Methodological review to develop a list of bias items used to assess reviews incorporating network meta-analysis: protocol and rationale

Carole Lunny; Andrea C Tricco; Areti-Angeliki Veroniki; Sofia Dias; Brian Hutton; Georgia Salanti; James M Wright; Ian White; Penny Whiting

INTRODUCTION
Reviews with network meta-analysis (NMAs) have gained popularity due to their ability to provide comparative effectiveness of multiple treatments for the same condition. Reviews with NMA have grown in number. Between 1997 and 2015, 771 NMAs were published in 336 journals from 3459 authors and 1258 institutions in 49 countries. More than three-quarters (n=625; 81%) of these NMAs were published in the last 5 years. Many organisations such as the National Institute for Health and Care Excellence (NICE) in the UK, the World Health Organisation (WHO) and the Canadian Agency for Drugs and Technologies in Health (CADTH) conduct NMAs as they represent the best available evidence to inform clinical practice guidelines.3-5 We adopt a broad definition of NMAs, specifically: a review that intends to, simultaneously synthesise more than two health care interventions of interest. Reviews that intend to compare multiple treatments with an NMA but then find that the assumptions are violated (eg, a disconnected network, or studies are too heterogeneous to combine) and that NMA is not feasible, will also be included in our definition.

Evidence shows that biased results from poorly designed and reported studies can...
mislead decision-making in healthcare at all levels.6-9 If a review is at risk of bias and inappropriate methods are used, the validity of the findings can be compromised.10-12 Evaluating how well a review has been conducted is essential to determining whether the findings are relevant to patient care and outcomes. Several empirical studies have shown that bias can obscure the real effects of a treatment.13-16 Being able to appraise reviews with NMA is central to informed decision-making in patient care.

The systematic procedures required to conduct a systematic review help mitigate the risk of bias. However, bias can also be introduced when interpreting the reviews findings. For example, review’s conclusions may not be supported by the evidence presented, the relevance of the included studies may not have been considered by review authors, and reviewers may inappropriately emphasise results on the basis of their statistical significance.17 A well-conducted systematic review draws conclusions that are appropriate to the included evidence and can therefore be free of bias even when the primary studies included in the review have high risk of bias.

Tools are available for most study designs to make risk of bias assessment easier for a knowledge user (eg, healthcare practitioners, policymakers, patients18). Many tools and checklists can be used either when conducting a systematic review (quality of conduct), when assessing how well a study has been described (reporting), or when knowledge users want to assess the risk of bias in the conclusions of a review. The methodological quality of studies (ie, how well the study is conducted) is often confused with reporting quality (ie, how well authors describe their methodology and results). A risk of bias assessment is an assessment of review limitations, which focus on the potential of those methods to bias the study findings.17

More than 40 tools have been identified19 20 for critically appraising the quality of reviews with pairwise meta-analysis. AMSTAR (A MeaSurement Tool to Assess the methodological quality of systematic Reviews)21 and the OQAQ (Overview Quality Assessment Questionnaire22) have been identified as the most commonly used, and they follow a simple checklist format.20 23 AMSTAR has been recently updated to AMSTAR 2, which aims to evaluate how reviews are planned and conducted.24 The ROBIS (Risk Of Bias In Systematic reviews) tool is designed to assess the risk of bias in systematic reviews with or without pairwise meta-analysis.17 The ROBIS tool involves assessment of methodological features in reviews known to increase the risk of bias in review conclusions. Domain-based assessment tools require a careful reading and thoughtful analysis of the study to adequately rate risk of bias, instead of simply identifying keywords reported in the article, as usually made in a checklist type of assessment.

For critically appraising reviews with NMA, several checklists exist. To assess reporting quality, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement extension for reviews incorporating network meta-analysis (PRISMA-NMA) 25 or the National Institute for Health and Care Excellence Decision Support Unit checklist (NICE-DSU)26 can be used. To assess quality of conduct, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) checklist27 can be used. However, many quality assessment tools are not created rigorously. To be rigorous, they must follow a series of methodological steps.28 29 As a quality of conduct tool, the ISPOR checklist27 did not follow the methodology proposed by Whiting29 for creating a systematically developed quality tool. Due to important methodological advances in the field of NMA, the ISPOR, published in 2014, is also outdated. As table 1 shows, several tools are designed with different purposes; some for assessing reporting quality, and some for assessing quality of conduct but none are designed to assess risk of bias in NMAs.

A comprehensive and systematic process should be used to develop a rigorous risk of bias tool for assessing NMAs, as outlined in Whiting et al’s29 Framework for Developing Quality Assessment Tools. The first step is to: (1) conduct a systematic search of biases that can inform the assessment of the validity and reliability of NMAs and prepare a pilot list of items, (2) create a draft tool, (3) obtain expert opinion on the draft tool and inclusion of items through Delphi exercises, and (4) pilot test and refine the tool.29 No review has comprehensively and systematically listed and categorised all items related to bias in NMAs. Such a list will inform a new tool to assess the risk of bias in NMAs, and potentially other reporting or quality tools which are being updated.

OBJECTIVE
Our objective for this protocol paper is to plan the conduct of a methodological review to develop a list of items relating to bias in NMAs. This is the first step in the goal of developing a risk of bias tool to assess NMAs. Further steps will involve conducting a series of Delphi surveys to select, refine and compile items into a tool; pilot test and then refine the draft tool with different user groups; and finally develop and evaluate an evidence-based (knowledge translation) KT strategy to disseminate the tool. This protocol pertains to our first objective to systematically search for and identify a list of bias items for NMAs.

METHODS AND ANALYSIS
We will follow the methodology proposed by Whiting29 and Sanderson30 and Page7 for creating systematically developed lists of quality items. Although this protocol is for a methodological review, and not a health intervention review, our protocol was described and reported in accordance with the preferred reporting items for systematic reviews and meta-analysis protocol (PRISMA-P) checklist with not applicable indicated for items not pertaining to methods reviews31 (online supplemental appendix 1).
Table 1  Tools and checklists to aid in systematic review conduct, or to assess the reporting or methodological quality of a review

<table>
<thead>
<tr>
<th>Tool purpose</th>
<th>Examples of tools or checklists</th>
<th>Description of the example checklist or tool</th>
<th>Available tool for reviews with NMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guidance for conducting systematic reviews</td>
<td>MECIR</td>
<td>Detailed methodological guidance on how to conduct a systematic review with or without pairwise meta-analysis of effectiveness, diagnostic test accuracy, individual patient data reviews and reviews in public health and health promotion</td>
<td>No</td>
</tr>
<tr>
<td>Assess the quality of published reviews</td>
<td>AMSTAR-2, OQAQ</td>
<td>AMSTAR-2 is a critical appraisal tool to assess the quality of conduct of reviews of randomised controlled trials of interventions</td>
<td>ISPOR</td>
</tr>
<tr>
<td>Assess the risk of bias of published reviews</td>
<td>ROBIS</td>
<td>ROBIS is a tool for assessing the risk of bias in reviews. It is aimed at four broad categories of reviews mainly within healthcare settings: interventions, diagnosis, prognosis, and aetiology.</td>
<td>No</td>
</tr>
<tr>
<td>Assess the certainty in evidence and the strength of recommendations</td>
<td>GRADE</td>
<td>GRADE approach defines the certainty of a body of evidence as the extent to which one can be confident that a pooled effect estimate is close to the true effect of the intervention. Five domains assessed: risk of bias, inconsistency, indirectness, imprecision and publication bias.</td>
<td>GRADE-NMA, CINeMA, Threshold method</td>
</tr>
<tr>
<td>Guidelines for the complete reporting of published reviews</td>
<td>PRISMA</td>
<td>PRISMA is an evidence-based minimum set of items for reporting in reviews and meta-analyses. PRISMA focuses on the reporting of reviews evaluating randomised trials, but can also be used as a basis for reporting reviews of other types of research, particularly evaluations of interventions.</td>
<td>PRISMA-NMA, NICE-DSU</td>
</tr>
</tbody>
</table>

Eligibility criteria

There will be two types of studies included. Study type 1 are articles that present and describe items related to bias, reporting or methodological quality of reviews with NMA. Items related to reporting will be retained because they can potentially be translated into a risk of bias item. For example, in the PRISMA-P guideline, one item asks whether study PICO (Population, Interventions, Comparisons, Outcomes) characteristics were used as criteria for determining study eligibility. Reporting of all outcomes in a protocol may prevent authors from only selecting outcomes that are statistically significant when publishing their systematic review. This PRISMA-P reporting item can then be translated into a bias item related to the ‘selective reporting’ of outcomes. Study type 2 are studies that assess the methodological quality in a sample of reviews with NMA.

Study type 1 will meet any of these inclusion criterion

- Articles describing items related to bias or methodological quality in reviews with NMA (e.g., Dias 2018); tools that only assess general aspects of systematic reviews without focusing specifically on NMA will be excluded (e.g., AMSTAR, AMSTAR 2, or ROBIS). Articles describing editorial standards for reviews with NMA (e.g., similar to the Cochrane MECIR (Methodological standards for the conduct of new Cochrane Intervention Reviews) standards for systematic reviews).
- Articles describing items related to reporting quality in reviews with NMA (e.g., PRISMA-NMA).
- Articles identifying or addressing sources of bias and variation in NMA and published after PRISMA-NMA in 2014.

Study type 2 will meet any of these inclusion criterion

- Articles assessing the methodological quality (or risk of bias) of reviews with NMA (i.e., a sample of NMA are assessed for methodological quality; e.g., Chambers 2015) using criteria that focus specifically on aspects of NMA not just on general aspects of systematic reviews.

We will include articles with any publication status and in any language, and where the coauthors are not fluent in the language, Google Translate will be used. If through our main search, we identify a systematic review encompassing the eligible articles, or one aspect of the eligible article, we will use the results of the systematic review and only include primary studies published subsequent to the systematic review. For example, a review by Laws et al in 2019 identified all guidance documents for conducting an NMA from countries throughout the world. We therefore would not search for guidance documents published before the last search date of this review.

AMSTAR-2, A Measurement Tool to Assess Systematic Reviews 2; CiNeMA, Confidence in Network Meta-Analysis; GRADE, Grading of Recommendations Assessment, Development and Evaluation; GRADE-NMA, Grading of Recommendations Assessment, Development and Evaluation for Network Meta-Analysis; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; MECIR, Methodological Expectations of Cochrane Intervention Reviews; NICE-DSU, National Institute for Health and Care Excellence Decision Support Unit checklist; OQAQ, Overview Quality Assessment Questionnaire; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA-NMA, PRISMA statement extension for reviews incorporating NMA; ROBIS, Risk Of Bias In Systematic reviews.
Search strategy
We will search Ovid MEDLINE (January 1946 to June 2020), the Cochrane library as well as the following grey literature databases: the EQUATOR Network (http://www.equator-network.org/reportingguidelines/), Dissertation Abstracts, websites of evidence synthesis organisations (Campbell Collaboration Cochrane Multiple Treatments Methods Group, CADTH, NICE-DSU, Health Technology Assessment International (HTAi), Pharmaceutical Benefits Advisory Committee, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, European Network for Health Technology Assessment, Guidelines International Network, ISPOR, International Network of Agencies for Health Technology Assessment, and JBI) as well as methods collections (ie, Cochrane Methodology Register, AHRQ Effective Healthcare Programme). We will validate the MEDLINE strategy by using the PubMed IDs of 10 included studies (identified by experts prior to our eligibility screening) and evaluating whether the strategy identified the PMIDs (online supplemental appendix 2).

A systematic search strategy will be developed by two methodologists (CL and PW) without limitations to publication type, status, language or date to identify existing tools or articles. An information specialist will check the search strategy for MEDLINE Ovid and assess it using the PRESS (Peer Review Electronic Search Strategies) guidance. The full search strategies for all databases and websites can be found in Appendix 2. To identify other potentially relevant studies, we will examine the reference lists of included studies. We will ask experts in methods for NMA to identify articles missed by our search. We will contact authors of abstracts to retrieve the full report or poster.

We will search the reference section of a bibliometric study of reviews with NMA and extract the name of the journals that publish NMAs. We will then contact their editors in chief and ask if they have any in-house editorial standards for reviews with NMA.

Process for screening, data extraction and analysis
The eligibility criteria will be piloted in Microsoft Excel by two reviewers independently on a sample of 25 citations retrieved from the search to ensure consistent application. After high agreement (>70%) is achieved, the Covidence web-based tool (https://www.covidence.org) will be used by two reviewers to independently screen the citations based on the eligibility criteria. Disagreements will be discussed until consensus is reached. A third reviewer (CL) will arbitrate if disagreements cannot be resolved.

The data extraction form will be piloted by reviewers independently on a sample of five included papers to extract data on the characteristics of the studies and items. Any disagreements will be arbitrated by a third author.

Data extraction
The sources will first be categorised by the type of article coded as per our inclusion criteria. A table of tool characteristics will be developed with the following headings: first author, year; type of tool (tool, scale, checklist or domain-based tool); whether the tool is designed specific topic areas (specify); number of items; domains within the tool; whether the item relates to reporting or methodological quality (or other concepts such as precision, acceptability); how items and domains within the tool are rated; methods used to develop the tool (eg, review of items, Delphi study, expert consensus meeting) and the availability of an ‘explanation and elaboration’.

Data will be extracted on items that are potentially relevant to the risk of bias or quality of reviews with NMAs. Items will be initially extracted verbatim.

Data analysis
The following steps will be used when analysing items:

1. Map to ROBIS domains
    Items will be mapped to ROBIS domains (study eligibility criteria; identification and selection of studies; data collection and study appraisal; and synthesis and findings) and specific items within the domains. The rationale for mapping items to ROBIS is that it is the only tool to assess risk of bias in reviews. Items that do not clearly map to the existing ROBIS domains will be listed separately and grouped by similar concept. New domains may be created if items do not fit well into the established ROBIS domains.

2. Split items so that each item only covers a single concept
    Two or more concepts grouped in one item will be split so that each item covers a single concept. A rationale as to why the item was split will be described. For example, PRISMA-NMA item 15 (‘Specify any assessment of risk of bias that may affect the cumulative evidence (eg, publication bias, selective reporting within studies)’) will be split into two items because this item is represented by two items in ROBIS in the synthesis and findings domain, namely ‘4.5 Were the findings robust, for example, as demonstrated through funnel plot or sensitivity analyses?’ and ‘4.6 Were biases in primary studies minimal or addressed in the synthesis?’.

3. Group similar items
    Items that are conceptually similar will be grouped together and noted with the source. We will classify items as relating to bias or other aspect of quality (eg, reporting). When relevant, items related to reporting will be reworded into items related to bias in NMA review conclusions.

4. Omit duplicate items (but keep these in a column in the table for transparency)
    If items are worded vaguely or are unexplained, we will use an iterative process to interpret the item and ensure that there is a mutual understanding of the item between authors when coding. The process will be iterative, and
if any gaps in items related to bias in reviews of NMA are identified, a new item will be inferred.

The final list of items deemed unique will be retained. We will reword items as signalling questions, where an answer of ‘yes’ suggests the absence of bias. We will provide examples to illustrate the items and write a rationale and description of each item. These items will be submitted in a multiround Delphi exercise by NMA experts who will give their opinion about each item’s potential inclusion in the tool.

We will count the number of sources and unique items included. We will summarise the characteristics of included tools in tables and figures. We will calculate the median and IQR of items across all tools and tabulate the frequency of different biases identified in the tools.

**Patient and public involvement**

Patients or the public were not involved in the design of our research protocol.

**ETHICS AND DISSEMINATION**

No ethics approval was required as no human subjects were involved. Our research aims to develop a list of items related to bias in the goal of developing the first tool for assessing risk of bias in the findings of reviews with NMA. We plan to publish the full study open access in a peer-reviewed journal, and disseminate the findings via social media (Twitter, Facebook and author affiliated websites).

Patients, healthcare providers and policy-makers need the highest quality evidence to make decisions about which treatments should be used in healthcare practice. Being able to critically appraise the findings of reviews with NMA is central to evidenced-based decision-making in patient care.

**Author affiliations**

1Knowledge Translation Program, Li Ka Shing Knowledge Institute of St Michael’s Hospital, Toronto, Ontario, Canada
2Dalila Lana School of Public Health & Institute of Health Policy, Management, and Evaluation, University of Toronto, University of Toronto, Toronto, Ontario, Canada
3Queen’s Collaboration for Health Care Quality Joanna Briggs Institute Centre of Excellence, Queen’s University, Kingston, Ontario, Canada
4School of Education, University of Ioannina, Ioannina, Greece
5Centre for Reviews and Dissemination, University of York, York, UK
6Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
7School of Epidemiology and Public Health, Ottawa University, Ottawa, Ontario, Canada
8Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
9Anesthesiology, Pharmacology & Therapeutics, Cochrane Hypertension Review Group and the Therapeutics Initiative, University of British Columbia, Vancouver, BC, Canada
10MRC Clinical Trials Unit, University College London, London, UK
11Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

**Twitter** Carole Lunny @carole_lunny

**Contributors** CL conceived of the study; all authors contributed to the design the study; CL wrote the draft manuscript; CL, PW, ACT, BH, SD, GS, A-AV, IW and JW revised the manuscript; all authors edited the manuscript; and all authors read and approved the final manuscript.

**Funding** We have received a CIHR Spring Project Grant (ID 433402) for $360 116 for this project (https://webapps.cihr-irsc.gc.ca/decisions/p/project_details.html?appid=433402&lang=en). The funders played no role in the conduct of this project. Andrea Tricco currently holds a Tier 2 Canada Research Chair in Knowledge Synthesis. Brian Hutton has previously received honoraria from Eversana Incorporated for the provision of methodologic advice related to the conduct of systematic reviews and meta-analysis.

**Competing interests** None declared.

**Patient consent for publication** Not required.

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

**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 iDs**

Carole Lunny http://orcid.org/0000-0002-7825-6765
Andrea C Tricco http://orcid.org/0000-0002-4114-8971
Sofia Dias http://orcid.org/0000-0002-2172-0221
Brian Hutton http://orcid.org/0000-0001-5662-8647

**REFERENCES**

1 Ioannidis JPA. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ* 2009;181:488–93.
15 Pham BA, Klassen TP, Lawson ML, et al. Language of publication restrictions in systematic reviews gave different results depending on
whether the intervention was conventional or complementary. J Clin Epidemiol 2005;58:769–76.


18 Canadian Institutes of Health Research. Knowledge user engagement; 2016.


38 Veritas Health Innovation Melbourne. Covidence systematic review software; 2016.


