

# Assessments of risk of bias in systematic reviews of observational nutritional epidemiologic studies are often not appropriate or comprehensive: a methodological study

Dena Zeraatkar <sup>1,2</sup>, Alana Kohut,<sup>3</sup> Arti Bhasin,<sup>2</sup> Rita E Morassut,<sup>4</sup> Isabella Churchill,<sup>2</sup> Arnav Gupta,<sup>5</sup> Daeria Lawson,<sup>2</sup> Anna Miroshnychenko,<sup>2</sup> Emily Sirotych,<sup>2</sup> Komal Aryal,<sup>2</sup> Maria Azab,<sup>3</sup> Joseph Beyene,<sup>2</sup> Russell J de Souza<sup>2,6</sup>

**To cite:** Zeraatkar D, Kohut A, Bhasin A, *et al.* Assessments of risk of bias in systematic reviews of observational nutritional epidemiologic studies are often not appropriate or comprehensive: a methodological study. *BMJ Nutrition, Prevention & Health* 2021;4:e000248. doi:10.1136/bmjnph-2021-000248

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjnph-2021-000248>).

For numbered affiliations see end of article.

## Correspondence to

Dr Russell J de Souza, Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON L8S 4L8, Canada; [desouzrj@mcmaster.ca](mailto:desouzrj@mcmaster.ca)

Received 30 January 2021

Accepted 2 August 2021

Published Online First

7 December 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

## ABSTRACT

**Background** An essential component of systematic reviews is the assessment of risk of bias. To date, there has been no investigation of how reviews of non-randomised studies of nutritional exposures (called ‘nutritional epidemiologic studies’) assess risk of bias.

**Objective** To describe methods for the assessment of risk of bias in reviews of nutritional epidemiologic studies.

**Methods** We searched MEDLINE, EMBASE and the Cochrane Database of Systematic Reviews (Jan 2018–Aug 2019) and sampled 150 systematic reviews of nutritional epidemiologic studies.

**Results** Most reviews (n=131/150; 87.3%) attempted to assess risk of bias. Commonly used tools neglected to address all important sources of bias, such as selective reporting (n=25/28; 89.3%), and frequently included constructs unrelated to risk of bias, such as reporting (n=14/28; 50.0%). Most reviews (n=66/101; 65.3%) did not incorporate risk of bias in the synthesis. While more than half of reviews considered biases due to confounding and misclassification of the exposure in their interpretation of findings, other biases, such as selective reporting, were rarely considered (n=1/150; 0.7%).

**Conclusion** Reviews of nutritional epidemiologic studies have important limitations in their assessment of risk of bias.

## BACKGROUND

Due to the challenges of conducting randomised controlled trials (RCTs) of dietary interventions, most of the evidence in nutrition comes from non-randomised, observational studies of nutritional exposures, hereon referred to as ‘nutritional epidemiologic studies’.<sup>1–4</sup> Clinicians, guideline developers, policymakers and researchers use systematic reviews of these studies to advise patients on optimal dietary habits, formulate recommendations and policies, and plan future research.<sup>2,5,6</sup>

## What this study adds

- An essential component of systematic reviews is the assessment of risk of bias.
- To date, there has been no empirical assessment of how systematic reviews of nutritional epidemiologic studies assess risk of bias.
- We show that reviews of nutritional epidemiologic studies have important limitations in their assessment of risk of bias and produce recommendations for review authors.

Bias may arise in nutritional epidemiologic studies, and other non-randomised studies, due to confounding, inappropriate criteria for selection of participants, error in the measurement of the exposure or outcome, departures from the intended exposure, missing outcome data and selective reporting.<sup>7–10</sup> The assessment of the validity of studies included in a systematic review and the extent to which that they might overestimate or underestimate the true effects—called risk of bias—is a critical component of the systematic review process.<sup>11,12</sup> The assessment of risk of bias informs the evaluation of the certainty of evidence and the interpretation of review findings and failure to appropriately consider risk of bias using appropriate criteria may lead to erroneous conclusions.<sup>13–18</sup> Prevailing guidance dictates for systematic reviews to present a rigorous and comprehensive assessment of the risk of bias of primary studies and to incorporate risk of bias assessments in the synthesis and interpretation of findings.<sup>11</sup>

While methods for the assessment of risk of bias of RCTs have been well established, criteria for the assessment of risk of bias in

non-randomised studies are less clear.<sup>17–21</sup> Further, there are unique and complex challenges to assessing the risk of bias of nutritional epidemiologic studies, such as making judgments about the validity and reliability of dietary measures.

The objective of this study was to describe and evaluate methods for the assessment of risk of bias in systematic reviews of nutritional epidemiologic studies and to propose guidance addressing major limitations. This study capitalises on the methods and data of our previously published meta-epidemiological study of systematic reviews of nutritional epidemiologic studies.<sup>6</sup>

## METHODS

We registered the protocol for this study at the Open Science Framework (<https://osf.io/wr6uy>).

### Search strategy

With the help of an experienced research librarian, we searched MEDLINE and EMBASE from January 2018 to August 2019 and the Cochrane Database of Systematic Reviews from January 2018 up to February 2019 for systematic reviews of nutritional epidemiologic studies (online supplemental material 1).<sup>6</sup>

### Study selection

We included systematic reviews if they investigated the association between one or more nutritional exposures and health outcomes and reported on one or more epidemiologic studies.<sup>6</sup> We defined systematic reviews as studies that explicitly described a search strategy (including at minimum databases searched) and eligibility criteria (including at minimum the exposure(s) and health outcome(s) of interest)<sup>1</sup>; epidemiologic studies as non-randomised, non-experimental studies (eg, cohort studies) that include a minimum of 500 participants<sup>2</sup>; nutritional exposures as macronutrients, micronutrients, bioactive compounds, foods, beverages or dietary patterns; and health outcomes as measures of morbidity, mortality and quality of life.<sup>6</sup> We did not restrict eligibility based on the language of publication.<sup>6</sup> We excluded scoping and narrative reviews, reviews of acute postprandial studies, and reviews of supplements and chemicals involuntarily consumed through the diet.<sup>6</sup>

Reviewers performed screening independently and in duplicate following calibration exercises. We resolve disagreements by discussion or by third-party adjudication. We estimated that 150 reviews will allow estimation of the prevalence of even uncommon review characteristics (ie, prevalence ~5% of studies) with acceptable precision (ie,  $\pm 3.5\%$ ).<sup>6,18</sup> Our sample of 150 eligible reviews was selected using a computer-generated random number sequence.

### Data collection

Following calibration exercises, reviewers, working independently and in duplicate, extracted the following

information from each review using a standardised and pilot-tested data collection form: research question; eligibility criteria; methods and criteria used for the assessment of risk of bias; presentation and reporting of risk of bias; details related to how assessments of risk of bias were incorporated in the analysis and the interpretation of findings. Items of the data collection form were drawn from authoritative sources that had published guidance on optimal practices for assessing risk of bias in systematic reviews, data collection forms of previous studies, and literature on methodological issues relevant to the assessment of risk of bias in non-randomised studies and nutritional epidemiologic studies.<sup>12 22–27</sup>

We collected information on any tools or criteria that included one or more items or domains that addressed the internal validity of studies or the likelihood of bias or were interpreted by review authors as indicators of bias or internal validity. In order to evaluate both appropriate and inappropriate methods by which reviews assessed risk of bias, we collected information on tools and criteria regardless of whether they were originally designed to address risk of bias or whether they were valid indicators of risk of bias. For example, some reviews applied and interpreted reporting checklists. In such cases, we still collected information on the reporting checklist if it was interpreted by the review authors as an indicator of internal validity or risk of bias. We did this because we were also interested in estimating the proportion of reviews that assess risk of bias using inappropriate methods. For reviews that also included RCTs or other experimental designs in addition to nutritional epidemiologic studies, we only collected data on the tools that were used to assess the risk of bias of nutritional epidemiologic studies.

We reviewed tools and ad hoc criteria and categorised their items and domains according to the type of biases that they addressed. We used the domains of the Cochrane ROBINS-I tool as a framework for categorisation and created additional categories as necessary.<sup>7</sup> We classified risk of bias criteria as ad hoc when a study developed a set of criteria de novo to assess risk of bias. We classified risk of bias tools as scales if each item was assigned a numerical score and the tool yielded an overall summary score, as checklists if judgements for each item were presented individually and not aggregated with other items, and domain based if judgements were presented across domains with at least one domain composed of more than one item.<sup>28</sup>

### Data synthesis and analysis

To synthesise the data, we present frequencies and percentages for dichotomous outcomes and median and IQRs for continuous outcomes.

## RESULTS

Online supplemental material 2 presents details of the selection of systematic reviews. We retrieved a total of 4267 unique records and screened a random sample of

2273 titles and abstracts and 184 full-text articles to identify a sample of 150 eligible reviews.

### General characteristics of systematic reviews

Table 1 presents general characteristics of systematic reviews. Reviews were most frequently published in general nutrition journals by authors from Europe or Asia. Only a small minority of reviews were conducted to inform a particular guideline or policy decision or to fulfil the needs of a specific evidence user. A very small minority of reviews were funded by marketing/advocacy organisations or food companies and most were funded by either government agencies or institutions. Reviews most frequently reported on cancer morbidity and mortality, foods or beverages, and included a median of 15 studies and 200 000 participants. Three quarters of reviews conducted meta-analysis. Nearly all reviews included cohort studies and more half included case-control studies. More than three quarters of reviews attempted to assess the risk of bias of included studies.

### Risk of bias methods and reporting

Table 2 presents details on the methods by which risk of bias was assessed and reported in reviews. The most commonly used tool was the Newcastle-Ottawa scale. Among reviews that used modified versions of published tools, nearly all used modifications of the Newcastle-Ottawa scale, which was modified either to include alternative response options or to be applicable to cross-sectional studies (the original Newcastle-Ottawa scale is designed only for cohort and case-control studies).<sup>29</sup> Three reviews used modified versions of the Critical Appraisal Skills Programme (CASP) cohort study checklist,<sup>30</sup> the NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies,<sup>31</sup> and the American Diabetes Association Quality Criteria Checklist<sup>32</sup>; the former two were modified to include only a subset of the original items,<sup>33 34</sup> and the latter included a subset of the original items as well as several additional items.<sup>35</sup> Nearly all risk of bias tools were scales and only a minority were checklists or domain based. Only half of all reviews reported assessing risk of bias in duplicate.

Nearly a quarter of reviews only reported the range of risk of bias ratings across studies (eg, “The [Newcastle-Ottawa scale] scores ranged from 6 to 9.”<sup>36</sup>) rather than risk of bias ratings for each study. More than a third of reviews failed to report judgements for each risk of bias item or domain. Among reviews that reported on more than one outcome across which risk of bias could conceivably differ, risk of bias was seldom assessed separately for each outcome.

### Characteristics of risk of bias tools and ad hoc criteria

Table 3 presents characteristics of the tools and ad hoc criteria that were used to assess risk of bias. One review reported using the ‘cross-sectional study quality assessment criteria’ but did not provide a reference to the tool or report any other details on the tool and so it

**Table 1** General characteristics of systematic reviews

	Number of reviews (%) N=150
<b>Journal</b>	
General nutrition journal (journals with only a nutrition focus) (eg, <i>American Journal of Clinical Nutrition</i> )	61 (40.7%)
Specialised nutrition journal (journals with a focus on nutrition and a specific disease area) (eg, <i>Nutrition, Metabolism and Cardiovascular Diseases</i> )	7 (4.7%)
General medical journal (eg, <i>Lancet</i> )	28 (18.7%)
Specialised medical journal (eg, <i>Clinical Breast Cancer</i> )	54 (36%)
<b>Country of corresponding author’s affiliation</b>	
North America	14 (9.3%)
Europe	43 (28.7%)
Oceania	13 (8.7%)
Middle East	28 (18.7%)
Asia	49 (32.7%)
South America	3 (0.7%)
<b>Was the review conducted to inform a particular guideline or policy decision or to fulfil the needs of a particular evidence user?</b>	
Yes	6 (4%)
No	144 (96%)
<b>Funding*</b>	
Government support	56 (37.3%)
Institutional support	34 (22.7%)
Private not-for-profit foundation	20 (13.3%)
Food marketing/advocacy organisations	4 (3.3%)
Food companies	2 (1.3%)
No funding	32 (21.3%)
Not reported	34 (22.7%)
<b>Did the authors declare any conflicts of interest?</b>	
Yes	10 (6.7%)
No	135 (90%)
Not reported	5 (3.3%)
<b>Exposure(s)*</b>	
Micronutrient	27 (18%)
Macronutrient	24 (16%)
Bioactive compounds	15 (10%)
Food or beverage	60 (40%)
Food group	21 (14.0%)

Continued

**Table 1** Continued

	Number of reviews (%) N=150
Dietary pattern	49 (32.7%)
Non-nutritive components of foods/beverages	25 (18.7%)
Outcome(s)*	
Cardiometabolic morbidity or mortality	26 (17.3%)
Cancer morbidity or mortality	54 (36%)
Diseases of the digestive system	10 (6.7%)
All-cause mortality	9 (6%)
Anthropometric measures	8 (5.33%)
Surrogate outcomes	17 (11.3%)
Other	55 (36.7%)
Eligible study designs*	
Cohort	146 (97.3%)
Case-control	97 (64.7%)
Cross-sectional	80 (53.3%)
Randomised controlled trials	74 (49.3%)
Median no of primary studies (IQR)	15 (11 to 23)
Median no of participants (IQR)	208 117 (84951 to 510954)
Method for the synthesis of results	
Meta-analysis	115 (76.7%)
Narrative	21 (14%)
Tabular/graphical summary of quantitative results without meta-analysis	14 (9.3%)
Did the review assess risk of bias?	
Yes	131 (87.3%)
No	19 (12.6%)

\*Each review can be classified in more than one category.

was excluded from our analysis.<sup>37</sup> The majority of tools addressed biases due to confounding, classification of the exposure and measurement of the outcome. Biases due to selection of the participants and missing data were addressed by approximately half of the tools and biases due to departures from the intended exposure and selection of the reported results were rarely addressed. Nearly all tools included one or more constructs unrelated to risk of bias, such as reporting, generalisability (external validity), or precision.

#### Incorporation of risk of bias in the synthesis of results

Table 4 presents details on how reviews incorporated assessments of risk of bias in the synthesis of results. Two reviews excluded studies at high risk of bias from

**Table 2** Risk of bias methods and reporting

	Number of reviews (%) N=131 (reviews that assessed risk of bias)
Tools used to assess risk of bias*	
Newcastle-Ottawa Scale <sup>38</sup>	92 (70.2%)
AHRQ for Cross-Sectional Studies <sup>60</sup>	3 (2.3%)
Quality in Prognosis Studies <sup>61</sup>	3 (2.3%)
ROBINS-I <sup>7</sup>	3 (2.3%)
SIGN checklist for cohort studies <sup>62</sup>	2 (1.5%)
STROBE <sup>63</sup>	2 (1.5%)
Cochrane risk of bias tool <sup>9</sup>	1 (0.7%)
Critical Appraisal Skills Programme tools for cohort studies <sup>30</sup>	1 (0.7%)
Cross-sectional study quality assessment criteria	1 (0.7%)
Data collection instrument and procedure for systematic reviews in the guide to community preventative services <sup>64</sup>	1 (0.7%)
Effective Public Health Practice Project <sup>65</sup>	1 (0.7%)
Mixed Methods Appraisal Tool <sup>66</sup>	1 (0.7%)
NICE Methodological Checklist for Cohort Studies <sup>67</sup>	1 (0.7%)
NICE Methodological Checklist for Case-Control Studies <sup>67</sup>	1 (0.7%)
NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies <sup>31</sup>	1 (0.7%)
Qualitative assessment	1 (0.7%)
Quality Assessment Tool for Systematic Reviews of Observational Studies (QATSO) <sup>68</sup>	1 (0.7%)
Research Triangle Institute Item Bank on Risk of Bias <sup>69</sup>	1 (0.7%)
SIGN checklist for case-control studies <sup>62</sup>	1 (0.7%)
Modified version of an existing tool	18 (13.7%)
Ad hoc criteria	10 (7.6%)
Not reported	2 (1.5%)
Type of risk of bias tool/ad hoc criteria*	
Scale	116 (88.5%)
Checklist	28 (21.3%)
Domain based	45 (34.3%)
Method for the assessment of risk of bias	
Completed in duplicate or more	69 (52.7%)
Completed by one reviewer and verified by a second reviewer	1 (0.7%)
Not reported	61 (46.7)
Median proportion of studies rated at high risk of bias (IQR) among reviews that assigned an overall rating of risk of bias to each study (n=81; 61.8%)	0 [0 to 25.9]
Median proportion (IQR) of studies rated as unclear risk of bias for one or more items/domains	0 (0 to 0)
Did the review report the risk of bias of each study?	
Yes	105 (80.2%)

Continued

Table 2 Continued

	Number of reviews (%) N=131 (reviews that assessed risk of bias)
No, only the range of risk of bias across studies or the proportion of studies at low or high risk of bias is reported	16 (12.2%)
The review reports that risk of bias was assessed but presents no additional information on risk of bias	10 (7.6%)
Did the review report judgements for all risk of bias items/domains?	
Yes	74 (56.4%)
No; only the overall study risk of bias is presented	47 (35.8%)
The review reports that risk of bias was assessed but presents no additional information on risk of bias	10 (7.6%)
Among reviews that reported on more than one outcome across which risk of bias may differ (n=29; 22.1%), is risk of bias presented for each outcome separately?	
Yes	4 (14.3%)
No	22 (78.6%)
The review reports that risk of bias was assessed but presents no additional information on risk of bias	3 (10.7%)

\*Each review can be classified in more than one category. AHRQ, Agency for Healthcare Research and Quality; NICE, National Institute for Health and Care Excellence; NIH, National Institutes of Health; SIGN, Scottish Intercollegiate Guidelines Network; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

meta-analysis.<sup>13 14</sup> Less than half of reviews explored potential differences in results between studies at higher versus lower risk of bias. Among those that did, nearly all reviews conducted either subgroup analyses or meta-regressions based on the overall rating of study risk of bias. Reviews rarely detected any statistically significant differences between the results of studies at lower versus higher risk of bias. None of the reviews weighted studies in meta-analyses based on risk of bias, implemented credibility ceilings, or attempted to adjust the results of studies for bias.

### Incorporation of risk of bias in the interpretation of review findings

Table 5 presents details related to how risk of bias assessments informed the interpretation of findings. Less than one-fifth of reviews described the overall risk of bias of the body of evidence. While reviews frequently considered biases due to confounding and misclassification of the exposure as potential limitations, biases due to the selection of the participants in the study, departures from the intended exposure, missing outcome data, measurement of the outcome and selective reporting were rarely considered. Reviews rarely described or hypothesised the direction in which results may have been biased. Among the few that did, most hypothesised that results for studies at high risk of bias are likely to have been biased towards the null. Sixteen reviews evaluated the certainty of evidence using a formal system, all of which included

considerations related to risk of bias, among which five downgraded the certainty of evidence due to risk of bias.

## DISCUSSION

### Main findings

Our investigation provides a comprehensive summary of how systematic reviews of nutritional epidemiologic studies assess and report risk of bias and how risk of bias assessments inform the synthesis of results and the interpretation of review findings.

We found that while most reviews attempted to assess risk of bias, the tools and criteria which were used often had serious limitations. For example, commonly used tools frequently neglect to address biases related to departures from the intended exposure and selective reporting and often conflate risk of bias with other study characteristics, such as reporting quality, generalisability and precision. Tools that conflate other study characteristics with risk of bias—often referred to as study quality tools—are poorly suited for the assessment of risk of bias.<sup>27</sup> Some reviews even used reporting checklists, like the STROBE checklist, but interpreted these measures as indicators of internal validity or risk of bias. Furthermore, tools often only partially addressed certain biases. The Newcastle-Ottawa scale, for example, includes items related to the selection of participants, but it does not address all potential issues that may arise due to the suboptimal selection of participants (eg, immortal time bias, inception bias).<sup>38 39</sup>

We also found that existing tools did not provide sufficient guidance to facilitate application, particularly for nutritional epidemiologic studies. While many tools, for example, addressed bias due to classification of the exposure, none provided sufficient guidance for reviewers to make judgements regarding whether tools for measuring dietary exposures are sufficiently valid and reliable, which highlights the need for additional nutrition-specific guidance for applying risk of bias tools.

We identified serious limitations related to how reviews assessed risk of bias. Despite the possibility for risk of bias to vary across outcomes, reviews seldom assessed risk of bias for each outcome individually.<sup>7 11</sup> Further, most reviews assigned a numerical rating of risk of bias to each study—a practice that is discouraged because it requires arbitrary assumptions about the relative weights of risk of bias items and domains.<sup>40–42</sup>

We often found the assessment of risk of bias in reviews to be of questionable validity. For example, reviews rated a median of only 0% (IQR 0% to 25.9%) of studies at high risk of bias. This finding is consistent with previous evidence suggesting that common risk of bias tools, such as the Newcastle-Ottawa Scale, poorly discriminate between studies at lower versus higher risk of bias<sup>43</sup> but is striking since risk of bias issues are ubiquitous in nutritional epidemiology.<sup>44–46</sup> Commonly used dietary measures in nutritional epidemiologic studies, for example, have very serious limitations.<sup>46–49</sup> Furthermore, nutritional epidemiologic studies are usually at risk of selective reporting

**Table 3** Characteristics of risk of bias tools and ad hoc criteria

Tool	Confounding	Bias in selection of participants into the study	Bias in classification of the exposure	Bias due to departures from intended exposures			Bias in measurement of the outcome	Bias in selection of the reported results	Reporting quality	Generalisability	Precision/sample size	Other
				Bias due to missing data	Bias in measurement of the outcome	Bias in selection of the reported results						
Newcastle-Ottawa Scale <sup>38</sup>	X	X	X	X	X	X	X	X	X		Duration of follow-up	
AHRQ Checklist for Cross-Sectional Studies <sup>30</sup>		X	X			X		X				
Quality in Prognosis Studies <sup>51</sup>	X	X	X	X	X	X	X	X	X		Appropriateness of the statistical methods	
ROBINS-1 <sup>7</sup>	X	X	X	X	X	X	X					
SIGN checklist for cohort studies <sup>62</sup>	X	X	X	X	X	X	X	X				
SIGN checklist for case-control studies <sup>62</sup>	X	X	X	X	X	X	X	X				
STROBE <sup>63</sup>								X				
Cochrane risk of bias tool <sup>9</sup>		X		X	X	X	X					
Critical Appraisal Skills Programme tools for cohort studies <sup>30</sup>	X	X	X	X	X	X	X	X	X	X	Magnitude of effect, Bradford-Hill Criteria	
Data collection instrument and procedure for systematic reviews in the guide to community preventative services <sup>64</sup>	X	X	X	X	X	X	X	X			Appropriateness of the statistical methods, control for design effects, accounting for different levels of exposure in segments of the study population in the analysis	
Effective Public Health Practice Project <sup>65</sup>	X	X	X	X	X	X	X		X		Appropriateness of the statistical methods	
Mixed Methods Appraisal Tool <sup>66</sup>		X	X	X	X	X	X	X	X			
NICE Methodological Checklist for Cohort Studies <sup>67</sup>	X		X	X	X	X	X				Duration of follow-up	
NICE Methodological Checklist for Case-Control Studies <sup>67</sup>	X	X	X			X		X				
NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies <sup>51</sup>	X	X	X	X	X	X	X	X		X	Duration of follow-up	
Quality Assessment Tool for Systematic Reviews of Observational Studies (QATSO) <sup>68</sup>	X					X			X		Privacy and sensitive nature of the question considered	

Continued

**Table 3** Continued

Tool	Confounding	Bias in selection of participants into the study	Bias in classification of the exposure	Bias due to departures from intended exposures	Bias due to missing data	Bias in measurement of the outcome	Bias in selection of the reported results	Reporting quality	Generalisability	Precision/sample size	Other
Research Triangle Institute Item Bank on Risk of Bias <sup>69</sup>	X	X	X	X	X	X	X	X	X	X	Study design, prespecification of the outcomes, duration of follow-up, missing outcomes, appropriateness of statistical methods, believability of the results, reporting of funding
Ad hoc criteria											
Theal <i>et al.</i> 2018 <sup>70</sup> (combined two sets of ad hoc criteria)	X	X	X	X	X	X	X	X	X	X	Appropriateness of statistical methods
Beydoun <i>et al.</i> 2019 <sup>71</sup>	X					X				X	Study design
Gianfredi <i>et al.</i> 2018	X	X	X		X	X					
Gianfredi <i>et al.</i> 2018 <sup>72,73</sup>											
Asgari-Taee <i>et al.</i> 2018 <sup>74</sup>	X			X				X			Study design, appropriateness of statistical methods, validity of findings
Padilha <i>et al.</i> 2018 <sup>75</sup>											
Dandamudi <i>et al.</i> 2018 <sup>76</sup> (combined two sets of ad hoc criteria)	X	X	X	X							
Dobbels <i>et al.</i> 2019 <sup>77</sup>		X	X	X	X	X	X	X	X	X	Study design
Dallacker <i>et al.</i> 2018 <sup>78</sup>		X	X			X			X		Study design

AHRQ, Agency for Healthcare Research and Quality; NICE, National Institute for Health and Care Excellence; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

**Table 4** Incorporation of risk of bias in the synthesis of results

	Number of reviews (%) n=131 (reviews that assessed risk of bias)
Did the review exclude studies at high risk of bias from either the review or the synthesis?	
Studies at high risk of bias were excluded from the review	1 (0.7)
Studies at high risk of bias were excluded from the meta-analysis	2 (1.5)
Studies were not excluded from either the review or synthesis based on risk of bias	128 (97.7)
Among reviews that assessed performed meta-analysis (n=101; 77.1%), was a subgroup analysis or meta-regression based on risk of bias conducted?	
Yes; based on overall risk of bias	28 (27.7)
Yes; based on specific risk of bias items/domains	7 (6.9)
No	66 (65.3)
Among reviews that conducted a subgroup analysis or meta-regression based on risk of bias (n=35; 26.7%), was the subgroup analysis or meta-regression statistically significant?	
Yes	5 (14.2)
No	22 (62.8)
Not reported	8 (22.8)

bias due to the virtual absence of standard practices for the registration of protocols and statistical analysis plans.<sup>44 50–52</sup> Our findings also suggest that review authors may disregard biases that they consider to be inherent to the design of nutritional epidemiologic studies.

We identified important deficiencies related to the reporting of risk of bias. Among reviews that assessed risk of bias, for example, nearly half did not report risk of bias judgements for each item or domain of the risk of bias tool. Further, reviews rarely described the criteria that were used to judge each risk of bias item or domain. For example, while almost all tools included an item or domain addressing risk of bias related to the measurement of the exposure, criteria for classifying measures as sufficiently valid and reliable and at low risk of bias were seldom described. Such deficiencies in reporting prevent evidence users from understanding the nature and extent of biases in studies.

We found most reviews did not sufficiently address risk of bias in their synthesis of results or interpretation of findings. Only half of reviews, for example, incorporated risk of bias assessments in statistical analyses, which is important to detect potential differences in results between studies at higher versus lower risk of bias.<sup>16</sup> While review authors often discussed the possibility of confounding and misclassification of the exposure, other important biases, such as biases due to missing data and selective reporting, were rarely discussed. Finally, review

**Table 5** Incorporation of risk of bias in the interpretation of review findings

Did the review consider the overall risk of bias across the body of evidence in the interpretation of findings?	
Yes, overall high risk of bias is described as a limitation	11 (7.3%)
Yes, overall low risk of bias is used to support findings	14 (9.3%)
No	125 (83.3%)
Did the review consider bias due to the selection of participants in the interpretation of findings?	
Yes, potential for selection bias is acknowledged as a limitation	14 (9.3%)
Yes, selection bias is described as unlikely	10 (6.7%)
No	126 (84.0%)
Did the review consider bias due to confounding in the interpretation of findings?	
Yes, potential for confounding is acknowledged as a limitation	77 (51.3%)
Yes, confounding is described as unlikely	2 (1.3%)
No	71 (47.3%)
Did the review consider bias due to the misclassification of the exposure in the interpretation of findings?	
Yes, potential for bias due to misclassification of the exposure is acknowledged as a limitation	85 (56.7%)
Yes, bias due to misclassification of the exposure is described as unlikely	6 (4.0%)
No	59 (39.3%)
Did the review consider bias due to departures from the intended exposure in the interpretation of findings?	
Yes, potential for bias due to departures from the intended exposure is acknowledged as a limitation	23 (15.3%)
Yes, bias due to departures from the intended exposure is described as unlikely	0 (0%)
No	127 (84.7%)
Did the review consider bias due to missing outcome data in the interpretation of findings?	
Yes, potential for bias due to missing outcome data is acknowledged as a limitation	2 (1.3%)
Yes, bias due to missing outcome data is described as unlikely	0 (0.0%)
No	148 (98.6%)
Did the review consider bias in the measurement of the outcome in the interpretation of findings?	
Yes, potential for bias in the measurement of the outcome is acknowledged as a limitation	23 (15.3%)
Yes, bias in the measurement of the outcome is described as unlikely	4 (2.7%)
No	123 (82.0%)
Did the review consider bias due to selective reporting in the interpretation of findings?	
Yes, potential for selective reporting bias is acknowledged as a limitation	1 (0.6%)

Continued



**Table 5** Continued

Yes, selective reporting bias is described as unlikely	0 (0.0%)
No	149 (99.3%)
Did the review hypothesise about the likely direction of bias?	
Yes, the authors hypothesise that effects for studies at high risk of bias are likely to have been biased away from the null	6 (4.0%)
Yes, the authors hypothesise that effects for studies at high risk of bias are likely to have been biased towards the null	14 (9.3%)
No	130 (86.7%)
Did the review evaluate the certainty of evidence using a formal system?	
Yes, using GRADE <sup>15</sup>	9 (6%)
Yes, using NutriGRADE <sup>79</sup>	2 (1.3%)
Yes, using SIGN <sup>80</sup>	1 (0.7%)
Yes, using the NHMRC FORM methodology <sup>81</sup>	1 (0.7%)
Yes, using a modified version American Diabetes Association system <sup>32</sup>	1 (0.7%)
Yes, using a modified version of the National Osteoporosis Foundation evidence grading system <sup>82</sup>	1 (0.7%)
Yes, using an ad hoc system	1 (0.7%)
No	134 (89.3%)
Among reviews that used a formal system to evaluate the certainty of evidence (n=16; 10.7%), was the certainty of evidence downgraded due to risk of bias?	
Yes	5 (31.3%)
No	11 (73.3%)

authors often neglected to make a judgement regarding the overall risk of bias of the body of the evidence, which is a critical step in evaluating the overall certainty of evidence.<sup>14</sup>

We hypothesise that reviews of RCTs addressing nutrition interventions also have limitations related to the assessment and interpretation of risk of bias. We restricted the scope of this research to reviews of nutritional epidemiologic studies because methods for the assessment of risk of bias of RCTs are better established than for non-randomised studies and there are unique challenges to assessing the risk of bias of nutritional epidemiologic studies, such as assessing the validity and reliability of dietary measures.

### Relation to previous work

To our knowledge, our study is the first to evaluate methods for the assessment of risk of bias in systematic reviews of nutritional epidemiologic studies. Previous studies that have addressed the assessment of risk of bias in general biomedical reviews have also found that reviews use a range of different tools to assess the risk of bias of non-randomised studies,<sup>53–56</sup> existing risk of bias tools do not address all important sources of bias in non-randomised studies and often include constructs that are unrelated to

risk of bias,<sup>54 56</sup> and that reviews often fail to incorporate assessments of risk of bias in the synthesis of results and interpretation of findings.<sup>57–59</sup> Our findings add to the body of evidence that suggests advancements in methods for the assessment of risk of bias—both in nutritional epidemiology and in other fields comprised primarily of non-randomised studies—are urgently needed.

### Implications and recommendations

Evidence users should be aware that risk of bias assessments in reviews of nutritional epidemiologic studies often have important limitations due to which findings from such reviews may be misleading.<sup>13–18</sup> We have compiled a list of recommendations for review authors that describe optimal methods for the assessment, reporting and interpretation of risk of bias in reviews of nutritional epidemiologic studies (box 1). We acknowledge, however, that there is great uncertainty in optimal tools and methods for the assessment of risk of bias in nutritional epidemiology. Our recommendations provide guidance on accepted best practice in the interim until further advancements.

### Strengths and limitations

This study summarises current methods for the assessment of risk of bias in reviews of nutritional epidemiologic studies and presents recommendations for review authors to improve risk of bias assessments in future reviews. Other strengths of this study include the duplicate assessment of review eligibility and data collection, which reduces the risk for errors.

This study also has limitations. While we identified many deficiencies and errors in the assessment of risk of bias in reviews, it is unclear the extent to which such issues may have impacted the conclusions drawn by reviewers and the implementation of evidence. Empirical evidence suggests that the failure to appropriately consider risk of bias can reduce the interpretability of review findings and even lead to misleading conclusions.<sup>13–18</sup> It is possible, however, that evidence users use the most rigorous systematic reviews—which apply rigorous methods for the assessment of the risk of bias—in which case the impact of the issues described in our study may be negligible.

Our analysis is also limited by the possibility that review authors did not address certain biases in the assessment of risk of bias or in the interpretation of review findings because they were deemed to not be a concern in the included studies. However, failure to use comprehensive risk of bias tools combined with failure to adequately address risk of bias in the interpretation of findings leaves evidence users unable to gauge the overall risk of bias of the evidence.

### CONCLUSION

Systematic reviews of nutritional epidemiologic studies often have important limitations related to their assessment of risk of bias. Review authors can improve risk

## Box 1 Recommendations for authors of systematic reviews addressing the assessment of risk of bias of nutritional epidemiologic studies

### 1. Assess the risk of bias of included studies using an appropriate tool or set of criteria.

Review authors should assess the risk of bias of all included studies. Our investigation shows that there is currently no consensus among review authors on the optimal tool for the assessment of risk of bias of nutritional epidemiologic studies and that many commonly used tools have important limitations. Review authors should select a tool that addresses all potential sources of bias in non-randomised studies, including biases due to confounding, inappropriate criteria for selection of participants, error in the measurement of the exposure and outcome, departures from the intended exposure, missing outcome data and selective reporting, and that does not include constructs unrelated to risk of bias, such as reporting, precision or generalisability.<sup>11 12</sup> Review authors should avoid using quality tools that combine risk of bias with other constructs since such tools are poor indicators of risk of bias.

The selected tool should assign studies a qualitative category, and not a quantitative score, representing the degree of risk of bias (eg, 'low risk', 'moderate risk', 'serious risk' and 'critical risk') (42). The overall study risk of bias should reflect the highest rated risk of bias item or domain (ie, a single limitation in a crucial aspect of the study should be considered sufficient to put the study at high risk of bias). We direct review authors to the Cochrane-endorsed ROBINS-I tool, which addresses all established sources of bias in non-randomised studies, does not include unrelated constructs and which is accompanied by an additional guidance document for its implementation.<sup>7</sup> A similar tool, called the ROBINS-E, for the assessment of risk of bias of non-randomised studies of exposures, modelled after the ROBINS-I, is currently under development.<sup>83</sup> A preliminary version of the ROBINS-E tool is available for piloting (<https://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/robins-e/>), which shares much of the same structure and guiding questions of the ROBINS-I. Despite concerns about their complexity and low inter-rater reliability, the ROBINS-E and ROBINS-I tools appear to be the most rigorous and comprehensive tools available for the assessment of risk of bias of non-randomised studies.<sup>84 85</sup>

Researchers in nutrition and environmental sciences are often concerned that evidence from non-randomised studies may be discounted in favour of randomised trials despite feasibility concerns with conducting rigorous trials in these fields.<sup>86</sup> ROBINS-I, however, may be the least likely among available risk of bias tools for non-randomised studies to discount evidence from non-randomised studies since it accommodates situations in which non-randomised studies may provide high or moderate certainty evidence, similar to RCTs.<sup>87</sup>

ROBINS-I's consideration of the magnitude of bias as ranging from low risk to critical risk may be considered another advantage to other tools that simply classify studies at low or high risk of bias. We note, however, that judgements related to the magnitude of bias or importance of bias are complex, most often not justified by empirical evidence, and difficult to make for users.<sup>40–42</sup>

### 2. Assess risk of bias in comparison to a 'target' RCT.

Bias can arise due to the actions of study investigators (eg, failure to follow up all study participants) or may be unavoidable due to constraints on how studies addressing a particular question can be designed.<sup>11</sup> Our findings suggest that the latter category of bias may

## Box 1 Continued

often be neglected by review authors. The assessment of risk of bias relative to a target RCT—a hypothetical RCT that may or may not be feasible, which addresses the question of interest without any features putting it at risk of bias—provides a benchmark against which risk of bias can be assessed and can ensure that biases that are inherent to the design of studies addressing particular questions are also accounted for.<sup>7 39 88</sup> This approach is also incorporated in the ROBINS-I tool.<sup>7</sup>

Some nutritional and environmental science researchers have expressed concern with this approach since a trial of nutritional or environmental exposures may not be feasible.<sup>86 89</sup> We emphasise that the 'target' RCT need not be feasible or practical. This is important because 'target' RCTs will not be limited by typical limitations of dietary trials such as poor adherence and attrition due to the need for long follow-up.

### 3. Report all criteria that were used to judge each risk of bias item or domain.

Review authors will need to make judgements regarding which study design features sufficiently protect against bias and which design features may lead to bias. For example, review authors will need to identify factors that may act as confounders for the question being addressed and will need to determine which methods for the measurement of the exposure and outcome of interest are valid and reliable. Ideally, review authors should develop criteria to make these judgements a priori to avoid risk of bias assessments from being influenced by the results of studies.

Authors of nutritional epidemiology reviews may find making judgements related to biases due to confounding and the classification of the exposure to be particularly challenging. When deciding on the list of potential confounders that should be controlled for a study to be considered at low risk of confounding bias, review authors should consider the evidence on prognostic factors of the outcome of interest and correlates of the exposure. The list of confounders should not be generated solely on the basis of confounders considered in primary studies (at least, not without some form of independent confirmation).<sup>90</sup> We refer the reader to other sources that describe optimal methods for the selection of confounders.<sup>90–93</sup>

In making judgements regarding the risk of bias associated with the classification of the exposure, review authors can typically rate well-established biomarkers of nutritional exposures at lowest risk of bias (eg, 24-hour urinary sodium excretion for sodium intake).<sup>94</sup> Dietary records and diaries can usually be considered more valid than recall-based methods, although all self-reported methods suffer from serious limitations.<sup>47 48 95 96</sup> The validity of food frequency questionnaires and other recall-based methods also depends on the results of validation studies.<sup>48</sup> A questionnaire may be sufficiently valid for some exposures and may not have been validated or may not be valid for other exposures and so review authors should look for results of validation studies specific to the exposure being investigated. Review authors should also consider food composition databases that are used to derive nutrient intake levels from food intake data. For these databases to be considered valid, they should represent the nutrient composition of the foods at the time they were consumed and account for variations.<sup>97</sup> For example, there are wide variations in the nutrient content of different brands of fruit juices and breakfast cereals and food composition databases that do not include brand-specific information may not yield accurate nutrient intake data from these foods.

Continued

Continued

## Box 1 Continued

**4. Conduct risk of bias assessments in duplicate.**

To reduce the risk for errors, review authors should conduct risk of bias assessments in duplicate.<sup>98 99</sup> Review authors should resolve discrepancies by discussion or, when discussion is insufficient, adjudication by a third party with expertise in methodology and nutritional epidemiology.

**5. Conduct risk of bias assessments for each result used in the synthesis.**

A single non-randomised study may report several numerical results representing the effect of a single nutritional exposure on a health outcome. For example, a study may report the association between multiple eligible measures of the exposure and outcome, at multiple eligible timepoints, or using several analytical specifications, which may all vary in their risk of bias. Hence, review authors should perform risk of bias assessments separately for each numerical result that is extracted and used in the synthesis.<sup>11</sup>

**6. Report risk of bias judgements for all items or domains of the risk of bias tool.**

For all studies, review authors should report risk of bias judgements for all items or domains of the risk of bias tool. One way in which this information can be presented is through traffic light plots that use a colour-coded system to represent risk of bias ratings across items or domains.<sup>11</sup> Traffic light plots may also be presented adjacent to forest plots to allow evidence users to simultaneously visualise the results from studies, their relative contributions in the meta-analysis and their risk of bias. This approach allows evidence users to identify the risk of bias of the most influential studies and to identify variations in the results of studies based on risk of bias ratings.

**7. Incorporate risk of bias in the synthesis of results.**

Authors should address risk of bias in their synthesis of results using one or more of the following methods: (1) restricting the eligibility of studies for inclusion in the synthesis to only those that are at low risk of bias and conducting a sensitivity analysis including all studies (typically when there is sufficient evidence available from studies at low risk of bias); (2) performing subgroup analyses or meta-regressions to explore differences in results of studies at lower or higher risk of bias; or (3) reporting results based on all available studies, along with a description of the risk of bias and how and to what degree bias may have influenced the results—the latter of which is the least informative of the three strategies but is the only plausible approach when there is no variability in risk of bias across studies.<sup>11</sup> Review authors should use either the first or third method when the synthesis is narrative rather than quantitative.

Review authors may also attempt to adjust the results of studies in an attempt to remove bias or in attempt to incorporate the uncertainty of results from non-randomised studies as part of the study precision.<sup>100</sup>

The operationalisation of these approaches, however, is difficult because it requires review authors to make assumptions about the direction and/or magnitude of biases, either based on expert opinion, which is often arbitrary, or based on empirical evidence, which is limited.

**8. Incorporate risk of bias in the interpretation of review findings.**

Review authors should make a judgement regarding the overall risk of bias across the body of evidence to inform the evaluation of the overall certainty of evidence. For example, the application of the GRADE approach, the most widely endorsed system for evaluating the certainty of evidence, requires review authors to make a judgement regarding

## Box 1 Continued

whether the certainty of evidence should be downgraded for risk of bias.<sup>15</sup> In making this judgement, review authors should consider the relative contribution of studies at higher versus lower risk of bias (ie, larger studies or studies with a greater number of events have a more significant contribution than smaller studies or studies with few events) and whether there are appreciable differences in the results of studies at higher versus lower risk of bias.<sup>101</sup> When studies at high risk of bias contribute substantially and when there is insufficient evidence from studies at low risk of bias, review authors should express less certainty in the effect estimate. Alternatively, when studies at high risk of bias have only a small contribution or when studies at high and low risk of bias report consistent results, review authors may not need to be concerned about risk of bias. When there are appreciable differences in the results of studies at higher versus lower risk of bias, estimates from studies at lower risk of bias may be considered more credible and review conclusions may be primarily based on these lower risk of bias studies. Review authors should justify their decision whether to rate down the certainty of evidence for risk of bias and justify their concerns, if any, related to risk of bias. Review authors must also be cognisant of personal biases, including perceptions of 'landmark' studies by established names in the field as automatically highly credible.

**9. Involve researchers with substantive knowledge of the topic.**

Researchers with substantive knowledge of the topic can suggest important criteria for the assessment of each risk of bias domain (eg, the validity and reliability of dietary measures) and should be consulted in the assessment of risk of bias.

of bias assessments in future reviews by using tools that address all potential sources of bias in non-randomised studies without including unrelated constructs, by transparently reporting all risk of bias judgements, and by incorporating risk of bias assessments in the synthesis of results and the interpretation of findings. Additional guidance for applying existing risk of bias tools to non-randomised studies, particularly nutritional epidemiologic studies, is needed.

**Author affiliations**

<sup>1</sup>Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA

<sup>2</sup>Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

<sup>3</sup>McMaster University, Hamilton, Ontario, Canada

<sup>4</sup>Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

<sup>5</sup>Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada

<sup>6</sup>Population Health Research Institute, Hamilton Health Sciences Corporation, Hamilton, Ontario, Canada

**Twitter** Dena Zeraatkar @dena.zera

**Contributors** DZ, JB and RJdS designed the study. DZ, AK, AB, REM, IC, AG, DL, AM, ES, KA and MA collected data. DZ and AK analysed data. DZ, AK, JB and RJdS interpreted the data. DZ produced the first draft of the article. DZ, JB and RJdS provided critical revision of the article for important intellectual content. All authors approved the final version of the article. DZ and RJdS are the guarantors.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** RJ de Souza has served as an external resource person to the World Health Organization's Nutrition Guidelines Advisory Group on trans

Continued

fats, saturated fats, and polyunsaturated fats. The WHO paid for his travel and accommodation to attend meetings from 2012–2017 to present and discuss this work. He has also done contract research for the Canadian Institutes of Health Research's Institute of Nutrition, Metabolism, and Diabetes, Health Canada, and the World Health Organization for which he received remuneration. He has received speaker's fees from the University of Toronto, and McMaster Children's Hospital. He has held grants from the Canadian Institutes of Health Research, Canadian Foundation for Dietetic Research, Population Health Research Institute, and Hamilton Health Sciences Corporation as a principal investigator, and is a co-investigator on several funded team grants from the Canadian Institutes of Health Research. He serves as a member of the Nutrition Science Advisory Committee to Health Canada (Government of Canada), a co-opted member of the Scientific Advisory Committee on Nutrition (SACN) Subgroup on the Framework for the Evaluation of Evidence (Public Health England), and as an independent director of the Helderleigh Foundation (Canada).

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository (<https://osf.io/WYQHE/>).

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Dena Zeraatkar <http://orcid.org/0000-0003-4287-0541>

## REFERENCES

- Boeing H. Nutritional epidemiology: new perspectives for understanding the diet-disease relationship? *Eur J Clin Nutr* 2013;67:424–9.
- Zeraatkar D, Johnston BC, Guyatt G. Evidence collection and evaluation for the development of dietary guidelines and public policy on nutrition. *Annu Rev Nutr* 2019;39:227–47.
- Ortiz-Moncada R, González-Zapata L, Ruiz-Cantero MT, et al. Priority issues, study designs and geographical distribution in nutrition journals. *Nutr Hosp* 2011;26:784–91.
- Hébert JR, Frongillo EA, Adams SA, et al. Perspective: randomized controlled trials are not a panacea for diet-related research. *Adv Nutr* 2016;7:423–32.
- Brannon PM, Taylor CL, Coates PM. Use and applications of systematic reviews in public health nutrition. *Annu Rev Nutr* 2014;34:401–19.
- Zeraatkar D, Bhasin A, Morassut RE, et al. Characteristics and quality of systematic reviews and meta-analyses of observational nutritional epidemiology: a cross-sectional study. *Am J Clin Nutr* 2021;113:1578–1592.
- Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
- Blair A, Stewart P, Lubin JH, et al. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. *Am J Ind Med* 2007;50:199–207.
- Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;359:248–52.
- Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.0*. Cochrane, 2019.
- Institute of Medicine Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. *Finding what works in health care: standards for systematic reviews*. Washington (DC): National Academies Press (US), 2011.
- Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6.
- Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- Büttner F, Winters M, Delahunt E, et al. Identifying the 'incredible'! Part 2: Spot the difference – a rigorous risk of bias assessment can alter the main findings of a systematic review. *Br J Sports Med* 2020;54:801–808.
- Losilla J-M, Oliveras I, Marin-Garcia JA, et al. Three risk of bias tools lead to opposite conclusions in observational research synthesis. *J Clin Epidemiol* 2018;101:61–72.
- Voss PH, Rehfuess EA. Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. *J Epidemiol Community Health* 2013;67:98–104.
- Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36:666–76.
- Kiss N, Tongbram V, Fortier KJ. Quality assessment of observational studies for systematic reviews. *Value Health* 2013;16:A614.
- Moja LP, Telaro E, D'Amico R, et al. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 2005;330:1053.
- Dekkers OM, Vandenbroucke JP, Cevallos M, et al. COSMOS-E: guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. *PLoS Med* 2019;16:e1002742.
- Page MJ, Shamsler L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;13:e1002028.
- Salam RA, Welch V, Bhutta ZA. Systematic reviews on selected nutrition interventions: descriptive assessment of conduct and methodological challenges. *BMC Nutr* 2015;1:9.
- Naude CE, Durao S, Harper A, et al. Scope and quality of Cochrane reviews of nutrition interventions: a cross-sectional study. *Nutr J* 2017;16:22.
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539–58.
- Viswanathan M, Ansari MT, Berkman ND. *AHRQ methods for effective health care: assessing the risk of bias of individual studies in systematic reviews of health care interventions. methods guide for effectiveness and comparative effectiveness reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US), 2008.
- Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open* 2018;8:e019703.
- Herzog R, Álvarez-Pasquin MJ, Díaz C, et al. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. *BMC Public Health* 2013;13:154.
- Critical Appraisal Skills Programme. Cohort study checklist, 2018. Available: [https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist\\_2018.pdf](https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist_2018.pdf)
- National Institutes of Health. Study quality assessment tools. Bethesda, MD: National Institutes of Health. Available: <https://www.nlm.nih.gov/health-topics/study-quality-assessment-tools>
- Introduction: The American Diabetes Association's (ADA) evidence-based practice guidelines, standards, and related recommendations and documents for diabetes care. *Diabetes Care* 2012;35 Suppl 1:S1–2.
- Ismail SR, Maarof SK, Siedar Ali S, Ali SS, et al. Systematic review of palm oil consumption and the risk of cardiovascular disease. *PLoS One* 2018;13:e0193533.
- Picasso MC, Lo-Tayrac JA, Ramos-Villanueva JM, et al. Effect of vegetarian diets on the presentation of metabolic syndrome or its components: a systematic review and meta-analysis. *Clin Nutr* 2019;38:1117–1132.
- Mijatovic-Vukas J, Capling L, Cheng S, et al. Associations of diet and physical activity with risk for gestational diabetes mellitus: a systematic review and meta-analysis. *Nutrients* 2018;10. doi:10.3390/nu10060698. [Epub ahead of print: 30 May 2018].

- 36 Panahande B, Sadeghi A, Parohan M. Alternative healthy eating index and risk of hip fracture: a systematic review and dose-response meta-analysis. *J Hum Nutr Diet* 2019;32:98–107.
- 37 Hu D, Cheng L, Jiang W. Fruit and vegetable consumption and the risk of postmenopausal osteoporosis: a meta-analysis of observational studies. *Food Funct* 2018;9:2607–16.
- 38 Wells G. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analysis, 2004. Available: [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp)
- 39 Hernán MA, Sauer BC, Hernández-Díaz S, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016;79:70–5.
- 40 Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
- 41 Emerson JD, Burdick E, Hoaglin DC, et al. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin Trials* 1990;11:339–52.
- 42 Jüni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
- 43 Bae J-M. A suggestion for quality assessment in systematic reviews of observational studies in nutritional epidemiology. *Epidemiol Health* 2016;38:e2016014..
- 44 Trepanowski JF, Ioannidis JPA. Perspective: limiting dependence on nonrandomized studies and improving randomized trials in human nutrition research: why and how. *Adv Nutr* 2018;9:367–77.
- 45 Ioannidis JPA. Unreformed nutritional epidemiology: a lamp post in the dark forest. *Eur J Epidemiol* 2019;34:327–31.
- 46 Mendez MA. Invited commentary: Dietary misreporting as a potential source of bias in diet-disease associations: future directions in nutritional epidemiology research. *Am J Epidemiol* 2015;181:234–6.
- 47 Archer E, Marlow ML, Lavie CJ. Controversy and debate: memory-based methods paper 1: the fatal flaws of food frequency questionnaires and other memory-based dietary assessment methods. *J Clin Epidemiol* 2018;104:113–24.
- 48 Kirkpatrick SI, Baranowski T, Subar AF, et al. Best practices for conducting and interpreting studies to validate self-report dietary assessment methods. *J Acad Nutr Diet* 2019;119:1801–16.
- 49 Archer E, Lavie CJ, Hill JO. The failure to measure dietary intake engendered a fictional discourse on diet-disease relations. *Front Nutr* 2018;5:105.
- 50 Silberzahn R, Uhlmann EL, Martin DP, et al. Many analysts, one data set: making transparent how variations in analytic choices affect results. *Adv Methods Pract Psychol Sci* 2018;1:337–56.
- 51 Gelman A, Loken E. *The garden of forking paths: why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Department of Statistics, Columbia University, 2013.
- 52 Thomas L, Peterson ED. The value of statistical analysis plans in observational research: defining high-quality research from the start. *JAMA* 2012;308:773–4.
- 53 Farrah K, Young K, Tunis MC, et al. Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols. *Syst Rev* 2019;8:280.
- 54 Quigley JM, Thompson JC, Halpenny NJ, et al. Critical appraisal of nonrandomized studies—a review of recommended and commonly used tools. *J Eval Clin Pract* 2019;25:44–52.
- 55 Seehra J, Pandis N, Koletsi D, et al. Use of quality assessment tools in systematic reviews was varied and inconsistent. *J Clin Epidemiol* 2016;69:179–84.
- 56 Shamlilyan T, Kane RL, Dickinson S. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *J Clin Epidemiol* 2010;63:1061–70.
- 57 Hopewell S, Boutron I, Altman DG, et al. Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study. *BMJ Open* 2013;3:e003342.
- 58 Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. *J Epidemiol Community Health* 2015;69:189–95.
- 59 Babic A, Vuka I, Saric F, et al. Overall bias methods and their use in sensitivity analysis of Cochrane reviews were not consistent. *J Clin Epidemiol* 2020;119:57–64.
- 60 Rostam A, Dubé C, Cranney A. *104 celiac disease: summary*. Rockville, MD, 2004.
- 61 Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158:280–6.
- 62 Scottish Intercollegiate Guidelines Network. *SIGN 50: a guideline developers' handbook*. Scottish Intercollegiate Guidelines Network, 2001.
- 63 von Elm E, Altman DG, Egger M, et al. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg* 2014;12:1495–9.
- 64 Zaza S, Wright-De Agüero LK, Briss PA, et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000;18:44–74.
- 65 Thomas BH, Ciliska D, Dobbins M, et al. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. *Worldviews Evid Based Nurs* 2004;1:176–84.
- 66 Pace R, Pluye P, Bartlett G, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *Int J Nurs Stud* 2012;49:47–53.
- 67 National Institute for Health and Care Excellence. *The social care guidance manual*, 2013.
- 68 Wong WCW, Cheung CSK, Hart GJ. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. *Emerg Themes Epidemiol* 2008;5:23.
- 69 Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65:163–78.
- 70 Theal R, Tay VXP, Hickman IJ. Conflicting relationship between dietary intake and metabolic health in PTSD: a systematic review. *Nutr Res* 2018;54:12–22.
- 71 Beydoun MA, Chen X, Jha K, et al. Carotenoids, vitamin A, and their association with the metabolic syndrome: a systematic review and meta-analysis. *Nutr Rev* 2019;77:32–45.
- 72 Gianfredi V, Salvatori T, Nucci D, et al. Can chocolate consumption reduce cardio-cerebrovascular risk? A systematic review and meta-analysis. *Nutrition* 2018;46:103–14.
- 73 Gianfredi V, Nucci D, Abalsamo A, et al. Green tea consumption and risk of breast cancer and recurrence—A systematic review and meta-analysis of observational studies. *Nutrients* 2018;10. doi:10.3390/nu10121886. [Epub ahead of print: 03 Dec 2018].
- 74 Asgari-Taee F, Zerfati-Shoae N, Dehghani M, et al. Association of sugar sweetened beverages consumption with non-alcoholic fatty liver disease: a systematic review and meta-analysis. *Eur J Nutr* 2019;58:1–11.
- 75 Dos Reis Padilha G, Sanches Machado d'Almeida K, Ronchi Spillere S, et al. Dietary patterns in secondary prevention of heart failure: a systematic review. *Nutrients* 2018;10. doi:10.3390/nu10070828. [Epub ahead of print: 26 Jun 2018].
- 76 Dandamudi A, Tommie J, Nommsen-Rivers L, et al. Dietary patterns and breast cancer risk: a systematic review. *Anticancer Res* 2018;38:3209–22.
- 77 Dobbels F, Denhaerynck K, Klem ML, et al. Correlates and outcomes of alcohol use after single solid organ transplantation: a systematic review and meta-analysis. *Transplant Rev* 2019;33:17–28.
- 78 Dallacker M, Hertwig R, Mata J. The frequency of family meals and nutritional health in children: a meta-analysis. *Obes Rev* 2018;19:638–53.
- 79 Schwingshackl L, Knüppel S, Schwedhelm C, et al. Perspective: NutriGrade: a scoring system to assess and judge the meta-evidence of randomized controlled trials and cohort studies in nutrition research. *Adv Nutr* 2016;7:994–1004.
- 80 Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;323:334–6.
- 81 Hillier S, Grimmer-Somers K, Merlin T, et al. Form: an Australian method for formulating and grading recommendations in evidence-based clinical guidelines. *BMC Med Res Methodol* 2011;11:23.
- 82 Wallace TC, Bauer DC, Gagel RF, et al. The National Osteoporosis Foundation's methods and processes for developing position statements. *Arch Osteoporos* 2016;11:22.
- 83 Morgan RL, Thayer KA, Santesso N, et al. A risk of bias instrument for non-randomized studies of exposures: a users' guide to its application in the context of GRADE. *Environ Int* 2019;122:168–84.
- 84 Jeyaraman MM, Rabbani R, Copstein L, et al. Methodologically rigorous risk of bias tools for nonrandomized studies had low reliability and high evaluator burden. *J Clin Epidemiol* 2020;128:140–147.
- 85 Bero L, Chartres N, Diong J, et al. The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures. *Syst Rev* 2018;7:242.

- 86 Steenland K, Schubauer-Berigan MK, Vermeulen R, *et al.* Risk of bias assessments and evidence syntheses for observational epidemiologic studies of environmental and occupational exposures: strengths and limitations. *Environ Health Perspect* 2020;128:95002.
- 87 Schünemann HJ, Cuello C, Akl EA, *et al.* GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol* 2019;111:105–14.
- 88 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.
- 89 Arroyave WD, Mehta SS, Guha N, *et al.* Challenges and recommendations on the conduct of systematic reviews of observational epidemiologic studies in environmental and occupational health. *J Expo Sci Environ Epidemiol* 2021;31:21–30.
- 90 Greenland S, Pearce N. Statistical foundations for model-based adjustments. *Annu Rev Public Health* 2015;36:89–108.
- 91 Lederer DJ, Bell SC, Branson RD, *et al.* Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals. *Ann Am Thorac Soc* 2019;16:22–8.
- 92 Sauer BC, Brookhart MA, Roy J, *et al.* A review of covariate selection for non-experimental comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2013;22:1139–45.
- 93 Zeraatkar D, Cheung K, Milio K, *et al.* Methods for the selection of covariates in nutritional epidemiology studies: a meta-epidemiological review. *Curr Dev Nutr* 2019;3:nzz104.
- 94 Cogswell ME, Maalouf J, Elliott P, *et al.* Use of urine biomarkers to assess sodium intake: challenges and opportunities. *Annu Rev Nutr* 2015;35:349–87.
- 95 Prentice RL. Dietary assessment and the reliability of nutritional epidemiology research reports. *J Natl Cancer Inst* 2010;102:583–5.
- 96 Martín-Calvo N, Martínez-González Miguel Ángel. Controversy and debate: memory-based dietary assessment methods paper 2. *J Clin Epidemiol* 2018;104:125–9.
- 97 Marconi S, Durazzo A, Camilli E, *et al.* Food composition databases: considerations about complex food matrices. *Foods* 2018;7. doi:10.3390/foods7010002. [Epub ahead of print: 01 Jan 2018].
- 98 Buscemi N, Hartling L, Vandermeer B, *et al.* Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol* 2006;59:697–703.
- 99 Waffenschmidt S, Knelangen M, Sieben W, *et al.* Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. *BMC Med Res Methodol* 2019;19:132.
- 100 Ioannidis JPA. Commentary: adjusting for bias: a user's guide to performing plastic surgery on meta-analyses of observational studies. *Int J Epidemiol* 2011;40:777–9.
- 101 Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.