



A pulmonologist's guide to understanding network meta-analysis

Claudia C. Dobler ¹, Michael E. Wilson^{1,2} and M. Hassan Murad ¹

Affiliations: ¹Evidence-Based Practice Center, Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA. ²Division of Pulmonary and Critical Care Medicine, Dept of Medicine, Mayo Clinic, Rochester, MN, USA.

Correspondence: Claudia C. Dobler, Evidence-Based Practice Center, Mayo Clinic, Rochester MN 55905, USA. E-mail: dobler.claudia@mayo.edu

 @ERSpublications

Network meta-analysis allows comparisons between interventions, even in the absence of head-to-head trials <http://ow.ly/ex1630kfB4m>

Cite this article as: Dobler CC, Wilson ME, Murad MH. A pulmonologist's guide to understanding network meta-analysis. *Eur Respir J* 2018; 52: 1800525 [<https://doi.org/10.1183/13993003.00525-2018>].

Introduction

Decision-makers require systematic reviews in which a specific question is answered by collecting and appraising relevant studies following an established protocol. Systematic reviews often include a meta-analysis, which is a method for pooling (aggregating) results from multiple trials to produce a combined estimate. Traditional meta-analyses compare two interventions. However, it is common in practice to have more than two options, and decision-makers have to choose between numerous possible interventions, some of which may or may not have head-to-head comparisons. The technique that allows such comparison is called network meta-analysis. Readers of the medical literature have become quite familiar with traditional meta-analyses, but are less familiar with network meta-analyses. In this guide, we present a simplified description of network meta-analyses for practising pulmonologists.

The basic premise

A common scenario is that drugs A and B have been compared to placebo in several trials; but no or very few trials have compared A to B directly (*i.e.* head-to-head studies). If we subtract the effect of (A *versus* placebo) from (B *versus* placebo), we can conclude the effect of A *versus* B. This effect is called an indirect estimate. If there was an actual head-to-head study that provided a direct estimate, then we can combine the direct estimate with the indirect estimate to produce a network meta-analysis estimate. Thus, a "network" of multiple interventions can be created. This network provides relative effectiveness estimates comparing all the possible permutations of the comparisons between interventions. Network meta-analysis also allows ranking of the interventions to help decision makers choose the most effective therapies [1].

Example

We are interested in determining which treatment to use for a patient with pulmonary arterial hypertension. A network meta-analysis is available and it included 31 randomised controlled trials with 6565 patients [2].

The systematic review conducted for the network meta-analysis aimed to assess the comparative efficacy and tolerability of pharmacological interventions for pulmonary arterial hypertension and searched five databases for relevant publications (between January 1, 1990 and March 3, 2016). The two main outcomes were mortality and clinical worsening. The available trials reporting each outcome can be depicted in a

Received: March 15 2018 | Accepted after revision: May 19 2018

Copyright ©ERS 2018

network graph (for example, for mortality see figure 1. In these graphs each intervention (also called node) is shown as a circle and each direct comparison as a solid line.

Mortality

Mortality was ranked lowest for riociguat (risk ratio (RR) 0.36, 95% CI 0.07–1.81; compared with placebo), followed by endothelin receptor antagonist (ERA) (RR 0.70, 95% CI 0.45–1.07; compared with placebo), and highest with selexipag (RR 1.52, 95% CI 0.86–2.70; compared with placebo). In this network, no trials directly compared ERA to selexipag (*i.e.* looking at the figure, no line connects the two interventions). However, using their effect on placebo (the common comparator), analysis produced a RR of 0.46 (95% CI 0.22–0.93); suggesting significantly reduced mortality with ERA compared to selexipag (figure 2).

Clinical worsening

Compared with a median placebo rate of 14.5%, clinical worsening was estimated in 2.8% of patients with riociguat (RR 0.19, 95% CI 0.05–0.76), in 3.9% with an ERA combined with a phosphodiesterase-5 inhibitor (PDE5i) (RR 0.27, 95% CI 0.14–0.52), in 5.7% with a PDE5i (RR 0.39, 95% CI 0.24–0.62), and in 7.7% with an ERA (RR 0.53, 95% CI 0.36–0.78). The results suggested that riociguat and ERA combined with a PDE5i were ranked as being most effective in reducing clinical deterioration, followed by PDE5i and ERA.

Assumptions

The following three assumptions have to be met for a network meta-analysis to be valid [3].

1) Homogeneity of direct evidence

Heterogeneity (unexplained differences in treatment effect) across trials decreases the confidence in the pooled estimate [4], and is often the consequence of differences in study protocols (differences in included population, duration/intensity of intervention, *etc.*). Thus, each direct comparison in a network needs to be sufficiently homogeneous. For example, the trials that are pooled to generate the direct estimate of A *versus*

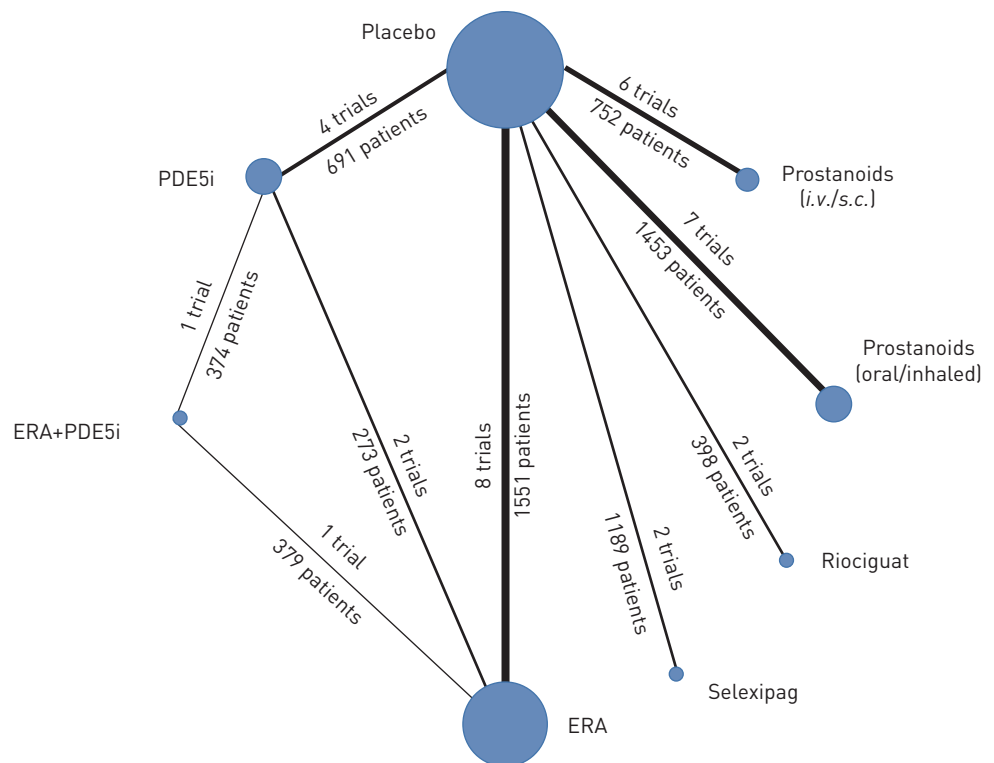


FIGURE 1 Network diagram of comparisons of treatments for the outcome “mortality”. Solid lines indicate direct comparisons between treatments. Thickness of lines is proportional to the number of available trials. Each node (circle) represents a different treatment, and its size is proportional to the number of patients in which it is directly examined. ERA: endothelin receptor antagonist; PDE5i: phosphodiesterase-5 inhibitor. Produced from data reported in the study by JAIN *et al.* [2].

Efficacy in reducing clinical worsening as risk ratio (95% CI)								
Efficacy in reducing mortality as risk ratio (95% CI)	ERA	1.37 [0.86–2.18]	1.98 [1.10–3.59]	0.82 [0.42–1.60]	2.85 [0.66–12.31]	0.71 [0.39–1.29]		0.53 [0.36–0.78]
	1.02 [0.36–2.86]	PDE5i	1.45 [0.79–2.66]	0.60 [0.29–1.22]	2.09 [0.47–9.20]	0.52 [0.27–1.00]		0.39 [0.24–0.62]
	1.08 [0.32–3.57]	1.05 [0.4–2.78]	ERA+PDE5i	0.41 [0.18–0.97]	1.43 [0.30–6.67]	0.36 [0.16–0.79]		0.27 [0.14–0.52]
	0.46 [0.22–0.93]	0.45 [0.14–1.47]	0.43 [0.11–1.67]	Selexipag	3.45 [0.77–16.67]	0.87 [0.43–1.79]		0.65 [0.38–1.12]
	1.92 [0.36–10.00]	1.89 [0.28–12.50]	1.80 [0.24–13.77]	4.22 [0.76–23.33]	Riociguat	0.25 [0.06–1.10]		0.19 [0.05–0.76]
	0.79 [0.35–1.75]	0.78 [0.22–2.70]	0.74 [0.18–3.04]	1.73 [0.71–4.20]	0.41 [0.07–2.35]	Prostanoid (oral/inhaled)		0.75 [0.47–1.19]
	1.25 [0.59–2.70]	1.23 [0.36–4.17]	1.17 [0.29–4.71]	2.75 [1.18–6.40]	0.65 [0.12–3.66]	1.59 [0.63–4.00]	Prostanoid (i.v./s.c.)	
	0.70 [0.45–1.07]	0.68 [0.24–1.95]	0.65 [0.19–2.25]	1.52 [0.86–2.70]	0.36 [0.07–1.81]	0.88 [0.45–1.73]	0.55 [0.30–1.03]	Placebo

FIGURE 2 Comparative effectiveness of pharmacological treatments for pulmonary arterial hypertension to reduce clinical worsening and mortality. Comparisons should be read from left (active agent) to right (comparator agent or placebo). Lower numbers indicate a more favourable outcome (reduction of the relative risk of clinical worsening or mortality); darker shading indicates statistical significance. ERA: endothelin receptor antagonist; PDE5i: phosphodiesterase-5 inhibitor.

placebo need to be sufficiently homogeneous among each other. The same should apply to B versus placebo, and so on.

The decision to pool estimates for pairwise comparisons should be made *a priori*. If potential heterogeneity due to certain differences in study characteristics is suspected (e.g. due to differences in length of treatment, disease severity of study participants), estimates can usually still be pooled, but subgroup analysis for these characteristics should be planned. If significant heterogeneity between trials is suspected (e.g. between studies in adults and children), estimates should not be pooled and a network meta-analysis including these trials should not be performed.

2) Transitivity, similarity or exchangeability

Patients enrolled in different trials in a network of comparisons need to be sufficiently similar, *i.e.* they were equally likely to have received any of the treatments in the network [5]. For example, patients in trials of A versus placebo need to be similar to patients in trials of B versus placebo. Another way to think of similarity is that the distribution of effect modifiers should be similar across populations included in the various comparisons. Therefore, network meta-analysis is rarely done using observational studies because the transitivity assumption is unlikely to be fulfilled in observational studies because these studies by definition will have different protocols, methods of patient selection and settings.

For randomised controlled trials, the transitivity assumption may, for example, not be fulfilled when comparing trials of pharmacological and non-pharmacological interventions. Patients who received pharmacological treatment might not have been equally likely to receive, for example, an exercise intervention due to their physical condition. In these cases, network meta-analysis should not be performed.

Outcomes also have to measure the same or a similar construct in trials included in the network meta-analysis. The use of different measurement scales for an outcome does not preclude network meta-analysis, as long as they measure the same construct. For example, clinical worsening in the pulmonary arterial hypertension network meta-analysis was a composite of death, pulmonary arterial hypertension-related hospitalisation, lung transplantation, atrial septostomy, initiation of rescue therapy and deterioration of functional class or worsening of 6-min walk distance. There were some variations across trials in the way this composite outcome was determined. However, the authors of the network meta-analysis judged these differences in definitions to be not important and thought that all of them fell under the same construct, of clinical worsening. As in traditional meta-analysis, not all trials in a network meta-analysis need to include all outcomes of interest.

3) Consistency between direct and indirect evidence

The network is considered consistent if the pooled effect size for a direct comparison is similar to the pooled effect size from indirect comparison; otherwise, the network is inconsistent or incoherent [6, 7]. Causes for inconsistency can include differences in patients, treatments and study setting, and also differences in the study methodological quality (*i.e.* risk of bias).

Consistency between direct and indirect evidence can only be explored after network meta-analysis has been performed. Effect estimates from direct and indirect comparisons need to be calculated separately. If there are significant differences between the two, it is inappropriate to make conclusions based on the combined effect size. However, exploring reasons for differences in trials that informed direct and indirect comparisons in these instances could reveal important information, *e.g.* longer duration of treatment conveys more benefits than shorter treatment.

Analysis and results

Network meta-analyses rank interventions in terms of efficacy and report these rankings as a rank (*i.e.* drug A ranks first as most effective), a probability (drug A has 80% probability of being most effective) or as a surface under the cumulative ranking area, or SUCRA (*e.g.* SUCRA for drug A is 80%). For each treatment the probability that it has a certain rank can be derived from the distributions of all treatment effects based on analysis of combined direct and indirect evidence. The treatments can then be ranked by SUCRA [8]. Network meta-analyses also present “pairwise” results (*i.e.* comparison of each possible permutation between every drug and all others in the network). These comparisons are commonly presented as odds ratios or relative risks in a table that is called a staircase or league table (figure 2). These estimates are ideally presented as direct, indirect and network meta-analysis (combined) estimates to allow readers determine consistency and quality of the evidence. The analysis is done either following a Bayesian or frequentist approach. The choice between the two approaches depends on the training and experience of the statistical team conducting the network meta-analysis. No one approach of the two is preferred [3]. For the most part, the two approaches yield similar results. The exception to this is when the Bayesian analysis includes informative priors, which means that assumptions are made about the magnitude of heterogeneity or of the network meta-analysis effect size using outside data (data from other studies or literature aside from the trials being analysed in the network meta-analysis).

How to read and apply network meta-analysis results

Readers of network meta-analysis should first verify that a rigorous systematic review has been performed before trusting estimates reported in the analysis. Readers should then look at risk of bias tables in the review to get an overall sense of how well performed the individual studies were. Next, they should scrutinise the three assumptions and make their judgement about whether these assumptions have been met [5]. If so, they can consider whether the rankings make clinical sense by evaluating each individual pairwise comparison and determine whether important differences exist between interventions. Readers should also determine the network has sufficient direct evidence from head-to-head comparisons (*versus* being solely based on indirect comparisons). If only indirect evidence is available for comparison of two interventions, results must be interpreted with caution. Sparse direct evidence can make the analysis unstable and less reliable. Sophisticated frameworks such as GRADE (Grading of Recommendations, Assessment, Development and Evaluation) allow a formal evaluation of the quality of evidence (also called certainty in estimates) [9, 10].

Limitations of network meta-analysis

The results of network meta-analyses that do not meet the assumptions of homogeneity of direct evidence and/or transitivity (*e.g.* because of different target populations among included studies) can be potentially misleading.

Ranking probabilities need to be interpreted with caution and they can be misleading. For example, a drug may be ranked most effective but the difference in efficacy between this drug and the second-ranked drug could be trivial. Such probabilities are difficult to interpret clinically, and should be interpreted in the context of the magnitude of difference between each pair of interventions being compared.

The evidence from indirect comparisons is often not reported separately in network meta-analyses. A systematic review of 121 published network meta-analyses found that 89% of studies did not report indirect evidence [11]. It is important to check direct *versus* indirect evidence for each comparison, rather than depending on global tests of inconsistency [3].

36% of network meta-analyses, especially industry-sponsored studies, have been found not to present the full matrix of head-to-head treatment comparisons [12]. Readers of network meta-analyses need to be aware that such studies may not include all treatment comparisons for a certain condition, *e.g.* when there is a narrow research question.

Current and future applications of network meta-analysis in respiratory medicine

The major application of network meta-analysis is for the evaluation of comparative effectiveness of interventions that have not been assessed in head-to-head trials, *e.g.* different combinations of long-acting

inhaled therapy (beta-agonists, anticholinergics and steroids) for chronic obstructive pulmonary disease [13]. Network meta-analysis has also been used in situations where the duration of an intervention varies, but data on different treatment length are not available from direct comparisons, e.g. isoniazid for preventive tuberculosis treatment for 6, 9 or 12 months [14]. There is a lack of comparative effectiveness studies that have compared pharmacological and non-pharmacological treatments in a wide spectrum of respiratory and sleep disorders, where network meta-analysis can potentially provide some indirect evidence.

Network meta-analysis can be used to calculate the required sample size for trials and to identify missing evidence on important comparisons. It can be expected that network meta-analysis will increasingly play a role in informing research priorities (which treatments should be compared against each other) [15, 16]. Network meta-analysis will also become increasingly important to inform policy-makers who often need to make decisions about interventions for clinical conditions for which a number of interventions are available, but only few have been studied in head-to-head comparisons.

Conflict of interest: None declared.

Support statement: This work was supported by the Department of Health, Australian Government (National Health and Medical Research Council fellowship for C.C. Dobler; grant number APP1123733). Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 Benkhadra K, Wang Z, Murad MH. Network meta-analysis: introduction and an example that compares devices for PFO closure. *Eur Heart J* 2015; 36: 80–82.
- 2 Jain S, Khera R, Girotra S, et al. Comparative effectiveness of pharmacologic interventions for pulmonary arterial hypertension: a systematic review and network meta-analysis. *Chest* 2017; 151: 90–105.
- 3 Morton SC, Murad MH, O'Connor E, et al. Quantitative Synthesis – An Update. *Methods Guide for Comparative Effectiveness Reviews*. Rockville, Agency for Healthcare Research and Quality, 2018. Available from: <https://doi.org/10.23970/AHRQEPCCMETHGUIDE3>
- 4 Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence – inconsistency. *J Clin Epidemiol* 2011; 64: 1294–1302.
- 5 Mills EJ, Ioannidis JP, Thorlund K, et al. How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA* 2012; 308: 1246–1253.
- 6 Higgins JP, Jackson D, Barrett JK, et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012; 3: 98–110.
- 7 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.
- 8 Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; 64: 163–171.
- 9 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.
- 10 Puhan MA, Schunemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014; 349: g5630.
- 11 Bafeta A, Trinquart L, Seror R, et al. Reporting of results from network meta-analyses: methodological systematic review. *BMJ* 2014; 348: g1741.
- 12 Chambers JD, Naci H, Wouters OJ, et al. An assessment of the methodological quality of published network meta-analyses: a systematic review. *PLoS ONE* 2015; 10: e0121715.
- 13 Kew KM, Dias S, Cates CJ. Long-acting inhaled therapy (beta-agonists, anticholinergics and steroids) for COPD: a network meta-analysis. *Cochrane Database Syst Rev* 2014; 3: CD010844.
- 14 Zenner D, Beer N, Harris RJ, et al. Treatment of latent tuberculosis infection: an updated network meta-analysis. *Ann Intern Med* 2017; 167: 248–255.
- 15 Nikolakopoulou A, Mavridis D, Salanti G. Using conditional power of network meta-analysis (NMA) to inform the design of future clinical trials. *Biom J* 2014; 56: 973–990.
- 16 Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *BMJ* 2013; 346: f2914.