Journal of Clinical Epidemiology 115 (2019) 1−13

ORIGINAL ARTICLE

Development of the summary of findings table for network meta-analysis

Juan José Yepes-Nuñez[a,b], Shelly-Anne Li[c], Gordon Guyatt[a,d], Susan M. Jack[a,e], Jan L. Brozek[a,d], Joseph Beyene[a], M. Hassan Murad[f], Bram Rochwerg[a,d], Lawrence Mbuagbaw[a], Yuan Zhang[a], Ivan D. Flórez[a,g], Reed A. Siemieniuk[a], Behnam Sadeghirad[a], Reem Mustafa[a,h], Nancy Santesso[a], Holger J. Schünemann[a,d,*]

[a]*Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada*
[b]*School of Medicine, Universidad de los Andes, Carrera 7 No. 116 − 05, Bogotá, D.C., Colombia*
[c]*University of Toronto, 155 College Street, Toronto, Ontario, Canada*
[d]*Department of Medicine, McMaster University, Hamilton, Canada*
[e]*School of Nursing, McMaster University, Hamilton, Ontario, Canada*
[f]*Mayo Clinic, Evidence-Based Practice Center, 200 1st Street SW, Rochester, MN 55905, USA*
[g]*Department of Pediatrics, University of Antioquia, Calle 70 No. 52 − 21, Medellín, Colombia*
[h]*Division of Nephrology and Hypertension, University of Kansas Medical Center, Kansas City, KS, USA*

Accepted 24 April 2019; Published online 2 May 2019

Abstract

Objectives: The aim of the study was to develop a Grading of Recommendations, Assessment, Development and Evaluation (GRADE) summary of findings (SoF) table format that displays the critical information from a network meta-analysis (NMA).

Study Design and Setting: We applied a user experience model for data analysis based on four rounds of semistructured interviews.

Results: We interviewed 32 stakeholders who conduct or use MA. Four rounds of interviews produced six candidate NMA-SoF tables. Users found a final NMA-SoF table that included the following components highly acceptable: (1) details of the clinical question (PICO), (2) a plot depicting network geometry, (3) relative and absolute effect estimates, (4) certainty of evidence, (5) ranking of treatments, and (6) interpretation of findings.

Conclusion: Using stakeholder feedback, we developed a new GRADE NMA-SoF table that includes the relevant components that facilitate understanding NMA findings and health decision-making.

Keywords: Network meta-analysis; Decision-making; Systematic reviews; Summary of findings; GRADE; Certainty of evidence

1. Introduction

Multiple treatments are often available for patients with the same condition. Credible systematic reviews constitute the most trustworthy source of evidence to determine the effectiveness of interventions. Most systematic reviews focus on pair-wise direct comparisons of interventions, often with placebo as the comparator group [1]. In the face of multiple available interventions, a lack of direct comparisons between active interventions can make determination of their relative desirability challenging [1].

When investigators have studied multiple interventions for the same condition, network meta-analysis (NMA) can provide information regarding how each option compares to the alternatives. Using direct and indirect evidence when deemed appropriate, NMA offers estimates of the relative impact of all the interventions even when head-to-head comparisons are unavailable [2].

Conflict of interest: The authors of this study declare no financial conflict of interest. They are members of the GRADE working group.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. It was internally supported by the McMaster GRADE center.

* Corresponding author. Tel.: +1 905 525 9140; fax: +1 905 522 9507.
E-mail address: schuneh@mcmaster.ca (H.J. Schünemann).

https://doi.org/10.1016/j.jclinepi.2019.04.018
0895-4356/

What is new?

Key findings

- User testing methodology provided key information to develop a Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) summary of findings (SoF) table that displays network meta-analysis (NMA) findings.
- Fundamental components of the new NMA-SoF table include details of the question and interventions for a specific outcome, the relative effect estimate for each intervention, the anticipated absolute effects, the GRADE certainty of evidence, the rank probability of the intervention, and the interpretation of findings.

What this adds to what was known?

- NMA publications are increasingly popular. However, the optimal presentation of their synthesized findings remains uncertain. The new NMA-SoF table provides relevant information in a simple and user-friendly format.

What are the implication and what should change now?

- NMA authors should consider presenting results in NMA-SoF tables based on the structure that our users have found informative and appealing.

Summaries of evidence for health professionals exist in different formats and for different purposes. The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) Summary of Findings (SoF) tables present the main findings of conventional pair-wise systematic reviews in a transparent and simple form [3,4]. They provide key information concerning the certainty of the evidence and the magnitude of treatment effects for all patient-important outcomes [3−7].

Expert guidance [2,8−14] suggests that the presentation of NMA findings in an NMA-SoF table should report (1) the relative and absolute effects estimates of the interventions included in the NMA; (2) the certainty of the evidence (also known as confidence in effect estimates or quality of the evidence); (3) rank probabilities; and (4) geometry of the NMA. NMA authors do not always adhere to this guidance.

Irrespective of adherence to standards, the optimal presentation of NMA remains uncertain and challenging, particularly when the alternatives are numerous [13]. Although SoF have proved extremely helpful for summarizing evidence from conventional systematic reviews, creating a similarly useful SoF for NMAs presents many challenges, and no one has thus far suggested a structure. Therefore, in response to the challenge of optimally NMA findings presentation, we developed and tested a novel format for GRADE NMA-SoF tables.

2. Methods

The principles of fundamental qualitative description guided this study for data collection and research inquiry. This approach allowed for comprehensive summaries of the users' perspectives about the NMA-SoF tables, and it may be useful for refining existing programs and projects [15].

We conducted this study in three phases: (1) an initial development of two draft NMA-SoF tables with feedback from GRADE working group members; (2) iterative modification of NMA-SoF table formats through four brainstorming meetings with researchers; and (3) user testing. The brainstorming meetings and the user testing phases were carried out alternately (Fig. 1). A steering committee (J.Y., H.J.S., J.L.B., J.B., and G.G.) considered feedback from brainstorming meetings and user testing and made decisions regarding response to feedback and modification of NMA-SoF table formats. Appendix A includes the protocol for the user testing.

2.1. Development of initial NMA-SoF table format

Initially, a steering committee member (H.J.S.) developed two draft versions of GRADE NMA-SoF tables (see Supplement Appendix B) that he presented during a GRADE working group meeting in 2012, modified on the basis of feedback, and presented a second time at a workshop in 2013.

2.2. Brainstorming meetings

From 2013 to 2016, we conducted four brainstorming meetings with 42 health care researchers with experience in conducting meta-analysis and NMAs. We documented the information retrieved from each meeting. At the first two brainstorming meetings, participants viewed the two initial draft NMA-SoF tables and provided their opinions on how they could be improved. At the third and fourth sessions, the participants viewed revised versions that the steering committee had developed on the basis of feedback from prior user testing rounds.

2.3. User testing

User testing (or usability testing) is a method of formative evaluation "with the goal of learning about the design from a user's perspective to improve its next iteration" in which products are tested by end users, as opposed to developers or experts [11,16]. The primary focus of our user testing was to explore the extent to which the NMA-SoF

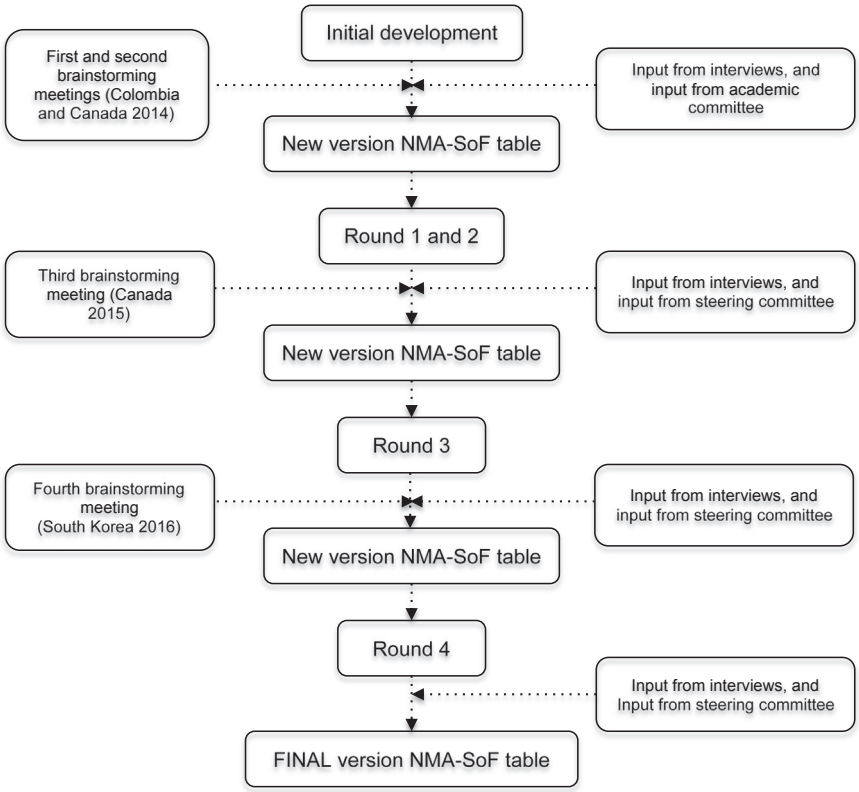


Fig. 1. Phases conducted to develop the GRADE NMA-SoF table.

table formats meet the needs of the users and optimizes their understanding of the information reported.

We conducted four rounds of user testing. The first two rounds were developed after the first two brainstorming meetings. The third and fourth rounds were conducted after the third and fourth brainstorming meetings. Our final goal was to develop one NMA-SoF table for dichotomous outcomes and another table for continuous outcomes. We used an NMA-SoF table that displayed findings for dichotomous outcomes during the brainstorming meetings and user testing. We finally developed an NMA-SoF table for continuous outcomes based on the information retrieved from the brainstorm meetings and user testing.

2.3.1. Rounds

We defined "rounds" as a group of 10 interviews using the same NMA-SoF table format with different users. After each round, the steering committee modified the NMA-SoF table format based on the comments and reflections provided by the users. The new version of the NMA-SoF table was presented to a new group of users during each new round.

2.3.2. Participants

2.3.2.1. Selection of participants. We used purposeful sampling to select participants who could help us answer our research questions during the brainstorming meetings and user testing [16]. We applied criterion sampling [17,18] to ensure that the final pool of participants represented a wide range of NMA-SoF tables users.

Individuals were eligible if they considered themselves meta-analysis (MA) or NMA users. Three target populations participated in the user testing: (1) systematic review authors, (2) methodologists, and (3) clinicians or healthcare-related professionals as defined by the World Health Organization [16]. We considered *systematic review authors* as participants if they published a systematic review in the last 2 years and read one NMA report in the year preceding participation. We defined *methodologists* as participants if they declared dedicating more than 70% of their time to conducting MA or NMA research. We classified *clinicians* as participants if they reported at least 50% of their total time dedicated to clinical practice.

2.3.2.2. Setting and recruitment. We recruited users who met our definition through international networks of the study team and from the GRADE working group. Potential users received an e-mail invitation. The project coordinator (J.Y.) carried out recorded interviews face-to-face or by video chat and voice calls (e.g., Skype) with those who provided informed consent. None of the users participated in more than one round.

2.3.2.3. Sample size. We estimated our sample size to be 20 participants, as there is evidence that this number should

identify at least 95% of problems of the prototype in usability tests, and other researchers proposed any number of participants ranging between 3 and 20 participants [19]. We anticipated that approximately half would be healthcare professionals, and half researchers and both groups would be represented during each round of interviews. We planned to conduct interviews until we gathered sampling to redundancy [20]. We extended the sample size information in Appendix A.

2.3.2.4. Interviews. We used a pilot test to first explore the impressions and understanding of the NMA-SoF table as a whole, and then of each NMA-SoF table elements using a user experience model (Honeycomb model) developed by Morville [21]. The Honeycomb model addresses seven separate facets of user experience: findability, accessibility, usability, usefulness, credibility, desirability, and value [21]. All semistructured interviews took approximately 1 hour and included three sections: (1) background information of the users; (2) users' reactions to the NMA-SoF table including their first impressions, their needs, and understanding of the information; and (3) an overall evaluation. The interviews ended with questions about how the interview process could be improved.

2.3.2.5. Data collection and transcription. The project coordinator made notes that summarized the discussion at the brainstorm meetings. In the user testing interviews, users viewed either a paper or electronic version of an NMA-SoF table. An individual with no other role in the study transcribed all audio-recorded data verbatim and, after transcription, deleted the recording.

2.3.2.6. Data Analysis. Data collection and analysis occurred concurrently so that questions from the interview guide could be modified to allow for better capture of themes. The analysis was done using MAXQDA 12 software (VERBI GmbH Software 2015, Berlin, Germany) [22].

We used both deductive and inductive approaches to analyze the data in duplicate. First, we used Hsieh and Shannon's [23] directed content analysis (deductive approach) to determine the initial coding scheme based on Morville's Honeycomb model [21], a model used successfully in previous research evaluating SoF tables for conventional pairwise meta-analysis [24,25]. The seven facets of the Honeycomb model served as the initial coding categories. We also included an understandability code, as it had been tested in other pairwise meta-analysis SoF studies [6,26]. Next, we developed operational definitions for each category based on the model. Coding began with reading and then highlighting all segments of data related to the categories. Next, we coded all the highlighted passages using the list of predetermined codes. Using inductive content analysis, we searched for new information that was not captured using the initial coding scheme. Any content that did not fit into any of the categories was assigned a new code according to Elo et al. [27]. We used inductive content analysis because no other studies have investigated health professionals' experience with NMA-SoF tables [27]. For each interview transcription, two researchers (J.Y. and S.A.L.) independently organized the information by open coding, creating categories, and abstraction.

2.4. Ethical considerations

The Hamilton Integrated Research Ethics Board approved the research conducted in this project. All users provided both oral and written consent before the interview (review reference 2016-0956-GRA).

3. Results

We conducted four brainstorming meetings and four rounds of interviews.

3.1. Initial NMA-SoF table formats

GRADE working group members and 11 clinicians and methodologists first viewed and commented on the initial two versions of the NMA-SoF. They reported concerns including distracting box colors, failure to clearly highlight the certainty of evidence, omission of ranking treatment information and network geometry, failure to clearly identify the reference comparator intervention, and deficiencies in reporting the absolute effects and their interpretation in the NMA context. GRADE working group members also reported missing information regarding the direct and indirect estimates. At the end of the meeting, some users suggested seeing all the interventions compared with a single reference intervention in only one two-by-two table (or league table). Based on this feedback, we developed two modified NMA-SoF table formats (see Supplement Appendix C and D), which we discussed at the brainstorming meetings.

3.2. Brainstorming meetings

We conducted the brainstorming meetings in three different countries: one in Colombia (Medellín, June 2014), two in Canada (Hamilton, July 2014 to July 2015), and a final one in South Korea (Seoul, October 2016). In total, 42 users provided information on five NMA-SoF table formats. Two of the brainstorm meetings in which participants commented on four NMA-SoF table formats (Colombian June 2014 and Canadian July 2014) took place before the first of the user testing. In the Colombian brainstorming meeting, we explored format limitations in the two initial NMA-SoF tables (see Supplement Appendix

B). In the Canadian brainstorming meeting in 2014, we received feedback using two additional NMA-SoF table formats (see Supplement Appendix C and D). For the third and fourth brainstorming meetings, we obtained feedback on a single NMA-SoF table format (see Supplement Appendix E). We used this format during the first three rounds of interviews.

The steering committee decided which elements should be included and which elements were less important to include. We restructured the NMA-SoF table to display the NMA relevant information for one outcome, including relative and absolute effect estimates, certainty of the evidence, and ranking probabilities for all the comparisons using a single comparator with the interventions displayed in rows.

3.3. User testing

We interviewed 32 users: 21 methodologist, 5 MA users, and six clinicians. We report the characteristics of these participants in Table 1.

3.3.1. User testing: first impressions

Users reported that the NMA-SoF table format (see Supplement Appendix E) was similar to other GRADE SoF tables. Users found the relative and the absolute effects, the certainty of evidence, and the interpretation of findings the most engaging sections.

Users found inclusion of the network geometry and different colors for each intervention appealing because they easily identified each intervention in the network geometry and the table. They found that displaying this information by outcome was clear and attractive considering a high number of interventions. The users found that presenting a single outcome per NMA-SoF table more readable than alternatives that included more than one outcome per NMA-SoF table. Users noted the drawback, however, in a clinical decision context, of having to refer to several NMA-SoF tables rather than a single one, one for each outcome.

3.3.2. User testing: usability

Users, in general, found it easy to use the NMA-SoF table (see Supplement Appendix E) when we presented the information with the clinical context. Presenting the information of NMA relative and absolute effects without the direct and indirect estimates was attractive for most users:

> "I don't know. It is not possible to see all of the indirect or direct NMA estimate I believe, right?"

The steering committee therefore attempted to add the direct and indirect effect estimates and their confidence intervals to the NMA-SoF table, but including this additional data added further complexity. The steering committee ultimately decided only to present the highest certainty of the available effect estimates, whether it was the network estimate, the direct estimate, or the indirect estimate.

3.3.3. Accessibility

Users found the information in the NMA-SoF table (PICO information, network geometry, NMA relative effects, and so on) easily accessible, and they did not report any substantial difficulties accessing to the information:

> "Well, I think it's helpful you have the question at the very top and kind of deliberations on the actual PICO components. And at the top I think that's helpful."

For NMA-SoF tables with fewer than six comparisons, a single page was sufficient, and users worked on the table without any accessibility issues. However, in the examples with more than six interventions, the table did not fit onto a single page, and users expended more time understanding the information in the NMA-SoF tables.

Some users expressed concerns about the chosen colors and suggested that we should consider the needs of users with color visual impairment. In response, the final NMA-SoF table format will not include red and green to avoid issues with the most common color blindness.

3.3.4. Usefulness

In a clinical decision scenario, users reported that a single NMA-SoF table was useful:

> "I think it would be useful. The key challenge is always to match the patient, [that] the patient matches what is in there. I think the rank would be helpful to decide [on the option]."

Users found all the components of the NMA-SoF table helpful, but some aspects, including the network geometry and the certainty of evidence assessment, were particularly helpful.

3.3.5. Credibility

In general, users found the information reported in the NMA-SoF table trustworthy because they were able to identify information about the NMA relative effect, the absolute effect, and the certainty of evidence in a same NMA-SoF table format:

> "*…So that does, why I would trust this information because I see the anticipated absolute effect. That's all relative effects and the certainty in the evidence. That's all clearly shown here.*"

Users made suggestions about how to increase the value of the table by adding some information. For instance, the number of randomized controlled trials (RCTs) included in direct estimates of a specific intervention and the numbers of participants in a specific pairwise comparison were two additional characteristics added in response to this feedback. We added information in the NMA relative effect

Table 1. Characteristics of participants

Formal education users (n = 32)	
Healthcare-related professional	62.5%
Master	53.3%
PhD	59.3%
Years of experience[a]	
Healthcare-related professional (n = 19)	15.2 (8.7)
Researcher (n = 25)	12.8 (5.5)
Frequency of using/reading scientific literature, n (%)	
Never	0
Less than once per year	0
1–4 times per year	0
5–10 times per year	0
More than 10 times per year	29 (100)
Frequency in using NMA literature (decision-making or academia), n (%)	
Never	3 (10.3)
Less than once per year	2 (6.9)
1–4 times per year	10 (34.5)
5–10 times per year	6 (20.7)
More than 10 times per year	8 (27.6)
Familiarity with GRADE, n (%)	
Very familiar	23 (79.3)
Somewhat familiar	2 (6.9)
A little familiar	2 (6.9)
Not familiar at all	2 (6.9)

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation; NMA, network meta-analysis.
[a] Mean (SD).

estimate column that explains if the estimate was from the direct, indirect, or both bodies of evidence.

3.3.6. Findability

Users did not report challenges locating the information in the NMA-SoF table that included the PICO section and the network graph that oriented users to the reference comparator. Because some users were unfamiliar with NMA terminology, we added a section with NMA definitions and explanations for all abbreviations.

3.3.7. Desirability and value

Users did not report negative emotional reactions and expressed a need to have SoF tables accompanying NMA reports. Users noted that this NMA-SoF table can be part of an NMA publication but also can be used in a clinical practice guideline context:

> "…*Yes, these tables would probably be in the appendix of the guideline, it would be in the original report of the NMA meta-analysis. And eventually an abbreviated version could be actually in the text but then I don't think that you would necessarily.*"

3.3.8. Understandability

In the first round of interviews, we received feedback to improve the understanding of the information reported in the network diagram, the ranking probabilities, and the interpretation of findings. For the network diagram, we increased the size of the circles (which represent individual interventions) and added an abbreviation label for the interventions. We kept the solid lines linking the interventions that were directly compared. We reported, for each intervention, along with the absolute effect for the population in the studies, an additional row with baseline risk information taken from an observational study. Users did not completely understand the report of two rows with baseline risk information for one intervention in the context of an NMA.

We presented two approaches to reporting NMA-SoF rankings: the probability of being the best intervention for each outcome, and SUCRA. In general, users had issues understanding the information reported about the ranking, regardless of the approach used. SUCRA was the most appealing format for most; users suggested adding the uncertainty of the information with corresponding credible intervals. The last column of the NMA-SoF table was labeled as "what happens." This term is used in a regular current pairwise comparison GRADE SoF table. Users had some difficulty understanding the label, which we changed to "Interpretation of Findings."

During the second round of interviews, we added information including the setting population (PICO component), the number of RCTs, as well as the number of participants for each direct pairwise comparison, the GRADE domain for downgrading the certainty of evidence, the median rank with the credible intervals for each intervention, and a section of NMA definitions in the lowest part of the table. Users understood each of these elements of the NMA-SoF table. We presented two baseline risk information in two rows for each intervention. Users, although they expressed their familiarity with baseline risk information reported in pairwise SoF tables, did not understand the presence of two baseline risk information for each intervention in the context of an NMA. The NMA-specific definitions in the footnotes facilitated understanding of NMA nomenclature.

In the third round, users easily understood the source of the NMA relative effect estimate and the information regarding the anticipated absolute effects when we kept only the information for the population in the studies, that is, one baseline risk for each intervention. However, they suggested modifying the labels of the columns that display the anticipated absolute effects as: "with [reference comparator]," "with intervention," and "difference."

3.3.9. Suggestions

In addition to the seven facets of the Honeycomb model, we identified another theme named "suggestions." Some components in the NMA-SoF table elicited additional

Table 2. NMA-SoF table final format

Estimates of effects, credible intervals, and certainty of the evidence for in XXXXX

Bayesian NMA-SoF table

BENEFITS

Patient or population:

Interventions:

Comparator (reference):

Outcome:

Setting:

Network Meta-analysis geometry plot

Total studies: Total Participants:		Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of the evidence	Ranking**** (95% CrI)	Interpretation of Findings
			Without intervention	*With* intervention	Difference			

NMA-SoF table definitions

* Lines represent direct comparisons

** Estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals (9) since a Bayesian analysis has been conducted.

*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risks of the intervention group with the risk of the control group.

**** Surface under the cumulative (SUCRA) ranking and credible intervals for efficacy are presented. Rank statistics is defined as the probabilities that a treatment out of *n* treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

GRADE Working Group grades of evidence (or certainty in the evidence)

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanatory Footnotes

suggestions from the users. They suggested adding explanations of the GRADE assessment judgments in the context of NMA, using plain language that facilitates optimal understanding:

"*I think generally yes they are helpful, and we should include, I would propose to include footnotes asking for standard SoF but I think in this specific case it would need to be slightly more detailed because I am interested to know kind of the amount of heterogeneity and what is the rationale for downgrading precision for example.*"

3.4. Final NMA-SoF table

The final NMA-SoF table is composed of three sections (Tables 2 and 3).

3.4.1. Upper section

The upper section displays information regarding the PICO components. One outcome needs to be chosen for each NMA-SoF table. Also, one intervention needs to be selected as a "reference comparator," and the other interventions are listed under the "intervention" label. A network graph for the entire network is included in this section.

3.4.2. Middle section

This section is composed of eight columns that report the following information: (1) interventions for a specific outcome. Below each intervention is a description of the number of participants included in the direct comparison and whether the contribution for the relative effect comes from the direct evidence, indirect evidence, or both; (2) the relative effect estimate for each intervention, which is calculated relative to the reference comparator; (3) three columns, two of them that report the anticipated absolute effects information relative to the reference comparator, for each of the interventions, and one that reports the risk difference; (4) the certainty of evidence with the rationale for downgrading the body of the evidence, (5) the ranking, which can be expressed as a median or SUCRA with the corresponding credible interval, (6) and the interpretation of findings that describes the level of superiority or inferiority of each intervention compared with the reference comparator after combination of the relative effect estimate, the certainty of evidence, and the rank probability components.

3.4.3. Lower section

This section has three cells with (1) definitions about NMA terminology and abbreviations, (2) description of each of the final GRADE certainty of evidence judgments,

Table 3. NMA-SoF table template for dichotomous outcomes

Bayesian NMA-SoF table

BENEFITS

Estimates of effects, credible intervals, and certainty of the evidence for chemoprevention of colorectal cancer in individualswith previous colorectal neoplasia

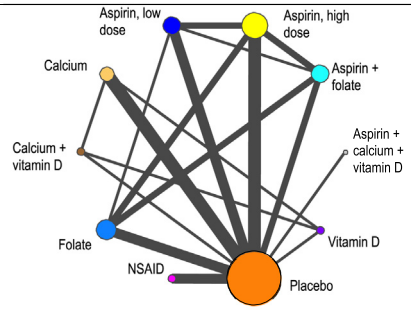
Patient or population: Individuals with previous colorectal neoplasia

Interventions: Low and high dose aspirin, nonaspirin non-steroidal anti-inflammatory drugs (NSAIDs), calcium, vitamin D, folic acid

Comparator (reference): Placebo

Outcome: Prevention of advanced neoplasia; range of follow up between three to five years

Setting: Outpatient



Total studies: 21 RCT **Total Participants**: 12088		**Relative effect** **(95% CrI)**	**Anticipated absolute effect*** (95% CrI)**			**Certainty of evidence**	**Ranking**** **(95% CrI)**	**Interpretation of Findings**
			***Without* intervention**	***With* intervention**	**Difference**			
●	Aspirin + calcium + vitamin D (1 RCT; 427 participants)	**OR 0.71** (0.18 to 2.49) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (61 fewer to 110 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	3 (1 to 10)	-
●	Calcium + vitamin D (1 RCT; 1028 participants)	**OR 0.91** (0.52 to 1.63) Network estimate	74 per 1000[1]	67 per 1000	7 fewer per 1000 (36 fewer to 47 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	6 (1 to 10)	-
●	Aspirin + folate (2 RCT; 916 participants)	**OR 0.73** (0.43 to 1.19) Network estimate	74 per 1000[1]	54 per 1000	20 fewer per 1000 (42 fewer to 14 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	4 (2 to 8)	-
●	Aspirin, high dose (3 RCT; 917 participants)	**OR 0.81** (0.50 to 1.28) Network estimate	74 per 1000[1]	60 per 1000	14 fewer per 1000 (37 fewer to 21 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	5 (2 to 9)	-
●	Aspirin, low dose (3 RCT; 823 participants)	**OR 0.71** (0.41 to 1.23) Network estimate	74 per 1000[1]	53 per 1000	21 fewer per 1000 (44 fewer to 17 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	3 (2 to 9)	-
●	Nonaspirin NSAIDs (4 RCT; 3486 participants)	**OR 0.37** (0.24 to 0.53) Network estimate	74 per 1000[1]	27 per 1000	47 fewer per 1000 (56 fewer to 35 fewer)	⊕⊕⊕⊕ **High**[5]	1 (1 to 2)	-
●	Vitamin D (1 RCT; 764 participants)	**OR 1.19** (0.65 to 2.15) Network estimate	74 per 1000[1]	88 per 1000	14 more per 1000 (26 fewer to 85 more)	⊕⊕◯◯ **Low** Due to Imprecision[3, 5]	9 (3 to 10)	-
●	Calcium (3 RCT; 2503 participants)	**OR 1.00** (0.66 to1.52) Network estimate	74 per 1000[1]	74 per 1000	0 fewer per 1000 (25 fewer to 38 more)	⊕⊕◯◯ **Low** Due to Imprecision[4, 5]	7 (3 to 10)	-
●	Folate (3 RCT; 1224 participants)	**OR 1.32** (0.85 to 2.00) Network estimate	74 per 1000[1]	51 per 1000	23 more per 1000 (11 fewer to 74 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 5]	9 (5 to 10)	-
●	Placebo	Reference comparator	No estimable	No estimable	No estimable	Reference comparator	7 (4 to 9)	-

NMA-SoF table definitions

* Lines represent direct comparisons

** Estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals since a Bayesian analysis has been conducted.

*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risks of the intervention group with the risk of the control group.

**** Median rank and credible intervals for efficacy outcome are presented. Rank statistics is defined as the probabilities that a treatment out of *n* treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

GRADE Working Group grades of evidence (or certainty in the evidence)
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanatory Footnotes
[1] Baseline risks (assumed control risk) obtained from the National Cancer Institute pooling project
[2] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting high possibility of harm.
[3] Very serious imprecision since RR>1 (suggesting greater likelihood of harm than benefit), and with wide credible intervals).
[4] Very serious imprecision since RR is one (suggesting no evidence of benefit) and wide credible intervals suggesting high possibility of harm.
[5] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

Estimates of effects, credible intervals, and certainty of the evidence for chemoprevention of colorectal cancer in individuals with previous colorectal neoplasia

Bayesian NMA-SoF table

HARMS

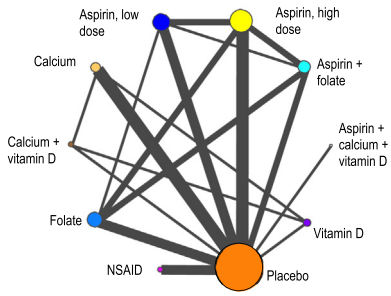
Patient or population: Individuals with previous colorectal neoplasia

Interventions: Low and high dose aspirin, nonaspirin non-steroidal anti-inflammatory drugs (NSAIDs), calcium, vitamin D, folic acid

Comparator (reference): Placebo

Outcome: Serious adverse events; range of follow up between three to five years

Setting: Outpatient



Total studies: 21 RCT Total Participants: 14135	Relative effect** (95% CrI)	Anticipated absolute effect*** (95% CrI)			Certainty of evidence	Ranking**** (95% CrI)	Interpretation of Findings
		Without intervention	*With* intervention	Difference			

and (3) the explanatory footnotes that describe in detail the rationale for the certainty of evidence assessments.

3.5. Considerations of benefit and harms and continuous outcomes in NMA-SoF tables

Users received two NMA-SoF tables simultaneously during the fourth round of interviews. The purpose of presenting these two tables was to obtain feedback about both beneficial and harmful outcomes. In general, users felt it was useful to have access to beneficial and harmful outcomes. They understood the information reported in the tables without relevant issues. Although users mentioned the two NMA-SoF tables looked clean and well organized, they also felt worried due to high number of interventions reported in each NMA-SoF table and its interpretation for decision-making. The steering committee also developed an NMA-SoF table for continuous outcomes based on this final NMA-SoF table format (see Supplement Appendix F).

3.6. A second final NMA-SoF table

The final table displays summary of evidence for one outcome at a time. Thus, to use the information for clinical or policy decisions, stakeholders would need multiple separate SoF tables to present summary of evidence for each patient-important outcomes. This constitutes an important limitation of our final table.

To overcome this challenge, the steering committee developed an alternative NMA-SoF table (Table 4) using one of the first two NMA-SoF table formats as a template, but dealing with its shortcomings identified in the brainstorming meetings and user testing. This format is focused on a smaller number of interventions that matter based on the highest certainty of the evidence for the beneficial outcome, ideally the top three options against the chosen key comparator. It presents information about both beneficial and harmful outcomes for these options to facilitate decision making.

4. Discussion

4.1. Main findings

We conducted a descriptive qualitative study to develop an NMA-SoF table format using strategies to obtain information from diverse constituencies including clinicians and researchers. Our suggested NMA-SoF tables include the main aspects that NMA reports address and capture the complexity of the information of NMAs while maximizing simplicity to achieve a user-friendly presentation.

4.1.1. The evolution of NMA-SoF formats

We tested six different NMA-SoF table formats. The first two NMA-SoF table formats (see Supplement Appendix B) provided an advantage over the final NMA-

	Aspirin + calcium + vitamin D (1 RCT; 714 participants)	**OR 0.90** (0.54 to1.51) Network estimate	187 per 1000[1]	89 per 1000	15 more per 1000 (71 more to 77 fewer)	⊕⊕◯◯ **Low** Due to Imprecision[2, 3]	4 (2 to 7)	-
	Calcium + vitamin D (1 RCT; 1125 participants)	**OR 1.11** (0.76 to 1.70) Network estimate	187 per 1000[1]	203 per 1000	16 more per 1000 (38 fewer to 94 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 3]	2 (1 to 7)	-
	Aspirin + folate (3 RCT; 1017 participants)	**OR 1.21** (0.83 to 1.77) Network estimate	187 per 1000[1]	218 per 1000	31 more per 1000 (27 fewer to 102 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 3]	10 (6 to 10)	-
	Aspirin, high dose (3 RCT; 1507 participants)	**OR 1.06** (0.76 to 1.49) Network estimate	187 per 1000[1]	196 per 1000	9 more per 1000 (38 fewer to 68 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 3]	6 (1 to 10)	-
	Aspirin, low dose (2 RCT; 794 participants)	**OR 0.78** (0.43 to 1.38) Network estimate	187 per 1000[1]	152 per 1000	35 fewer per 1000 (54 more to 97 fewer)	⊕⊕◯◯ **Low** Due to Imprecision[2, 3]	8 (3 to 10)	-
	Nonaspirin NSAIDs (3 RCT; 3964 participants)	**OR 1.23** (0.95 to 1.64) Network estimate	187 per 1000[1]	221 per 1000	34 more per 1000 (8 fewer to 87 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 3]	2 (1 to 9)	-
	Vitamin D (1 RCT; 835 participants)	**OR 1.10** (0.74 to 1.70) Network estimate	187 per 1000[1]	212 per 1000	25 more per 1000 (20 fewer to 78 more)	⊕⊕◯◯ **Low** Due to Imprecision[2, 3]	5 (2 to 10)	-
	Calcium (4 RCT; 2669 participants)	**OR 1.38** (1.07 to 1.89) Network estimate	187 per 1000[1]	238 per 1000	51 more per 1000 (22 more to 82 more)	⊕⊕⊕⊕ **High**[3]	8 (3 to 10)	-
	Folate (3 RCT; 1511 participants)	**OR 0.85** (0.59 to 1.22) Network estimate	187 per 1000[1]	165 per 1000	22 fewer per 1000 (21 more to 59 fewer)	⊕⊕◯◯ **Low** Due to Imprecision[2, 3]	6 (2 to 10)	-
	Placebo	Reference comparator	No estimable	No estimable	No estimable	Reference comparator	3 (1 to 10)	-

NMA-SoF table definitions

* Lines represent direct comparisons

** Estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals since a Bayesian analysis has been conducted.

*** Anticipated absolute effect. Anticipated absolute effect compares two risks by calculating the difference between the risks of the intervention group with the risk of the control group.

**** Median rank and credible intervals for harm outcome are presented. Rank statistics is defined as the probabilities that a treatment out of *n* treatments in a network meta-analysis is the best, the second, the third and so on until the least effective treatment.

GRADE Working Group grades of evidence (or certainty in the evidence)

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanatory Footnotes

[1] Based on assumed control risk of 18.7% (corresponding to pooled 18.7% risk of SAEs in placebo-treated patients of included trials)

[2] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting uncertainty in the estimate.

[3] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

SoF table format, as they displayed multiple outcomes in a single presentation. However, when a large number of interventions were compared, their format became unwieldy.

The third and fourth NMA-SoF tables format offered the advantage of reporting all estimates of all comparison of one or more interventions to each other for a specific outcome. Users still felt, however, confused and overwhelmed due to the large amount of information from multiple interventions. We, therefore, included evidence for one main comparator vs. other interventions for a single outcome in our final format.

The selection of the comparator remains challenging. We suggest the following options to choose the reference comparator for the NMA-SoF table: (1) a placebo intervention, (2) a ‘‘gold’’ standard treatment for the condition under review, (3) or the most cost-effective intervention, (4) or the least effective intervention. We suggest presenting the interventions by row in the NMA-SoF table based on the ranking order.

We initially reported two different baseline risk estimates. Understanding risks is challenging, and misinterpretation is common [28]. Therefore, we ultimately included only one baseline risk. This modification enhanced accessibility of the information and increased understanding of NMA findings.

The resulting NMA-SoF table (see Supplement E and Table 3) captures the complexity of the information reported in an NMA publication. The network graph proved understandable for users. Previous work validated the importance of presenting relative estimates in a consistent format for understanding the effect of an intervention

Table 4. NMA-SoF table format reporting information about multiple treatment comparisons and multiple outcomes

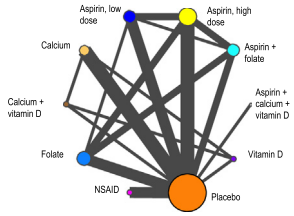
Estimates of effects, credible intervals, and certainty of the evidence for chemoprevention of colorectal cancer in individualswith previous colorectal neoplasia

Patient or population: Individuals with previous colorectal neoplasia

Interventions: Low and high dose aspirin, nonaspirin non-steroidal anti-inflammatory drugs (NSAIDs), calcium, vitamin D, folic acid

Comparison: Placebo

Settings: Outpatient, range of follow up between three to five years



Outcome	Effects and confidence in the estimate of effects						Comments
	Nonaspirin NSAIDs		Aspirin, low dose		Aspirin + calcium + vitamin D		
Prevention of neoplasia Follow up: range from 24 months to 60 months							
Placebo Comparator 74 per 1000[1] (7.4%)	**OR 0.37** (0.24 to 0.53) Network estimate	**47 fewer per 1000** (56 fewer to 35 fewer)	**OR 0.71** (0.41 to 1.23) Network estimate	**21 fewer per 1000** (44 fewer to 17 more)	**OR 0.71** (0.18 to 2.49) Network estimate	**21 fewer per 1000** (61 fewer to 110 more)	None of the ranking treatments between placebo versus other NSAIDs, calcium, vitamin D, or folic acid were highest from the ones we reported. Therefore, we did not include other comparisons in the table.
	⊕⊕⊕⊕ **High** Confidence in estimate		⊕⊕◯◯ **Low** Confidence in estimate due to Imprecision [2, 3]		⊕⊕◯◯ **Low** Confidence in estimate due to Imprecision [2, 3]		
Rank 7 (4 to 9)	**Rank**[4] 1 (1 to 2)		**Rank** 3 (2 to 9)		**Rank** 3 (1 to 10)		
	Based on 3,486 participants (4 RCT)		Based on 823 participants (3 RCT)		Based on 427 participants (1 RCT)		
Serious adverse events Follow up: range from 24 months to 60 months							
Placebo Comparator 74 per 1000[1] (7.4%)	**OR 1.23** (0.95 to 1.64) Network estimate	**34 more per 1000** (8 fewer to 87 more)	**OR 0.78** (0.43 to 1.38) Network estimate	**35 fewer per 1000** (54 more to 97 more)	**OR 0.90** (0.54 to 1.51) Network estimate	**15 more per 1000** (71 more to 77 fewer	Interventions reported for harm outcome were chosen based on the interventions included for beneficial outcome. Therefore, we did not include other comparisons in the table.
	⊕⊕◯◯ **Low** Confidence in estimate due to Imprecision [2, 3]		⊕⊕◯◯ **Low** Confidence in estimate due to Imprecision [2, 3]		⊕⊕◯◯ **Low** Confidence in estimate due to Imprecision [2, 3]		
Rank 4 (2 to 7)	**Rank** 2 (1 to 9)		**Rank** 8 (3 to 10)		**Rank** 4 (2 to 7)		
	Based on 3,964 participants (3 RCT)		Based on 12,098 participants (1 RCT)		Based on 714 participants (1 RCT)		

NMA-SoF table definitions

Lines in the network graphic represent direct comparisons

Estimates are reported as odds ratio. CrI: credible interval. Results are expressed in credible intervals as opposed to the confidence intervals (9) since a Bayesian analysis has been conducted.

The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; OR: Odds ratio;

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

FOOTNOTES

[1] Baseline risks (assumed control risk) obtained from the National Cancer Institute pooling project

[2] Very serious imprecision since 95% CrI crosses unity, and with wide credible intervals suggesting high possibility of harm.

[3] Conceptually, there was no significant intransitivity, with comparable distribution of plausible effect modifiers across trials of different chemopreventive agents.

[4] Ranking is shown as median rank (rank 1-10) and 95% credible intervals.

[25,29]. Reporting absolute effect estimates is also needed to make trade-offs between interventions with beneficial and harmful effects [30,31]. We also included natural frequencies to report the difference between two absolute effects, as it improves understandability of communicating findings [32]. To support the interpretation of findings, additional details of estimates and ratings such as those suggested in our other work should complement the NMA-SoF [33]. We included a final GRADE NMA certainty of evidence judgment. We suggest following the GRADE NMA approach reported in two recent publications [33,34]. The domains that lead to rating down the certainty include risk of bias, inconsistency, indirectness, imprecision, publication bias, intransitivity, and incoherence.

We included ranking information in our NMA-SoF table. Ranking probabilities need to be integrated into the context of healthcare decision-making, as different factors can influence the rank probabilities such as the sample size of individual studies, network geometry, number of studies included for each treatment comparison, and the estimated treatment effects [35].

Ranking can, for several reasons, also be misleading: (1) the evidence on which the rankings are based may be of very low quality; (2) a treatment that is best in one outcome (e.g., a benefit outcome) may be the worst in another outcome (e.g., a harm outcome); or (3) issues such as cost and a clinician's familiarity with the use of a particular treatment may also bear consideration [36]. We encourage users to interpret ranking information with caution. Based on our findings, we suggest presenting the ranking as the

median ranking or the SUCRA with—given that rankings generated in an NMA can have a high degree of imprecision [37]—the associated credible intervals.

4.1.2. An alternative to our final NMA-SoF table

Because of the important limitation of our suggested tables including the necessity for multiple tables, one for each outcome, we created an alternative that is based on the initial approach. Thus, according to primary interest, NMA authors can produce one of two NMA-SoF table formats. One format presents multiple tables reports on all the available interventions with one table for each outcome, whereas the other presents a single NMA-SoF table displaying the findings for the top three interventions across multiple outcomes. For healthcare decision-making, we suggest to display the same top three interventions for all beneficial and harmful outcomes.

4.1.3. Strengths and limitations

Our qualitative design allowed transparency in the data analysis. Two investigators conducted the content analysis of all transcripts independently and in duplicate, which enhances the reliability of findings. We reached saturation in the fourth round of interviews, suggesting that we achieved a sufficient sample size to draw conclusions about users' experience. We used triangulation of multiple data sources to identify areas of convergence. In addition, we included a broad range of users with different backgrounds and experience in research and clinical fields. Finally, we integrated statistical and clinical approaches in a single NMA-SoF table [38].

This study has limitations. We did not test an NMA-SoF table format that displays findings of continuous outcomes. However, based on the findings of the user testing, we designed a format to report results of continuous outcomes (see Supplement Appendix F). Although we tested our NMA-SoF table with findings from studies using the Bayesian approach, the NMA-SoF table can also be used in the context of the frequentist approach. The NMA-SoF table was not evaluated in a randomized control trial.

5. Conclusion and further research

Our NMA-SoF table represents a model for presenting NMA results to diverse stakeholders. In addition to the content of a pairwise comparison SoF table, we suggest to report the following elements in an NMA-SoF table:

- Network geometry
- Relative effect estimates for the highest certainty of the evidence
- Baseline risk information
- Certainty of the evidence for the NMA estimates with judgments and explanations about rating the body of the evidence
- Ranking treatment and its uncertainty
- Include text with definitions of NMA topics (e.g., ranking, absolute effects)

Although this work has been conducted by members of the GRADE working group and includes elements of GRADE, this article does not yet provide official GRADE guidance. Further work includes research about how to present NMA findings for continuous outcomes and the development of interactive SoF tables for inclusion in evidence-to-decision frameworks.

CRediT authorship contribution statement

Juan José Yepes-Nuñez: Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration, Funding acquisition. **Shelly-Anne Li:** Methodology, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing. **Gordon Guyatt:** Methodology, Writing - original draft, Writing - review & editing, Supervision. **Susan M. Jack:** Methodology, Writing - review & editing. **Jan L. Brozek:** Conceptualization, Investigation, Supervision. **Joseph Beyene:** Methodology, Investigation, Supervision. **M. Hassan Murad:** Investigation, Writing - review & editing. **Bram Rochwerg:** Investigation, Writing - review & editing. **Lawrence Mbuagbaw:** Investigation, Writing - review & editing. **Yuan Zhang:** Investigation, Writing - review & editing. **Ivan D. Flórez:** Investigation, Writing - review & editing. **Reed A. Siemieniuk:** Investigation, Writing - review & editing. **Behnam Sadeghirad:** Investigation, Writing - review & editing. **Reem Mustafa:** Formal analysis, Investigation, Writing - review & editing. **Nancy Santesso:** Formal analysis, Investigation, Writing - review & editing. **Holger J. Schünemann:** Conceptualization, Methodology, Formal analysis, Resources, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgments

The authors would like to thank to all participants of the interviews that with their knowledge and feedback help us to create the GRADE NMA-SoF table. This work has been funded by research grant available to Dr. Schunemann in part through an endowed Michael Gent Chair in Healthcare Research.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.04.018.

References

[1] Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897−900.

[2] Bafeta A, Trinquart L, Seror R, Ravaud P. Reporting of results from network meta-analyses: methodological systematic review. BMJ 2014;348:g1741.
[3] Schünemann HJ, Oxman A, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, Guyatt GH. Chapter 12: interpreting results and drawing conclusions. In: Higgins JPTGS, editor. Cochrane handbook for systematic reviews of interventions version 510 (updated March 2011). The Cochrane Collaboration; 2008:2008. Available at www.cochrane-handbook.org.
[4] Schünemann HJ, Oxman A, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: presenting results and 'Summary of findings' tables. In: Higgins JPTGS, editor. Cochrane handbook for systematic reviews of interventions version 510 (updated March 2011). The Cochrane Collaboration; 2008:2008. Available at www.cochrane-handbook.org.
[5] Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. J Clin Epidemiol 2013;66:173−83.
[6] Carrasco-Labra A, Brignardello-Petersen R, Santesso N, Neumann I, Mustafa RA, Mbuagbaw L, et al. Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format. J Clin Epidemiol 2016;74:7−18.
[7] Langendam M, Carrasco-Labra A, Santesso N, Mustafa RA, Brignardello-Petersen R, Ventresca M, et al. Improving GRADE evidence tables part 2: a systematic survey of explanatory notes shows more guidance is needed. J Clin Epidemiol 2016;74:19−27.
[8] Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 2009;181(8):488−93.
[9] Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med 2013;159:130−7.
[10] Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. Med Decis Making 2013;33(5):679−91.
[11] Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ispor task force on indirect treatment comparisons good research practices: part 2. Value Health 2011;14(4):429−37.
[12] Tan SH, Bujkiewicz S, Sutton A, Dequen P, Cooper N. Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons. J Health Serv Res Pol 2013;18:224−32.
[13] Sullivan SM, Coyle D, Wells G. What guidance are researchers given on how to present network meta-analyses to end-users such as policymakers and clinicians? A systematic review. PLoS One 2014;9:e113277.
[14] Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015;162:777−84.
[15] Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000;23:334−40.
[16] Miles M, Huberman A. Qualitative data analysis: an expanded sourcebook. 2th ed. Thousand Oaks, CA: SAGE Publications; 1994.
[17] Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Pol Ment Health 2015;42(5):533−44.
[18] Patton MQ. Chapter 5. Designing qualitative studies. In: Qualitative research & evaluation methods. 3rd ed. Thousand Oaks, CA: Sage Publications; 2002:230−47.
[19] Virzi R. Refining the test phase of usability evaluation: how many subjects is enough. Hum Factors 1992;34(4):457−68.
[20] Creswell JW. Five qualitative approached to inquiry. In: Creswell J, editor. Quality inquiry & Research Design. 3rd ed. California: SAGE; 2013:69−110.
[21] Morville P. User experience design. 2004. Available at http://www.semanticstudios.com/publications/semantics/000029.php. Accessed June 21, 2004.
[22] VERBI Software GmbH. MAXQDA. In: Gerson J, editor. The art of data analysis. Berlin: VERBI GmbH; 2017.
[23] Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15:1277−88.
[24] Rosenbaum SE, Glenton C, Cracknell J. User experiences of evidence-based online resources for health professionals: user testing of the cochrane library. BMC Med Inform Decis making 2008;8:34.
[25] Rosenbaum SE, Glenton C, Nylund HK, Oxman AD. User testing and stakeholder feedback contributed to the development of understandable and useful summary of findings tables for cochrane reviews. J Clin Epidemiol 2010;63:607−19.
[26] Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in cochrane reviews improved understanding and rapid retrieval of key information. J Clin Epidemiol 2010;63:620−6.
[27] Elo S, Kyngas H. The qualitative content analysis process. J Adv Nurs 2008;62:107−15.
[28] Reyna VF. How people make decisions that involve risk. A dual-processes approach. Curr Dir Psychol Sci 2004;13(2):60−6.
[29] Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Med Decis Making 2007;27(5):696−713.
[30] Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. BMJ 2003;327:741−4.
[31] Paling J. Strategies to help patients understand risks. BMJ 2003;327:745−8.
[32] Trevena LJ, Davey HM, Barratt A, Butow P, Caldwell P. A systematic review on communicating with patients about evidence. J Eval Clin Pract 2006;12:13−23.
[33] Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochwerg B, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. J Clin Epidemiol 2018;93:36−44.
[34] Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014;349:g5630.
[35] Kibret T, Richer D, Beyene J. Bias in identification of the best treatment in a Bayesian network meta-analysis for binary outcome: a simulation study. Clin Epidemiol 2014;6:451−60.
[36] Mbuagbaw L, Rochwerg B, Jaeschke R, Heels-Andsell D, Alhazzani W, Thabane L, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. Syst Rev 2017;6(1):79.
[37] Trinquart L, Attiche N, Bafeta A, Porcher R, Ravaud P. Uncertainty in treatment rankings: reanalysis of network meta-analyses of randomized trials. Ann Intern Med 2016;164:666−73.
[38] Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? BMJ 2008;336:1472−4.