutrition. It is the hope that this information will be helpful to producers, reviewers, and consumers in the field of nutrition.

### Acknowledgements

No funding was received for this work.

G. A. K. was responsible for the conception and design, acquisition of data, analysis and interpretation of data, drafting the initial manuscript and revising it critically for important intellectual content. K. S. K. was responsible for the conception and design, acquisition of data, drafting the initial manuscript and revising all drafts critically for important intellectual content. Both authors read and approved the final manuscript.

There are no conflicts of interest.

### Patient consent

Not required.

### Data sharing statement

All data are available upon request from the corresponding author.

### References

1. Sacks HS, Berrier J, Reitman D, *et al.* (1987) Meta-analysis of randomized controlled trials. *N Engl J Med* **316**, 450–455.
2. Zhang Y, Akl EA & Schunemann HJ (2018) Using systematic reviews in guideline development: the GRADE approach. *Res Synth Methods* (publication ahead of print version 14 July 2018).

3. Wasserstein RL, Schirm AL & Lazar NA (2019) Moving to a world beyond “ $p < 0.05$ ”. *Am Stat* **73**, 1–19.
4. Daudt HM, van Mossel C & Scott SJ (2013) Enhancing the scoping study methodology: a large, inter-professional team’s experience with Arksey and O’Malley’s framework. *BMC Med Res Methodol* **13**, 48.
5. Amouzandeh C, Fingland D & Vidgen HA (2019) A scoping review of the validity, reliability and conceptual alignment of food literacy measures for adults. *Nutrients* **11**, E801.
6. Higgins JPT & Green S (editors) (2011) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration. [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
7. Tricco AC, Lillie E, Zarin W, *et al.* (2018) PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation: the PRISMA-ScR statement. *Ann Intern Med* **169**, 467–473.
8. Agostoni C, Guz-Mark A, Marderfeld L, *et al.* (2019) The long-term effects of dietary nutrient intakes during the first 2 years of life in healthy infants from developed countries: an umbrella review. *Adv Nutr* **10**, 489–501.
9. Ioannidis JPA (2016) The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q* **94**, 485–514.
10. Kelley GA & Kelley KS (2018) Systematic reviews and cancer research: a suggested stepwise approach. *BMC Cancer* **18**, 9.
11. Cochrane (2016) Editorial and publishing policy resource. <http://community.cochrane.org/editorial-and-publishing-policy-resource> (accessed November 2017).
12. Shojania KG, Sampson M, Ansari MT, *et al.* (2007) *Updating Systematic Reviews: Technical Review No. 16*. Rockville, MD: Agency for Healthcare Research and Quality.
13. Garner P, Hopewell S, Chandler J, *et al.* (2016) When and how to update systematic reviews: consensus and checklist. *BMJ* **354**, i3507.
14. Shea BJ, Reeves BC, Wells G, *et al.* (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* **358**, j4008.
15. Whiting P, Savovic J, Higgins JP, *et al.* (2016) ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* **69**, 225–234.
16. Guyatt G, Oxman AD, Akl EA, *et al.* (2011) GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* **64**, 383–394.
17. Whiting PF, Rutjes AW, Westwood ME, *et al.* (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* **155**, 529–536.
18. McKenzie JE & Brennan SE (2017) Overviews of systematic reviews: great promise, greater challenge. *Syst Rev* **6**, 185.
19. Lunny C, Brennan SE, McDonald S, *et al.* (2017) Toward a comprehensive evidence map of overview of systematic review methods: paper 1-purpose, eligibility, search and data extraction. *Syst Rev* **6**, 231.
20. Lunny C, Brennan SE, McDonald S, *et al.* (2018) Toward a comprehensive evidence map of overview of systematic review methods: paper 2-risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Syst Rev* **7**, 31.
21. Ballard M & Montgomery P (2017) Risk of bias in overviews of reviews: a scoping review of methodological guidance and four-item checklist. *Res Synth Methods* **8**, 92–108.
22. Gates A, Gates M, Duarte G, *et al.* (2018) Evaluation of the reliability, usability, and applicability of AMSTAR, AMSTAR 2, and ROBIS: protocol for a descriptive analytic study. *Syst Rev* **7**, 85.

23. Pieper D, Waltering A, Holstiege J, *et al.* (2018) Quality ratings of reviews in overviews: a comparison of reviews with and without dual (co-)authorship. *Syst Rev* **7**, 63.
24. Hunt H, Pollock A, Campbell P, *et al.* (2018) An introduction to overviews of reviews: planning a relevant research question and objective for an overview. *Syst Rev* **7**, 39.
25. Fusar-Poli P & Radua J (2018) Ten simple rules for conducting umbrella reviews. *Evid Based Ment Health* **21**, 95–100.
26. Pollock A, Campbell P, Brunton G, *et al.* (2017) Selecting and implementing overview methods: implications from five exemplar overviews. *Syst Rev* **6**, 145.
27. Pieper D, Pollock M, Fernandes RM, *et al.* (2017) Epidemiology and reporting characteristics of overviews of reviews of healthcare interventions published 2012–2016: protocol for a systematic review. *Syst Rev* **6**, 73.
28. Pollock M, Fernandes RM, Becker LA, *et al.* (2016) What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. *Syst Rev* **5**, 190.
29. Calder PC, Campoy C, Eilander A, *et al.* (2019) A systematic review of the effects of increasing arachidonic acid intake on PUFA status, metabolism and health-related outcomes in humans. *Br J Nutr* **121**, 1201–1214.
30. Liberati A, Altman DG, Tetzlaff J, *et al.* (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* **151**, W65–W94.
31. Zhang S, Zhang F, Du M, *et al.* (2019) Efficacy and safety of iron supplementation in patients with heart failure and iron deficiency: a meta-analysis. *Br J Nutr* **121**, 841–848.
32. Dekkers OM, Vandenbroucke JP, Cevallos M, *et al.* (2019) COSMOS-E: guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. *PLoS Med* **16**, e1002742.
33. Gaksch M, Jorde R, Grimnes G, *et al.* (2017) Vitamin D and mortality: individual participant data meta-analysis of standardized 25-hydroxyvitamin D in 26916 individuals from a European consortium. *PLOS ONE* **12**, e0170791.
34. Riley RD, Lambert PC & Abo-Zaid G (2010) Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* **340**, c221.
35. Riley RD (2010) Commentary: like it and lump it? Meta-analysis using individual participant data. *Int J Epidemiol* **39**, 1359–1361.
36. Kelley GA, Kelley KS & Tran ZV (2002) Retrieval of individual patient data for an exercise meta-analysis. *Am J Med Sport* **4**, 350–354.
37. Riley RD, Simmonds MC & Look MP (2007) Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *J Clin Epidemiol* **60**, 431–439.
38. Kelley GA & Kelley KS (2016) Retrieval of individual participant data for exercise meta-analyses may not be worth the time and effort. *Biomed Res Int* **2016**, 5059041.
39. Polanin JR (2018) Efforts to retrieve individual participant data sets for use in a meta-analysis result in moderate data sharing but many data sets remain missing. *J Clin Epidemiol* **98**, 157–159.
40. Riley RD, Lambert PC, Staessen JA, *et al.* (2008) Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med* **27**, 1870–1893.
41. Steinberg KK, Smith SJ, Stroup DF, *et al.* (1997) Comparison of effect size estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. *Am J Epidemiol* **145**, 917–925.
42. Cooper H & Patall EA (2009) The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychol Methods* **14**, 165–176.
43. Olkin I & Sampson A (1998) Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* **54**, 317–322.
44. Mathew T & Nordstrom K (1999) On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics* **55**, 1221–1223.
45. Tudur Smith C, Marcucci M, Nolan SJ, *et al.* (2016) Individual participant data meta-analyses compared with meta-analyses based on aggregate data. *Cochrane Database Syst Rev*, issue 9, MR000007.
46. Smelt AF, Gussekloo J, Bermingham LW, *et al.* (2018) The effect of vitamin B<sub>12</sub> and folic acid supplementation on routine haematological parameters in older people: an individual participant data meta-analysis. *Eur J Clin Nutr* **72**, 785–795.
47. Stewart LA, Clarke M, Rovers M, *et al.* (2015) Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* **313**, 1657–1665.
48. Tierney JF, Vale C, Riley R, *et al.* (2015) Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med* **12**, e1001855.
49. Schwingshackl L, Buyken A & Chaimani A (2019) Network meta-analysis reaches nutrition research. *Eur J Nutr* **58**, 1–3.
50. Galaviz KI, Weber MB, Straus A, *et al.* (2018) Global diabetes prevention interventions: a systematic review and network meta-analysis of the real-world impact on incidence, weight, and glucose. *Diabetes Care* **41**, 1526–1534.
51. Hutton B, Salanti G, Caldwell DM, *et al.* (2015) The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* **162**, 777–784.
52. Laws A, Kendall R & Hawkins N (2014) A comparison of national guidelines for network meta-analysis. *Value Health* **17**, 642–654.
53. Rouse B, Chaimani A & Li TJ (2017) Network meta-analysis: an introduction for clinicians. *Intern Emerg Med* **12**, 103–111.
54. Riley RD, Jackson D, Salanti G, *et al.* (2017) Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ* **358**, j3932.
55. Doi SAR & Barendregt JJ (2018) A generalized pairwise modelling framework for network meta-analysis. *Int J Evid Based Healthc* **16**, 187–194.
56. Brittain EH, Fay MP & Follmann DA (2012) A valid formulation of the analysis of noninferiority trials under random effects meta-analysis. *Biostatistics* **13**, 637–649.
57. Schmidli H, Wandel S & Neuenschwander B (2013) The network meta-analytic-predictive approach to non-inferiority trials. *Stat Methods Med Res* **22**, 219–240.
58. Acuna SA, Chesney TR, Ramjist JK, *et al.* (2019) Laparoscopic versus open resection for rectal cancer: a noninferiority meta-analysis of quality of surgical resection outcomes. *Ann Surg* **269**, 849–855.
59. Acuna SA, Chesney TR, Amarasekera ST, *et al.* (2018) Defining non-inferiority margins for quality of surgical resection for rectal cancer: a Delphi consensus study. *Ann Surg Oncol* **25**, 3171–3178.

60. Liberati A & D'Amico R (2010) Commentary: the debate on non-inferiority trials: 'when meta-analysis alone is not helpful'. *Int J Epidemiol* **39**, 1582–1583.
61. Beller EM, Glasziou PP, Altman DG, *et al.* (2013) PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med* **10**, e1001419.
62. Saint S, Christakis DA, Saha S, *et al.* (2000) Journal reading habits of internists. *J Gen Intern Med* **15**, 881–884.
63. Yamamoto JM, Kellett JE, Balsells M, *et al.* (2018) Gestational diabetes mellitus and diet: a systematic review and meta-analysis of randomized controlled trials examining the impact of modified dietary interventions on maternal glucose control and neonatal birth weight. *Diabetes Care* **41**, 1346–1361.
64. Page MJ, Shamseer L & Tricco AC (2018) Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Syst Rev* **7**, 32.
65. Stewart L, Moher D & Shekelle P (2012) Why prospective registration of systematic reviews makes sense. *Syst Rev* **1**, 7.
66. Asghari G, Farhadnejad H, Hosseinpanah F, *et al.* (2018) Effect of vitamin D supplementation on serum 25-hydroxyvitamin D concentration in children and adolescents: a systematic review and meta-analysis protocol. *BMJ Open* **8**, e021636.
67. Shamseer L, Moher D, Clarke M, *et al.* (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* **349**, g7647.
68. Denova-Gutierrez E, Mendez-Sanchez L, Munoz-Aguirre P, *et al.* (2018) Dietary patterns, bone mineral density, and risk of fractures: a systematic review and meta-analysis. *Nutrients* **10**, E1922.
69. Hidayat K, Chen GC, Zhang R, *et al.* (2016) Calcium intake and breast cancer risk: meta-analysis of prospective cohort studies. *Br J Nutr* **116**, 158–166.
70. van Driel ML, De Sutter A, De Maeseneer J, *et al.* (2009) Searching for unpublished trials in Cochrane reviews may not be worth the effort. *J Clin Epidemiol* **62**, 838–844.
71. Bramer WM, Rethlefsen ML, Kleijnen J, *et al.* (2017) Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev* **6**, 245.
72. Vine R (2006) Google Scholar. *J Med Libr Assoc* **94**, 97–99.
73. Burnham JF (2006) Scopus database: a review. *Biomed Digit Libr* **3**, 1.
74. Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* **70**, 213–220.
75. Pedder H, Sarri G, Keeney E, *et al.* (2016) Data extraction for complex meta-analysis (DECIMAL) guide. *Syst Rev* **5**, 212.
76. Sanderson S, Tatt ID & Higgins JP (2007) Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* **36**, 666–676.
77. Seehra J, Pandis N, Koletsis D, *et al.* (2016) Use of quality assessment tools in systematic reviews was varied and inconsistent. *J Clin Epidemiol* **69**, 179–184.
78. Higgins JPT, Sterne JAC, Savović J, *et al.* (2016) A revised tool for assessing risk of bias in randomized trials. *Cochrane Database Syst Rev* **10**, Suppl. 1, 29–31.
79. Sterne JA, Hernán MA, Reeves BC, *et al.* (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* **355**, i4919.
80. Firth J, Marx W, Dash S, *et al.* (2019) The effects of dietary improvement on symptoms of depression and anxiety: a meta-analysis of randomized controlled trials. *Psychosom Med* **81**, 265–280.
81. Borenstein M, Hedges L, Higgins J, *et al.* (2009) *Introduction to Meta-analysis*. Chichester, West Sussex: John Wiley & Sons.
82. Sacks HS, Chalmers TC, Smith H (1982) Randomized versus historical controls for clinical trials. *Am J Med* **72**, 233–240.
83. Schulz KF, Chalmers I, Hayes R, *et al.* (1995) Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J Am Med Assoc* **273**, 408–412.
84. DerSimonian R & Kacker R (2007) Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* **28**, 105–114.
85. Dersimonian R & Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* **7**, 177–188.
86. Dersimonian R & Laird N (2015) Meta-analysis in clinical trials revisited. *Contemp Clin Trials* **45**, 139–145.
87. Biggerstaff BJ & Tweedie RL (1997) Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Stat Med* **16**, 753–768.
88. Sidik K & Jonkman JN (2002) A simple confidence interval for meta-analysis. *Stat Med* **21**, 3153–3159.
89. Sidik K & Jonkman JN (2007) A comparison of heterogeneity variance estimators in combining results of studies. *Stat Med* **26**, 1964–1981.
90. Zeng D & Lin DY (2015) On random-effects meta-analysis. *Biometrika* **102**, 281–294.
91. Poole C & Greenland S (1999) Random-effects meta-analyses are not always conservative. *Am J Epidemiol* **150**, 469–475.
92. Doi SA, Barendregt JJ, Khan S, *et al.* (2015) Advances in the meta-analysis of heterogeneous clinical trials I: the inverse variance heterogeneity model. *Contemp Clin Trials* **45**, 130–138.
93. Doi SA, Barendregt JJ, Khan S, *et al.* (2015) Advances in the meta-analysis of heterogeneous clinical trials II: the quality effects model. *Contemp Clin Trials* **45**, 123–129.
94. Doi SAR, Furuya-Kanamori L, Thalib L, *et al.* (2017) Meta-analysis in evidence-based healthcare: a paradigm shift away from random effects is overdue. *Int J Evid Based Healthc* **15**, 152–160.
95. Barendregt JJ & Doi SA (2016) *Meta XL*, 5.3 ed. Queensland, Australia: EpiGear International Pty Ltd.
96. Amrhein V, Greenland S & McShane B (2019) Scientists rise up against statistical significance. *Nature* **567**, 305–307.
97. Higgins JP, Thompson SG & Spiegelhalter DJ (2009) A re-evaluation of random-effects meta-analysis. *J R Stat Soc Series A* **172**, 137–159.
98. Partlett C & Riley RD (2017) Random effects meta-analysis: coverage performance of 95% confidence and prediction intervals following REML estimation. *Stat Med* **36**, 301–317.
99. Kriston L (2013) Dealing with clinical heterogeneity in meta-analysis. Assumptions, methods, interpretation. *Int J Methods Psychiatr Res* **22**, 1–15.
100. Cariolou M, Cupp MA, Evangelou E, *et al.* (2019) Importance of vitamin D in acute and critically ill children with subgroup analyses of sepsis and respiratory tract infections: a systematic review and meta-analysis. *BMJ Open* **9**, e027666.
101. Cochran WG (1954) The combination of estimates from different experiments. *Biometrics* **10**, 101–129.
102. Higgins JPT, Thompson SG, Deeks JJ, *et al.* (2003) Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560.
103. Ioannidis JP, Patsopoulos NA & Evangelou E (2007) Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* **335**, 914–916.
104. Kelley GA, Kelley KS, Roberts S, *et al.* (2012) Comparison of aerobic exercise, diet or both on lipids and lipoproteins in

- adults: a meta-analysis of randomized controlled trials. *Clin Nutr* **31**, 156–167.
105. Egger M, Davey Smith G, Schneider M, *et al.* (1997) Bias in meta-analysis detected by a simple graphical test. *BMJ* **315**, 629–634.
  106. Sterne JAC, Gavaghan D & Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* **53**, 1119–1129.
  107. Furuya-Kanamori L, Barendregt JJ & Doi SAR (2018) A new improved graphical and quantitative method for detecting bias in meta-analysis. *Int J Evid Based Healthc* **16**, 195–203.
  108. Lau J, Ioannidis JP, Terrin N, *et al.* (2006) The case of the misleading funnel plot. *BMJ* **333**, 597–600.
  109. Sterne JA, Sutton AJ, Ioannidis JP, *et al.* (2011) Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* **343**, d4002.
  110. Mallett S & Clarke M (2002) The typical Cochrane review. How many trials? How many participants? *Int J Technol Assess Health Care* **18**, 820–823.
  111. Clarke M, Brice A & Chalmers I (2014) Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLOS ONE* **9**, e102670.
  112. Xu C & Doi SAR (2018) The robust error meta-regression method for dose-response meta-analysis. *Int J Evid Based Healthc* **16**, 138–144.
  113. Lopez-Lopez JA, Van den Noortgate W, Tanner-Smith EE, *et al.* (2017) Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: a Monte Carlo simulation. *Res Synth Methods* **8**, 435–450.
  114. Higgins J, Thompson S, Deeks J, *et al.* (2002) Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy* **7**, 51–61.
  115. Fu R, Gartlehner G, Grant M, *et al.* (2011) Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol* **64**, 1187–1197.
  116. Morze J, Schwedhelm C, Bencic A, *et al.* (2019) Chocolate and risk of chronic disease: a systematic review and dose-response meta-analysis. *Eur J Nutr* (epublication ahead of print version 25 February 2019).
  117. Greenland S & Longnecker MP (1992) Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* **135**, 1301–1309.
  118. Hartemink N, Boshuizen HC, Nagelkerke NJ, *et al.* (2006) Combining risk estimates from observational studies with different exposure cutpoints: a meta-analysis on body mass index and diabetes type 2. *Am J Epidemiol* **163**, 1042–1052.
  119. da Costa BR, Rutjes AW, Johnston BC, *et al.* (2012) Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. *Int J Epidemiol* **41**, 1445–1459.
  120. Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
  121. Hasselblad V & Hedges LV (1995) Meta-analysis of screening and diagnostic tests. *Psychol Bull* **117**, 167–178.
  122. Stamler J, Rose G, Stamler R, *et al.* (1989) INTERSALT study findings. Public health and medical care implications. *Hypertension* **14**, 570–577.
  123. Guyatt GH, Oxman AD, Vist GE, *et al.* (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336**, 924–926.
  124. Baranski M, Srednicka-Tober D, Volakakis N, *et al.* (2014) Higher antioxidant and lower cadmium concentrations and lower incidence of pesticide residues in organically grown crops: a systematic literature review and meta-analyses. *Br J Nutr* **112**, 794–811.
  125. Rucker G & Schumacher M (2008) Simpson's paradox visualized: the example of the rosiglitazone meta-analysis. *BMC Med Res Methodol* **8**, 34.