Journal of Clinical Epidemiology 173 (2024) 111445

ORIGINAL RESEARCH

A systematic survey of 200 systematic reviews with network meta-analysis (published 2020−2021) reveals that few reviews report structured evidence summaries

Per Olav Løvsletten[a,b,*], Xiaoqin Wang[c,d], Tyler Pitre[e], Marte Ødegaard[f], Areti Angeliki Veroniki[g,h], Carole Lunny[g,i], Andrea C. Tricco[g,j,k], Thomas Agoritsas[l,m,n], Per Olav Vandvik[a,b,n]

[a]Department of Medicine, Lovisenberg Diaconal Hospital, Oslo, Norway
[b]Faculty of Medicine, Institute of Health and Society, University of Oslo, Oslo, Norway
[c]Michael G. DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada
[d]University of Ottawa Heart Institute Research Corporation, Ottawa, Ontario, Canada
[e]Department of Respirology, University of Toronto, Toronto, Ontario, Canada
[f]Library of Medicine and Science, University of Oslo Library, University of Oslo, Oslo, Norway
[g]Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Ontario, Toronto, Canada
[h]Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, Ontario, Canada
[i]Cochrane Hypertension Group and the Therapeutics Initiative, University of British Columbia, Vancouver, Canada
[j]Queen's Collaboration for Health Care Quality: A JBI Centre of Excellence, Queen's School of Nursing. Queen's University, Kingston, Ontario, Canada
[k]Epidemiology Division and Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
[l]Department of Internal Medicine, University Hospitals of Geneva, Geneva, Switzerland
[m]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
[n]MAGIC Evidence Ecosystem Foundation, Oslo, Norway

Accepted 24 June 2024; Published online 26 June 2024

Abstract

Objectives: To map whether and how systematic reviews (SRs) with network meta-analysis (NMA) use presentation formats to report (a) structured evidence summaries − here defined as *reporting of effects estimates in absolute effects with certainty ratings and with a method to rate interventions across one or more outcome(s)* − and (b) NMA results in general.

Study Design and Setting: We conducted a systematic survey, searching MEDLINE (Ovid) for SRs with NMA published between January 1, 2020, and December 31, 2021. We planned to include a random sample of publications, with predefined mechanisms in place for saturation, and included SRs that met prespecified quality criteria and extracted data on presentation formats that reported: (a) estimates of effects, (b) certainty of the evidence, or (c) rating of interventions.

Results: The 200 eligible SRs, from 158 unique Journals, utilized 1133 presentation formats. We found structured evidence summaries in 10 publications (5.0%), with 3 (1.5%) reporting structured evidence summaries across all outcomes, including benefits and harms. Sixteen of the 133 SRs (11.7%) reporting dichotomous outcomes included estimates of absolute effects. Seventy-six SRs (38.0%) reported both benefits and harms and 26 SRs (13.0%) reported certainty ratings in presentation formats, 20 (76.9%) used *Grading of Recommendations Assessment, Development and Evaluation* and 6 (23.1%) used *Confidence In Network Meta-analysis*. *Surface Under the Cumulative Ranking Curve* was the most common method to rate interventions (69 SRs, 34.5%). NMA results were most often reported using forest plots (108 SRs, 54.0%) and league tables (93 SRs, 46.5%).

Conclusion: Most SRs with NMA do not report structured evidence summaries and only rarely do such summaries include reporting of both benefits and harms; those that do offer effective user-friendly communication and provide models for optimal NMA presentation practice.

Keywords: Systematic reviews; Network meta-analysis; Evidence summaries; Presentation formats; Summary of findings tables; Certainty of the evidence

Funding: This work was supported by a PhD grant from South-Eastern Norway Regional Health Authority (grant number 2017015). The funding body was not involved in planning, execution, or reporting of this study.

* Corresponding author. Department of Medicine, Lovisenberg Diaconal Hospital, Lovisenberggata 17C, Oslo 0456, Norway.
E-mail address: perolav@magicevidence.org (P.O. Løvsletten).

https://doi.org/10.1016/j.jclinepi.2024.111445
0895-4356/

Plain Language Summary

In this study, we aimed to explore how results are reported in overviews (systematic reviews) summarizing results from multiple primary studies that altogether compare 3 or more treatments for a given medical condition. Such overviews are key resources for the development of clinical guideline recommendations, but the amount and complexity of information can be overwhelming for users.

To allow well-informed decisions in health policy and practice, results from such overviews should be reported in a certain way. Users need to know the absolute effect of both the potential benefits and harms of the treatments. Additionally, they need to know the extent to which they can have confidence in the results. Finally, with comparison of multiple treatment options, it may be helpful that treatments are ranked/rated from best to worst.

We identified that in a sample of 200 such overviews, only 3 provided results according to the guidance above. Thus, to enhance the value of such overviews, the reporting of results needs to improve. We have provided examples of how this could be done, but also suggest emerging solutions to reduce the risk of information overload. If the reporting practices improve, it is likely that more people can understand and benefit from the latest medical knowledge.

1. Introduction

Systematic reviews (SRs) with network meta-analysis (NMA) are essential in informing treatment decisions in health policy and practice [1−3]. By combining direct and indirect evidence across a network of existing primary studies, NMAs facilitate comparisons for multiple treatment options for a range of health conditions [4−6].

The large volume of results generated from NMAs can, however, be overwhelming; raising daunting challenges in reporting results in comprehensible tables and figures (presentation formats) for end-users like policymakers and physicians without inducing information overload [7]. A recently published SR with NMA on type II diabetes drugs serves as an illustration; comparisons of 13 different drug classes on 12 patient-important outcomes from 816 trials generated approximately 10,000 estimates of effect [6].

To guide end-users through such complex bodies of evidence requires *structured evidence summaries.* For pairwise comparisons, structured evidence summaries, most often presented as summary of findings (SoFs) tables, have facilitated understanding and efficient use of the evidence [8]. Structured evidence summaries include factors essential for decision making in clinical practice; reporting absolute differences in effects across benefits and harms with corresponding certainty (of the evidence) ratings [9−12]. In addition, multiple comparisons necessitates rating of interventions to help users draw conclusions [7,13,14]. Currently there is no clear consensus on how authors can best report structured evidence summaries and NMA results visually [7,15−17]. Investigators, aware of the challenges, continue to propose new presentation formats and approaches, providing authors with a range of options [13,14,18−25].

Regarding methods, there are several approaches for (a) rating the certainty of NMA evidence, and (b) rating interventions in the network. Both the *Grading of Recommendations Assessment, Development and Evaluation* (GRADE) and *Confidence In Network Meta-Analysis* (CINeMA) are well-established methods for assessing and rate the certainty [26−31]. In addition, threshold analysis has been proposed as an alternative method in the context of guideline recommendations [32]. In rating interventions, numbers of approaches are available with *Surface Under the Cumulative Ranking Curve* (SUCRA) as 1 example [13,14,33,34]. However, regardless of methods used, the information must be incorporated in presentation formats.

This diversity in methods and presentation formats raises questions regarding how authors are currently reporting NMA results, and more specifically to what extent and how they are using structured evidence summaries to communicate results to end-users. Previous studies have examined the reporting of NMA results in SRs and Health Technology Assessments, but they are a decade old [35,36]. In this study we: 1) mapped the reporting of structured evidence summaries and NMA results in presentation formats in a sample of recently published SRs with NMA and 2) categorized the presentation formats identified.

2. Material and methods

2.1. Protocol registration

We developed our protocol using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) guideline [37] and it is available on Open Science Framework (https://doi.org/10.17605/OSF.IO/Y3JER).

2.2. Search strategy

Our plan was to extract and analyze a random sample of eligible publications. We conducted our literature search in MEDLINE using the platform Ovid (https://www.ovid.com) (Supplementary Table 1). Given our sample strategy,

What is new?

Key findings

- The reporting of structured evidence summaries in systematic reviews with network meta-analysis is poor.

What this adds to what was known?

- Few systematic reviews with network meta-analysis have applied presentation formats developed to display structured evidence summaries for multiple outcomes.
- Most systematic reviews with network meta-analysis have not applied systematic and transparent methods for rating the certainty of the evidence.

What is the implication and what should change now?

- Innovations to enhance efficiency of GRADEing processes for network meta-analysis are necessary.
- Incorporation of existing best practices on reporting of structure evidence summaries in systematic reviews with network meta-analysis would greatly improve their usefulness.

we did not examine or retrieve additional publications from reference lists of included studies and did not search the gray literature for unpublished studies. An external librarian peer reviewed the search using the Peer Review of Electronic Search Strategies Checklist [38].

2.3. Eligibility criteria

We included peer-reviewed SRs with NMA written in English, published between January 1, 2020, and December 31, 2021, evaluating any health-care interventions in humans with an NMA based on individual participant data and/or aggregate data. We excluded SRs of diagnostic test accuracy or prognostic studies.

The SRs needed to fulfill 5 quality criteria to be included: (a) the term ‘‘systematic review’’ was clearly stated either in title, abstract or used as keyword/label, (b) the review had clear eligibility criteria, (c) the literature search included at least 2 databases, (d) reviewers assessed risk of bias in all individual primary studies, (e) the network contained 3 or more nodes with a larger number of studies than nodes.

We considered both NMAs based on randomized controlled trials and observational studies eligible to include.

2.4. Study selection

Using Covidence SR software (Veritas Health Innovation, Melbourne, VIC, Australia; https://www.covidence.org) 2 reviewers independently conducted title and abstract screening. Any paper identified as possible eligible by either reviewer [39] underwent independent full-text review by 3 reviewers working in pairs (POL, XW, and TP). Reviewers resolved disagreements by consensus or, if necessary, adjudication by the third reviewer.

We estimated we needed to include 250 publications to reach saturation, defined as identification of 5 or fewer new presentation formats in the last 50 publications extracted, with a plan to expand the sample if saturation was not reached. The RAND-function in Microsoft Excel facilitated random ordering of publications that remained eligible for full-text review. We retrieved full texts for the first 350 publications on the list with a plan to, if necessary, expand.

2.5. Data extraction

Three reviewers conducted data extraction independently in pairs (POL, XW, and TP). We used Microsoft Excel for data extraction. To ensure consistency in interpretation of data items, we calibrated extraction between reviewers on a sample of 25 records. Disagreements were resolved by discussion. However, the judgment of the third reviewer was necessary to resolve remaining disagreements on 2 data items. Communication with the corresponding author was also required to reach a final decision on whether to include 7 presentation formats (from 2 publications).

2.5.1. Data items

We extracted data items on 2 levels, per publication and per presentation format.

With our focus on visual reporting practices from and end-user perspective, data items were restricted to the main manuscript, excluding - with 3 exceptions - data reported in supplementary files. These exceptions were: (a) the total number of outcomes with NMA in both main text and supplementary file (n =), (b) rating of interventions in supplementary files (Y/N) and (c) reporting of certainty of evidence in supplementary files (Y/N).

We included presentation formats applied in the main article which at minimum reported: (a) estimates of effect for at least 1 outcome (benefit or harm) across all interventions against at least 1 comparator, or (b) certainty ratings or factor(s) for certainty of the evidence, or (c) rating of interventions (across all interventions) for 1 outcome.

Informed by GRADE guidance on SoF tables for pairwise comparisons [40], we defined a structured evidence summary as a presentation format *reporting the absolute effects for dichotomous and/or continuous outcomes across all interventions for at least 1 outcome against at least 1 comparator with their respective rating of certainty using*

GRADE or CINeMA as well as rating of interventions. Structured evidence summaries reported in figures and tables in the supplementary material was not considered.

We defined rating of interventions as *an approach to ordering interventions (one-by-one or in categories) across 1 variable or a combination of variables*. For example, the most beneficial to the least beneficial treatment, the least harmful to the most harmful, the highest quality evidence to the lowest [41]. We judged rating based solely on distinction between statistically significant vs not statistically significant results to fall outside our definition.

For each publication we extracted data on: (a) general characteristics of the SR (title, authors, journal, issue, publication date, impact factor of journal, topic under study), (b) number and type of presentation formats (figures or tables) reported, (c) number and type of outcomes reported (benefits and harms, categorical/continuous variables), (d) certainty of the evidence, (e) rating of interventions, and (f) presentation formats. Supplementary Table 2 displays the complete list of all data items.

2.6. Data analysis

We applied descriptive statistics for data analysis. The statistical analysis was conducted in STATA Standard Edition version 17.0 (StataCorp, College Station, TX). For categorical variables, we reported frequency and proportions. For continuous or discrete variables, we reported the median and interquartile range (IQR) or full range.

Based on which of the 3 eligibility criteria they met (Supplementary Table 3), we gave presentation formats tags. We also categorized presentation formats as "tables" and "figures" with several subcategories within both labels. To distinguish between presentation formats within a category, we considered both design and function of the presentation format. For example, a table only reporting SUCRA values for an outcome would be labeled "ranking table". If the table reported both effect estimates and SUCRA values, it would be labeled a "SoF table" with sublabel "SoF table without GRADE or CINeMA certainty ratings" (Supplementary Table 9).

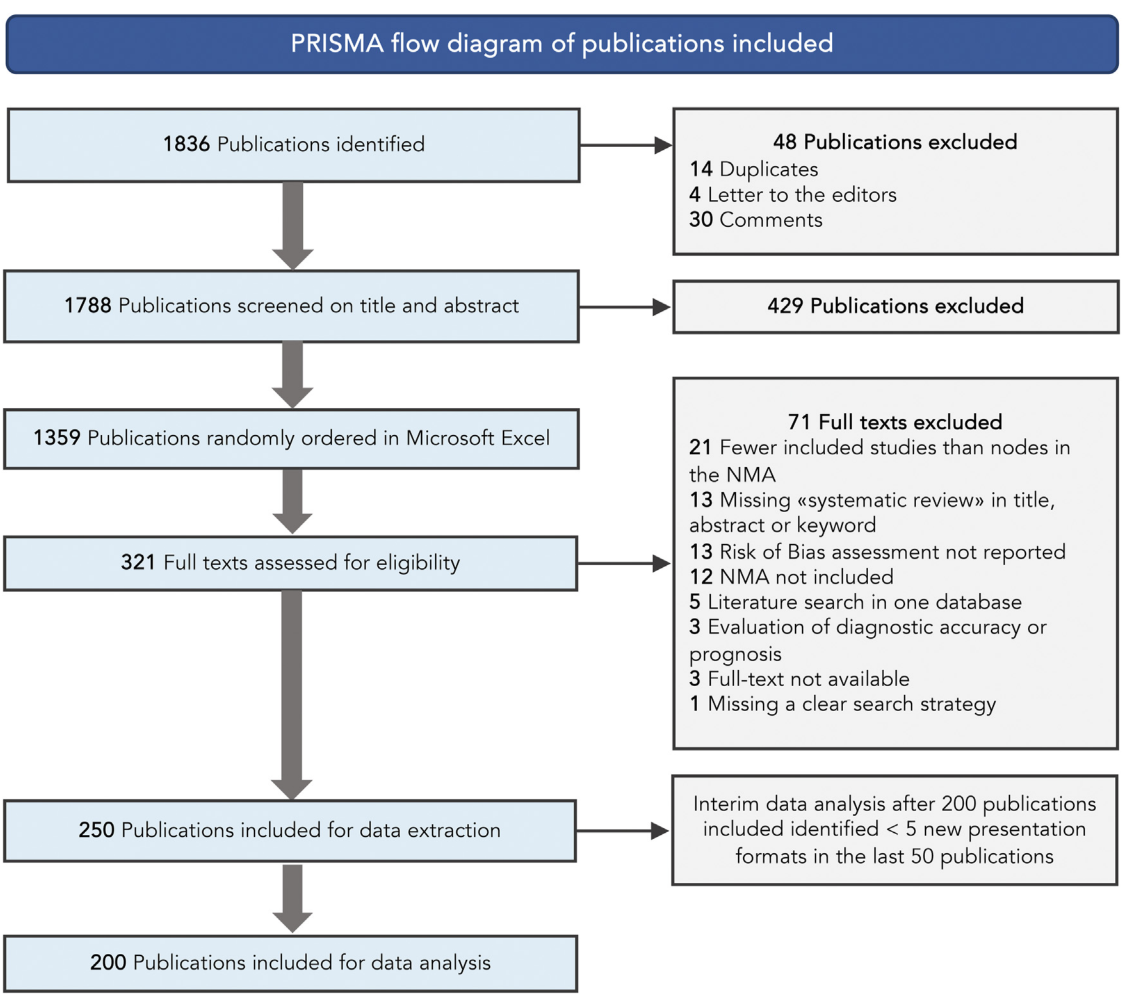

Figure 1. PRISMA Flow diagram of publications included in this article. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

3. Results

3.1. Description of sample

We retrieved 1836 records of which 1788 publications proved eligible for screening. Data-analysis after 200 publications demonstrated saturation of the sample and we stopped inclusion (Fig 1). The sample includes SRs published in 158 journals with a median impact factor of 5.17 (IQR 3.61−7.28) (Supplementary Tables 4−6). The most frequent topics were Internal medicine (23.0%), Surgery (19.0%), Oncology (18.0%) and Psychiatry (9.5%) (Supplementary Tables 7 and 8).

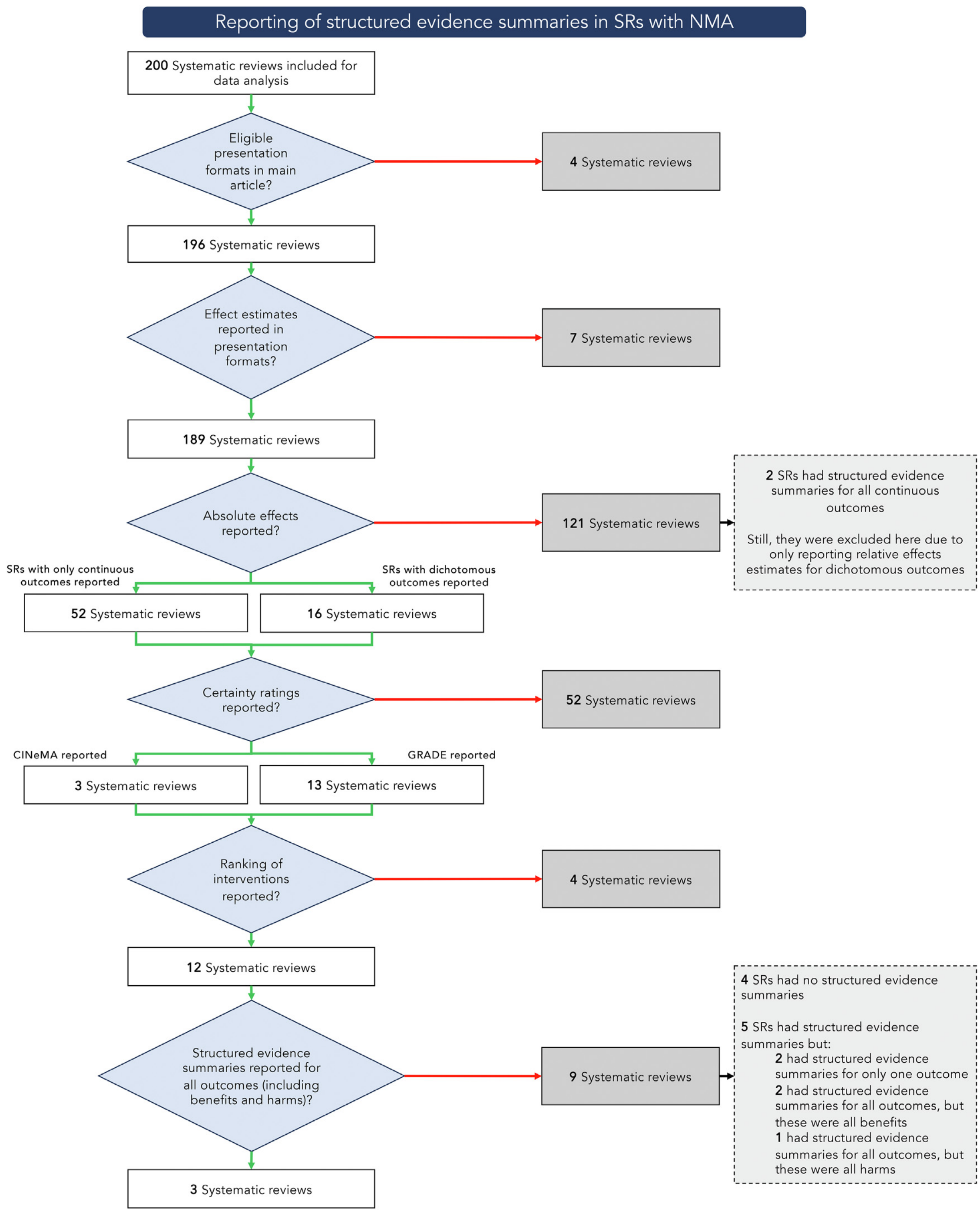

Figure 2. Flow chart displaying reporting of structured evidence summaries in systematic reviews with network meta-analysis. CINeMA, Confidence in network meta-analysis; GRADE, Grading of recommendations, assessment, development, and evaluation; NMA, Network meta-analysis; SR, Systematic review.

3.2. Reporting of structured evidence summaries

Among the 200 SRs, 3 (1.5%) provided structured evidence summaries across all outcomes, including reporting of both benefits and harms [42–44] (Fig 2, full details in Supplementary Fig 1). In total, 10 SRs (5.0%) reported structured evidence summaries for at least 1 outcome (Fig 2, Supplementary Figs 1 and 2).

Across these 10 publications we identified 24 presentation formats (Supplementary Fig 2) in the form of 15 SoF tables (from 8 SRs) and 9 forest plots (from 2 SRs). Four SoF tables (from 4 SRs) reported multiple outcomes (median 6, range 3–11) with 3 SoF tables reporting both benefits and harms [42–44].

3.3. Reporting of effect estimates and benefits and harms

Of the 200 SRs, 137 (66.5%) reported dichotomous outcomes and 99 (49.5%) reported continuous outcomes in presentation formats. Of the SRs reporting dichotomous outcomes, 16 (11.7%) reported results in absolute effects. Of the 89 SRs (44.5%) that did evaluate benefits and harms of interventions, 76 SRs (85.4%) reported the results for both benefits and harms in presentation formats (Supplementary Fig 6).

3.4. Application and reporting of GRADE, CINeMA, and threshold analysis

We found 59 SRs (29.5%) reporting use of either GRADE (n = 41) or CINeMA (n = 18). Of these, 26 SRs (44.1%) reported the certainty ratings in presentation formats – 20 (76.9%) applied GRADE and 6 (23.1%) applied CINeMA (Fig 3). None reported use of Threshold Analysis [32].

3.5. Rating of interventions

Of the 200 SRs, 69 (34.5%) rated interventions using SUCRA values, 51 (25.5%) used ranking probability (including both individual ranking probabilities per rank and cumulative ranking probabilities) and 42 (21.0%) rated by estimates of effect (Table 1). Four SRs (2.0%) used rating methods including certainty of the evidence, such as an approach suggested by *GRADE* [13]. In total, 144 (72.0%) SRs reported rating of interventions in presentation formats.

3.6. Current use of presentation formats in SRs with NMA

Of the 1133 eligible presentation formats, we classified 774 (68.3%) as figures and 359 (31.7%) as tables. Figures occurred in 157 SRs (78.5%) and tables in 152 (76.0%). Median number of outcomes per presentation format were 1 (IQR 1–1), and interventions 7 (IQR 4–10) (full distribution displayed in Supplementary Fig 4A,B).

We classified the presentation formats into 34 categories and subcategories (Supplementary Table 9). Figure 4 shows the most frequently occurring categories, which included forest plots (108 SRs, 54.0%) and league tables (93, 46.5%) as the top 2 categories.

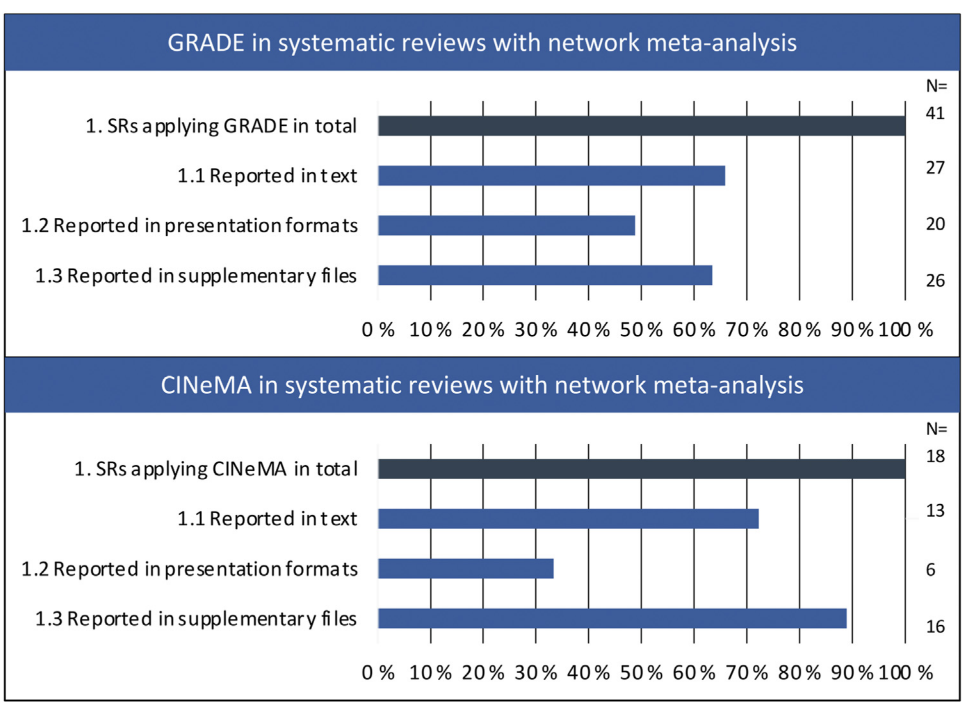

Figure 3. Reporting of GRADE and CINeMA in systematic reviews with network meta-analysis. CINeMA, Confidence in network meta-analysis; GRADE, Grading of recommendations assessment, development, and evaluation; SR, systematic review.

Table 1. Rating methods used in systematic reviews with network meta-analysis

Rating methods or methods	N of 200	%
1. SUCRA values	69	34.5
1.1 Rating per outcome	63	31.5
1.2 Comparison of values across 2 outcomes	8	4.0
1.2.1 Cluster analysis	3	1.5
1.2.2 Not further specified	5	2.5
1.3 Comparison of values across 3 outcomes	1	0.5
2. Rating probability	51	25.5
2.1 Ranking probability (“the probability that an intervention is at a specific rank”)	28	14.0
2.2 Cumulative ranking probabilities	27	13.5
3. Estimate of effect	42	21.0
3.1 Rating per outcome	39	19.5
3.2 Comparison of effects across 2 outcomes	3	1.5
3.3 Effect thresholds	3	1.5
4. P-scores	24	12.0
5. Mean or median rank	15	7.5
6. Probability of being the best	8	4.0
7. Thresholds for effect and certainty	4	2.0
8. Posterior probability	1	0.5
9. Probability of being better than comparator	1	0.5

SUCRA, surface under the cumulative ranking curve.

4. Discussion

Our mapping of a recent sample of SR with NMAs, meeting 5 key quality standards and thus potentially excluding the lowest quality SRs, demonstrates that even among these selected SRs, important gaps in presentation remain. Only 10 of 200 publications provided structured evidence summaries for at least 1 outcome; 3 provided such summaries for all outcomes including benefits and harms (supplementary material not included). SRs seldom reported dichotomous outcomes in absolute effects; fewer than half reported both benefits and harms outcomes; and only one-third applied GRADE or CINeMA to rate certainty of which less than half of those provided certainty ratings in presentation formats. Of the 144 SRs that did rate the interventions, most (47.9%) used SUCRA. Forest plots (54.0%) and league tables (46.5%) proved the most frequently applied presentation formats.

4.1. Strengths and limitations

Strengths of our study include rigorous methods to identify and describe reporting of results in a representative sample of NMA publications from a variety of journals and topics of interest that met 5 key quality standards and thus potentially excluding the publications of the lowest quality. We captured more than 30 presentation format categories, suggesting that we have captured most if not all of those in common use.

Our results complement previous research on presentation of NMA results. In 2020, Kossmeier et al published a comprehensive overview with categorization of more than 200 graphical formats applicable for use in meta-analysis, with graphical formats for NMA as 1 subgroup. However, this overview did not focus on reporting of structured evidence summaries or include tabular formats such as league tables [50].

We also recognize some general and specific limitations. Searching only 1 database meant we retrieved fewer publications. Although we cannot exclude that this has affected the generalizability of our results, we consider this less likely given our approach of extracting a random sample. Secondly, while restricting SRs to those that met 5 quality standards ensures our sample exclude the least rigorous reviews, we did not further assess adherence to methods authors claimed to have used (eg, GRADE). Thus, our insight into the methodological quality of the included SRs is limited.

Furthermore, some may consider our definition of a structured evidence summary as being too stringent. We did not include structured evidence summaries that were reported in supplementary files. However, the likelihood of most end-users of SRs accessing supplementary files is likely to be, at best, limited. Moreover, structured evidence summaries that only displayed interventions considered as most relevant were also not included; however, such examples were only seen in a couple of SRs and thus would not have impacted the results [51,52].

Finally, because we restricted our inquiry to PDFs of manuscripts, we did not capture interactive presentation formats. There are some examples within our sample which in web-versions include or link to such formats [42,43,53], but our identification of only 3 such examples suggest they are uncommon.

4.2. Implications for practice and research

Our findings underscore the need to improve reporting of NMA results and optimize the use of presentation formats. Reporting of structured evidence summaries for multiple outcomes within 1 format, including benefits and harms (achieved in only 3 of our 200 SRs) — rather than 1 outcome per figure — will allow users to quickly grasp a broad overview of the findings. Investigators user-testing one such format found that clinicians experienced the format as both visually appealing and providing an easily grasped overview of interventions across both benefits and harms [19].

To further help readers digest complex NMA evidence, investigators should continue to explore approaches to reduce information overload. Interactive and multilayered presentation formats can help browse complex evidence by facilitating selection of the

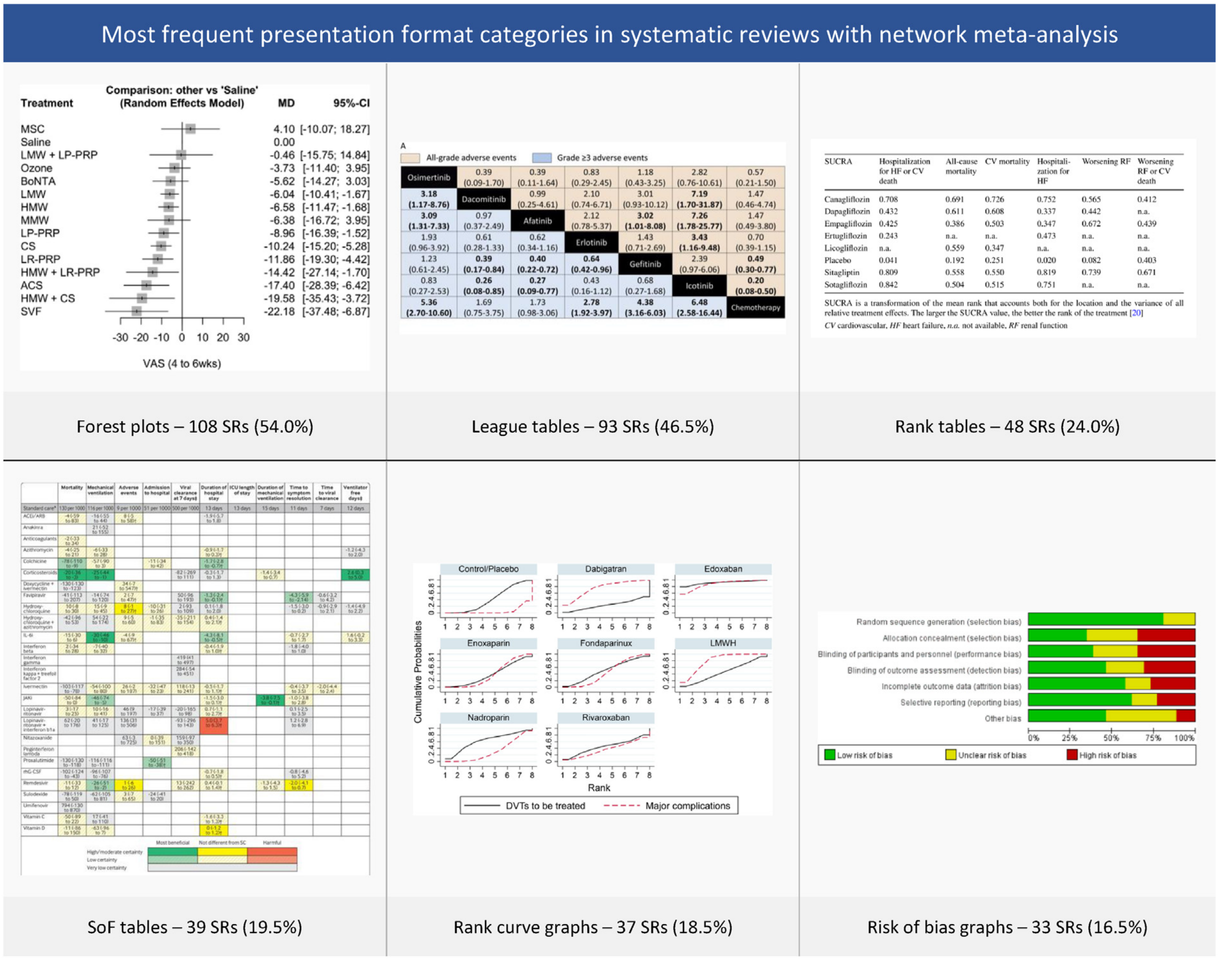

Figure 4. Illustration of the most frequently used presentation formats categories. SoF, Summary of Findings. All figures and tables are reprinted with permission according to creative common license (https://creativecommons.org/licenses/by/4.0/), no changes were made. Citations in order from left to right first row, then second row [43,45–49].

comparisons and outcomes a particular user finds most relevant [7]. Such interactivity has already proved useful in GRADE evidence summaries for pairwise comparisons [54]. Investigators have also developed interactive presentation formats for NMA results (The Kilim plot, Rank-Heat Plot, Statin ranking tool) [23,25,55,56], with some reporting structured evidence summaries in the form of interactive SoF tables or graphical figures [6,42,43,57]. Examples of interactive SoF tables include the MATCH-IT tool — published for diabetes type 2 drugs and physical exercise for patients with Parkinson disease — and a prototype format for metastatic renal cell cancer, while BMJ infographics on covid-19 treatments serves as an example of a graphical figure. Whereas few of these formats have undergone user-testing, exposing MATCH-IT to health care professionals have helped to optimize tool performance [58]. Similar studies on other formats suggest the same [18,19,54,59] and underscore the importance of applying user-centered design principles in development of new presentation formats.

Rating certainty of evidence is a key element in SRs and NMAs enhances the complexity of this process. We found that only a third (29.5%) of SRs applied systematic and transparent processes for rating and reporting certainty of evidence (GRADE or CINeMA approaches). Those that did often failed (55.9%) to present certainty results with the prominence warranted by their importance in treatment decisions. As the number of interventions increases, rating certainty of evidence becomes increasingly resource-demanding and requires high-level competence in methods and biostatistics. Achieving the necessary certainty ratings will require both creating teams large and expert enough to conduct certainty rating processes, and innovations to improve the efficiency of the processes [60].

SUCRA was the most common method for rating interventions in our sample but drawing conclusions solely from this method can be misleading. Indeed, while SUCRA as a method of rating has a number of limitations, its most important is the failure to consider certainty of evidence: large effect sizes and a stellar-looking SUCRA can come from very low certainty evidence [17,61]. A shift toward applying new emerging methods for ranking of interventions also assessing certainty of the evidence as a factor, such as GRADE contextualized approaches, is supported [13,14].

5. Conclusion

Our findings demonstrate that user-friendly, visually compelling presentation formats for NMAs that report structured evidence summaries across benefits and harms are now available but are severely underused [42−44]. Further refinements will be welcome, and innovations to enhance efficiency of GRADEing processes are necessary, but existing approaches, if widely incorporated in SRs including NMAs would greatly improve their usefulness. Interactive presentation formats provide a promising approach that may further help reduce information overload, but extensive user testing and subsequent monitoring of use will be required to optimize their presentation and demonstrate their practical usefulness.

CRediT authorship contribution statement

Per Olav Løvsletten: Writing − review & editing, Writing − original draft, Visualization, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xiaoqin Wang:** Writing − review & editing, Validation, Resources, Investigation, Data curation. **Tyler Pitre:** Writing − review & editing, Validation, Resources, Investigation, Data curation. **Marte Ødegaard:** Writing − review & editing, Resources, Methodology, Formal analysis, Data curation. **Areti Angeliki Veroniki:** Writing − review & editing, Methodology, Conceptualization. **Carole Lunny:** Writing − review & editing, Methodology, Conceptualization. **Andrea C. Tricco:** Writing − review & editing, Methodology, Conceptualization. **Thomas Agoritsas:** Writing − review & editing, Methodology, Conceptualization. **Per Olav Vandvik:** Writing − review & editing, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Data availability

We confirm that the raw data from this study will be made available upon request.

Declaration of competing interest

T.A. reports a relationship with MAGIC Evidence Ecosystem Foundation that includes: board membership and employment. P.O.V. reports a relationship with MAGIC Evidence Ecosystem Foundation that includes: board membership and employment. P.O.V. and T.A. are CEO and deputy CEO of the nonprofit organization MAGIC Evidence Ecosystem Foundation (https://www.magicevidence.org). In collaboration with P.O.L. at Lovisenberg Diaconal Hospital, MAGIC is conducting the research project Making Alternative Treatment Choices Intuitive and Trustworthy (MATCH-IT). The overarching aim of MATCH-IT is to develop a new interactive SoF table and decision support tool, displaying multiple structured evidence summaries from NMAs. The tool may be integrated as part of MAGICapp (https://app.magicapp.org/#/guidelines), an online author and publication platform for guidelines, evidence summaries and decision aids. A.A.V. is Statistical Associate Editor for the Journal of Clinical Epidemiology, but had no involvement with the peer review process or decision for publication. A.C.T. is Coeditor-in-Chief for the Journal of Clinical Epidemiology, but had no involvement with the peer review process or decision for publication. X.W., T.P., M.Ø., and C.L. declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We want to thank Gordon Guyatt for his valuable contribution in preparing this manuscript for submission.

References

[1] Li L, Catalá-López F, Alonso-Arroyo A, Wang Q, Yang K. The global research collaboration of network meta-analysis: a social network analysis. PLoS One 2016;11:e0163239. https://doi.org/10.1371/journal.pone.0163239.

[2] Kanters S, Ford N, Druyts E, Thorlund K, Mills EJ, Bansback N. Use of network meta-analysis in clinical guidelines. Bull World Health Organ 2016;94(10):782−4. https://doi.org/10.2471/BLT.16.174326.

[3] Laws A, Tao R, Wang S, Padhiar A, Goring S. A comparison of national guidelines for network meta-analysis. Value Health 2019; 22(10):1178−86. https://doi.org/10.1016/j.jval.2019.05.013.

[4] Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018;391(10128):1357−66. https://doi.org/10.1016/S0140-6736(17)32802-7.

[5] Cohen AT, Hamilton M, Mitchell SA, Phatak H, Liu X, Bird A, et al. Comparison of the novel oral anticoagulants apixaban, dabigatran, edoxaban, and rivaroxaban in the initial and long-term treatment and prevention of venous thromboembolism: systematic review and network meta-analysis. PLoS One 2015;10:e0144856. https://doi.org/10.1371/journal.pone.0144856.

[6] Shi Q, Nong K, Vandvik PO, Guyatt GH, Schnell O, Rydén L, et al. Benefits and harms of drug treatment for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. BMJ 2023;381:e074068. https://doi.org/10.1136/bmj-2022-074068.

[7] Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 6.4

(Updated August 2023). Cochrane; 2023. Available at: https://training.cochrane.org/handbook#how-to-cite. Accessed July 13, 2024.
[8] 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. https://doi.org/10.1016/j.jclinepi.2009.12.014.
[9] Tikkinen KAO, Guyatt GH. Understanding of research results, evidence summaries and their applicability—not critical appraisal—are core skills of medical curriculum. BMJ Evid Based Med 2021231−3. https://doi.org/10.1136/bmjebm-2020-111542.
[10] Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Chapter 14: completing 'summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page Mj, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 6.4 (Updated August 2023). Cochrane; 2023. Available at: https://training.cochrane.org/handbook#how-to-cite. Accessed July 13, 2024.
[11] Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. J Clin Epidemiol 2013;66:158−72. https://doi.org/10.1016/j.jclinepi.2012.01.012.
[12] 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. https://doi.org/10.1016/j.jclinepi.2012.08.001.
[13] Brignardello-Petersen R, Florez ID, Izcovich A, Santesso N, Hazlewood G, Alhazanni W, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. BMJ 2020;371:m3900. https://doi.org/10.1016/j.jclinepi.2012.08.001.
[14] Brignardello-Petersen R, Izcovich A, Rochwerg B, Florez ID, Hazlewood G, Alhazanni W, et al. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. BMJ 2020;371:m3907. https://doi.org/10.1136/bmj.m3907.
[15] 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. https://doi.org/10.7326/M14-2385.
[16] 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. https://doi.org/10.1371/journal.pone.0113277.
[17] 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;6:79. https://doi.org/10.1186/s13643-017-0473-z.
[18] Yepes-Nuñez JJ, Li S-A, Guyatt G, Jack SM, Brozek JL, Beyene J, et al. Development of the summary of findings table for network meta-analysis. J Clin Epidemiol 2019;115:1−13. https://doi.org/10.1016/j.jclinepi.2019.04.018.
[19] Phillips MR, Sadeghirad B, Busse JW, Brignardello-Petersen R, Cuello-Garcia CA, Kenji Nampo F, et al. Development and design validation of a novel network meta-analysis presentation tool for multiple outcomes: a qualitative descriptive study. BMJ Open 2022;12(6):e056400. https://doi.org/10.1016/j.jclinepi.2019.04.018.
[20] Tan SH, Cooper NJ, Bujkiewicz S, Welton NJ, Caldwell DM, Sutton AJ. Novel presentational approaches were developed for reporting network meta-analysis. J Clin Epidemiol 2014;67:672−80. https://doi.org/10.1016/j.jclinepi.2013.11.006.
[21] Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in Stata. PLoS One 2013;8:e76654. https://doi.org/10.1371/journal.pone.0076654.
[22] Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: the network graphs package. Stata J 2015;15(4):905−50. https://doi.org/10.1177/1536867X1501500402.
[23] Veroniki AA, Straus SE, Fyraridis A, Tricco AC. The rank-heat plot is a novel way to present the results from a network meta-analysis including multiple outcomes. J Clin Epidemiol 2016;76:193−9. https://doi.org/10.1016/j.jclinepi.2016.02.016.
[24] Law M, Alam N, Veroniki AA, Yu Y, Jackson D. Two new approaches for the visualisation of models for network meta-analysis. BMC Med Res Methodol 2019;19:61. https://doi.org/10.1186/s12874-019-0689-9.
[25] Seo M, Furukawa TA, Veroniki AA, Pillinger T, Tomlinson A, Salanti G, et al. The Kilim plot: a tool for visualizing network meta-analysis results for multiple outcomes. Res Synth Methods 2021;12(1):86−95. https://doi.org/10.1002/jrsm.1428.
[26] Puhan MA, Schünemann 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. https://doi.org/10.1136/bmj.g5630.
[27] 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. https://doi.org/10.1016/j.jclinepi.2017.10.005.
[28] Brignardello-Petersen R, Murad MH, Walter SD, McLeod S, Carrasco-Labra A, Rochwerg B, et al. GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks. J Clin Epidemiol 2019;105:60−7. https://doi.org/10.1016/j.jclinepi.2018.08.022.
[29] Brignardello-Petersen R, Mustafa RA, Siemieniuk RAC, Murad MH, Agoritsas T, Izcovich A, et al. GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence. J Clin Epidemiol 2019;108:77−85. https://doi.org/10.1016/j.jclinepi.2018.11.025.
[30] Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Del Giovane C, Egger M, Salanti G. CINeMA: an approach for assessing confidence in the results of a network meta-analysis. PLoS Med 2020;17(4):e1003082. https://doi.org/10.1371/journal.pmed.1003082.
[31] Papakonstantinou T, Nikolakopoulou A, Higgins JPT, Egger M, Salanti G. CINeMA: software for semiautomated assessment of the confidence in the results of network meta-analysis. Campbell Syst Rev 2020;16(1):e1080. https://doi.org/10.1002/cl2.1080.
[32] Phillippo DM, Dias S, Welton NJ, Caldwell DM, Taske N, Ades AE. Threshold analysis as an alternative to GRADE for assessing confidence in guideline recommendations based on network meta-analyses. Ann Intern Med 2019;170:538−46. https://doi.org/10.7326/m18-3542.
[33] Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011;64:163−71. https://doi.org/10.1016/j.jclinepi.2010.03.016.
[34] Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol 2015;15:58. https://doi.org/10.1186/s12874-015-0060-8.
[35] 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 Policy 2013;18:224−32. https://doi.org/10.1177/1355819613498379.
[36] Bafeta A, Trinquart L, Seror R, Ravaud P. Reporting of results from network meta-analyses: methodological systematic review. BMJ 2014;348:g1741. https://doi.org/10.1136/bmj.g1741.
[37] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4(1):1−9. https://doi.org/10.1136/bmj.g1741.

[38] McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. Press peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016;75:40−6. https://doi.org/10.1016/j.jclinepi.2016.01.021.

[39] Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012;1:10. https://doi.org/10.1186/2046-4053-1-10.

[40] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383−94. https://doi.org/10.1016/j.jclinepi.2010.04.026.

[41] GRADE Working Group, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490. https://doi.org/10.1136/bmj.328.7454.1490.

[42] Riaz IB, He H, Ryu AJ, Siddiqi R, Naqvi SAA, Yao Y, et al. A living, interactive systematic review and network meta-analysis of first-line treatment of metastatic renal cell carcinoma. Eur Urol 2021;80(6):712−23. https://doi.org/10.1016/j.eururo.2021.03.016.

[43] Siemieniuk RAC, Bartoszko JJ, Ge L, Zeraatkar D, Izcovich A, Kum E, et al. Drug treatments for Covid-19: living systematic review and network meta-analysis. BMJ 2020;370:m2980. https://doi.org/10.1136/bmj.m2980.

[44] Bartoszko JJ, Siemieniuk RAC, Kum E, Qasim A, Zeraatkar D, Martinez JPD, et al. Prophylaxis against Covid-19: living systematic review and network meta-analysis. BMJ 2021;373:n949. https://doi.org/10.1136/bmj.n949.

[45] Anil U, Markus DH, Hurley ET, Manjunath AK, Alaia MJ, Campbell KA, et al. The efficacy of intra-articular injections in the treatment of knee osteoarthritis: a network meta-analysis of randomized controlled trials. Knee 2021;32:173−82. https://doi.org/10.1016/j.knee.2021.08.008.

[46] Zhao Y, Cheng B, Chen Z, Li J, Liang H, Chen Y, et al. Toxicity profile of epidermal growth factor receptor tyrosine kinase inhibitors for patients with lung cancer: a systematic review and network meta-analysis. Crit Rev Oncol Hematol 2021;160:103305. https://doi.org/10.1016/j.critrevonc.2021.103305.

[47] Täger T, Frankenstein L, Atar D, Agewall S, Frey N, Grundtvig M, et al. Influence of receptor selectivity on benefits from SGLT2 inhibitors in patients with heart failure: a systematic review and head-to-head comparative efficacy network meta-analysis. Clin Res Cardiol 2022;111(4):428−39. https://doi.org/10.1007/s00392-021-01913-z.

[48] Xu S, Chen JY, Zheng Q, Shi L, Tan AHC. Is chemoprophylaxis required after total knee and total hip arthroplasty in the Asian population? A systematic review and network meta-analysis. Thromb Res 2021;198:86−92. https://doi.org/10.1016/j.thromres.2020.11.033.

[49] Lian J, Fu J. Efficacy of various hypoglycemic agents in the treatment of patients with nonalcoholic liver disease with or without diabetes: a network meta-analysis. Front Endocrinol (Lausanne) 2021;12:649018. https://doi.org/10.3389/fendo.2021.649018.

[50] Kossmeier M, Tran US, Voracek M. Charting the landscape of graphical displays for meta-analysis and systematic reviews: a comprehensive review, taxonomy, and feature analysis. BMC Med Res Methodol 2020;20:26. https://doi.org/10.1186/s12874-020-0911-9.

[51] Woolnough T, Axelrod D, Bozzo A, Koziarz A, Koziarz F, Oitment C, et al. What is the relative effectiveness of the various surgical treatment options for distal radius fractures? A systematic review and network meta-analysis of randomized controlled trials. Clin Orthop Relat Res 2021;479(2):348−62. https://doi.org/10.1097/corr.0000000000001524.

[52] Sridharan K, Sivaramakrishnan G. Interventions for oral lichen planus: a systematic review and network meta-analysis of randomized clinical trials. Aust Dent J 2021;66:295−303. https://doi.org/10.1111/adj.12835.

[53] Palmer SC, Tendal B, Mustafa RA, Vandvik PO, Li S, Hao Q, et al. Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. BMJ 2021;372:m4573. https://doi.org/10.1136/bmj.m4573.

[54] Heen AF, Vandvik PO, Brandt L, Achille F, Guyatt GH, Akl EA, et al. Decision aids linked to evidence summaries and clinical practice guidelines: results from user-testing in clinical encounters. BMC Med Inform Decis Mak 2021;21:202. https://doi.org/10.1186/s12911-021-01541-7.

[55] Veroniki AA, Bender R, Glasziou P, Straus SE, Tricco AC. The number needed to treat in pairwise and network meta-analysis and its graphical representation. J Clin Epidemiol 2019;111:11−22. https://doi.org/10.1016/j.jclinepi.2019.03.007.

[56] Naci H. Communication of treatment rankings obtained from network meta-analysis using data visualization. Circ Cardiovasc Qual Outcomes 2016;9(5):605−8. https://doi.org/10.1161/circoutcomes.116.002874.

[57] Ernst M, Folkerts A-K, Gollan R, Lieker E, Caro-Valenzuela J, Adams A, et al. Physical exercise for people with Parkinson's disease: a systematic review and network meta-analysis. Cochrane Database Syst Rev 2024;4(4):CD013856. https://doi.org/10.1002/14651858.cd013856.pub2.

[58] Løvsletten PO, Hunskaar B, Bekkering G, et al. hysicians found an interactive tool displaying structured evidence summaries for multiple comparisons understandable and useful: a qualitative user testing study. J Clin Epidemiol 2024;172:111399. https://doi.org/10.1016/j.jclinepi.2024.111399.

[59] Treweek S, Oxman AD, Alderson P, Bossuyt PM, Brandt L, Brożek J, et al. Developing and evaluating communication Strategies to support informed decisions and practice based on evidence (DECIDE): protocol and preliminary results. Implement Sci 2013;8:6. https://doi.org/10.1186/1748-5908-8-6.

[60] Izcovich A, Chu DK, Mustafa RA, Guyatt G, Brignardello-Petersen R. A guide and pragmatic considerations for applying GRADE to network meta-analysis. BMJ 2023;381:e074495. https://doi.org/10.1136/bmj-2022-074495.

[61] Faltinsen EG, Storebø OJ, Jakobsen JC, Boesen K, Lange T, Gluud C. Network meta-analysis: the highest level of medical evidence? BMJ Evid Based Med 2018;23(2):56−9. https://doi.org/10.1136/bmjebm-2017-110887.