



Interpreting and assessing confidence in network meta-analysis results: an introduction for clinicians

Alan Yang¹ · Petros Pechlivanoglou¹ · Kazuyoshi Aoyama^{1,2} 

Received: 16 December 2021 / Accepted: 26 April 2022 / Published online: 1 June 2022
© The Author(s) 2022

Abstract

Purpose We aimed to provide clinicians with introductory guidance for interpreting and assessing confidence in on Network meta-analysis (NMA) results.

Methods We reviewed current literature on NMA and summarized key points.

Results Network meta-analysis (NMA) is a statistical method for comparing the efficacy of three or more interventions simultaneously in a single analysis by synthesizing both direct and indirect evidence across a network of randomized clinical trials. It has become increasingly popular in healthcare, since direct evidence (head-to-head randomized clinical trials) are not always available. NMA methods are categorized as either Bayesian or frequentist, and while the two mostly provide similar results, the two approaches are theoretically different and require different interpretations of the results.

Conclusions We recommend a careful approach to interpreting NMA results and the validity of an NMA depends on its underlying statistical assumptions and the quality of the evidence used in the NMA.

Keywords Network meta-analysis · Indirect treatment comparisons · Multiple treatment comparisons · Credible intervals · Confidence intervals

Introduction

The highest level of evidence for the comparative effectiveness of different clinical interventions generally comes from systematic reviews of randomized controlled trials (RCTs) [1–3]. The most conventional and widely used method for synthesizing the results of different RCTs is pairwise meta-analysis [4, 5]. While this statistical approach is useful, it is limited as it can only compare two interventions at a time, and only head-to-head RCTs that involve the comparison of interest [6].

Network meta-analysis (NMA) is a statistical method that extends the principles of pairwise meta-analysis to the evaluation of multiple interventions in a single process, which is achieved by combining both direct

and indirect evidence [4, 5, 7, 8]. Direct evidence represents evidence obtained from head-to-head RCTs [4]. For example, in an RCT comparing interventions A and B, the estimate of relative effectiveness of A versus B counts as direct evidence. Indirect evidence represents evidence obtained from one or more common comparators; for example, in the absence of RCTs that evaluate interventions A and B directly, interventions A and B can be indirectly compared if both have been compared to a common intervention C in existing trials [4]. The combination of direct and indirect evidence is at the core of a network meta-analysis [5, 7, 8].

Network meta-analysis is a statistical method for synthesizing direct and indirect evidence from a network of clinical trials to concurrently compare multiple clinical interventions in a single process [4, 5, 7–9]. Synonymous names of NMA include multiple treatment meta-analysis, indirect treatment comparisons, and mixed treatment comparisons [1, 10]. NMA has become attractive among clinicians and health-care researchers in recent years because of its ability to evaluate the comparative clinical effectiveness of different clinical interventions based on clinical evidence through a robust quantitative

✉ Kazuyoshi Aoyama
kazu.aoyama@utoronto.ca

¹ Program in Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, Canada

² Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, 555 University Ave, #2211, Toronto, ON M5G 1X8, Canada

framework [3, 8, 11]. However, due to its complex structure and methodological requirements, a careful approach is required when interpreting NMA results, to avoid drawing biased or incorrect conclusions [3, 12]. This article aims to provide clinicians with introductory guidance for interpreting and assessing confidence in NMA results.

Interpretation of NMA results

NMA has matured over the recent years and NMA models are available for different types of individual-level and trial-level data and summary effect measures (e.g., odds ratio, risk difference) and are being implemented in both frequentist and Bayesian frameworks [2, 13, 14]. Typically, interventions are displayed in the form of a network, called a network diagram. Statistical approaches to NMA are broadly classified as frequentist and Bayesian frameworks [1, 2, 15]. The Bayesian framework allows for a more logical analysis of indirect and multiple comparisons, which are essential for an NMA; therefore, 60–70% of NMA studies have adopted a Bayesian approach [16, 17]. The differences between the two methodological frameworks are further outlined below. While these two methodological frameworks have different fundamental concepts for approaching the NMA model, they produce almost identical results if the sample size is large [17, 18]. Table 1 explains the common terms used in an NMA with plain words as much as possible, to help readers navigate through the following paragraphs [1–5, 8, 11, 13, 17–28].

The Bayesian method combines the known information obtained in the past (prior information) with the present data (likelihood) to calculate the posterior (“post” data observation) probability where the research hypothesis holds [29]. Therefore, the Bayesian method takes a probabilistic approach that allows us to calculate the probability that the research hypothesis holds true, the probability that the true effect size falls within a range—the 95% *credible interval* (CrI), and the ranking probabilities of interventions [8, 29, 30]. Moreover, these probabilities can change depending on prior information [30]. The frequentist method calculates the P value or the 95% *confidence interval* (CI) for rejecting the research hypothesis based solely on present data [7, 8, 17]. Table 2 also highlights differences and similarities between frequentist and Bayesian approaches for NMA [4, 5, 15, 17, 18, 26, 31].

Illustration of interpretation of NMA results through a recent publication in the Journal of Anesthesia

The Journal of Anesthesia has recently published several NMAs [32–36]. We illustrate the interpretation of NMA results through published studies in the journal. One NMA examined the comparative effectiveness of interventions for managing postoperative catheter-related bladder discomfort (CRBD) [33]. A Bayesian Table 3 NMA including 29 trials with 2841 participants was performed for this study. A total of 14 interventions including placebo were included in the evidence network. The effect sizes of interest were the odds ratio (OR) of CRBD at 0 and 1 h after surgery. The results of a Bayesian NMA are usually presented as estimates of relative effect sizes accompanied by 95% CrI. Relative effect sizes are often ratios (e.g., OR, risk ratio, hazard ratio), and in such cases if the credible interval contains 1, then the comparators are not considered as different in the effect size. If the credible interval lies entirely above or below 1, then the comparators are considered as different in the effect size, and the direction (positive or negative) depends on the nature of the effect size associated with the outcome of interest [5, 37]. For example, the estimated OR of CRBD at 0 h after surgery for ketamine versus placebo is 0.17 with a 95% CrI of (0.04, 0.82), which means the odds of CRBD at 0 h after surgery of ketamine is significantly lower than that of placebo. The 95% CrI also implies the true odds ratio of CRBD at 0 h after survey of ketamine versus placebo has a 95% *probability* of being between 0.04 and 0.82. The estimated OR of CRBD at 0 h after surgery of tramadol versus placebo is 0.26 with a 95% CrI of (0.04, 1.73). Since this 95% CrI contains 1, OR of CRBD at 0 h after surgery of tramadol versus placebo has a 95% probability of not being different. A 95% CI under the frequentist approach does not have the same intuitive and practical interpretation, but can only conclude whether the two interventions are statistically different in the effect size at 5% level of significance [37, 38]. A significance level of 5% indicates that there is a 5% risk of concluding that there is a difference when there is actually no difference. That is, if a result is statistically significant, it means it is unlikely to have occurred solely by chance or random factors.

We illustrate the interpretation of results of a frequentist NMA through a study that examined the effects of individualized positive end-expiratory pressure (PEEP) combined

Table 1 Network meta-analysis concepts and definitions

	Framework	Concept/definition
Indirect treatment comparison (ITC)	Bayesian and frequentist	A comparison of the relative effectiveness across different clinical interventions using data from separate non-head-to-head RCTs
Fixed effects model (FE)	Bayesian and frequentist	The fixed-effect model assumes that there is a true effect size that underlies all the RCTs for each comparison in the network, and that all differences in the observed effect sizes are due to sampling error
Random effects model (FE)	Bayesian and frequentist	The random-effects model assumes that the true effect size can differ from trial to trial
Likelihood function	Frequentist	The likelihood function characterizes the joint probability of the observed data as a function of the parameters of the statistical model
<i>P</i> value	Frequentist	The <i>P</i> value is the probability of finding a result that is more extreme than the observed result if the null hypothesis was true. <i>P</i> values are used to help determine whether to reject the null hypothesis. The smaller the <i>P</i> value, the more likely will the null hypothesis be rejected. If the <i>P</i> value is smaller than a pre-specified significance level (usually 5%), then the null hypothesis is rejected at this significance level
Confidence interval	Frequentist	A confidence interval provides an estimated range of values that is likely to include an unknown population parameter; it is calculated from the observed data. The confidence level of a confidence interval is the probability that the interval produced by the method used to calculate the confidence interval includes the true value of the parameter; it is usually 95%
Prior distribution	Bayesian	A prior distribution, or prior, of an unknown parameter, usually the mean effect size, is the probability distribution that represents one's beliefs about this parameter before considering any evidence or observed data
Posterior distribution	Bayesian	The posterior distribution encapsulates all information about an unknown parameter, usually effect sizes, after evidence and observed data are considered. It combines information from the prior distribution and the likelihood function
Posterior summaries	Bayesian	Summary statistics of a posterior distribution; often the mean, median, maximum, minimum, and standard deviation are reported
Credible intervals	Bayesian	A credible interval is an interval within which an unknown parameter value, usually an effect size, falls with a specific probability. It is an interval within a posterior distribution
Ranking probabilities; probability of best treatment; surface under the cumulative ranking area (SUCRA)	Bayesian and Frequentist	Ranking probability is the probability that an intervention is at a specific rank (first, second, etc.) when compared with the other interventions based on a statistic (e.g., mean odds, mean risk, median survival probability). The probability of best treatment is the probability that an intervention is ranked first. The surface under the cumulative ranking curve (SUCRA) is a single number that summarizes the overall ranking of each intervention. Ranking probabilities and SUCRA range from 0 to 100%
Predictive distributions	Bayesian	The predictive distribution is the distribution of possible unobserved (new/ forecasted) values given the observed values
Akaike information criterion (AIC) and Bayesian information criterion (BIC)	Frequentist	The AIC and the BIC are model fit assessments that attempt to explicitly balance model complexity with fit to the observed data. The BIC tends to penalize complex models more compared to the AIC

Table 1 (continued)

	Framework	Concept/definition
Deviance information criterion (DIC)	Bayesian	The DIC compares the relative fit of a set of Bayesian models. Like the AIC and the BIC, it is a model selection method which tries to explicitly balance model complexity with fit to the data
Network geometry	Bayesian and Frequentist	The geometry of the network, usually presented as a network plot, consists of a number of nodes (i.e., interventions), a number of edges (i.e., direct comparison evidence), and number of included studies (thickness of the edges)
Transitivity, similarity or exchangeability	Bayesian and Frequentist	The selection of RCTs to formulate the NMA should be based on rigorous criteria and therefore the included RCTs should be similar such that there are no systematic differences between them other than the interventions. That is, the trials in comparison do not differ with respect to the distribution of effect modifiers
Heterogeneity	Bayesian and Frequentist	The variation in trial outcomes between RCTs within the same comparison
Consistency	Bayesian and Frequentist	The degree of agreement between estimates of effect sizes from direct and indirect evidence
Convergence	Bayesian	Samples from the fitted posterior distributions tend to the theoretical posterior distributions as the number of samples becomes adequately large
Effect modifiers	Bayesian and Frequentist	Characteristics that impact the relative clinical intervention effects
Meta-regression	Bayesian and Frequentist	A regression model that models trial-level or arm-level effect sizes with trial-level covariates. It is often used to reduce heterogeneity and inconsistency between RCTs in the network
I^2	Frequentist	The I^2 statistic is the percentage of variation across RCTs that is due to unexplained heterogeneity rather than randomness
T^2	Frequentist	T^2 is the between-studies variance (the variance of the true effect size parameters across all RCTs) parametrized in the random effects model
τ^2	Bayesian	τ^2 is the precision parameter and also the inverse of the between-trial variance parameter in the random effects model. The lower the between-trial variance, the higher is the precision

Table 2 Differences and similarities between frequentist and Bayesian approaches for network meta-analysis

	Frequentist framework	Bayesian framework
Prior information	Prior information is informally introduced often in the form of supplementary text and is underemphasized	Incorporated within user-specified prior distributions
Basic interpretation	How likely is it to observe the data given a specific parameter value?	How likely is a specific parameter value given the observed data?
Presentation of results	P values, confidence intervals, ranking probabilities	Posterior distributions, credible intervals, ranking probabilities
Caveat	P values are often misinterpreted as probability that the alternative hypothesis is true. Confidence intervals are often misinterpreted as the probability that the true effect size lies in a particular interval	Priors may be difficult to choose Readers often uncritically overemphasize the subjective component induced by the prior and therefore undermine the quality of the analysis. More complex to conduct
Additional features	Model fit and quality assessed with Akaike information criteria or other similar criteria	Model fit and quality assessed with deviance information criterion

Table 3 Available software and statistical packages for network meta-analysis as of December 13, 2021

Statistical package	Framework	Pros	Cons	URL
R	Bayesian and frequentist	Great flexibility, high-quality customizable graphs, free access	Limited user friendliness, steep learning curve, requiring extensive programming knowledge	https://www.rproject.org
WinBUGS/ OPENBUGS/ JAGS	Bayesian	Great flexibility, free access, accessible through other software (e.g., R)	Limited user friendliness, steep learning curve, requiring extensive programming knowledge, limited graphical functionality	https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs https://www.mrc-bsu.cam.ac.uk/software/bugs/openbugs https://mcmc-jags.sourceforge.io
SAS	Bayesian and Frequentist	Great flexibility	Limited user friendliness, requiring fundamental programming knowledge, cost	https://www.sas.com
Stata	Bayesian and Frequentist	High-quality graphs, variety of analyses available	Limited user friendliness, cost	https://www.stata.com
ADDIS/GeMTC	Bayesian	User friendliness, embeds well-developed methods and techniques that are ready to use	Limited modeling capabilities, limited graphical options	https://gemtc.drugis.org

with recruitment maneuver (RM) on intraoperative oxygenation during abdominal surgery [32]. A frequentist NMA including 15 trials with 3634 participants was performed for this study. A total of eight interventions were included in the evidence network. The main effect size of interest was the mean difference in oxygenation index. The results of a frequentist NMA are usually presented as estimates of absolute or relative effect sizes accompanied by 95% CI. If the CI does not contain the equalization threshold (e.g., 0 for difference-type effect sizes, 1 for ratio-type effect sizes), the comparators are statistically different in the effect size, and the direction (positive or negative) depends on the nature of the effect size associated with the outcome of interest. For example, the estimated mean difference in oxygenation index between interventions is 145.0 with 95% CI (87.0, 202.9), which means the oxygenation index of Individualized PEEP + RM is 145.0 higher than that of High PEEP at a 5% significance level. The difference is statistically significantly as the lower edge of 95% CI (i.e., 87.0) is greater than 0.

It is worthwhile to discuss the interpretation of ranking probabilities such as surface under the cumulative ranking area (SUCRA), since these often tend to be misinterpreted in the literature [27, 28, 39]. Table 1 also provides an explanation of these terms. When interpreting these ranking statistics, one should also consider (1) the quality of evidence used in the NMA; (2) confidence in NMA results (further described in the next session); (3) the magnitude of differences in intervention effects; and (4) random chance that may explain any apparent differences between intervention rankings [3, 26, 27, 40]. That is, clinicians and decision makers should not assume an intervention as being “best”

simply because it is ranked first, unless the aforementioned aspects of the NMA are fully considered.

Confidence in NMA results

NMA inherits all challenges present in a conventional pairwise meta-analysis, but with magnified complexity due to the large number of comparisons within the evidence network [37]. To cope with these challenges, NMA adopts a set of assumptions that should be satisfied. The assumptions are (1) similarity or exchangeability, (2) homogeneity and (3) transitivity or consistency [8, 22, 23]. Definitions and concepts of these assumptions are described in detail in Table 1. Typically, if the trial population, trial design and outcome measures are similar for trials that compose the NMA, and that the trials are comparable on effect modifiers (Table 1), these assumptions are adequately satisfied [22, 23]. If one or more assumptions are not satisfied, the NMA becomes inherently biased and in turn yields biased and inaccurate results [41]. To prevent this, remedial measures and adjustments should be applied if appropriate. Methods for assessing NMA assumptions and remedial measures have been developed and widely adopted over the past few years [22, 23]. In addition to these more statistical assumptions, the characteristics of trials in the evidence network that affect the certainty of evidence should be evaluated [42]. These characteristics include risk of bias and publication bias and are often part of the systematic review. These biases usually increase the level of uncertainty of individual trial evidence and subsequently the synthesized evidence in an NMA [3].

In summary, violation of the similarity, homogeneity and consistency assumptions, as well as the presence of any risk of bias and publication bias, affect the overall confidence in the results of an NMA. Therefore, when reviewing a published NMA, one should examine if these issues were identified and how they were dealt with and base one's confidence in the NMA on these factors. GRADE (Grading of Recommendations, Assessment, Development and Evaluations) is a transparent framework for developing and presenting summaries of evidence [42, 43]. It is the most widely adopted tool for grading the quality of evidence with over 100 organizations worldwide officially endorsing GRADE [42]. GRADE provides a tool to assess the aforementioned statistical assumptions and evidence characteristics for any NMA [42–44]. We recommend reviewing the GRADE assessment of a published NMA if it is available. Other tools to assess the quality of an NMA include checklists published by the National Institute for Health and Care Excellence (NICE), the Professional Society for Health Economics and Outcomes Research (ISPOR), PRISMA and Medical Decision Making (MDM) [3, 26, 40, 45].

Using individual patient data in a network meta-analysis

Nowadays, as data become easier to collect and assess, we enter an era of “big data” with big data analysis emerging as a new analysis technique in clinical research [46]. We can utilize big data to improve precision of an NMA. An NMA can turn into a big data analysis through incorporating individual patient data (IPD) into its evidence synthesis process [47, 48]. There are benefits of conducting an NMA using IPD over a usual NMA using aggregated trial-level data. If there is interest in patient-specific covariates, either to explain between-study inconsistency or to explore intervention effects in subgroups of patients, using IPD can have much more statistical power than using aggregated trial-level covariates [48]. Furthermore, several studies have shown that the use of IPD in NMA will considerably improve the precision of estimates of intervention effects and regression coefficients in most scenarios [49, 50]. However, IPD may not provide significant improvement to NMAs that have large and dense intervention networks, since the amount of data and evidence are already large and using IPD on top of these will not much improve the precision in the intervention effect estimates [47]. In most NMAs, since IPD may not be available from all eligible RCTs, techniques for combining IPD and aggregated trial-level data into the NMA have been developed [47, 50].

Conclusions

Network meta-analysis has become increasingly popular for synthesizing multiple sources of clinical evidence. It provides the ability to compare multiple clinical interventions where head-to-head trials are not always available by combining direct and indirect evidence from a network of clinical trials. By doing so, it produces less biased and more precise intervention efficacy estimates. While Bayesian and frequentist methods often yield similar results, the two approaches are fundamentally different in theoretical principles and more importantly require different interpretation of the results. The major limitation of NMA is that NMA results hinge on the inherent statistical assumptions of the NMA and the quality of the evidence used in the NMA. The inherent statistical assumptions are strict and often difficult to satisfy, and the quality of evidence used in the NMA are often difficult to uphold. Multiple requirements need to be met for the results to be sound and useful. Therefore, we recommend a thorough, careful, and conservative approach to interpreting and evaluating the results of an NMA. We also recommend using big data analysis techniques to integrate IPD into the NMA to improve the overall quality and precision of the NMA.

Acknowledgements KA acknowledges the Department of Anesthesiology and Pain Medicine, University of Toronto, and the Hospital for Sick Children, for secured academic time to conduct the current work as a recipient of Merit Awards.

Author contributions AY and KA: conceived this paper. AY and KA: wrote the initial and subsequent draft of the manuscript, which PP: commented on and edited. All authors approved the final version.

Funding This work was supported by the Perioperative Services Facilitator Grant Program 2020–2021, and Outcomes Research Award 2020–2021, Department of Anesthesia and Pain Medicine, Hospital for Sick Children, and Project Grants (#179857) 2022–2023, Canadian Institutes of Health Research.

Declarations

Conflict of interest Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. *Med Decis Making*. 2013;33(5):1–5. <https://doi.org/10.1177/0272989X13487604>.
- Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making*. 2013;33(5):1–4. <https://doi.org/10.1177/0272989X12458724>.
- 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):1–3. <https://doi.org/10.1177/0272989X13485156>.
- Tonin FS, Rotta I, Mendes AM, Pontarolo R. Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. *Pharm Pract*. 2017;15(1):1–5. <https://doi.org/10.18549/PharmPract.2017.01.943>.
- Rouse B, Chaimani A, Li T. Network meta-analysis: an introduction for clinicians. *Intern Emerg Med*. 2017;12(1):1–6. <https://doi.org/10.1007/s11739-016-1583-7>.
- Greco T, Biondi-Zoccai G, Saleh O, Pasin L, Zangrillo A, Landoni G. The attractiveness of network meta-analysis: a comprehensive systematic and narrative review. *Heart Lung Vessel*. 2015;7(2):1–2.
- Dias S, Caldwell DM. Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed*. 2019;104(1):1–6. <https://doi.org/10.1136/archdischild-2018-315224>.
- Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network Meta-Analysis for Decision-Making*. WILEY: Chichester; 2018. p. 200–349.
- Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR task force on indirect treatment comparisons good research practices: part 1. *Value Health*. 2011;14(4):1–2. <https://doi.org/10.1016/j.jval.2011.04.002>.
- Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(2):3–6. <https://doi.org/10.1002/jrsm.1037>.
- Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Making*. 2013;33(5):1–4. <https://doi.org/10.1177/0272989X13485155>.
- Cooper NJ, Peters J, Lai MC, Juni P, Wandel S, Palmer S, Paulden M, Conti S, Welton NJ, Abrams KR, Bujkiewicz S, Spiegelhalter D, Sutton AJ. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value Health*. 2011;14(2):1–3. <https://doi.org/10.1016/j.jval.2010.09.001>.
- Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. *Med Decis Making*. 2013;33(5):1–5. <https://doi.org/10.1177/0272989X13487257>.
- 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(1):1–3. <https://doi.org/10.1186/s12874-019-0689-9>.
- Chaimani A. Undertaking network meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley Online Library; 2021. <https://doi.org/10.1002/9781119536604.ch11>.
- Bhatnagar N, Lakshmi PV, Jeyashree K. Multiple treatment and indirect treatment comparisons: an overview of network meta-analysis. *Perspect Clin Res*. 2014;5(4):1–4. <https://doi.org/10.4103/2229-3485.140550>.
- Hong H, Carlin BP, Shamliyan TA, Wyman JF, Ramakrishnan R, Sainfort F, Kane RL. Comparing Bayesian and Frequentist approaches for multiple outcome mixed treatment comparisons. *Med Decis Making*. 2013;33(5):1–3. <https://doi.org/10.1177/0272989X13481110>.
- Brignardello-Petersen R, Sadeghirad B, Johnston BC, Guyatt GH, Beyene J. Comparing Bayesian and Frequentist approaches for network meta-analysis: an empirical study colloquium abstracts. *Cochrane Database Syst Rev*. 2017;9:1–3.
- Guobing Lu, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc*. 2012. <https://doi.org/10.1198/016214505000001302>.
- Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat Med*. 2003;22(19):1–3. <https://doi.org/10.1002/sim.1566>.
- Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29(7–8):2–5. <https://doi.org/10.1002/sim.3767>.
- Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity—subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Making*. 2013;33(5):1–6. <https://doi.org/10.1177/0272989X13485157>.
- Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making*. 2013;33(5):1–6. <https://doi.org/10.1177/0272989X12455847>.
- Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS ONE*. 2013;8(10):1–3. <https://doi.org/10.1371/journal.pone.0076654>.
- Batson S, Webb N, Greenall G. Meta-analysis to support technology submissions to health technology assessment authorities: criticisms by nice and evidence review groups in the UK. *Value Health*. 2015;18(7):1–2. <https://doi.org/10.1016/j.jval.2015.09.2729>.
- Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, Ioannidis JP, Straus S, Thorlund K, Jansen JP, Mulrow C, Catalá-López F, Gøtzsche PC, Dickersin K, Boutron I, Altman DG, Moher D. 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(11):1–2. <https://doi.org/10.7326/M14-2385>.
- Mbuagbaw L, Rochweg B, Jaeschke R, Heels-Andsell D, Alhazani W, Thabane L, Guyatt GH. Approaches to interpreting and choosing the best treatments in network meta-analyses. *Syst Rev*. 2017;6(1):1–5. <https://doi.org/10.1186/s13643-017-0473-z>.
- Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol*. 2015;15(1):1–9. <https://doi.org/10.1186/s12874-015-0060-8>.
- Uhlmann L, Jensen K, Kieser M. Hypothesis testing in Bayesian network meta-analysis. *BMC Med Res Methodol*. 2018;18(1):1–11. <https://doi.org/10.1186/s12874-018-0574-y>.
- Hu D, O'Connor AM, Wang C, Sargeant JM, Winder CB. How to conduct a bayesian network meta-analysis. *Front Vet Sci*. 2020;7:1–4. <https://doi.org/10.3389/fvets.2020.00271>.
- Reginster J, Bianic F, Campbell R, Martin M, Williams SA, Fitzpatrick LA. Abaloparatide for risk reduction of nonvertebral and vertebral fractures in postmenopausal women with osteoporosis: a network meta-analysis. *Osteoporos Int*. 2019;30(7):1465–73. <https://doi.org/10.1007/s00198-019-04947-2>.
- Li X, Ni ZL, Wang J, Liu XC, Guan HL, Dai MS, Gao X, Zhou Y, Hu XY, Sun X, Zhou J, Zhao Q, Zhang QQ, Liu H, Han Y, Cao JL. Effects of individualized positive end-expiratory pressure

- combined with recruitment maneuver on intraoperative ventilation during abdominal surgery: a systematic review and network meta-analysis of randomized controlled trials. *J Anesth*. 2021. <https://doi.org/10.1007/s00540-021-03012-9>.
33. Hur M, Park SK, Yoon HK, Yoo S, Lee HC, Kim WH, Kim JT, Ku JH, Bahk JH. Comparative effectiveness of interventions for managing postoperative catheter-related bladder discomfort: a systematic review and network meta-analysis. *J Anesth*. 2019;33(2):1–6. <https://doi.org/10.1007/s00540-018-2597-2>.
 34. Hong B, Bang S, Oh C, Park E, Park S. Comparison of PECS II and erector spinae plane block for postoperative analgesia following modified radical mastectomy: Bayesian network meta-analysis using a control group. *J Anesth*. 2021;35(5):1–6. <https://doi.org/10.1007/s00540-021-02923-x>.
 35. Yoshihiro S, Hongo T, Ohki S, Kaneko T, Ishikawa J, Ihara S, Taito S, Sakaguchi M, Yatabe T. Steroid treatment in patients with acute respiratory distress syndrome: a systematic review and network meta-analysis. *J Anesth*. 2021. <https://doi.org/10.1007/s00540-021-03016-5>.
 36. Seki H, Shiga T, Mihara T, Hoshijima H, Hosokawa Y, Hyuga S, Fujita T, Koshika K, Okada R, Kurose H, Ideno S, Ouchi T. Effects of intrathecal opioids on cesarean section: a systematic review and Bayesian network meta-analysis of randomized controlled trials. *J Anesth*. 2021;35(6):1–3. <https://doi.org/10.1007/s00540-021-02980-2>.
 37. Watt J, Tricco AC, Straus S, Veroniki AA, Naglie G, Drucker AM. Research techniques made simple: network meta-analysis. *J Investig Dermatol*. 2019;139(1):1–3. <https://doi.org/10.1016/j.jid.2018.10.028>.
 38. Cote MP, Lubowitz JH, Brand JC, Rossi MJ. Understanding network meta-analysis (NMA) conclusions requires scrutiny of methods and results: introduction to nma and the geometry of evidence. *Arthroscopy*. 2021;37(7):1–4. <https://doi.org/10.1016/j.arthro.2021.04.070>.
 39. Rosenberger KJ, Duan R, Chen Y, Lin L. Predictive P-score for treatment ranking in Bayesian network meta-analysis. *BMC Med Res Methodol*. 2021;21(1):1–10. <https://doi.org/10.1186/s12874-021-01397-5>.
 40. Batson S, Greenall G. Network meta-analysis for health technology submissions worldwide: a report checklist for network meta analysis best practices globally. *Value Health*. 2015;18(7):1–2. <https://doi.org/10.1016/j.jval.2015.09.2736>.
 41. de Souza IS, Robert Allen R, Thode HC. Assessing the confidence in network meta-analysis results. *Am J Emerg Med*. 2021. <https://doi.org/10.1016/j.ajem.2021.01.009>.
 42. Brignardello-Petersen R, Murad MH, Walter SD, McLeod S, Carrasco-Labra A, Rochweg B, Schünemann HJ, Tomlinson G, Guyatt GH. GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks. *J Clin Epidemiol*. 2019;105:1–8. <https://doi.org/10.1016/j.jclinepi.2018.08.022>.
 43. Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochweg B, Hazlewood GS, Alhazzani W, Mustafa RA, Murad MH, Puhan MA, Schünemann HJ, Guyatt GH. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *J Clin Epidemiol*. 2018;93:1–7. <https://doi.org/10.1016/j.jclinepi.2017.10.005>.
 44. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, Brignardello-Petersen R, Carrasco-Labra A, De Beer H, Hultcrantz M, Kuijpers T, Meerpohl J, Morgan R, Mustafa R, Skoetz N, Sultan S, Wiyongse C, Guyatt G, Schünemann HJ. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol*. 2020;119:1–3. <https://doi.org/10.1016/j.jclinepi.2019.10.014>.
 45. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, Boersma C, Thompson D, Larholt KM, Diaz M, Barrett A. 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):1–2. <https://doi.org/10.1016/j.jval.2011.01.011>.
 46. Müller-Wirtz LM, Volk T. Big data in studying acute pain and regional anesthesia. *J Clin Med*. 2021;10(7):1–2. <https://doi.org/10.3390/jcm10071425>.
 47. Kanters S, Karim ME, Thorlund K, Anis A, Bansback N. When does the use of individual patient data in network meta-analysis make a difference? A simulation study. *BMC Med Res Methodol*. 2021;21(1):1–13. <https://doi.org/10.1186/s12874-020-01198-2>.
 48. Chaimani A. Conduct and reporting of individual participant data network meta-analyses need improvement. *BMC Med*. 2020;18(1):1–2. <https://doi.org/10.1186/s12916-020-01630-w>.
 49. Aoyama K, Pinto R, Ray JG, Hill A, Scales DC, Fowler RA. Determining associations and estimating effects with regression models in clinical anesthesia. *Anesthesiology*. 2020;133(3):1–3. <https://doi.org/10.1097/ALN.0000000000003425>.
 50. Leahy J, O'Leary A, Afdhal N, Gray E, Milligan S, Wehmeyer MH, Walsh C. The impact of individual patient data in a network meta-analysis: an investigation into parameter estimation and model selection. *Res Synth Methods*. 2018;9(3):1–4. <https://doi.org/10.1002/jrsm.1305>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.