

Theory and practice of Bayesian and frequentist frameworks for network meta-analysis

Behnam Sadeghirad ^{1,2,3} Farid Foroutan,^{2,4}
Michael J Zoratti,² Jason W Busse,^{1,2,3}
Romina Brignardello-Petersen,² Gordon Guyatt,²
Lehana Thabane^{1,2,5,6}

10.1136/bmjebm-2022-111928

¹Department of Anesthesia,
McMaster University,
Hamilton, Ontario, Canada

²Department of Health
Research Methods, Evidence,
and Impact, McMaster
University, Hamilton, Ontario,
Canada

³Michael G. DeGroot National
Pain Centre, McMaster
University, Hamilton, Ontario,
Canada

⁴Ted Rogers Centre for Heart
Research, University Health
Network, Toronto, Ontario,
Canada

⁵Biostatistics Unit, St Joseph's
Healthcare - Hamilton,
Hamilton, Ontario, Canada

⁶Faculty of Health Sciences,
University of Johannesburg,
Johannesburg, South Africa

Correspondence to:
Dr Behnam Sadeghirad,
Department of Anesthesia,
McMaster University,
Hamilton, ON L8S 4k1,
Canada; sadeghb@mcmaster.
ca



© Author(s) (or their
employer(s)) 2023. No
commercial re-use. See
rights and permissions.
Published by BMJ.

To cite: Sadeghirad B,
Foroutan F, Zoratti MJ,
et al. *BMJ Evidence-
Based Medicine*
2023;28:204–209.

Abstract

Network meta-analysis (NMA) is an increasingly popular statistical method of synthesising evidence to assess the comparative benefits and harms of multiple treatments in a single analysis. Several automated software packages facilitate conducting NMA using either of two alternative approaches, Bayesian or frequentist frameworks. Researchers must choose a framework for conducting NMA (Bayesian or frequentist) and select appropriate model(s), and those conducting NMA need to understand the assumptions and limitations of different approaches. Bayesian models are more frequently used and can be more flexible but require checking additional assumptions and greater statistical expertise that are often ignored. The present paper describes the important theoretical aspects of Bayesian and frequentist models for NMA and the applications and considerations of contrast-synthesis and arm-synthesis NMAs. In addition, we present evidence from a limited number of simulation and empirical studies that compared different frequentist and Bayesian models and provide an overview of available automated software packages to perform NMA. We will conclude that when analysts choose appropriate models, there are seldom important differences in the results of Bayesian and frequentist approaches and that network meta-analysts should therefore focus on model features rather than the statistical framework.

Introduction

In recent years, evidence synthesis using network meta-analysis (NMA) methods has gained popularity,¹ becoming an attractive methodology for researchers, clinicians and decision-makers across the clinical spectrum. A quick search in PubMed shows an exponential increase in the number of NMA publications in the last few years (see [figure 1](#)). While conventional pairwise meta-analysis is limited to providing relative effects for two interventions, NMA offers the ability to simultaneously estimate the relative benefits and harms of multiple interventions or diagnostic tests, thus better supporting complex decision-making processes.² By combining direct evidence with indirect evidence, NMA improves the precision of relative effect estimates. Further, its results

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Network meta-analysis (NMA) has become a popular method to combine results from several studies comparing multiple treatments or interventions.
- ⇒ Statistical models to perform NMA models have been developed both in Bayesian and frequentist frameworks.

WHAT THIS STUDY ADDS

- ⇒ We present important theoretical aspects of Bayesian and frequentist models for NMA, review evidence from simulation and empirical studies that compared different frequentist and Bayesian models, and provide an overview of available automated software packages to perform NMA.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE AND/OR POLICY

- ⇒ Network meta-analysts should focus on model features and their applications and assumptions rather than the statistical framework.

can provide guidance on rating treatment options and reduce the uncertainty of parameters for cost-effectiveness models.^{3 4} The past few years have seen important advances in the statistical methods, software development and methodologies to facilitate interpretation and decision-making.^{3 5–7}

Despite the appealing advantages, NMA presents challenges. The two key assumptions of NMA, transitivity and coherence, rely on the agreement of different sources of evidence (direct and indirect evidence for the same treatment comparison and their similarity), which may be challenging to justify in practice.^{1 2 8} Despite all advancements that has made it easier to produce NMAs, many published systematic reviews with NMA are of poor quality.⁹ Limitations in the practice of NMA abound and include inappropriate network configuration and node selection,^{10 11} using probability ranking to draw conclusions from NMA results,^{12 13} relying on low power statistical tests to assess coherence assumption,¹⁴ using network

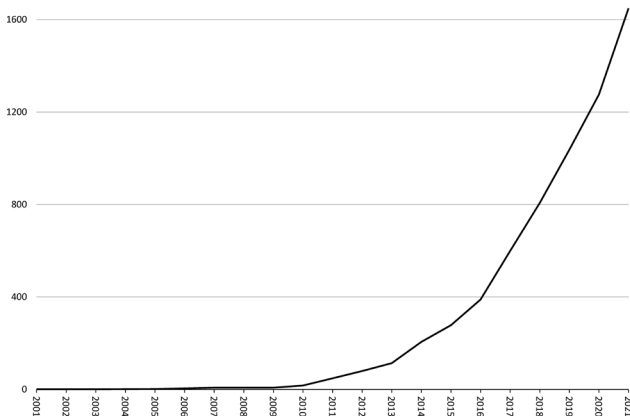


Figure 1 Number of PubMed indexed network meta-analysis publications. PubMed search query: ('Network Meta-Analysis' (Mesh)) or ('network meta-analysis' (title)) or ('multiple treatment meta-analysis' (Title)) or ('multiple treatment comparison' (Title)) or ('adjusted indirect comparison' (Title)).

meta-regression with a limited number of studies to assess effect modification, and failing to assess certainty of evidence for effect estimates.^{12 13}

In recent years, while statisticians and methodologists have made great attempts to improve understanding of NMA methods, there remain methodological topics with limited guidance for those conducting NMA. In the present article, we provide a comprehensive overview of statistical models for NMA, review currently available software packages, and provide guidance for model selection to perform NMA.

Fundamentals of statistical models for NMA

NMA models have been developed both in Bayesian and frequentist frameworks. From a broader perspective, the statistical inference framework differentiates the two approaches.¹⁵ In frequentist statistics, the parameters that represent the characteristics of the population are fixed, but an unknown constant can be inferred using the likelihood of the observed data. In other words, the probability that the research hypothesis is true within the observed data is specified; thus, the frequentist framework can only help decide whether to accept or reject a hypothesis based on the statistical significance level—based on estimation.^{15 16} The results from an analysis using the frequentist approach are given as a point estimate (eg, OR, relative risk or mean difference) with a 95% CI.

Bayesian statistics have a different perspective on uncertainty that mostly involves conditional probability—the probability of an event A, given event B. Unlike the frequentist approach that only uses the likelihood from the observed data, the Bayesian framework relies on the probability distribution of the model parameters given the observed data and the prior beliefs from external information about the values of the parameters. Combining these two using Markov Chain Monte Carlo (MCMC) simulations, which intends to reproduce the model many times until it stabilises and converges, generates a posterior probability.^{15–17} The results of the Bayesian framework are presented as a point estimate with a 95% credible interval (CrI), which is interpreted as the interval in which there is a 95% probability that the values of the point estimate will lie. For ratio measures (eg, OR, relative risk or HR), medians are used as point estimate, whereas either the mean or median can be reported for the pooled mean difference or standardised mean difference.

Frequentist models for NMA

Bucher *et al* proposed the first frequentist approach for a network of three interventions.¹⁸ In this approach, which is also known as *adjusted indirect comparison*, an estimate for the indirect treatment effect of A versus B can be obtained through the direct comparison of A versus C and B versus C (eg, as the difference in log relative effects of the two direct comparisons). The variance of the indirect estimate is the sum of the variances of the two direct effect estimates. In a closed triangular loop of evidence (or evidence cycle) the indirect effect estimate can be combined with the direct effect estimate to obtain the network (or mixed) effect estimate.

The most popular frequentist NMA models are the graph-theoretical approach (proposed by Rücker *et al*¹⁹), the meta-regression approach (proposed by Lumley²⁰ and then further developed by Salanti *et al*²¹), and multivariate meta-analysis model (suggested by White *et al*.²²). The theoretical aspects of these models and their advantages and limitations are discussed in detail elsewhere.^{3 23–25}

All the above models use estimates obtained at contrast or study level – whether NMA is performed under a fixed-effect model (assumes there is one true effect size underlying the trials for each comparison) or a random-effects model (assumes the true effect size can differ from trial to trial). We will go over the contrast-based and arm-based models later in the text. In addition, while theoretically all these models can be modified to use between-study heterogeneity variance (τ^2) from each comparison—informed by more than one study, their associated automated packages (tables 1 and 2) only use the between-study heterogeneity (τ^2) that is assumed to be equal between and constant for all treatment comparisons in the network (ie, a homogenous variance assumption). Heterogeneity variance in NMA models is discussed further down in the text.

Bayesian models for NMA

All Bayesian models suggested for NMA are hierarchical or multi-level models (ie, written in multiple levels or hierarchical form and that the estimates from submodels are dependent). Smith *et al*²⁶ proposed a hierarchical model for conventional meta-analysis that lays the foundation for the popular NMA model of Lu and Ades ('MTC' model),^{27 28} which then extended within the generalised linear mixed-modelling framework and was automated to create the 'GeMTC' model.²⁹ Within this framework, additional models were proposed to account for arm-level estimands (quantities of interest)—for example, Dias *et al* and Hong *et al* models.³⁰

For meta-analysis and NMA in a Bayesian framework, the effects of interventions to be estimated are given prior distributions. In general, we want the observed data to have the most influence on the posterior effect estimates and, thus, often non-informative or minimally informative prior distributions (in which the posterior distribution is determined as completely or as minimally as possible by the observed data) are used.^{23 31 32} The prior distribution should be within the range of plausible values for the pooled intervention effect and, if the prior is to be non-informative (or vague or uniform), it should be very large, which means that the distribution is essentially flat over the plausible range of values for the treatment effect.^{31–33} Despite the importance of prior distributions, a recent review of 44 Bayesian NMAs published in leading general medical journals found that approximately half did not specify their choice of heterogeneity prior distributions and 84% failed to provide a rationale for their selected priors.³³

Research methods and reporting

Table 1 Automated software packages for performing frequentist network meta-analysis

Package (software)	Input data	NMA model	Type of heterogeneity variance	Summary effect measure	Other features
<i>Netmeta</i> (R)	Arm-level Contrast-level	Graph theory	Common	Mean difference, standardised mean difference, OR, risk ratio, log hazard ratio (for time-to-event data)	Generates weighted network map; traditional forest plots used for checking the consistency assumption; global and comparison-specific incoherence assessment; net heat plot to evaluate the inconsistency; league table (comparative effects estimates); probability rankings (visualised with rankogram), SUCRA values, P-scores (can be seen as the frequentist equivalent of SUCRA values)
<i>Network suite</i> (Stata)	Arm-level Contrast-level	Multivariate meta-analysis	Common	Mean difference, standardised mean difference, OR, risk ratio, log hazard ratio (for time-to-event data), risk difference	Advanced options to develop network map; forest plots to compare the results of consistency and inconsistency models; global, comparison-specific, and loop-specific (only first and second order loops) incoherence assessment; league table (comparative effects estimates); probability rankings (visualised with rankogram) and SUCRA values; ranking of treatments using multidimensional scaling; univariable and multivariable network meta-regression; handling of different missing data scenarios (<i>metamiss2</i>)

SUCRA, surface under the cumulative ranking curve.

In a random-effects model, a second set of priors is defined for the between-study heterogeneity variance (τ^2). The range of plausible values for informative and non-informative priors of τ^2 and their influence on NMA effect estimates are discussed

elsewhere.^{32–34} In brief, a typical non-informative prior for τ^2 for a binary outcome using logit link models ranges from 0 to 2; this represents a huge range of trial-specific treatment effects. For example, for a median treatment effect of OR=1.25 (Ln OR=0.22),

Table 2 Automated software packages for performing Bayesian network meta-analysis

Package (software)	Input data	NMA model	Type of heterogeneity variance	Summary effect measure	Other features
<i>GeMTC</i> (R)	Arm-level Contrast-level	Generalised linear models	Common	Mean difference, standardised mean difference, OR, risk ratio, log hazard ratio (for time-to-event data)	Relies on <i>JAGS</i> or <i>OpenBUGS</i> to develop a hierarchical generalised linear model; generates uniform network map; global and comparison-specific incoherence assessment; league table (comparative effects estimates); probability rankings (visualised with rankogram) and SUCRA values; network meta-regression
<i>BUGSnet</i> (R)	Arm-level Contrast-level	Generalised linear models	Common	Mean difference, standardised mean difference, OR, risk ratio	Relies on <i>JAGS</i> to develop a hierarchical generalised linear model; advanced options to develop network map; global and comparison-specific incoherence assessment; league table (comparative effects estimates); probability rankings (visualised with rankogram) and SUCRA values; network meta-regression
<i>bnma</i> (R)	Arm-level Contrast-level	Generalised linear models	Common	Mean difference, standardised mean difference, OR, risk ratio, log hazard ratio (for time-to-event data)	Relies on <i>JAGS</i> to conduct MCMC simulations; generates uniform network map; network.inconsistency.plot function for visual to compare the difference between consistency and inconsistency models; global and comparison-specific incoherence assessment; probability rankings (visualised with rankogram) and SUCRA values
<i>pcnetmeta</i> (R)	Arm-level	Multivariate methods	Common Comparison-specific	Mean difference, standardised mean difference, OR, risk ratio, log hazard ratio (for time-to-event data), risk difference	Relies on <i>JAGS</i> to conduct MCMC simulations; limited advanced options to develop network map; handling of different missing data scenarios; incapable of assessing global or comparison-specific incoherence, or incorporating an inconsistency model, or using meta-regression to explore incoherence; can generate a plot of treatment rank probabilities but cannot estimate SUCRA values
<i>Multinma</i> (R)	Aggregate and individual patient data at arm-level and/or contrast-level	Multilevel network meta-regression	Common	Allows integrating continuous and binary outcomes or Poisson outcome count as mean difference, OR, risk ratio, HR (for time-to-event data)	Models are estimated in a Bayesian framework using ‘Stan’; extends the NMA framework to synthesise mixtures of individual patient and aggregate data; limited options to develop network map; population-adjusted treatment effects can be produced for any study population in the network; capable of producing relative effects, rankings, and absolute predictions. Capable of checking inconsistency using unrelated mean effects (UME) model.

JAGS, Just Another Gibbs Sampler; MCMC, Markov Chain Monte Carlo; NMA, network meta-analysis; SUCRA, surface under the cumulative ranking curve.

we would expect 95% of trials to have a true ORs between 0.17 and 9.24 (calculated as $\exp[0.22 \pm 2]$). It is important to ensure the posterior distribution of τ^2 is sufficiently different from the prior distribution, otherwise the prior distribution will dominate the data and no posterior updating will occur. An alternative, but outdated, approach would be to use a vague gamma prior distribution for the effect estimate to ensure lower prior weights for large values of τ^2 . The limitation of this approach is that it becomes unreasonably informative when low values of τ^2 are possible or if the evidence is described by a sparse network. A network of treatments that is informed by small number of trials (most comparisons informed by 1–2 trials) are typically called sparse network. A star-shape network does not have any closed-loop of evidence and typically consists of treatments that are mostly compared to the reference treatment (e.g., placebo or usual care) is a footnote sentence, a description for explaining “sparse network” term. If it can not be added as a footnote, please remove parenthesis from the beginning and the end. Similar to the frequentist model, the between-study heterogeneity variance is assumed to be zero in a fixed-effect model.

Contrast-synthesis and arm-synthesis NMAs

NMA can be performed using contrast-based or arm-based models. Currently, arm-based models are only available using a Bayesian framework.^{25 30 34} While contrast-based NMAs focus on modelling relative treatment effects (eg, log of the OR), arm-based models use absolute treatment effects from trial arms (eg, the log odds of success).^{25 30} Arm-based models have been criticised for their departure from standard meta-analysis practice by compromising randomisation in evidence, and that absolute effects tend to be highly variable compared to relative effects.^{25 30 35} The advantages of arm-based models are intriguing: they allow entry of arm-level covariates/effect modifiers into the model, data from single-arm studies can be included—this is helpful in cost-effectiveness analysis and decision-making models to have better performance in handling missingness, which is helpful in estimating comparative absolute effects.^{25 35–37} Arm-based models can also accommodate different heterogeneity variances as opposed to using a common-heterogeneity variance.³⁰

Heterogeneity variance in NMA models

All automated packages, with the exception of *pcnetmeta* which is used to perform arm-based NMA, assume a common heterogeneity variance for all treatments effects under the random-effects assumption. This means for all available direct comparisons in the network a single between-study heterogeneity (τ^2) is assumed.^{23 27 32} While theoretically both frequentist and Bayesian models are capable of estimating τ for each comparison informed by at least two studies, in practice it is challenging and mathematically complex.²⁷ Using a constant and common heterogeneity variance reduces the number of model parameters, simplifies the estimation and increases the precision of heterogeneity variance and as a result effect estimates from the NMA. Indeed, in order to simplify the NMA model, it is sometimes the preferred approach for performing arm-based synthesis.^{25 38}

In sparse networks and star-shaped networks informed by small number of studies, heterogeneity estimation across the network can have implausible results and using a common heterogeneity variance can lead to unexpected imprecision of network estimates (ie, the CI of the network estimate becomes wider than that of the direct or the indirect estimates).^{39 40} When the choice of priors for the heterogeneity variance is too wide, this can be more pronounced in Bayesian models than frequentist models.⁴⁰

In such instances, using a fixed-effect model ($\tau=0$) can resolve the problem.³⁹ Arm-based models also can be used, but the estimates using the arm-based models may not converge well if some comparisons are only informed by a few (say, less than three) studies.³⁰

Currently available software packages for NMA

GeMTC,^{41 42} *BUGSnet*⁴³ and *bnma*⁴⁴ are available R packages for performing Bayesian contrast-based NMA using a common heterogeneity variance. In addition, *pcnetmeta* package allows performing Bayesian arm-based NMA in R. Neupane *et al* provided a comprehensive review of three R packages—*GeMTC*, *pcnetmeta* and *netmeta*—for NMA.⁴⁵ *Multinma* is an additional automated package available in R for performing Bayesian NMA of individual patient and aggregate data using multilevel network meta-regression.^{46 47} R package *netmeta*¹⁹ and the Stata *network* suite^{48 49} are the available automated packages to perform frequentist contrast-based NMA. Further details about these packages are summarised in tables 1 and 2.

Empirical and simulation studies comparing Bayesian and frequentist models

Although a considerable amount of methodological research in NMA has been published recently, only a limited number of simulation and empirical studies have compared the existing methodological advances. Unlike other areas of biomedical research, methods are commonly evaluated on an empirical example, resulting in the notion of differences between candidate models rather than an evaluation against known theoretical values.

We conducted a scoping search of PubMed and Scopus to identify empirical and simulation studies comparing frequentist and Bayesian NMA models and found one study protocol,³⁸ two conference presentations^{50 51} and four peer-reviewed publications.^{24 25 52 53} The protocol describes an empirical study of three contrast-synthesis methods and two arm-synthesis methods.³⁸ Their group has published a simulation study comparing contrast-based model of Lu and Ades,²⁸ the arm-based model of Hong *et al*⁵⁴ and two intermediate contrast-based models with different study intercepts and applied these models to a real dataset.²⁵ They concluded that all these models can provide valid results and that the results of these models are comparable. The important difference between their selected models were not that they were arm-based or contrast-based but in other model features (ie, fixed vs random study intercepts, estimands, whether treatment effects relate to study intercepts and missing data assumptions).

Seide *et al* performed a simulation study for sparse networks of trials using two frequentist models (graph-theoretical approach and multivariate meta-analysis model) and the *GeMTC* hierarchical Bayesian model.²⁴ All models were contrast-based and used a common-heterogeneity variance. They concluded that all models performed well with respect to coverage, precision, bias and error, with minimal difference in estimated ranking probabilities. The important factor in the observed differences was the heterogeneity; for example, the credible intervals were wider compared with CIs when flatter priors were used for between-trial heterogeneity or bias, even though minimally, increased with heterogeneity. Kiefer *et al*⁵³ also performed a simulation study of networks with up to five interventions comparing the frequentist graph-theoretical approach (*netmeta*) with two Bayesian MTC hierarchical models, assuming a common heterogeneity variance (all random-effects contrast-based models). They concluded that Bayesian and frequentist consistency models perform similarly

and that with moderate or no incoherence and very low heterogeneity the latter had slightly better properties.

By re-analysing data from 14 published networks, Sadeghirad *et al* compared fixed and random-effects models from the GeMTC hierarchical Bayesian model with non-informative priors to a multivariate meta-analysis (*network suite*) frequentist model.⁵¹ They found no disagreement between network estimates from the two framework models. In a conference poster presentation, Harvey *et al*⁵⁰ compared Bayesian and frequentist approaches using a single real dataset but did not provide any details of NMA models or any of their findings in the abstract. Hong *et al*⁵² used Proc GLIMMIX in SAS to perform frequentist NMA and WinBUGS for Bayesian NMA (both random-effects contrast-based models) using a single real data set and concluded that Bayesian methods are more flexible and their results more clinically interpretable.

Conclusion

Some may argue that the Bayesian framework is more flexible than frequentist in accommodating alternative NMA models and construction of complicated models (eg, arm-based models, analysis of disconnected networks or longitudinal trial data, or using non-common heterogeneity variance) with less assumptions. Most of these advantages are unavailable in automated software packages and, because of their complexity, require greater statistical expertise and proficiency in Bayesian software (eg, WinBUGS, OpenBUGS, JAGS). Using automated software packages within Bayesian framework also has challenges that frequentist approaches avoid, including selection of appropriate prior distributions for treatment effects and between-study heterogeneity variance, and convergence of models derived from MCMC simulations.

Despite these potential challenges, Bayesian models have been used more frequently, perhaps because of the availability of free software packages and codes and previous shortcomings and complications of fitting frequentist NMA models. A limited number of simulation and empirical studies suggest no important difference in the performance, bias or errors between the frequentist and the Bayesian models, and highlight the importance of appropriate model selection. Theoretical aspects of different NMA models also support the idea that when appropriate NMA model is selected, it is unlikely to assume the results of Bayesian and frequentist NMA models would be different. Thus, network meta-analysts should focus on model features and their applications and assumptions rather than the statistical framework.

Contributors BS, FF, MJZ and LT did the planning and initiated the framework. All authors further developed the framework and participated in writing the article. BS is the guarantor of the article. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding Dr JWB is supported, in part, by a CIHR Canada Research Chair in the prevention and management of chronic pain.

Competing interests None declared.

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

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

ORCID iD

Behnam Sadeghirad <http://orcid.org/0000-0001-9422-5232>

References

- Petropoulou M, Nikolakopoulou A, Veroniki A-A, *et al*. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *J Clin Epidemiol* 2017;82:20–8.
- Dias S, Caldwell DM. Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed* 2019;104:F8–12.
- Efthimiou O, Debray TPA, van Valkenhoef G, *et al*. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods* 2016;7:236–63.
- Watt J, Tricco AC, Straus S, *et al*. Research techniques made simple: network meta-analysis. *J Invest Dermatol* 2019;139:4–12.
- Brignardello-Petersen R, Bonner A, Alexander PE, *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.
- Nikolakopoulou A, Higgins JPT, Papakonstantinou T, *et al*. Cinema: an approach for assessing confidence in the results of a network meta-analysis. *PLoS Med* 2020;17:e1003082.
- Yepes-Núñez JJ, Li S-A, Guyatt G, *et al*. Development of the summary of findings table for network meta-analysis. *J Clin Epidemiol* 2019;115:1–13.
- Rouse B, Chaimani A, Li T. Network meta-analysis: an introduction for clinicians. *Intern Emerg Med* 2017;12:103–11.
- Faltinsen EG, Storebø OJ, Jakobsen JC, *et al*. Network meta-analysis: the highest level of medical evidence? *BMJ Evid Based Med* 2018;23:56–9.
- James A, Yavchitz A, Ravaud P, *et al*. Node-making process in network meta-analysis of nonpharmacological treatment are poorly reported. *J Clin Epidemiol* 2018;97:95–102.
- Shi C, Westby M, Norman G, *et al*. Node-making processes in network meta-analysis of nonpharmacological interventions should be well planned and reported. *J Clin Epidemiol* 2018;101:124–5.
- Mbuagbaw L, Rochwerf B, Jaeschke R, *et al*. Approaches to interpreting and choosing the best treatments in network meta-analyses. *Syst Rev* 2017;6:79.
- Brignardello-Petersen R, Florez ID, Izcovich A, *et al*. Grade approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ* 2020;371:m3900.
- Riley RD, Jackson D, Salanti G, *et al*. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ* 2017;358:j3932.
- Pek J, Van Zandt T. Frequentist and Bayesian approaches to data analysis: evaluation and estimation. *Psychology Learning & Teaching* 2020;19:21–35.
- Bland JM, Altman DG. Bayesians and frequentists. *BMJ* 1998;317:1151.
- Kruschke JK, Liddell TM. The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon Bull Rev* 2018;25:178–206.
- Bucher HC, Guyatt GH, Griffith LE, *et al*. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683–91.
- Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012;3:312–24.
- Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21:2313–24.
- Salanti G, Higgins JPT, Ades AE, *et al*. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;17:279–301.
- White IR, Barrett JK, Jackson D, *et al*. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Methods* 2012;3:111–25.
- Dias S, Ades AE, Welton NJ. *Network meta-analysis for decision-making*. John Wiley & Sons, 2018.

- 24 Seide SE, Jensen K, Kieser M. A comparison of Bayesian and frequentist methods in random-effects network meta-analysis of binary data. *Res Synth Methods* 2020;11:363–78.
- 25 White IR, Turner RM, Karahalios A, *et al.* A comparison of arm-based and contrast-based models for network meta-analysis. *Stat Med* 2019;38:5197–213.
- 26 Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995;14:2685–99.
- 27 Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009;10:792–805.
- 28 Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006;101:447–59.
- 29 Dias S, Sutton AJ, Ades AE, *et al.* 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;33:607–17.
- 30 Lin L, Zhang J, Hodges JS, *et al.* Performing Arm-Based network meta-analysis in R with the pnetmeta package. *J Stat Softw* 2017;80. doi:10.18637/jss.v080.i05. [Epub ahead of print: 29 08 2017].
- 31 Ott M, Plummer M, Roos M. How vague is vague? how informative is informative? reference analysis for Bayesian meta-analysis. *Stat Med* 2021;40:4505–21.
- 32 Thorlund K, Thabane L, Mills EJ. Modelling heterogeneity variances in multiple treatment comparison meta-analysis--are informative priors the better solution? *BMC Med Res Methodol* 2013;13:2.
- 33 Rosenberger KJ, Xing A, Murad MH, *et al.* Prior choices of Between-Study heterogeneity in contemporary Bayesian network meta-analyses: an empirical study. *J Gen Intern Med* 2021;36:1049–57.
- 34 Wang Z, Lin L, Hodges JS, *et al.* The impact of covariance priors on arm-based Bayesian network meta-analyses with binary outcomes. *Stat Med* 2020;39:2883–900.
- 35 Dias S, Ades AE. Absolute or relative effects? Arm-based synthesis of trial data. *Res Synth Methods* 2016;7:23–8.
- 36 Dias S, Welton NJ, Sutton AJ, *et al.* Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Making* 2013;33:657–70.
- 37 Zhang J, Chu H, Hong H, *et al.* Bayesian hierarchical models for network meta-analysis incorporating nonignorable missingness. *Stat Methods Med Res* 2017;26:2227–43.
- 38 Karahalios AE, Salanti G, Turner SL, *et al.* An investigation of the impact of using different methods for network meta-analysis: a protocol for an empirical evaluation. *Syst Rev* 2017;6:119.
- 39 Brignardello-Petersen R, Murad MH, Walter SD, *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.
- 40 Lin L, Chu H, Hodges JS. On evidence cycles in network meta-analysis. *Stat Interface* 2020;13:425–36.
- 41 van Valkenhoef G, Dias S, Ades AE, *et al.* Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. *Res Synth Methods* 2016;7:80–93.
- 42 van Valkenhoef G, Lu G, de Brock B, *et al.* Automating network meta-analysis. *Res Synth Methods* 2012;3:285–99.
- 43 Béliveau A, Boyne DJ, Slater J, *et al.* BUGSnet: an R package to facilitate the conduct and reporting of Bayesian network meta-analyses. *BMC Med Res Methodol* 2019;19:196.
- 44 Seo M. *bnma: Bayesian Network Meta-Analysis using 'JAGS'*. Presentation at ESMAR Conf2021.
- 45 Neupane B, Richer D, Bonner AJ, *et al.* Network meta-analysis using R: a review of currently available automated packages. *PLoS One* 2014;9:e115065.
- 46 Phillippo DM, Dias S, Ades AE, *et al.* Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc Ser A Stat Soc* 2020;183:1189–210.
- 47 Phillippo DM, Dias S, Ades AE, *et al.* Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study. *Stat Med* 2020;39:4885–911.
- 48 Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: the network graphs package. *Stata J* 2015;15:905–50.
- 49 White IR. Network meta-analysis. *Stata J* 2015;15:951–85.
- 50 Harvey R, Charokopou M, Arteaga Duarte C, *et al.* PNS22 network meta-analysis: a comparison of Bayesian and Frequentist approaches to synthesis, and evaluation of their suitability for decision-making. *Value in Health* 2020;23:S647.
- 51 Sadeghirad B, Brignardello-Petersen R, Johnston B. Comparing Bayesian and Frequentist approaches for network meta-analysis: an empirical study. *Cochrane Database Syst Rev* 2017;9.
- 52 Hong H, Carlin BP, Shamlivan TA, *et al.* Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Med Decis Making* 2013;33:702–14.
- 53 Kiefer C, Sturtz S, Bender R. A simulation study to compare different estimation approaches for network meta-analysis and corresponding methods to evaluate the consistency assumption. *BMC Med Res Methodol* 2020;20:36.
- 54 Hong H, Chu H, Zhang J, *et al.* A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Synth Methods* 2016;7:6–22.