Knowledge user survey and Delphi process to inform development of a new risk of bias tool to assess systematic reviews with network meta-analysis (RoB NMA tool)

Carole Lunny,1,2 Areti Angeliki Veroniki,1,3 Brian Hutton,4,5 Ian White,6 JPT Higgins,7 James M Wright,2 Ji Yoon Kim,8 Sai Surabi Thirugnanasampathar,9 Shazia Siddiqui,1 Jennifer Watt,10 Lorenzo Moja,11 Nichole Taske,12 Robert C Lorenz,13 Savannah Gerrish,14 Sharon Straus,1,10 Virginia Minogue,15 Franklin Hu,16 Kevin Lin,16 Ayah Kapani,16 Samin Nagi,16 Lillian Chen,16 Mona Akbar-nejad,16 Andrea C Tricco1,3

ABSTRACT

Background  Network meta-analysis (NMA) is increasingly used in guideline development and other aspects of evidence-based decision-making. We aimed to develop a risk of bias (RoB) tool to assess NMAs (RoB NMA tool). An international steering committee recommended that the RoB NMA tool to be used in combination with the Risk of Bias in Systematic reviews (ROBIS) tool (i.e. because it was designed to assess biases only) or other similar quality appraisal tools (eg, A MeaSurement Tool to Assess systematic Reviews 2 [AMSTAR 2]) to assess quality of systematic reviews. The RoB NMA tool will assess NMA biases and limitations regarding how the analysis was planned, data were analysed and results were presented, including the way in which the evidence was assembled and interpreted.

Objectives  Conduct (a) a Delphi process to determine expert opinion on an item’s inclusion and (b) a knowledge user survey to widen its impact.

Design  Cross-sectional survey and Delphi process.

Methods  Delphi panellists were asked to rate whether items should be included. All agreed-upon item were included in a second round of the survey (defined as 70% agreement). We surveyed knowledge users’ views and preferences about the importance, utility and willingness to use the RoB NMA tool to evaluate evidence in practice and in policymaking. We included 12 closed and 10 open-ended questions, and we followed a knowledge translation plan to disseminate the survey through social media and professional networks.

Results  22 items were entered into a Delphi survey of which 28 respondents completed round 1, and 22 completed round 2. Seven items did not reach consensus in round 2. A total of 298 knowledge users participated in the survey (14% respondent rate). 75% indicated that their organisation produced NMAs, and 78% showed high interest in the tool, especially if they had received adequate training (84%). Most knowledge users and Delphi panellists preferred a tool to assess biases in an NMA, and about the concepts related to bias in NMAs to potentially include in the tool.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The development of new tools to inform evidence-based medicine requires the feedback of knowledge users and experts.

⇒ The purpose of the knowledge user survey and Delphi process was to ask respondents about the structure of a proposed tool for assessing biases in an network meta-analysis (NMA), and about the concepts related to bias in NMAs to potentially include in the tool.

WHAT THIS STUDY ADDS

⇒ The majority of knowledge users and Delphi respondents preferred a tool to assess the bias in an individual NMA’s results and in authors’ conclusions.

⇒ Delphi respondents agreed to potentially include 15 out of 22 concepts about bias in NMAs.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This Delphi and knowledge user surveys inform the development of a new risk of bias NMA tool.
more represented. One of the limitations of the Delphi process is that it depends on the purposive selection of experts and their availability, thus limiting the variability in perspectives and scientific disciplines.

Conclusions This Delphi process and knowledge user survey informs the development of the RoB NMA tool.

Introduction

Guidance about how to systematically develop quality and bias assessment tools is well established,1–4 multistaged and includes the involvement of knowledge users and experts. The benefits of engaging knowledge users in tool development is a key factor associated with knowledge translation and the reduction of research waste.3–5 Specifically, the benefits include: greater public acceptance6; identifying and prioritising topics for research10; providing feedback on the tool’s usability10; wider dissemination, uptake and communication of findings;8 and increased likelihood of impact.9,11 Identifying an external group of experts to obtain a multitude of perspectives will produce a more valid tool than a judgement given by an individual expert, or a group of experts heavily involved in the development process. Engaging with knowledge users and experts during development ensures that new tools will be relevant and applicable.

The risk of bias in network meta-analysis (RoB NMA) tool project aims to develop the first tool to assess risk of bias in a review using network meta-analyses (NMAs). We intended the RoB NMA tool to be used in combination with ROBIS12 (which we recommend as it was designed to assess biases specifically) or other similar tools (eg, AMSTAR 213) to assess quality of systematic reviews. The RoB NMA tool will assess NMA biases and limitations regarding how the analysis was planned, data were analysed and results were presented, including the way in which the evidence was assembled and interpreted. Our proposed RoB NMA tool has several uses. It can help knowledge users: (i) decide whether to believe the results from a single NMA and (ii) help choose between NMAs based on their risk of bias.

Development of the tool follows five stages. In the first stage, we conducted and published a methodological review to identify items related to bias in NMAs;14; second, the steering group made conceptual decisions about the type of tool that will be developed, and refined the items into concepts from the methodological review; third, a knowledge user survey was developed to solicit feedback on the structure of the tool from potential users and fourth, expert opinion was obtained through a Delphi process to select and define the concepts. The final stage will involve compiling the items into a tool; and conducting pilot testing to refine the items in the tool.

In this paper, we report on the knowledge user survey and Delphi process. We define ‘knowledge user’ as an individual who is likely to be able to use research results to make informed decisions about health policies, programmes and/or practices.15 A knowledge user can be, but is not limited to, a practitioner, a policy maker, an educator, a decision maker, a healthcare administrator, a community leader or an individual in a health charity, patient group, private sector organisation or media outlet.15 Our definition of experts is based on an individual’s scientific/professional expertise, in our case in methods for NMAs, bias in systematic reviews with or without NMAs and risk of bias tool development.

The purpose of the survey was to ask knowledge users about the structure of our proposed tool and about their potential use of the tool in evidence-informed practice, policymaking, guideline development or research. We also aimed to conduct a multiround Delphi process to solicit expert opinion about concepts to potentially include in the tool.

Methods

Management of the project

At the start of our project to develop a risk of bias tool for NMAs, we first convened a steering group of nine experts in NMA, bias and tool development (online supplemental appendix A).12 The steering group is responsible for the management of the project and has executive power over all decisions related to the proposed tool, which is still under development.

Protocol

We uploaded our study protocol on the Open Science Framework at https://osf.io/da4uy/. The knowledge user survey complied with the Checklist for Reporting Results of Internet E-Surveys (online supplemental appendix B).16 Important definitions are found in box 1.

A cross-sectional survey design was used for the knowledge user survey using Qualtrics.17,18 Unique site visitors were identified via IP address and personal information was collected on a voluntary basis from respondents (no incentives were offered). Knowledge users and experts were identified using a purposive sampling strategy.

Knowledge user survey

Design

An English-language survey with 15 (12 closed and 10 open-ended) questions was developed by the investigative team (online supplemental appendix C). Five authors piloted the survey and modified it iteratively to improve content validity. Respondents were allowed to skip questions they did not wish to answer. The knowledge user survey ran from June 28 to 1 August 2021.

There were two parts to the survey: (1) demographic information and information about whether the knowledge users’ organisation used or produced NMAs; (2) purpose of the RoB NMA tool, namely whether knowledge users preferred to assess the bias in the results, the authors’ conclusions of an NMA or both. Further sections asked about interest and engagement in development, piloting, dissemination and training.

Email list development

We created an email list of journal editors publishing NMAs, using one bibliometric study of NMAs.19 From this list, we extracted the journal names, and names of authors of NMAs. We also developed a list of organisations and institutions producing NMAs (online supplemental appendix D). We also included in the email list respondents from a UBC Methods Speaker Series on evidence synthesis methods (https://www.it.ubc.ca/2022/01/01/methods-speaker-series-2022/).

Dissemination

All potential survey respondents were sent an email describing the purpose of the study, requesting their participation and providing a link to the survey (online supplemental appendix E). A knowledge translation plan was followed to disseminate and advertise the survey (online supplemental appendix F). We used twitter cards (ie, advertisements with pictures) and targeted hashtags to increase awareness of the survey (see the Twitter Campaign in online supplemental appendix G). In addition, we advertised through the e-newsletters of Knowledge Translation Canada, SPOR Evidence Alliance and Therapeutics Initiative.
Network meta-analysis (NMA)
We adopted a broad definition of an NMA as a method that aims to, or intends to, synthesise simultaneously the evidence from multiple studies investigating more than two healthcare interventions of interest. We used the Cochrane Handbook definition of an NMA: ‘Any set of studies that links three or more interventions via direct comparisons forms a network of interventions. In a network of interventions there can be multiple ways to make indirect comparisons between the interventions. These are comparisons that have not been made directly within studies, and they can be estimated using mathematical combinations of the direct intervention effect estimates available’. A network is composed by at least three nodes (interventions or comparators) and these are connected (graphically depicted as lines/edges) when at least one study compares the underlying two interventions—that is the direct comparisons. Reviews that intend to compare multiple treatments with an NMA but then find that the expectations or assumptions are violated (eg, underlying assumptions of the method are not met), and hence an NMA is not possible or optimal, are also considered in our definition.

NMA risk of bias assessment
A risk of bias assessment would evaluate limitations in the way in which the NMA analysis was planned, analysed and presented. If these methods are inappropriate, the validity of the findings can be compromised. Our tool aims either/or to assess the biases in the individual results of the NMA, and the authors’ conclusions.

Bias in results of an NMA
NMA of effect estimates from primary studies can result in overestimation or underestimation of the effects of specific intervention comparisons. For example, Chaimani et al conducted a network meta-epidemiological study and found that, in the majority of the 32 networks they analysed, small studies tended to exaggerate the true effect estimate of the intervention, possibly due to small-study effects and publication bias. Our tool will focus on the results of an NMA (eg, network characteristics (including geometry, effect modifiers)). This is the approach taken in tools such as the Cochrane Risk of Bias 2 tool for assessing risk of bias in randomised trials.

Bias in the conclusions of an NMA
Bias may be introduced when interpreting the NMA results to draw conclusions. Conclusions may include ‘spin’ (eg, biased misrepresentation of the evidence, perhaps to facilitate publication) or (erroneous) misinterpretation of the evidence. Ideally, potential biases identified in the results of the NMA might be addressed appropriately when drawing conclusions. Similarly, a well-conducted systematic review draws conclusions that are appropriate to the included evidence and can therefore be free of bias even when

Data analysis
Questionnaires that were terminated early, where respondents did not go through all questionnaire pages, were included in analyses, but those that were entirely blank were excluded. We measured the time respondents took to fill in a questionnaire regardless of whether it was complete.

Descriptive statistics were calculated for each closed response question including count, frequency, with denominators taken as the number who provided a response to the question. One researcher coded the open-ended questions independently by identifying themes. Respondents’ comments on questions 9 and 10 were merged as they were similar in nature.

Delphi process
Design
Our published methodological review to identify items related to bias in NMAs identified 22 items related to bias in NMAs. These items were reworded into concepts by the steering committee because past Delphi groups focused on the wording of the item, and not its main idea. The concepts and questions about the structure of the tool (domains, signalling questions, rating scales) were entered into a Qualtrics survey platform.

Delphi panellists were asked to rate concepts based on a 5-point Likert scale of importance from 1 (not important—should be dropped as a concept to consider) to 5 (very important—must be included) or unable to score. If respondents did not provide a rating, the concept was recorded as missing. Respondents were asked to comment on whether they preferred to modify or reword the concepts. Free-text comment boxes allowed experts to provide additional comments. Non-responders or those failing to complete each round were sent three email reminders.

The respondents completed two survey rounds to reach a high level of agreement, defined as at least 70% scored 4 or above on the 5-point Likert scale (table 1). After round 1, we generated reports of group versus individual responses. Respondents were also provided with anonymised free-text comments from the last round.

Data analysis
We reported the number of respondents completing each round. An overall response rate was calculated as well as summary statistics for each concept. The qualitative data from the free-text questions was analysed through thematic analysis by one author (CL) and read by one or more of the other coauthors. The steering committee used their executive power to decide whether concepts excluded by the Delphi panel should be retained or not.

Results
Knowledge user survey

Recruitment results
A total of 2821 emails were sent out to advertise the survey, 87 failed to reach the recipients due to incorrect addresses, resulting in 2734 emails that reached the intended individual (online supplemental appendix H). Most respondents completed the survey through our Qualtrics email survey link (n=390, response
rate=14%) and 27 completed the survey through an anonymous link distributed over social media and e-newsletters (n=27).

After consolidating duplicates (using IP addresses, n=33) and blank responses (n=86), a total of 298 responses were included in the analysis. Of the 298 respondents, 252 (85%) answered all the survey questions and 46 (15%) completed half of the questions. The mean time to complete the survey was 2.27 min (SD 1.33).

Characteristics of respondents
Of the 298 respondents, 136 (45.6%) self-identified as a systematic review expert, 122 (40.9%) as a guideline developer, 98 (32.9%) as a healthcare professional (table 2). Half of the respondents had primary affiliations at a university (50.0%). Most respondents resided in Europe (53.6%) or North America (39.2%) (table 2). Three-quarters (75.1%) of respondents indicated that their organisation produced systematic reviews with NMAs, but only 54.2% of knowledge users said they used an NMA in their work (table 2).

Interest and type of tool preferred
Most knowledge users (84%) reported they would use the RoB NMA tool if they received adequate training on how to use it (figure 1). When asked about their level of interest in our tool, 182/298 (61.1%) had high interest, 53/298 (17.8%) had low interest and only one person had no interest. Many respondents said they would use the RoB NMA tool’s bias assessment when conducting an overview of reviews, health technology assessment (HTA) or guideline; and to distinguish between NMAs at higher or lower risk of bias.

When we asked knowledge users about the type of tool that might be useful to them or their organisation, half of the respondents (145/298) reported they preferred a tool to assess both the bias in individual NMA results and authors’ conclusions (figure 2).

Open-ended questions are summarised in online supplemental appendix 1 tables 1–4. We also report in online supplemental appendix 1 table 4 respondents’ interest in dissemination and engagement activities. The majority of respondents (153/231; 66%) said they would want to read the final study reports, receive updates (147/231; 64%) and receive training in using the new tool (140/231; 61%).

Delphi survey
Recruitment results
The steering committee invited 53 experts to participate in the Delphi surveys, and 19 emails failed for various reasons, resulting in 37 emails that reached the intended individual. Of these, 28 completed round 1 and 22 completed round 2 (flow chart in online supplemental appendix J). The response rate of panelists participating in our study was 28/37 (75.7%) in round 1 and 22/28 (78.6%) in round 2.

Characteristics of round 1 respondents
Of the 28 round 1 respondents, 15 (53.7%) self-identified as statisticians, 10 (35.7%) as academics, 4 (14.3%) as systematic review specialists or scientists, epidemiologists or graduate students/post-doctoral researchers (table 3). More than half of the respondents had a primary affiliation at a university (68%). Most respondents resided in Europe (53.6%) or North America (39.2%) (table 3). Most (96.4%) respondents indicated that their organisation produced systematic reviews with NMAs.

Rating of concepts
Of the 22 concepts, 7 did not reach consensus in round 2 (indicated in red in table 4). Table 4 lists all concepts in the left column that respondents rated from strongly disagree to strongly agree. The second column indicates whether the concept was included based on 70% of agreement (agree and strongly agree combined). The next columns indicate the number of respondents over the denominator (number of people who answered) for each rating, percentage responses and the group median. The list of concepts in table 4 is not intended to be used to assess biases in NMAs, but to inform the development of items to be included in our tool.

Structure of the RoB NMA tool
When asked about the structure of the RoB NMA tool, the majority of respondents agreed that a domain-based structure (25/28; 89.3%) with signalling questions (20/28; 71.4%) was preferred. The domain-based structure would be similar to that used in Cochrane Risk of Bias tool and the ROBIS tool. Signalling questions flag aspects of study design related to the potential for bias and aim to help reviewers judge risk of bias. They also agreed (19/28; 67.9%) that the steering committee should provide guidance on how to produce a risk of bias assessment for NMAs, outcomes within a network or authors’ conclusions of NMAs.

When asked about their preference for a tool to assess the risk of bias in NMA results and/or the authors’ conclusions, the majority of respondents (15/28; 53.6%) preferred a tool to assess bias in both results and conclusions, one-third (10/28; 35.7%) preferred a tool to assess bias in the results only and a minority (3/28; 10.7%) preferred to assess only the NMA authors’ conclusions.

Discussion
A majority of knowledge users responded that they had high interest in the RoB NMA tool if they received adequate training on how to use it and said they would use the tool to distinguish between NMAs at higher or lower risk of bias, and to assess an NMA in an overview of reviews, HTA or guideline. Delphi respondents articulated a clear preference for a tool that is domain-based with signalling questions which would be used to assess biases in the results and the authors’ conclusions. Seven out of 22 concepts did not reach consensus by the Delphi group, and these concepts...
If you have used one or more systematic reviews with NMA in your work, did you use:

- Both individual results of NMAs and conclusions: 127 (57.7%)
- Individual analysis results from the NMA to draw your own conclusions (eg, pooled effect estimate): 72 (32.7%)
- NMA authors’ conclusions: 21 (9.55%)
- Missing: 78 (26.2%)

*Percentages add to >100% because respondents could provide more than one response.

HTA, health technology assessment; NGO, non-governmental organisation; NMA, network meta-analysis.

and accompanying comments will be reviewed and considered by the steering committee for eligibility in the tool. The tool is still under development and the list of concepts is not intended to be used to assess biases in NMAs.

Respondents also indicated the need for guidance on how to use the tool to assess biases in the NMA. These results highlight the necessity for clear and easy to understand elaboration and explanation materials plus training, and perhaps the development of more structured guidelines for reaching domain-based risk of bias judgements (eg, algorithms).26 Many knowledge users erroneously thought the RoB NMA tool’s final assessment would be used in an evaluation of the certainty of the evidence (eg, CINeMA [Confidence in Network Meta-Analysis]26) or Grading of Recommendations, Assessment, Development and Evaluations27); even though we clearly stated that our proposed tool is intended for the assessment of the potential biases in an NMA. Only the quantitative results of an NMA (ie, the analysis) are used in a certainty of the evidence evaluation.

We aimed to engage knowledge users and NMA experts early in the tool development for multiple reasons. Engagement would ensure that our tool will be relevant, useable and accepted.5 The Delphi expert responses provided us with feedback on which concepts would be most relevant, and the knowledge user responses emphasised the desire for future training10 and desire to help us with dissemination and communication of findings.9 10

**Implications of this study**

Knowledge users5 7 26–31 and Delphi5 32–35 surveys have been successfully used to inform the development of other types of tools, systematic reviews and guidelines. Online surveys have...
A strength of our research was that we developed a protocol which we followed in the conduct of the research (https://osf.io/da4uy/). We aimed to engage knowledge users and NMA experts early in the tool development for multiple reasons. Engagement would ensure that our tool will be relevant, useable and accepted.9 The responses provided us with feedback on which concepts are most relevant, training needs36 and future dissemination of findings. We combined newsletter, email distribution lists and social media to reach a wide range of knowledge users from across the globe. We attempted to maximise the response rate by sending email reminders and repeating messages through social media. Response bias in our sample is a major limitation as knowledge users working in high-income countries were more represented, than one response.6

**Strengths and limitations**

A strength of our research was that we developed a tool to evaluate the reliability and face validity of tools, and the use of the Cochrane risk of bias tool in practice.25 However, we are not aware of similar published surveys conducted prior to the development of a risk of bias or quality appraisal tool, targeted specifically at knowledge users.

The responses provided us with feedback on which concepts are most relevant, training needs36 and future dissemination of findings. We combined newsletter, email distribution lists and social media to reach a wide range of knowledge users from across the globe. We attempted to maximise the response rate by sending email reminders and repeating messages through social media. Response bias in our sample is a major limitation as knowledge users working in high-income countries were more represented, and respondents (i.e., systematic reviewers [45.6%] or academics [40.9%]) may have been more likely to have responded to a survey about a new tool to assess the bias in NMAs.

A limitation is we did not ask knowledge users to define what their role was and whether they considered themselves: (i) decision makers; (ii) purchasers of services/pharma products; (iii) professional service providers; (iv) evidence generators or (v) advocates of health promotion. Another limitation is that our targeted emails and social media advertisement may have missed important knowledge users that use NMAs (e.g., members of the Canadian Institutes of Health Research, Drug Safety and Effectiveness Network, Methods and Applications Group for Indirect Comparisons Group).

A strength of the Delphi process was that by performing the surveys online, experts from around the world were able to participate. The response rate of Delphi panellists participating to our study was high in both rounds. All comments were thoroughly read by one of the authors (CL), and part of them were read by one or more of the other coauthors. In addition to a full feedback report, a summary of the comments in round 1 was also given in the next round to panellists. Delphi processes have many limitations, one of which is that they depend too much on the purposive selection of ‘experts’ and their availability, thus raising the question of whether all relevant perspectives and scientific disciplines have been taken into consideration. Our Delphi panel consisted of a small double digit number, which may risk collecting certain thought collectives and may be an issue of reliability.36

**Future research**

The results of the survey will inform a new tool to assess biases in NMAs. Our tool is not targeted at authors of NMAs, as it does not
<table>
<thead>
<tr>
<th>Domain</th>
<th>Concepts related to risk of bias in network meta-analysis</th>
<th>Round</th>
<th>Included (70% consensus)</th>
<th>Responses</th>
<th>Group median (Q1, Q3)</th>
<th>Unable to Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network characteristics/geometry</strong></td>
<td>1: Whether all interventions in the network (including comparators) were potentially suitable for all eligible study respondents</td>
<td>1 Y</td>
<td>0/27</td>
<td>0/27 (7%)</td>
<td>15/27 (56%)</td>
<td>1 (4, 5)</td>
</tr>
<tr>
<td></td>
<td>2: Whether any interventions were inappropriately excluded from the network (e.g., through eligibility criteria or after seeing the results)</td>
<td>1 Y</td>
<td>0/27</td>
<td>0/27 (7%)</td>
<td>12/27 (44%)</td>
<td>1 (4, 5)</td>
</tr>
<tr>
<td></td>
<td>3: Whether importantly different intervention strategies were kept as distinct nodes in the network (i.e., whether appropriate groupings were made of interventions—lumping vs splitting)</td>
<td>1 Y</td>
<td>0/28</td>
<td>0/28 (7%)</td>
<td>10/28 (37%)</td>
<td>1 (4, 5)</td>
</tr>
<tr>
<td><strong>Effect modifiers</strong></td>
<td>4: Whether effect-modifying participant characteristics are sufficiently similar across the whole network</td>
<td>1 Y</td>
<td>0/28</td>
<td>0/28 (7%)</td>
<td>15/28 (53%)</td>
<td>0 (4, 5)</td>
</tr>
<tr>
<td></td>
<td>5: Whether outcomes and time points are sufficiently similar across the whole network</td>
<td>1 Y</td>
<td>0/28</td>
<td>0/28 (7%)</td>
<td>10/28 (36%)</td>
<td>0 (4, 5)</td>
</tr>
<tr>
<td></td>
<td>6: Whether study-level risks of bias are sufficiently similar across the whole network</td>
<td>1 N</td>
<td>0/28</td>
<td>2/28 (7%)</td>
<td>15/28 (53%)</td>
<td>0 (4, 5)</td>
</tr>
<tr>
<td><strong>Statistical synthesis</strong></td>
<td>7: Whether other trial characteristics are sufficiently similar across the whole network</td>
<td>1 N</td>
<td>0/28</td>
<td>2/28 (7%)</td>
<td>10/28 (36%)</td>
<td>0 (4, 5)</td>
</tr>
<tr>
<td></td>
<td>8: Whether an appropriate prespecified approach was used in node making</td>
<td>1 Y</td>
<td>1/25 (6%)</td>
<td>0/25 (0%)</td>
<td>3/25 (20%)</td>
<td>4 (3, 4)</td>
</tr>
<tr>
<td></td>
<td>9: Whether a process was used to define nodes in the network (e.g., undertaken independently by two reviewers, following a preplanned node-making process)</td>
<td>1 N</td>
<td>0/25</td>
<td>2/25 (8%)</td>
<td>3/25 (12%)</td>
<td>3 (3, 4)</td>
</tr>
<tr>
<td></td>
<td>10: Whether effect metric(s) for each outcome (e.g., ORs, risk difference) in the network were presented with CIs/credible intervals</td>
<td>1 N</td>
<td>3/28 (11%)</td>
<td>4/28 (14%)</td>
<td>13/28 (46%)</td>
<td>0 (2.25, 5)</td>
</tr>
<tr>
<td></td>
<td>11: If disconnected networks were connected to perform the analysis, whether methods to do this were appropriate</td>
<td>1 Y</td>
<td>0/25</td>
<td>2/25 (8%)</td>
<td>6/25 (24%)</td>
<td>3 (3, 4)</td>
</tr>
<tr>
<td></td>
<td>12: Whether methods used to represent multi-arm studies in the dataset and in the analysis are appropriate</td>
<td>1 Y</td>
<td>0/28</td>
<td>4/28 (14%)</td>
<td>10/28 (36%)</td>
<td>4 (3, 25, 5)</td>
</tr>
<tr>
<td></td>
<td>13: Whether assumptions across the network about homogeneity/heterogeneity of effects within comparisons are appropriate</td>
<td>1 Y</td>
<td>0/27</td>
<td>1/27 (4%)</td>
<td>12/27 (44%)</td>
<td>4 (3, 25)</td>
</tr>
<tr>
<td></td>
<td>14: Whether a valid approach was used to determine whether there was conflict between direct and indirect sources of evidence on the same comparisons (often called inconsistency or incoherence)</td>
<td>1 Y</td>
<td>0/28</td>
<td>2/28 (7%)</td>
<td>18/28 (64%)</td>
<td>0 (4, 5)</td>
</tr>
<tr>
<td></td>
<td>15: If inconsistency detected, then whether methods such as re-evaluation of the choice of scale, effect modification and similarity of the contributing randomised controlled trials were investigated</td>
<td>1 N</td>
<td>0/28</td>
<td>4/28 (14%)</td>
<td>5/28 (18%)</td>
<td>0 (4, 3)</td>
</tr>
<tr>
<td></td>
<td>16: If a Bayesian analysis was conducted, whether the selection of prior distributions was justified</td>
<td>2 N</td>
<td>0/22</td>
<td>4/22 (18%)</td>
<td>5/22 (23%)</td>
<td>0 (4, 3)</td>
</tr>
<tr>
<td></td>
<td>17: Whether the analysis appropriately addressed any differences in effect modifiers across different parts of the network</td>
<td>1 Y</td>
<td>0/28</td>
<td>2/28 (7%)</td>
<td>5/28 (18%)</td>
<td>0 (4, 4)</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Domain</th>
<th>Concepts related to risk of bias in network meta-analysis</th>
<th>Responses</th>
<th>Group median (Q1, Q3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Included (70% consensus)</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>18: Whether there was evidence of conflicting results between direct and indirect evidence (often called inconsistency and incoherence in results)</td>
<td>Y</td>
<td>0/28</td>
<td>3/28 (21%)</td>
</tr>
<tr>
<td>19: If there were conflicting results between direct and indirect evidence was this addressed appropriately (e.g., meta-regression, data extraction errors, redefining the network)</td>
<td>Y</td>
<td>0/27</td>
<td>2/27 (74%)</td>
</tr>
<tr>
<td>20: Evidence that the statistical model, as it was used to get the key results, was not suitable for the data (e.g., from analysis of residuals or information criteria such as DIC)</td>
<td>N</td>
<td>0/28</td>
<td>7/28 (25%)</td>
</tr>
<tr>
<td>21: Whether sensitivity analyses demonstrate that findings were robust to the statistical model and estimation methods (including prior distributions if Bayesian methods were used)</td>
<td>N</td>
<td>0/14</td>
<td>1/14 (5%)</td>
</tr>
<tr>
<td>22: Whether limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias) were discussed</td>
<td>Y</td>
<td>0/28</td>
<td>2/28 (10%)</td>
</tr>
</tbody>
</table>

*Blank responses and 'unable to score' responses were not counted in the denominator.

DIC, deviance information criterion; N, no; Y, yes.

**Conclusions**

The surveys provided feedback from knowledge users and experts on their preferences for the structure, focus and concepts included in a proposed RoB NMA tool, which is under development. Both knowledge users and Delphi panelists preferred a tool to assess bias in individual NMA results and authors' conclusions. Both the bias in individual NMA results and authors' conclusions are relevant to the decision-making process of healthcare providers and researchers. Further research will involve reliability testing with different knowledge user groups and will also involve training providers and researchers in the use of the tool.

The steering committee will use the results of the knowledge user survey to determine preferences around the tools structure and design. This will be achieved through a series of workshops and qualitative interviews with knowledge users and Delphi panelists. The tool will be piloted with different knowledge user groups, and feedback will be collected to refine the tool. The final tool will be published and available for download.
Papakonstantinou, Isabelle Bourton, Sharon Strauss and Jenn Watt.

Contributors CL conceived of the study; ACT, AAV, BH, CL, IW, JPTH, and JMW contributed to the design of the study; CL drafted the survey; CL, LC and SST inputted the questions into Qualtrics; CL, SSI and SST analysed the data; CL wrote the draft manuscript; ACT, AAV, BH, CL, IW, JPTH, and JMW revised the manuscript; all authors edited the manuscript; and all authors read and approved the final manuscript. ACT is the guarantor.

Funding Funding was received from a 2020 CIHR Project Grant (2021–2024; # ID 451190). ACT currently holds a Tier 2 Canada Research Chair in Knowledge Synthesis. IW was supported by the Medical Research Council (programme MC_UU_00004/06). BH has previously received honoraria from Eversana for the provision of methodological advice related to the conduct of systematic reviews and meta-analysis. JPTH is a National Institute for Health Research (NIHR) Senior Investigator (NF-SI-0617-10145), is supported by NIHR Bristol Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, is supported by the NIHR Applied Research Collaboration West (ARC West) at University Hospitals Bristol and Weston NHS Foundation Trust, and is a member of the MRC Integrative Epidemiology Unit at the University of Bristol.

Competing interests AAV was an Associate Editor for the journal, but was not involved with the decision or peer-review process.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Approval from the University British Columbia Ethics Board was obtained for both the survey and the Delphi process, and consent was implied when respondents completed the online survey.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. The datasets used and/or analysed during the current study are available freely at https://osf.io/da4uy/.

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
Brian Hutton http://orcid.org/0000-0001-5662-8647
Jennifer Watt http://orcid.org/0000-0002-5296-6013
Robert G Lorenz http://orcid.org/0000-0003-3286-6792

References
15. CIHR. Knowledge user engagement. Canadian Institute for Health Research, 2016.


