Issues related to the conduct of systematic reviews: a focus on the nutrition field\textsuperscript{1–3}

David Moher and Andrea C Tricco

\textbf{ABSTRACT}

Systematic reviews (SRs) are an increasingly popular evidence-based tool and are often used to answer complex research questions across many different research domains. Early SR methodology was advanced by social scientists, and the term "meta-analysis" was coined by a social scientist who also conducted research in psychology. SRs have recently become popular in healthcare and are likely to be beneficial in any field. The aim of this report is to highlight issues in SR conduct with a focus on the field of nutrition and to make recommendations on improving SR conduct in this area. Development of the research question is probably the most important step in conducting an SR. The 4 main components of an answerable question are 1) the patient, population, or problem; 2) the intervention, independent variable, or exposure; 3) the comparators; and 4) the dependent variables or outcomes of interest. The question will be used to determine the optimal methods for conducting the SR. SRs often include study designs beyond randomized trials and do not always include a meta-analysis of the results. Other topics explored include understanding and interpreting discordant reviews and the importance of reporting tools [eg, QUality Of Reporting Of Meta-analyses (QUOROM Statement) or CONsolidated Standards Of Reporting of Trials (CONSORT Statement)]. Recommendations are then provided, such as developing a capacity-building program, searching the primary literature for research gaps, and extending reporting tools such as the QUOROM Statement to the field of nutrition. \textit{Am J Clin Nutr} 2008;88:1191–9.

\textbf{INTRODUCTION}

Synthesizing evidence from research dates back to the early 1900s, yet improved methods for evidence synthesis have been developed only within the past few years (1). Early systematic review (SR) methods were advanced by social scientists (1), and the term "meta-analysis" was coined in 1976 by a social scientist who also did research in the area of psychology (2). Although the importance of evidence synthesis in medicine was recognized in the 1970s (3), widespread use of these techniques did not occur until 2 decades later (1). Potentially contributing to this "move- ment" was evidence that the judgments and opinions of experts were often biased. A pivotal study by Antman et al (4) concluded that effective therapies were not mentioned, whereas ineffective and potentially harmful therapies often were recommended by experts. The term "evidence-based medicine" was coined only \approx 15 \text{y ago} (5).

Over the past 2 decades, SRs have become increasingly popular in evidence-based healthcare (6). Policy makers, clinicians, and others have little time to keep up with the published literature. SRs offer a convenient way for them to keep up-to-date with the current evidence. Furthermore, SRs are often used in the development of clinical practice guidelines (7) and are increasingly viewed as a useful tool for health decision makers (8, 9).

The increased utility of SRs has probably contributed to an increase in their publication rates. It was recently estimated that, each year, MEDLINE cites 2500 new SRs in English (10)—a rate 166 times that reported in 1991 (11).

Because of the utility of SRs, the SR approach is likely to benefit any field (12), including nutrition. The aims of this report are to highlight issues in SR conduct with a focus on the field of nutrition and to make recommendations for improving SR conduct in this area.

\textbf{METHODS}

\textbf{Terminology}

The terminology used to describe an SR and a meta-analysis has evolved over time. There is no consistent use of these terms, but we will adopt the definitions used by The Cochrane Collaboration (13). An SR consists of a clearly formulated question and explicit methods to identify, select, and critically appraise relevant research and collects and analyzes data from the studies that are included in the review. A meta-analysis is the use of statistical techniques in an SR, which integrates the results of included studies. Thus an SR does not necessarily include a meta-analysis.

Other types of reviews, such as narrative reviews, will not use the explicit methods outlined in the definition proposed by The Cochrane Collaboration (Table 1) (10). As such, these other types of reviews may be susceptible to bias (4). A properly conducted

\textbf{Table 1}

Perspective

---

\textsuperscript{1} From the Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Canada (DM and ACT) and the Department of Epidemiology & Community Medicine, Faculty of Medicine (DM), and the Institute of Population Health, University of Ottawa, Ottawa, Canada (ACT).

\textsuperscript{2} Supported by the International Life Sciences Institute Research Foundation, a University of Ottawa Research Chair (to DM), and a Canadian Institutes of Health Research Canada Graduate Scholarship and a University of Ottawa Excellence Scholarship (to ACT). The sponsor had no involvement in the writing of the manuscript or the decision to submit the manuscript for publication.

\textsuperscript{3} Reprints not available. Address correspondence to D Moher, Clinical Epidemiology Program, Ottawa Health Research Institute, Box 208, 501 Smyth Road, Ottawa, ON K1H 8L6, Canada. E-mail: dmoher@uottawa.ca.

Received April 4, 2008. Accepted for publication July 30, 2008. doi: 10.3945/ajcn.2008.26255.
SRs are currently being conducted in many different areas and have been used to address many types of research questions (Table 2). For example, an SR was used to answer the following question within the crime and justice domain: “Does a neighborhood watch reduce crime?” (15). Another SR was used to answer the question within education: “Which approaches to parental involvement improve the academic performance of elementary school–age children?” (16). SRs have been used to answer many complex questions within the field of nutrition, including “Which community-based nutrition and physical activity interventions are effective at preventing chronic disease in low-income populations?” (17); “What are the long-term effects of advice to restrict dietary sodium in adults with and without hypertension?” (18); and “What is the association between a diet rich in whole grain, bran, and germ and the risk of type 2 diabetes?” (19). Other real-life examples of questions answered by SRs are provided elsewhere (12).

After deciding who the SR team will comprise, developing a clear and concise question is probably the most important step in conducting an SR. The 4 components of an answerable question are: (1) the patient, population, or problem (P); (2) the intervention, independent variable, or exposure (I); (3) the comparators (C); and (4) the dependent variables or outcomes of interest (O) (27). Sometimes an additional component is added, namely, study design (S), which is used to limit the SR to certain types of studies, such as cohort studies. These components are known collectively as PICO (1). [Variations of PICO also exist, such as one that adds a “D” for study design (PICO-D) and one that incorporates a “T” for timing and an “S” for setting (PICO-TS).]

An example of the use of these components in creating a question is that, to examine the sodium intake in the United States, the question could be phrased as “What are the health-related effects (O) associated with a high-salt diet (I) compared with a low-salt diet (C) among persons living in the United States (P)?” To examine the association between fruit and vegetable intake and cardiovascular disease among diabetics, the question could be phrased as “Among persons with diabetes (P), what is the association of a high fruit and vegetable diet (I) compared with a diet low in fruits and vegetables (C) among persons living in the United States (P)?”

### Table 1

<table>
<thead>
<tr>
<th>Systematic review (minimum criteria)</th>
<th>Other types of reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>A protocol (ie, working plan for the systematic review) is developed, along with a clearly formulated question.</td>
<td>A protocol is not used, and the question may not be clear and concise.</td>
</tr>
<tr>
<td>The literature is searched broadly by using many different methods (eg, searching multiple relevant databases, contacting experts, or scanning reference lists).</td>
<td>The literature may not be searched, or only the literature that the authors are aware of is searched. The search strategy is not reported or is not fully reported in the review manuscript.</td>
</tr>
<tr>
<td>The search strategy, including databases and years searched and search terms employed, is transparently reported in the systematic review manuscript.</td>
<td>No eligibility criteria are set, and the authors are free to TAB and choose which studies should be included in the review. The inclusion criteria and the number of excluded articles are not reported in the manuscript.</td>
</tr>
<tr>
<td>To determine study eligibility, the literature is screened by using criteria set a priori. Ideally, 2 independent reviewers screen all material and resolve conflicts through discussion. The eligibility criteria, including the number of articles that were excluded and the reasons for exclusion, are transparently reported in the systematic review manuscript.</td>
<td>The quality of the included studies is not assessed, or unvalidated instruments are used to appraise the quality of included studies.</td>
</tr>
<tr>
<td>The risk of bias of the included studies is assessed by using validated and applicable study appraisal instruments to determine the validity of the study results. Ideally, 2 reviewers independently appraise the quality of all included studies and resolve conflicts through discussion.</td>
<td>The authors are free to TAB and choose which results from the included studies to report. Data may not be abstracted consistently from all included studies. The data abstraction form is not described.</td>
</tr>
<tr>
<td>The literature is searched broadly by using many different methods (eg, searching multiple relevant databases, contacting experts, or scanning reference lists).</td>
<td>Studies are summarized on the basis of the results that are most appealing to the authors. A meta-analysis may be performed with studies that are not homogenous.</td>
</tr>
<tr>
<td>The results are synthesized by using the totality of evidence. A meta-analysis may be conducted if the included studies are deemed homogenous in terms of study population, study design, exposure or intervention examined, comparators studied, and outcomes assessed.</td>
<td>The discussion section may not provide an overall summary of the weaknesses of included studies or of the weaknesses of the review itself.</td>
</tr>
<tr>
<td>The discussion section provides an overall summary of the strengths and weaknesses of the included studies and a summary of the strengths and weaknesses of the systematic review itself.</td>
<td></td>
</tr>
</tbody>
</table>

* One way to determine whether a study is a systematic review or another type of review, such as a narrative review, is to examine whether there is a methods section in the report. If a methods section is missing from the study report, chances are the study was not a systematic review, and, thus, it is susceptible to considerable bias in the results.
A distinctive characteristic of Cochrane reviews is that they often answer questions regarding the effectiveness or efficacy of an intervention (ie, “what works”). As such, The Cochrane Collaboration relies strongly on the synthesizing of evidence from randomized controlled trials (RCTs). However, there are some exceptions to this within Cochrane. An example is the Effective Practice and Organisation of Care (EPOC) group, which often addresses broader questions and includes other study designs.

RCTs are given high prominence within the Cochrane Collaboration because they minimize the influence of bias in their results. Whereas there is considerable evidence to support this perspective and the importance of randomized trials, there are also data indicating that such study designs account for only ≈10% of the healthcare literature (28, 29). In contrast, the Campbell Collaboration, the sister group to The Cochrane Collaboration, addresses broad questions related to crime and justice, education, and social welfare and includes all types of study designs (eg, RCTs, cross-sectional studies, case-control studies, cohort studies, and case series) (Table 2). Campbell reviews are conducted according to the guidance provided on the organization’s website (Internet: www.campbellcollaboration.org/guidelines.asp).

The EPC program (30), started in 1997, is funded by the Agency for Healthcare Research and Quality and consists of 14 centers throughout North America (Internet: http://www.ahrq.gov/clinic/epc指引.htm). The remit of this program is to complete SRs in a wide variety of healthcare and policy areas, and guides for conducting these reviews have been proposed (32). As do those of the Campbell Collaboration, these SRs address broader types of questions and thus are not limited to reports of RCTs. A more detailed discussion of the EPC program can be found elsewhere (31).

Whereas The Cochrane Collaboration, Campbell Collaboration, and EPC program are high-profile, ≥100 other groups conduct reviews, commission SR conduct, or do both (33). Examples include the Canadian Agency for Drugs and Technologies in Health (Internet: cadth.ca/index.php/en/hta/programs/health-technology-assessment/process), the Centre for Reviews and Dissemination (York University, York, United Kingdom; Internet: www.york.ac.uk/inst/erd/report4.htm), and, specific to the field of nutrition, the American Dietetic Association (Internet: www.eatright.org/cps/rde/xchg/ada/hs.xsl/education_13146_ENU_HTML.htm). All of these organizations conduct SRs by using different templates.

TABLE 2
Complex questions answered by systematic reviews within different domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nutrition</td>
<td>Which community-based nutrition and physical activity interventions are effective at preventing chronic disease in low-income populations? (17)</td>
</tr>
<tr>
<td></td>
<td>What are the long-term effects of advice to restrict dietary sodium in adults with and without hypertension? (18)</td>
</tr>
<tr>
<td></td>
<td>What is the association between a diet rich in whole grain, bran, and germ and the risk of type 2 diabetes? (19)</td>
</tr>
<tr>
<td>Crime and justice</td>
<td>Does a neighborhood watch reduce crime? (15)</td>
</tr>
<tr>
<td></td>
<td>What are the effects of closed-circuit television on crime? (20)</td>
</tr>
<tr>
<td></td>
<td>What is the effectiveness and cost-effectiveness of counter-terrorism strategies? (21)</td>
</tr>
<tr>
<td>Education</td>
<td>Which approaches to parental involvement improve the academic performance of elementary school–age children? (16)</td>
</tr>
<tr>
<td></td>
<td>To what extent and in what ways does access to after-school programs affect student context (ie, student location, supervision, and safety), participation in enriching activities, behaviors, social and emotional development, and academic outcomes for youth? (22)</td>
</tr>
<tr>
<td></td>
<td>What is the effectiveness of volunteer tutoring programs in improving the academic skills of students from kindergarten through grade 8 in the United States? (23)</td>
</tr>
<tr>
<td>Social welfare</td>
<td>What is the effectiveness of behavioral and cognitive behavioral training interventions in improving placement stability, the psychological well-being and functioning of persons providing foster care, and the behavioral and relationship problems of children receiving foster care? (24)</td>
</tr>
<tr>
<td></td>
<td>What are the effects of cognitive behavioral therapy and similar interventions on men’s physical abuse of their female partners? (25)</td>
</tr>
<tr>
<td></td>
<td>Does exercise alone or exercise as part of a comprehensive intervention improve self-esteem among children and young people? (26)</td>
</tr>
</tbody>
</table>

with a low fruit and vegetable diet (C) with respect to myocardial infarction, stroke, and cardiovascular disease–related mortality (O)?

The explicit methods involved in conducting an SR are dependent on the question that is being asked. Literature can be searched according to the PICOS components. The eligibility criteria (ie, inclusion and exclusion criteria) will be based on the question. The data abstraction instrument or instruments will be developed to abstract from the included studies the data that will be used to answer the question. The synthesis of results abstracted from the included studies will also help provide an answer to the question. The methods used in an SR may vary depending on the type of question and the requirements of the organization conducting or commissioning the review. Three examples of such an organization include the Cochrane Collaboration, the Campbell Collaboration, and the Evidence-based Practice Center (EPC) program.

The Cochrane Collaboration (13) is a global entity including ≈10 000 members organized into Clinical Review Groups (eg, Schizophrenia Group), Methods Groups (eg, Bias Methods Group), and Fields (eg, Diet and Nutrition Subfield of the Cochrane Primary Health Care Field). The mission of The Cochrane Collaboration is to conduct SRs in all areas of healthcare, including nutrition. Cochrane reviews are completed by using an approach described in the Cochrane Handbook (13). A distinctive characteristic of Cochrane reviews is

FINDINGS

Differences in questions or conduct may lead to differences in results and conclusions

Variations of the SR question or differences in the SR methods used may explain why SRs addressing a similar topic have different, sometimes startlingly different, results or conclusions, or both; these are called discordant reviews. An example of discordant reviews is apparent among 4 SRs examining the cardiovascular effects of vitamin E supplements (Table 3; 34–37). Although the SR questions were similar, variations are apparent. For example, one of the reviews (35) focused on effectiveness, whereas another (36) focused on efficacy. Furthermore, different methods were used to conduct the reviews. For example, one SR (36) searched multiple databases, whereas another (34) searched...
TABLE 3
Differences among 4 systematic reviews examining the cardiovascular effects of vitamin E

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Question posed</td>
<td>What effects do antioxidant vitamins (ie, vitamin E and (\beta)-carotene) have on all-cause mortality and CV death?</td>
<td>What is the effectiveness of vitamin E in the treatment and prevention of CVD?</td>
<td>What is the efficacy of vitamin E supplements in the prevention and treatment of CVD?</td>
<td>What effect do antioxidant vitamins (ie, vitamin E and (\beta)-carotene) have on all-cause mortality and CV death?</td>
</tr>
<tr>
<td>Methods</td>
<td>Searched a database and the references of included studies. Search dates NR.</td>
<td>“Computerized search.” Search dates NR.</td>
<td>Searched databases (inception to 2001 or 2002, depending on the database) and the references of included studies; contacted experts.</td>
<td>Searched databases (1996–August 2004), the references of included studies, and the references of previous SRs.</td>
</tr>
<tr>
<td>Databases searched</td>
<td>MEDLINE</td>
<td>NR</td>
<td>MEDLINE, EMBASE, MANTIS, Allied &amp; Complementary Medicine, Biosis Previews, CAB Health, CancerLit, The Cochrane Library, Social SciSearch, SciSearch Cited Ref Sci, TGG Health &amp; Wellness Database</td>
<td>MEDLINE, Cochrane Central Register of Controlled Trials (CENTRAL)</td>
</tr>
<tr>
<td>Inclusion criteria</td>
<td>Studies with &gt;1000 participants; developed countries without vitamin E deficiency; trials: study quality (trial size, randomization, ITT analysis)</td>
<td>Randomized trials; vitamin E therapy; treatment or prevention of CVD</td>
<td>Published or unpublished material; trials examining vitamin E</td>
<td>Random allocation; vitamin E alone or in combination; placebo group; men or nonpregnant women; study duration &gt;1 y; ≥10 deaths had to occur during the trial</td>
</tr>
<tr>
<td>Exclusion criteria</td>
<td>Trials without all-cause mortality data</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Language limits</td>
<td>NR</td>
<td>English only</td>
<td>No language restrictions</td>
<td>No language restrictions</td>
</tr>
<tr>
<td>Screening</td>
<td>2 independent reviewers</td>
<td>NR</td>
<td>2 independent reviewers</td>
<td>NR</td>
</tr>
<tr>
<td>Data abstraction</td>
<td>2 independent reviewers</td>
<td>NR</td>
<td>2 independent reviewers</td>
<td>3 independent reviewers</td>
</tr>
<tr>
<td>Outcomes</td>
<td>All-cause mortality; CV death; all-cause cerebrovascular accident; nonfatal MI</td>
<td>CV events: MI; nonfatal stroke; CV death</td>
<td>All-cause mortality; CV deaths</td>
<td>Adjusted and unadjusted all-cause mortality for high and low dosages</td>
</tr>
<tr>
<td>Analysis</td>
<td>Meta-analysis: odds ratios and 95% CIs</td>
<td>Meta-analysis: odds ratios and 95% CIs from the difference between observed minus expected number of events</td>
<td>Random effects meta-analysis: log risk ratios and 95% CIs</td>
<td>2-level hierarchical logistic regression model analyzed by low and high doses; analysis was transformed into risk differences and risk ratios</td>
</tr>
<tr>
<td>Included studies</td>
<td>Included studies (n)</td>
<td>7</td>
<td>84</td>
<td>19</td>
</tr>
<tr>
<td>Abbreviated trial name (if provided)</td>
<td>ATBC, CHAOS, GISSI, HOPE, HPS (MRC), AREDS, PPP</td>
<td>Linxian, ATBC, CHAOS, GISSI, HOPE, PPP, HPS (MRC)</td>
<td>Linxian, ATBC, PPP, ASAP, SPACE, HOPE, GISSI, CHAOS, MASI, HATS, MVP</td>
<td>MINVITAOX, Linxian, SUVIMAX, ATBC, Lingu, GISSI, PPP, HOPE, AREDS, PPS, VECAT, CHAOS, RACT, HPS (MRC), SPACE, WAVE, ADCS, DATATOP</td>
</tr>
<tr>
<td>Follow-up (y)</td>
<td>4.5 (1.4–6.3)</td>
<td>4.5 (1.5–6.1)</td>
<td>NR (2–7)</td>
<td>4.0 (1.4–8.2)</td>
</tr>
<tr>
<td>Participants (n)</td>
<td>4761 (1035–14 564)</td>
<td>11 324 (2002–29 584)</td>
<td>NR (NR)</td>
<td>3492 (256–32 704)</td>
</tr>
</tbody>
</table>

(Continued)

Downloaded from https://academic.oup.com/ajcn/article-abstract/88/5/1191/4648936 by guest on 18 April 2018
only one database. Different inclusion or exclusion criteria, language limitations (eg, including only reports in English), and outcomes of interest also were evident. The number of included studies varied from 7 (34, 35) to 84 (36).

Although the results of the reviews varied, consistency was observed in 3 reviews (34–36); no association between vitamin E and any cardiovascular endpoint was observed (Table 3). However, the fourth review (37) conducted a dose-response analysis, for which high doses of vitamin E (≥500 IU/d) were shown to significantly increase the risk of all-cause mortality by 9–14%. Two of the SRs (34, 36) concluded that vitamin E had no benefit with respect to cardiovascular events, and one SR (35) concluded that vitamin E had neither benefit nor harm with respect to cardiovascular events. However, the fourth review (37) concluded that a dose-response relation between vitamin E and all-cause mortality was observed (ie, that vitamin E at high doses is harmful).

For an understanding of discordant reviews, an assessment of the risk of bias (ie, study quality) of the SRs is likely to be beneficial (38). Such an exercise was completed for the 4 SRs discussed above by using the Oxman and Guyatt (39) instrument (Table 4). This validated instrument consists of 9 main criteria for assessing the scientific quality of review articles. The final item asks the assessor to rate the overall scientific quality of the SR by using a score ranging from 1 (ie, extensive flaws) to 7 (ie, minimal flaws). When the Oxman-Guyatt tool was applied to these 4 SRs, the scores ranged from 1 [indicating extensive flaws in the SR (35)] to 7 [indicating minimal flaws in the SR (36)]; the other 2 SRs fell between the extremes.

### TABLE 3 (Continued)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All-cause mortality</strong></td>
<td>1.02 (0.98–1.06)</td>
<td>NR</td>
<td>0.96 (0.84–1.10)</td>
<td></td>
</tr>
<tr>
<td><strong>Unadjusted results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adjusted results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CV event</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV death</td>
<td>NR</td>
<td>0.98 (0.94–1.03)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>All-cause stroke</td>
<td>6.0% vs 6.0% control (P = 0.94)</td>
<td>1.00 (0.92–1.05)</td>
<td>0.97 (0.80–1.19)</td>
<td>NR</td>
</tr>
<tr>
<td>CV death or nonfatal MI</td>
<td>3.6% vs 3.5% control (P = 0.7)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Nonfatal MI</td>
<td>9.8% vs 9.8% control (P = 0.93)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Nonfatal stroke</td>
<td>NR</td>
<td>1.00 (0.92–1.09)</td>
<td>0.72 (0.51–1.02)</td>
<td>NR</td>
</tr>
<tr>
<td>Ischemic stroke</td>
<td>NR</td>
<td>1.03 (0.93–1.14)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Hemorrhagic stroke</td>
<td>NR</td>
<td>1.01 (0.90–1.14)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Fatal MI</td>
<td>NR</td>
<td>1.24 (0.96–1.59)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Intermediate outcomes:</td>
<td>NR</td>
<td>NR</td>
<td>0.97 (0.74–1.27)</td>
<td>NR</td>
</tr>
<tr>
<td>total cholesterol, LDL, HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authors’ conclusions</td>
<td>“When used as secondary prevention, vitamin E did not reduce the risk of CV endpoints . . .”</td>
<td>“This overview . . . of vitamin E . . . provides conclusive evidence of a lack of statistically significant or clinically important benefit or harm regarding any important CV events or its components . . .”</td>
<td>“The available scientific studies offer little evidence that supplementation with vitamin E has any benefit on CVD prevention or treatment.”</td>
<td>“. . . [W]e identified a dose-response relationship between vitamin E supplementation and all-cause mortality . . . [A]ll-cause mortality progressively increased for dosages &gt;150 IU/d.”</td>
</tr>
</tbody>
</table>

1 CV, cardiovascular; CVD, CV disease; NR, not reported; ITT, intention-to-treat; MI, myocardial infarction; SR, systematic review.

2 Median; range in parentheses (all such values).

3 P < 0.05.
As it would be for any type of research, reporting SRs to the highest possible standard is critical, because the report is one of the most important ways of disseminating the results to the broader research community and beyond. Dissemination is likely to be more widespread if the report accurately and transparently describes the research methods and results. It is important that the report of a research endeavor is, in many instances, the only public record of the existence of the research itself. High-quality reports enable readers to critically interpret and use the results. There also are emerging data indicating that the use of reporting guidelines improves the quality of the research reports (40).

To help improve the quality of reporting of all research, the UK National Health Service National Library for Health and the UK National Institute for Health Research recently provided funds to set up the EQUATOR Network (Internet: www.equator-network.org). The objective of this network is to improve the quality of scientific publications by promoting transparent and accurate reporting of health research and by supporting the development of high-quality reporting guidelines. An early project completed by the network executive was an SR that identified existing reporting guidelines and a survey that identified issues faced by those developing such guidelines (41). More than 50 sets of reporting guidelines were identified; they can be accessed via the EQUATOR Network website.

Many studies have evaluated the quality of SR reports. In 1987, Mulrow (42) examined 50 review articles published in 1985 and 1986 in 4 leading medical journals; he found that none met all 8 explicit scientific criteria that they examined, such as a quality assessment of included studies. Also in 1987, Sacks et al (43) evaluated the adequacy of reporting on 23 characteristics in 6 domains in 83 meta-analyses. They found that reporting generally was poor; between 1 and 14 characteristics (\(\bar{x} \pm SD: 7.7 \pm 2.7\)) were adequately reported. A 1996 update of this study found little improvement (44). To address the suboptimal reporting of meta-analyses, an international group developed guidance on the quality of reporting of meta-analyses of RCTs, called the QUOROM Statement.

The QUOROM Statement consists of a 21-item checklist that documents the process of completing a meta-analysis and a flow diagram that details the number and status of included articles at each stage of the meta-analysis process (45). Since the QUOROM Statement was published in 1999, the evidence base underlying the conduct of SRs has matured. In addition, a more comprehensive understanding of some conceptual issues, advances in methodology, practical innovations in the conduct and reporting of SRs, and changes in terminology have occurred. It is for these reasons that the QUOROM Statement was updated in June 2005. It is now called the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement, and it applies not only to SRs of RCTs but also to any SR assessing the effectiveness of any intervention.

### Quality of reporting of systematic reviews within nutrition

We completed an extensive search to identify studies that have examined the general quality of reporting of SRs within nutrition; we found that few studies exist, perhaps because there are few individual studies (eg, cohort studies or RCTs) to be synthesized. More likely the reasons are multifactorial, such as a lack of capacity to conduct SRs on nutrition topics and possibly a lack of clarity as to how best to conduct and report such reviews.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Were the search methods used to find evidence (original research) on the primary questions(s) stated?</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2. Was the search for evidence reasonably comprehensive?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>3. Were the criteria used for deciding which studies to include in the overview reported?</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
</tr>
<tr>
<td>4. Was bias in the selection of studies avoided?</td>
<td>Y</td>
<td>C</td>
<td>Y</td>
<td>C</td>
</tr>
<tr>
<td>5. Were the criteria used for assessing the validity of the included studies reported?</td>
<td>P</td>
<td>C</td>
<td>Y</td>
<td>C</td>
</tr>
<tr>
<td>6. Was the validity of all of the studies referred to in the text assessed with the use of appropriate criteria (either in the selection of studies for inclusion or in analysis of the studies cited)?</td>
<td>C</td>
<td>C</td>
<td>Y</td>
<td>C</td>
</tr>
<tr>
<td>7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>8. Were the findings of the relevant studies combined appropriately with respect to the primary question the overview addressed?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>9. Were the conclusions made by the author or authors supported by the data or the analysis (or both) reported in the overview?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Score(^1)</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

\(^1\) Y, yes; N, no; P, partially; C, not clear.

\(^2\) The maximum score is 7. A score of 1 means the review has extensive flaws; a score of 2–3 represents major flaws; a score of 4–5 represents minor flaws; and a score of 6–7 represents minimal flaws. For further information on the scoring scheme, see Oxman and Guyatt (39).
There are limited data on the specific issues that systematic
reviewers in the nutrition domain should specifically address.
Gibson et al (46) noted that the following steps were involved in
nutrition assessment during the conduct of RCTs and meta-anal-
yses: measuring food intakes via qualitative or quantitative
means (eg, food diary or food recalls), converting food intakes to
observed intakes of nutrients by using food-composition tables;
and evaluating nutrient adequacy. Challenges encountered dur-
ing a series of reviews of omega-3 fatty acids were also docu-
mented (47), including difficulties in sorting through the numer-
ous endpoints reported in the included studies and variations in
the source of dietary intake (eg, supplement or dietary compo-
nent), study design, and study duration of the included studies.
These challenges highlight how important it is for those com-
pleting individual studies to ensure that these points are ade-
quately integrated into the conduct of such studies.

A systematic review is dependent on the quality of
included studies

The successful conduct of any SR is dependent to a large
degree on the quality of the included studies. Historians may well
view the first 50 y of reporting of RCTs with some surprise. They
will encounter what might be described as a cognitive disso-
nance: a disconnect between the increasing sophistication of
the design and the swelling cost of these studies and the apparent lack
of care—a disastrous lack in some cases—with which they have
been reported. A series of studies beginning in 1995 found em-
pirical evidence that results may be biased when trials use infe-
rior methods or are reported without adequate description of the
methods; notably, failure to conceal the allocation process is
associated with a ≥30% exaggeration in the effectiveness of an
intervention (48). The cause for concern is obvious: if the con-
duct or reporting of RCTs is poor, treatments may be introduced
that are less effective than was thought or that may even be
ineffective, and these reports are included in SRs.

This concern led to the Consolidated Standards of Report-
ing of Trials (CONSORT) Statement recommendations origi-
nally published in 1996 and updated in 2001; another update is
currently in progress (Internet: www.consort-statement.org). The
CONSORT Statement consists of a 22-item evidence-based
checklist that authors can use to guide the reporting of their trials
and a flow diagram to account for the flow of participants
throughout the trial process. The CONSORT Statement has been
extended in several directions, such as cluster trials, equivalence
and noninferiority trials, and nonpharmacologic treatments.
These extensions have been made because of inadequate reporting
of important study aspects not covered by the original CON-
SORT Statement. The CONSORT statement for herbal interven-
tions was developed to help improve the quality of reporting of
such studies, which may also have implications for how those
studies, as well as observational ones, are designed and con-
ducted. Currently, the CONSORT Statement has not been widely
implemented in the field of nutrition.

SUMMARY AND RECOMMENDATIONS

SR techniques have been used to answer complex research
questions across many different research domains. SRs have
recently become popular within healthcare and are likely to be
beneficial in any field, including nutrition. Developing the re-
search question is probably the most important step in conduct-
ing an SR. SRs often include study designs beyond randomized
trials and do not always include a meta-analysis of the results. As
with any other type of study, the optimal conduct and reporting
of SRs is necessary.

To strengthen the scientific rigor of nutrition science research
endeavors, we propose the following ideal recommendations,
recognizing that their adoption may be difficult without the in-
volvement of an expert panel akin to the Institute of Medicine
(Table 5).

First, we recommend the development of a capacity-building
program to conduct and report nutrition-related SRs, including a
comprehensive training portfolio. This program could be imple-
mented within universities and colleges that have nutrition pro-
grams or within organizations in the nutrition domain.

Second, to ascertain which nutrition topics are most likely to
benefit from synthesis (ie, an SR), we recommend that a com-
prehensive search of the primary literature (eg, cohort studies) be
conducted and that the resulting reports be categorized according
to the nutrition-related topic. When there is a sufficient evidence
base, we recommend that this priority list form the basis of SRs
that should be commissioned for immediate completion. Such an
exercise is also likely to provide evidence for locating the gaps in
the evidence base and thus to give a rationale for the areas in
which primary research could be earmarked for investment. To
identify gaps in the SR literature and generate new SR questions,
we also recommend the development of a repository of SRs
within the field of nutrition.

Third, SRs are most likely to be useful if the quality of report-
ing of the individual studies included is sufficiently high; this is
a longer-term undertaking than some of the other recommenda-
tions. We recommend that the CONSORT Statement be imple-
mented for nutrition-related clinical trials (Internet: www.
consort-statement.org). Such an initiative for herbal interven-
tions could build on the CONSORT Statement. Similarly,
where appropriate, we recommend that other reporting guides
be extended or implemented for this field (Internet:

<table>
<thead>
<tr>
<th>Number</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Develop a capacity-building program to conduct and report nutrition-related SRs, including a comprehensive training portfolio.</td>
</tr>
<tr>
<td>2</td>
<td>Conduct a comprehensive search to develop a list of priority topics on which sufficient evidence exists to warrant an SR, and identify gaps in primary research. Develop a repository of SRs in nutrition to identify gaps and generate new SR questions.</td>
</tr>
<tr>
<td>3</td>
<td>Implement CONSORT and other reporting guidelines within the field of nutrition research. Obtain endorsement of reporting guidelines by journals publishing nutrition-related research.</td>
</tr>
<tr>
<td>4</td>
<td>Extend PRISMA Statement to reporting of SRs evaluating nutrition. Obtain endorsement of SR reporting guidelines by journals publishing SRs on nutrition-related topics.</td>
</tr>
<tr>
<td>5</td>
<td>Develop a standard guide to the conduct of SRs in nutritional science.</td>
</tr>
<tr>
<td>6</td>
<td>Review the quality of nutrition literature every year.</td>
</tr>
</tbody>
</table>

TABLE 5
Recommendations with respect to systematic reviews (SRs)
www.equator-network.org). We also recommend that journals publishing nutrition-related research endorse evidence-based reporting guidelines to improve the quality of nutrition research in their journals.

Fourth, we recommend that the PRISMA Statement be extended to the reporting of SRs that evaluate nutrition (45). We also recommend that journals publishing SRs on nutrition-related topics endorse the PRISMA Statement to improve the quality of reporting of this research in their journals.

Fifth, because the conduct and reporting of SRs are so intertwined, we recommend that an evidence-based approach to developing a standard guide for conducting SRs within the nutritional science domain be undertaken. Preexisting SR conduct guides providing some insight into the nutrition field, such as the Cochrane Handbook (13), can be used as a starting point in the development of this standard guide. It would also be beneficial to include persons with expertise in many different areas to develop evidence-based guidelines for the best reporting and conduct of nutrition-related SRs. These persons may include a broad base of funders and those who commission nutrition-related SRs, those actively involved in obtaining empirical evidence to guide the conduct of nutrition-related SRs, journal editors who publish SRs, and persons with experience and expertise in conducting SRs. Such a consortium is likely to have a monumental effect in improving the reporting and conduct of nutrition-related SRs.

Sixth, we recommend that an annual review of the quality of nutrition literature be conducted. Such a review will provide important stakeholder groups with an evidence base as to the current quality of this important literature.

We thank the attendees of the Evidence-based Approaches to Developing Dietary Guidance Workshop, hosted by the International Life Sciences Institute Research Foundation in December 2007, for their useful feedback on a preliminary version of the manuscript. We also thank members of Chalmers Research Group for their feedback on Table 1.

The authors’ responsibilities were as follows—DM: conceptualized the study and wrote the manuscript; ACT: helped write the manuscript and edited it. Neither author had a personal or financial conflict of interest.

REFERENCES


