

EDUTORIAL

A Practical Guide for Application of Network Meta-Analysis in Evidence Synthesis

Meta-analysis is the statistical approach of synthesizing quantitatively the results of multiple studies addressing the same research hypothesis. A common limitation of conventional meta-analysis is the comparison of no more than two interventions at a time. Typically, the researcher is faced with a plethora of competing interventions and is interested in finding which of them are the most safe and effective. Network meta-analysis (NMA) addresses this problem by allowing multiple comparisons among interventions forming a connected network of evidence.

A network plot is the most common way to describe how interventions are connected through direct and indirect routes. In the hypothetical example presented in Fig. 1, a set of randomised controlled trials (RCTs) compared bare metal stent (BMS) with standard balloon angioplasty (SBA) and another set of RCTs compared drug eluting stent (DES) with SBA for the treatment of femoropopliteal disease. No RCTs compared BMS with DES, that is the network forms an open loop. The relative effectiveness for a pair of interventions may be assessed directly, using studies that compare these two interventions head to head (e.g. BMS to SBA and DES to SBA) and indirectly, if these interventions share at least a common comparator (e.g. BMS to DES). Had there been studies comparing BMS to DES, the network plot would be a closed loop and for each one of the three pairs of interventions, there would be two sources of evidence (direct and indirect). NMA synthesises direct and indirect evidence for all pairs of competing interventions to produce a mixed effects estimate, resulting in more precise and powerful estimates, providing a hierarchy of interventions according to the outcome of interest and estimating the relative effectiveness for any pair of interventions including those that have never been directly compared (Table 1).¹

The network geometry of a more complex network of treatments for venous thromboembolism is depicted in Fig. 2. In this network, comparison of aspirin vs. rivaroxaban can be informed through comparisons of aspirin-placebo-rivaroxaban, aspirin-placebo-vitamin K antagonist (VKA)-rivaroxaban, aspirin-placebo-dabigatran-VKA-rivaroxaban, etc. In general, in the absence of direct evidence, more treatment comparisons within a network of treatments provide more “routes” for calculation of indirect treatment estimates.

A consortium of experts has extended the Preferred Reporting Items for Systematic reviews and Meta-Analyses

(PRISMA) guidelines to systematic reviews with NMA.² The first steps in conducting a NMA are similar to those of a conventional meta-analysis, including forming a Problem/Patient/Population-Intervention-Comparison/Control-Outcome (PICO) question, searching multiple databases, identifying studies relevant to the research question, extracting data, and performing risk of bias assessment. When defining the inclusion criteria, it is important to consider studies that include a common comparator within the network, for example placebo drug or sham intervention. Although sham interventions may not be of clinical or research interest, they contribute to the treatment network and increase precision of effect estimates.

Steps unique to NMA are:

1. *Explore network geometry.* Visual inspection of network plots allows assessment of the presence of direct evidence and routes through which indirect evidence is informed. The size of the nodes and the thickness of the edges can be proportional to the number of participants randomised to that treatment and comparison, respectively, or any other characteristic (e.g. number of studies). Edges can have different colors to reflect differences across comparisons (e.g. comparisons that are at low/unclear/high risk of bias for allocation concealment may be presented in green/yellow/red color). Network graphs can be constructed using free (CINeMA, ADDIS, R) or commercially available software (Stata, Microsoft Excel).^{3–8}
2. *Assess the transitivity assumption.* This is the fundamental assumption of NMA and it refers to the validity of indirect comparisons for a given network.⁹ Rarely are data available to explore the network statistically, but it can be evaluated clinically and conceptually. NMA provides observational evidence and there is the risk of confounding. In the example presented in Fig. 1, if RCTs comparing SBA with DES used dual antiplatelet therapy whereas RCTs comparing SBA with BMS used single antiplatelet therapy, the transitivity assumption would probably not hold. The transitivity assumption requires that distribution of a priori defined effect modifiers be similar across treatment comparisons. It also requires that interventions do not differ substantially when they appear in different comparisons.
3. *Perform exploratory analyses and NMA.* Free (WinBUGS, R, JAGS) or commercially available software (Stata, SAS) can be used to perform NMA statistics. Pairwise meta-analyses are performed to explore direct treatment effects. NMA can be performed using either a Bayesian (WinBUGS, JAGS) or a frequentist method (R, Stata).

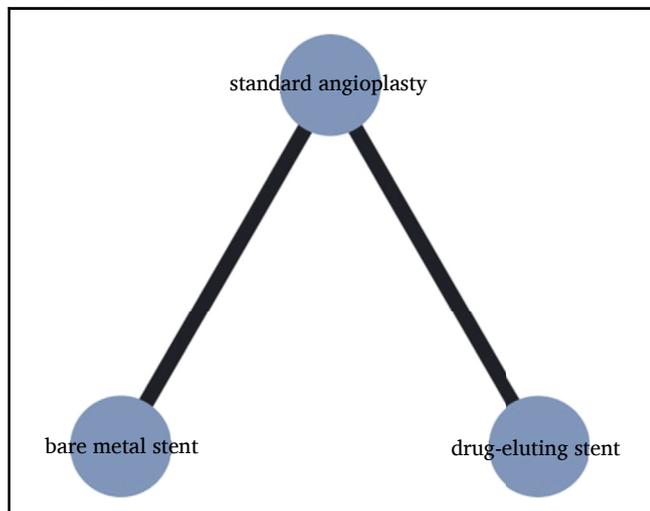


Figure 1. Example treatment network of standard balloon angioplasty, bare metal stent angioplasty, and drug eluting stent angioplasty forming an open loop. Nodes (circles) represent interventions and edges (lines) represent studies directly comparing the connected interventions. Produced using sham datasets and CINeMA: Confidence in Network Meta-analysis [Software]. Institute of Social and Preventive Medicine, University of Bern, 2017. Available from cinema.ispm.ch.

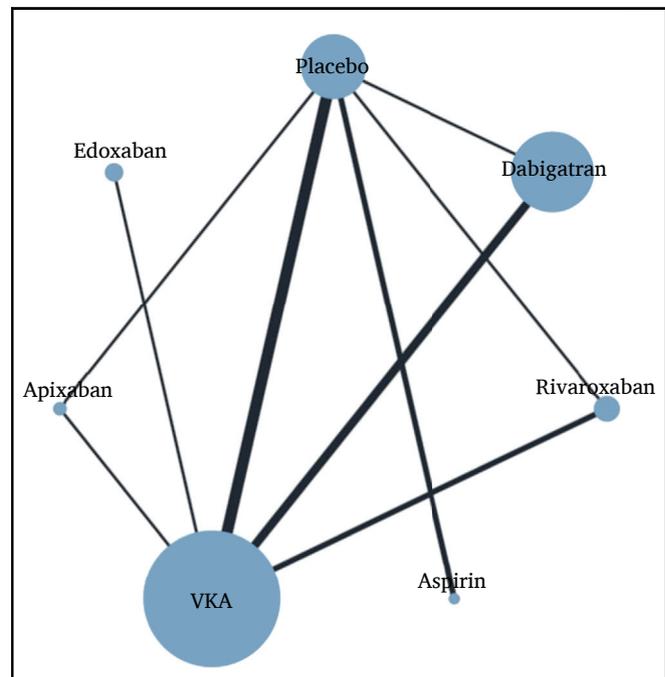


Figure 2. Example network of seven interventions with nine sets of comparisons, providing both direct and indirect evidence for each pair of treatments and forming both closed and open loops. VKA = vitamin K antagonist. Produced using sham datasets and CINeMA: Confidence in Network Meta-analysis [Software]. Institute of Social and Preventive Medicine, University of Bern, 2017. Available from cinema.ispm.ch.

Differences between the two approaches are mainly a matter of philosophy that revolves around the definition of probability. Comparative intervention effect estimates are presented using a league table (Table 2), which tabulates estimates for each combination of comparisons. Effect estimates can be graphically represented using forest plots with CIs and predictive intervals. Predictive intervals present the expected range of true effects in a future study and are helpful in the presence of substantial heterogeneity.

4. **Ranking measures.** Several ranking measures have been suggested. They all rely on estimating ranking probabilities. Ranking tables list the probability of each treatment assuming any rank (Table 3) and rankograms plot these values in graphs. Surface under the cumulative ranking curve (SUCRA) plots depict the cumulative ranking curve and show the percentage of effectiveness

an intervention achieves with reference to an imaginary ideal intervention. Recently, a frequentist counterpart of SUCRA values, called p-score, has been suggested.¹⁰ It would be misleading to focus only on the probability of being the best intervention, and the whole ranking distribution should be considered. Ranking measures should not be overinterpreted but always be examined along treatment effects and their corresponding 95% CIs. If DES has a larger SUCRA/p-score value than BMS but the relative effect is not large and shows no clinically important difference, it cannot be inferred with certainty that DES is better than BMS.

Table 1. Advantages of network meta-analysis over standard meta-analysis

Standard meta-analysis	Network meta-analysis
<ul style="list-style-type: none"> • Compares a pair of interventions at a time 	<ul style="list-style-type: none"> • Compares multiple interventions
<ul style="list-style-type: none"> • Compares interventions only when head to head comparisons are available 	<ul style="list-style-type: none"> • Allows indirect comparison of interventions sharing a common comparator, even when the interventions of interest have never been directly compared
<ul style="list-style-type: none"> • Intervention effect estimates are calculated on the basis of direct evidence 	<ul style="list-style-type: none"> • Intervention effect estimates are informed by both direct and indirect evidence and are thus more precise
	<ul style="list-style-type: none"> • Can rank multiple treatments and calculate the possibility of each treatment being ranked 1st, 2nd, etc.
	<ul style="list-style-type: none"> • Reaches conclusive evidence prior to conventional meta-analyses

Table 2. Example league table demonstrating the relative effectiveness for each pair of comparison

	VKA	Rivaroxaban	Dabigatran	Edoxaban
VKA (vitamin K antagonist)	90.8 (93.8)	0.89 (0.78–0.98)	0.53 (0.28–0.75)	0.13 (0.10–0.19)
Rivaroxaban	1.13 (0.98–1.26)	72.3 (3.2)	1.10 (0.95–1.23)	0.90 (0.75–1.08)
Dabigatran	–	0.91 (0.78–1.07)	23.6 (2.2)	0.98 (0.85–1.18)
Edoxaban	6.54 (5.12–13.10)	–	1.08 (0.81–1.23)	12.8 (0.8)

Estimates are presented as odds ratio (OR) with 95% confidence interval (CI) in parentheses. OR > 1 suggests that the treatment listed in the upper row is superior; OR < 1 suggests that the treatment listed in the left column is superior. Statistically significant values are given in bold. The transection point (diagonal in italic script) lists the surface under the cumulative ranking curve (SUCRA) value of each treatment in the upper row and the possibility of that treatment being the best in parentheses. ORs above the transection point (right upper half) represent mixed network meta-analysis outcomes. ORs below the transection point (left lower half) represent direct meta-analysis outcomes. VKA = vitamin K antagonist.

5. *Assess heterogeneity and inconsistency.* Statistical heterogeneity is assessed using modified versions of the Cochran’s Q statistic and quantified using the I² index. Lack of transitivity may manifest statistically as differences between direct and indirect evidence. This is called inconsistency in NMA terminology and is assessed by comparing effect estimates of direct and indirect evidence. In the example of the network depicted in Fig. 2, direct evidence suggests that VKA is more effective than edoxaban in the treatment of venous thromboembolism (OR 6.54, 95% CI 5.12–13.10; Table 2). If indirect treatment estimates suggest superiority of edoxaban over VKA, it can be judged that there is considerable inconsistency between direct and indirect evidence. Inconsistency is assessed statistically (inconsistency factor: difference between direct and indirect effect estimates accompanied by 95% CIs) and visually (inconsistency plots: plotting inconsistency factors along with their uncertainty). Other popular methods of exploring inconsistency include the node splitting approach¹¹ and the design by treatment interaction model.¹²
6. *Assess the contribution of each pair of direct comparison to the mixed effects and to the entire network using the contribution plot.* The magnitude of contribution of each pair of interventions is important to assess confidence on the mixed effect estimates. For example, if effect estimates from a pair of interventions are represented by studies at high risk of bias and this pair contributes 50% to the entire network, the confidence in the entire network must be downgraded. A bar chart of study limitations allows direct visual assessment of the risk of bias across comparisons (Fig. 3).
7. *Assess the quality of evidence using Grading of Recommendations, Assessment, Development and Evaluations (GRADE) for NMA.* In addition to standard GRADE

domains, assessment of transitivity, consistency and ranking probabilities across treatments is an important part of a NMA.¹³ GRADE assessments are required for each treatment comparison separately. Analysis of limitations and judicious assignment of evidence grade will inform confidence in intervention effects and guide decision making.

From the practical point of view, executing a NMA is a laborious process compared with conventional meta-analysis. The number of studies is usually much larger. Furthermore, the complex statistics and the parameters of transitivity and inconsistency which are unique in NMA require both clinical judgment and statistical expertise. It is strongly suggested that investigators planning to undertake a NMA attend relevant courses and get a statistician experienced in NMA involved.

The validity of results of a NMA, similar to any other statistical model, rely on the plausibility of the assumptions made. NMA provides observational evidence as, although participants are randomised within studies to receive one or the other treatment, they are not randomised across treatment comparisons. There is the risk that effect modifiers are not equally balanced across treatment comparisons (violation of transitivity) and potentially biased indirect estimates may contaminate the entire network.

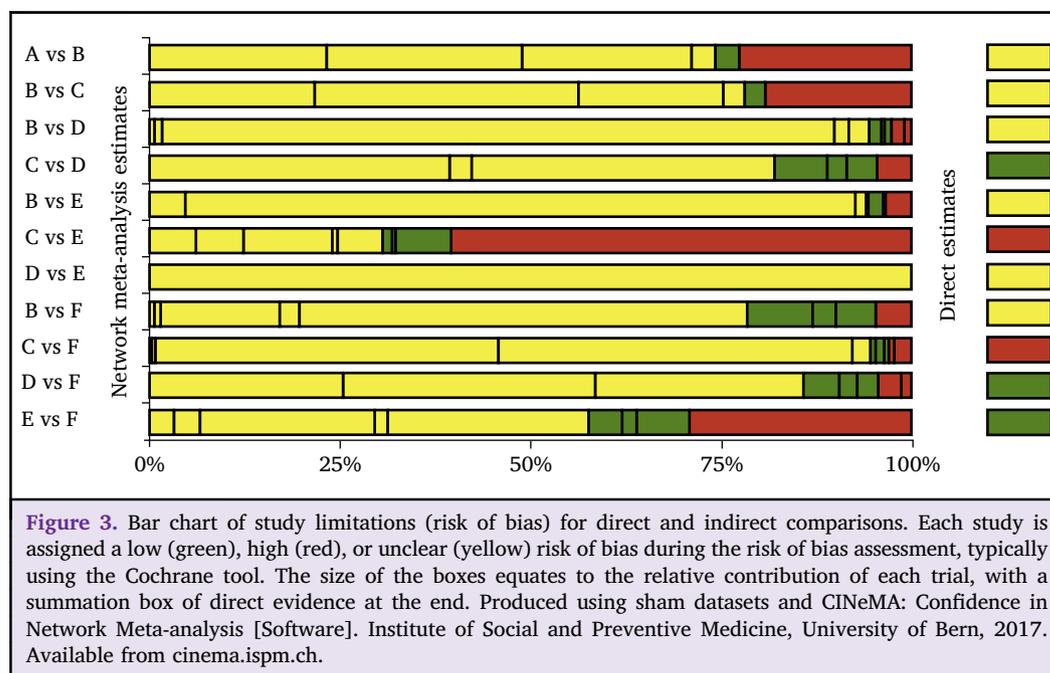
Transitivity needs to be assessed objectively under the prism of variations in study design and quality, patient population, interventions, and outcome assessment across included trials. It commonly happens that not enough studies per treatment comparison are available to explore transitivity or investigate for inconsistency with a statistical test. NMA is much more complex compared to pairwise meta-analysis and researchers often have difficulty conducting the analysis or interpreting the results. A typical mistake is that researchers often overinterpret their findings believing they produce a definite result regarding the hierarchy of interventions. Similar to relative effects, ranking probabilities are accompanied by some uncertainty and should always be evaluated along with the relative effects.

NMA is an exciting new tool in the armamentarium of clinical research and opens new horizons in evidence synthesis. Although pairwise meta-analysis can allow for reaching conclusive evidence earlier than individual RCTs, “living” NMA provides evidence of effect difference earlier than standard meta-analyses,¹⁴ thereby obviating the need

Table 3. Probability of each treatment being ranked best, 2nd, 3rd, or worse

Rank	VKA (vitamin K antagonist)	Rivaroxaban	Dabigatran	Edoxaban
Best	93.8%	3.2%	2.2%	0.8%
2nd	5.1%	50.3%	47.7%	0.8%
3rd	0.9%	43.7%	46.6%	1.8%
Worse	0.2%	2.8%	3.5%	96.6%

VKA = vitamin K antagonist.



to subject study participants to a potentially inferior or even harmful intervention.

CONFLICTS OF INTEREST

None.

FUNDING

None.

REFERENCES

- Mavridis D, Giannatsi M, Cipriani A, Salanti G. A primer on network meta-analysis with emphasis on mental health. *Evid Based Ment Health* 2015;18:40–6.
- 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. *Ann Intern Med* 2015;162:777–84.
- Xu C, Niu Y, Wu J, Gu H, Zhang C. Software and package applying for network meta-analysis: a usage-based comparative study. *J Evid Based Med* 2018;11:176–83.
- White IR. Network meta-analysis. *Stata J* 2015;15:951–85.
- Schwarzer G, Carpenter JR, Rücker G, editors. *Meta-analysis with R*. Springer International Publishing; 2015 [eBook].
- Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS One* 2013;8:e76654.
- van Valkenhoef G, Tervonen T, Zwinkels T, de Brock B, Hillege H. ADDIS: a decision support system for evidence-based medicine. *Decis Support Syst* 2013;55:459–75.
- Confidence in network meta-analysis [Software]*. Institute of Social and Preventive Medicine, University of Bern; 2017. Available from cinema.ispm.ch.
- 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;3:80–97.
- Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol* 2015;15:58.

- Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010;29:932–44.
- Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012;3:98–110.
- Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9:e99682.
- Nikolakopoulou A, Mavridis D, Furukawa TA, Cipriani A, Tricco AC, Straus SE, et al. Living network meta-analysis compared with pairwise meta-analysis in comparative effectiveness research: empirical study. *BMJ* 2018;360:k585.

Stavros A. Antoniou

Department of Surgery, School of Medicine, European University of Cyprus, Nicosia, Cyprus

Mark Koelemay

Department of Surgery, Amsterdam University Medical Centre, University of Amsterdam, The Netherlands

George A. Antoniou*

Department of Vascular and Endovascular Surgery, The Royal Oldham Hospital, Pennine Acute Hospitals NHS Trust, Manchester, UK
Division of Cardiovascular Sciences, School of Medical Sciences, University of Manchester, Manchester, UK

Dimitrios Mavridis

Department of Primary Education, School of Education, University of Ioannina, Ioannina, Greece

*Corresponding author. Surgical Offices, Phase 1, The Royal Oldham Hospital, Rochdale Road, Oldham OL1 2JH, UK.
Email-address: antoniou.ga@hotmail.com (George A. Antoniou)