Research Techniques Made Simple: Network Meta-Analysis

Jennifer Watt, Andrea C. Tricco, Sharon Straus, Areti Angeliki Veroniki, Gary Naglie, and Aaron M. Drucker

Version  Post-print/Accepted Manuscript


Copyright/License  This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

How to cite TSpace items

Always cite the published version, so the author(s) will receive recognition through services that track citation counts, e.g. Scopus. If you need to cite the page number of the author manuscript from TSpace because you cannot access the published version, then cite the TSpace version in addition to the published version using the permanent URI (handle) found on the record page.

This article was made openly accessible by U of T Faculty. Please tell us how this access benefits you. Your story matters.
Research Techniques Made Simple: Network Meta-Analysis

Jennifer Watt MD, PhD Candidate\textsuperscript{1,2,3} \hspace{1cm} jennifer.watt@mail.utoronto.ca
Andrea C. Tricco PhD, Associate Professor\textsuperscript{3,4} \hspace{1cm} triccoa@smh.ca
Sharon Straus MD, MSc, Professor\textsuperscript{1,2,3} \hspace{1cm} sharon.straus@utoronto.ca
Areti Angeliki Veroniki PhD, Scientist\textsuperscript{3,5,6} \hspace{1cm} averonik@cc.uoi.gr
Gary Naglie MD, Professor\textsuperscript{1,2,7} \hspace{1cm} gnaglie@baycrest.org
Aaron M. Drucker MD, ScM, Assistant Professor\textsuperscript{2,8,9*} \hspace{1cm} aaron.drucker@wchospital.ca

1. Division of Geriatric Medicine, Department of Medicine, University of Toronto, Toronto, Canada
2. Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada
3. Li Ka Shing Knowledge Institute, St. Michael’s Hospital, Toronto, Canada
4. Dalla Lana Faculty of Public Health, University of Toronto, Toronto, Canada
5. Department of Primary Education, School of Education, University of Ioannina, Ioannina, Greece
6. Institute of Reproductive and Developmental Biology, Department of Surgery & Cancer, Faculty of Medicine, Imperial College, London, England
7. Baycrest Health Sciences, Toronto, Canada
8. Division of Dermatology, Department of Medicine, Women’s College Hospital, Toronto, Canada
9. Women’s College Research Institute, Women’s College Hospital, Toronto, Canada

*Corresponding Author:
Dr. Aaron Drucker
6\textsuperscript{th} Floor, Women’s College Hospital
76 Grenville Street
Toronto, ON M5S1B2
Canada
aaron.drucker@wchospital.ca
Word count: 197 (abstract), 1890 (body); 6 figures, 4 tables
Abstract

When making treatment decisions, it is often necessary to consider the relative efficacy and safety of multiple potential interventions. Unlike traditional pairwise meta-analysis, which allows for a comparison between two interventions by pooling head-to-head data, network meta-analysis (NMA) allows for the simultaneous comparison of more than two interventions and for comparisons to be made between interventions that have not been directly compared in a randomized controlled trial (RCT). Given these advantages, NMAs are being published in the medical literature with increasing frequency. However, there are important assumptions that researchers and knowledge users (e.g. patients, clinicians, and policy-makers) must consider when conducting and evaluating a NMA: network connectivity, homogeneity, transitivity, and consistency. There are also multiple NMA outputs that researchers and knowledge users should familiarize themselves with in order to understand NMA results (e.g. network plots, mean ranks).

Our goals in this article are to: (1) demonstrate how NMAs differ from pairwise meta-analyses, (2) describe types of evidence in a NMA, (3) explain NMA model assumptions, (4) provide readers with an approach to interpreting a NMA, (5) discuss areas of ongoing methodological research, and (6) provide a brief overview of how to conduct a systematic review and NMA.

Key words: meta-analysis, network meta-analysis, systematic review
Introduction

A growing number of network meta-analyses (NMAs) are being published in the medical literature (Zarin et al., 2017). NMAs offer a way to make comparisons between many interventions simultaneously, helping to synthesize large amounts of data relating to clinical outcomes. NMAs can also make indirect comparisons between interventions that have not been compared in randomized controlled trials (RCTs) and rank interventions in terms of their relative efficacy or safety. While there are clear advantages to NMAs, their conduct and interpretation is more complex than that of pairwise meta-analyses. Therefore, it is important for those conducting and reading NMAs to learn how to understand and interpret the findings. In this article, we will: (1) delineate how NMAs differ from pairwise meta-analyses, (2) describe types of evidence in a NMA (3) explain NMA model assumptions, (4) provide readers with an approach to interpreting an NMA, (5) discuss areas of ongoing methodological research, and (6) provide a brief overview of how to conduct a systematic review and NMA. Two NMAs on treatments for psoriasis will be used to illustrate these concepts (Jabbar-Lopez et al., 2017, Reich et al., 2012).

1. Comparing Pairwise and Network Meta-Analysis

Pairwise and network meta-analyses are compared and contrasted in Table 1. Pairwise meta-analyses are applied when the desired endpoint is to derive a summary effect estimate across a number of studies that compare the same two interventions (Panel A, Figure 1) (Abuabara et al., 2012). However, for many comparative effectiveness questions, the goal is to understand the relative efficacy and safety of more than two interventions. For example, therapeutic decision making for a patient with moderate to severe chronic plaque psoriasis requires comparison of all possible interventions, including adalimumab, etanercept, other biologics, traditional systemic
medications, and small molecule targeted agents. This can be accomplished with NMA, from which summary effect estimates can be derived across more than two interventions, some of which have never been directly compared. Like pairwise meta-analyses, NMAs can be conducted in a frequentist or Bayesian framework (Chaimani et al., 2013, Dias et al., 2018, van Valkenhoef and Kuiper, 2016).

2. Direct and Indirect Evidence

Estimates of relative efficacy or safety from NMA models can be derived by combining both direct and indirect evidence from intervention comparisons that form a connected network (Figure 2) (see Assumptions of Network Meta-Analysis below). Direct evidence describes data taken from at least one RCT. Indirect evidence is derived from NMA models to describe the relative efficacy or safety for intervention comparisons that have not been studied in an RCT (Panel B, Figure 1). When a comparison is informed by both direct and indirect evidence, this is referred to as a mixed effect estimate (Dias et al., 2018). For example, in the NMA conducted by Jabbar-Lopez et al., on the evaluation of biologic therapies for psoriasis, there was no RCT evidence directly comparing adalimumab and etanercept for the outcome of ‘clear/nearly clear’; however, there were direct comparisons between (1) adalimumab and placebo and (2) etanercept and placebo. Authors were able to derive an indirect effect estimate comparing adalimumab and etanercept because each intervention had been compared to a common intervention (placebo) (Figure 2) (Jabbar-Lopez et al., 2017).

3. Assumptions of Network Meta-Analysis

There are 4 key assumptions of NMAs: (a) network connectivity (b) homogeneity, (c) transitivity, and (d) consistency (Table 2). The requirement for network connectivity is unique to
NMA. Interventions must be connected to the network to draw any conclusions about their direct and indirect relationships with other interventions. In Figure 2, each intervention is connected to at least one other intervention in each network. If a treatment comparison is not connected to any other treatments in the network, it cannot be a part of the NMA.

Readers are likely familiar with the concept of homogeneity: the true intervention effect should be sufficiently similar across all studies making a direct comparison between the same two intervention groups. Similar to pairwise meta-analyses, different potential sources of heterogeneity must be considered in studies included in NMAs: clinical, methodological, and statistical. If heterogeneity is anticipated between studies, then a random-effects as opposed to fixed-effects model should be implemented (Higgins and Green).

The assumptions of transitivity and consistency refer to our assessment of potential clinical and methodological effect modifiers across a network of interventions. In assessing transitivity, a judgment must be made about the distribution of effect modifiers and how they might influence direct and indirect effect estimates. For example, if all patients in one psoriasis intervention comparison have severe disease at baseline (Interventions 1 vs. 2) while all patients in the other two treatment comparisons in a loop have moderate disease at baseline (interventions 1 vs. 3 and 2 vs. 3), this violates the transitivity assumption. When there are imbalances in effect modifiers across the network, subgroup analyses or meta-regression could be used to explore their influence on NMA effect estimates, or perhaps the NMA should not be conducted.

Consistency is the statistical measure of transitivity. There may be inconsistency in a closed network loop if there is an imbalance of effect modifiers across treatment comparisons. In essence, direct and indirect effect estimates can be compared within a network to assess their
level of disagreement. There are tests that assess for consistency in a network as a whole (global tests) or at certain paths (e.g. closed loops) of a network (local tests) (Dias et al., 2018). For example, the results of a loop-specific approach to the assessment of inconsistency (local test) are presented in Figure 4. There is inconsistency in the closed loop containing three comparisons: placebo-methotrexate, placebo-infliximab, and methotrexate-infliximab. This means that the direct and indirect effect estimates of one of the treatment comparisons within this closed loop are significantly different from one another (the inconsistency factor’s 95% confidence interval does not cross zero). There is no inconsistency identified in the other closed loops. It is possible that statistical tests of consistency may fail to identify inconsistency; therefore, it is important to consider whether the transitivity assumption has been met prior to undertaking a NMA.

RCTs in a NMA are subject to the same biases as those included in pairwise meta-analyses. Critical appraisal of RCTS in a NMA is important because studies at high risk of bias can lead to violations of the homogeneity, transitivity and consistency assumptions. For example, if indirect evidence from a closed network loop of studies at low risk of bias in all aspects of critical appraisal did not show a significant benefit to receiving treatment, but one study (direct evidence), at high risk of bias from lack of participant and outcome assessor blinding, found a benefit to receiving treatment, this will violate the transitivity (and possibly the consistency) assumption. Similarly, between-study heterogeneity will be created if one study at high risk of bias due to lack of participant and outcome assessor blinding found a benefit to receiving a treatment, while a second study that was at low risk of bias on these aspects of critical appraisal did not find such a benefit.
4. Interpreting a Network Meta-Analysis

A number of different measures of intervention efficacy and safety can be derived from NMAs (Table 3) (Dias et al., 2018). Figures and explanations for network plots (Figure 2), surface under the cumulative ranking (SUCRA) curves (Figure 3), an inconsistency plot (Figure 4), and a comparison-adjusted funnel plot (Figure 5) are provided (Jabbar-Lopez et al., 2017). By convention, a higher mean rank or greater SUCRA value indicates that an intervention is either more efficacious or safer (Dias et al., 2018). While most people are familiar with the interpretation of a frequentist effect estimate, people may be less familiar with the interpretation of a Bayesian effect estimate. Reich et al. reported the mean relative risk (RR) (and 95% credible interval [(CrI]) of 50%, 75%, and 90% reductions in the Psoriasis Area and Severity Index (PASI) for patients with moderate to severe psoriasis receiving biologics (Reich et al., 2012). In this case, the RR value represents the mean of the RR posterior distribution for each relative treatment effect, and the 95% CrI represents the range of values within which there is a 95% probability that the true value of the RR is found, given the observed data. In contrast, Jabbar-Lopez et al. used a frequentist NMA approach (Jabbar-Lopez et al., 2017). In a frequentist framework, the 95% confidence interval means that there is a 95% chance of the true RR value being found within the intervals, given repeated randomized sampling. Frequentist modeling treats data as random and parameters as fixed unknown constants; whereas, Bayesian modeling treats data as fixed and parameters as random (Kadane, 1995).

Knowledge users can use the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) tool for interpreting NMAs in health care decision-making or the Journal of the American Medical Association (JAMA) Users’ Guide to the Medical Literature on NMAs for
interpreting and critically appraising a systematic review and NMA (Jansen et al., 2014, Mills et al., 2012). The GRADE (Grades of Recommendation, Assessment, Development and Evaluation) approach has also been extended to assess the certainty of NMA results. It provides a framework for determining the quality of evidence in NMA-derived effect estimates for each outcome (Brignardello-Petersen et al., 2018, Salanti et al., 2014).

5. Areas of Ongoing Methodological Research in Network Meta-Analysis

There remain a number of questions about how to apply NMA methods in clinical and policy decision-making. For example, what is the best way to present NMA results to knowledge users? In addition to reporting summary effect estimates, is it best to report all of the SUCRA curve values individually or should a method like a rank-heat plot be utilized (Veroniki et al., 2016b)? A rank-heat plot is a collated graphical representation of ranking statistics demonstrating the comparative effect of interventions on a number of outcomes (Figure 6). How can data from non-randomized studies be incorporated into NMAs? For adverse event data, in particular, this is an important topic because many RCTs are under-powered to detect the potential for harm. Several models have been proposed to include non-randomized studies in NMAs: (1) naïve pooling, (2) data from non-randomized studies as prior information, and (3) a three-level hierarchical model with an additional level of uncertainty to account for the inclusion of different study designs (Schmitz et al., 2013). Lastly, how can individual patient-level data best be included in NMAs to account for potential effect modifiers? Meta-analysts are using several methods to incorporate individual patient-level data, including one- and two-stage Bayesian hierarchical NMA models (Veroniki et al., 2016a).
6. Conducting a Systematic Review and Network Meta-Analysis

We provide an overview of the steps necessary to conduct a systematic review and network meta-analysis in Table 4. There are statistical packages available to conduct frequentist and Bayesian NMAs (Chaimani et al., 2013, van Valkenhoef and Kuiper, 2016). In conducting a Bayesian NMA, special consideration needs to be given to the choice of prior information for stochastic model parameters (Dias et al., 2018). Reich et al. implemented vague prior distributions for study-specific baselines in their NMA of biologic treatments for moderate to severe psoriasis, but minimally informative and informative priors are also used in Bayesian NMAs (Dias et al., 2018, Reich et al., 2012).

Summary

Researchers may wish to undertake a systematic review and NMA because they can: make indirect comparisons between interventions that have not been previously compared in RCTs, compare the relative efficacy or safety of more than two interventions simultaneously, and rank interventions in terms of their relative efficacy or safety. Much work has been done to improve the reporting and interpretability of NMA results; however, researchers and knowledge users must be cautious when reading NMA results and carefully consider many of the same limitations that face pairwise meta-analyses including potential threats to the validity of meta-analytic findings from systematic biases.

Summary Points

Comparing Pairwise and Network Meta-Analysis
• Pairwise meta-analyses allow evidence comparing two interventions to be synthesized; network meta-analyses (NMAs) are used to compare more than two interventions – some of which have not been directly compared in previous randomized controlled trials.

• NMAs can be used to rank interventions in terms of their relative efficacy or safety.

Limitations

• Assumptions underlying NMAs must be carefully considered, such as transitivity and consistency, because if these assumptions are not met, it may jeopardize the conclusions of NMA.

• RCTs included in a NMA are subject to the same biases as those included in pairwise meta-analyses and critical appraisal remains an important component of a well-conducted systematic review and NMA.

Conflict of interest statement: All authors have no conflicts of interest to declare related to this work.

Other disclosures: Dr. Drucker served as an investigator and has received research funding from Sanofi and Regeneron and has been a consultant for Sanofi, RTI Health Solutions, Eczema Society of Canada and Canadian Agency for Drugs and Technology in Health. He has received honoraria from Astellas Canada, Prime Inc, Spire Learning, CME Outfitters and Eczema Society of Canada. Dr. Watt is funded by a doctoral research award from the Canadian Institutes of Health Research and the University of Toronto Department of Medicine Eliot Phillipson Clinician Scientist Training Program. Dr. Tricco is funded by a Tier 2 Canada Research Chair in Knowledge Synthesis. Dr. Straus is funded by a Tier 1 Canada Research Chair in Knowledge Translation.

Author Roles: Dr. Watt drafted the manuscript. Drs. Watt, Tricco, Straus, Veroniki, Naglie and Drucker contributed to the conception, design, and critical revision of the manuscript, and approved the final manuscript.
# Tables

## Table 1. Comparing Pairwise and Network Meta-Analysis

<table>
<thead>
<tr>
<th></th>
<th><strong>Pairwise Meta-Analysis</strong></th>
<th><strong>Network Meta-Analysis</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Comparators</strong></td>
<td>2</td>
<td>&gt;2</td>
</tr>
<tr>
<td><strong>Questions Answered by Analysis Method</strong></td>
<td>(1) What is the efficacy or risk of harm associated with one intervention compared to another?</td>
<td>(1) Which interventions are efficacious and/or safe? (2) What intervention is the most efficacious and/or safe? (3) What is the comparative efficacy and/or safety between two interventions that haven’t been directly compared?</td>
</tr>
<tr>
<td><strong>Systematic Review Question Format</strong></td>
<td>PICO*</td>
<td>Modified PICO* to accommodate additional treatment comparisons</td>
</tr>
<tr>
<td><strong>Risk of Bias Appraisal</strong></td>
<td>Cochrane Risk of Bias Tool for RCTs</td>
<td>Cochrane Risk of Bias Tool for RCTs</td>
</tr>
<tr>
<td><strong>Assumptions</strong></td>
<td>(1) Homogeneity</td>
<td>(1) Network connectivity (2) Homogeneity (3) Transitivity (4) Consistency</td>
</tr>
<tr>
<td><strong>Influential Biases</strong></td>
<td>(1) Publication bias and small-study effects (2) Confounding (3) Selection bias (4) Information bias</td>
<td>(1) Publication bias and small-study effects (2) Confounding (3) Selection bias (4) Information bias</td>
</tr>
<tr>
<td><strong>Model Outputs</strong></td>
<td>(1) Summary effect estimates (e.g. OR, MD, SMD) and forest plot (2) Funnel plot</td>
<td>(1) Network plot (2) Transitivity plot or table (3) Summary effect estimates (e.g. OR, MD, SMD) and forest plot (4) Ranking statistic: mean rank, SUCRA value or P-score (5) Inconsistency plot (6) Comparison-adjusted funnel plot</td>
</tr>
<tr>
<td><strong>Limitations</strong></td>
<td>(1) Effect modifiers create heterogeneity (2) Biases can generate misleading results</td>
<td>(1) Effect modifiers create heterogeneity and/or inconsistency (2) Biases can generate misleading results</td>
</tr>
<tr>
<td><strong>Reporting Guidelines</strong></td>
<td>PRISMA</td>
<td>PRISMA-NMA</td>
</tr>
</tbody>
</table>
*Abbreviations: odds ratio (OR), mean difference (MD), standardized mean difference (SMD), surface under the cumulative ranking curve (SUCRA), preferred reporting guidelines for systematic reviews and meta-analyses (PRISMA), randomized controlled trial (RCT); PICO=population, intervention(s), comparator(s), outcome(s)
Table 2. Questions to Consider When Assessing the Assumptions of a Network Meta-Analysis

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Questions to Consider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogeneity</td>
<td>(1) Is there any clinical, methodological, or statistical heterogeneity between studies that compare the same interventions? (2) Are there effect modifiers (e.g. age, gender, illness severity) between studies making the same treatment comparison that could influence the summary effect estimate?</td>
</tr>
<tr>
<td>Network Connectivity</td>
<td>(1) Do all of the interventions form a connected network (as in Figure 2)?</td>
</tr>
<tr>
<td>Transitivity</td>
<td>(1) Is there an imbalance in effect modifiers among studies included in the network? (2) In theory, could any patient randomized in one study within a network have been randomized to any of the other studies in this same network?</td>
</tr>
<tr>
<td>Consistency</td>
<td>(1) Where possible to assess, are the direct and indirect effect estimates from closed loops in the network in agreement?</td>
</tr>
</tbody>
</table>
Table 3. Commonly Reported Network Meta-Analysis Outputs

<table>
<thead>
<tr>
<th>Network Meta-Analysis Output</th>
<th>Description</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Plot</td>
<td>a diagram depicting how interventions (nodes) are connected to one another through direct comparisons (lines) (see Figure 2)</td>
<td>provides an overview of the available evidence; a network estimate of an intervention's relative efficacy or safety compared to other interventions in the network can only be calculated if it is connected to the network</td>
</tr>
<tr>
<td>Transitivity Plot or Table</td>
<td>a table or plot summarizing potential effect modifiers across studies</td>
<td>studies in each network should appear sufficiently similar so that the observed treatment effects are the result of receiving each treatment and not an imbalance in effect modifiers</td>
</tr>
<tr>
<td>Summary Effect Estimate</td>
<td>estimate of the relative efficacy of interventions in the network (e.g. OR, MD, SMD, HR) compared to other network interventions, reported with a measure of uncertainty (e.g. confidence/credible intervals or predictive intervals)</td>
<td>same interpretation as a summary effect estimate in a pairwise meta-analysis</td>
</tr>
<tr>
<td>Ranking Statistics</td>
<td>frequently presented as a mean/median rank, SUCRA value (or P-Score) or probability of being the best treatment</td>
<td>an intervention with a higher treatment ranking, SUCRA value, or probability of being the best is more efficacious or more likely to cause harm</td>
</tr>
<tr>
<td>Inconsistency Plot</td>
<td>a plot reporting the inconsistency factors (absolute difference between direct and indirect effect estimates) for each comparison in a closed network loop (see Figure 4)</td>
<td>an inconsistency factor with a confidence interval that does not include zero indicates that there is significant inconsistency between direct and indirect effect estimates</td>
</tr>
<tr>
<td>Comparison-Adjusted Funnel Plot</td>
<td>similar to a funnel plot in pairwise meta-analyses; however, (1) the x-axis is the difference between each study-specific effect estimate and pooled effect estimate for each comparison and (2) comparisons have been ordered in a meaningful way (e.g. asymmetry in the plot indicates publication bias/small-study effects</td>
<td></td>
</tr>
</tbody>
</table>
chronological treatment order
(see Figure 5)

*Abbreviations: odds ratio (OR), mean difference (MD), standardized mean difference (SMD), hazard ratio (HR), network meta-analysis (NMA), SUCRA (surface under the cumulative ranking curve)
Table 4. Conducting a Systematic Review and Network Meta-Analysis

<table>
<thead>
<tr>
<th>Steps to follow when conducting a systematic review and network meta-analysis:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Follow a modified PICO (population, interventions, comparators, outcome(s)) format when developing clinical questions for systematic reviews and NMAs because you are considering multiple intervention and comparator groups.</td>
</tr>
<tr>
<td>2. Register your systematic review and NMA protocol with PROSPERO (International Prospective Register of Systematic Reviews) and consider publishing the protocol in a peer-reviewed journal.</td>
</tr>
<tr>
<td>3. Develop a comprehensive literature search strategy that will encompass all of the interventions and outcomes of interest.</td>
</tr>
<tr>
<td>4. Complete all steps relating to article screening, data abstraction, and risk of bias appraisal independently in duplicate.</td>
</tr>
<tr>
<td>5. Inspect network plots to ensure all interventions form a connected network.</td>
</tr>
<tr>
<td>6. Make judgments concerning the homogeneity and transitivity assumptions prior to conducting NMA. Be explicit about how you model heterogeneity in your NMA if you implement a random-effects model.</td>
</tr>
<tr>
<td>7. Describe any assessments of global and local inconsistency. If there is inconsistency in your NMA, state how this is addressed.</td>
</tr>
<tr>
<td>8. Assess for small-study effects and publication bias by using a plot such as the comparison-adjusted funnel plot.</td>
</tr>
<tr>
<td>9. Present summary effect estimates for interventions and an estimate of heterogeneity. You can also present ranking statistics such as a mean rank and a SUCRA value for each intervention.</td>
</tr>
<tr>
<td>10. Follow the recommendations of the PRISMA extension statement for the reporting of NMAs when submitting your systematic review and NMA for publication (Hutton et al., 2015).</td>
</tr>
</tbody>
</table>
Figures

Figure 1. Illustration of Intervention Comparisons in Pairwise and Network Meta-Analysis

Panel A demonstrates a pairwise comparison between interventions 1 and 2. Panel B demonstrates two direct comparisons (intervention 1 vs. 2 and intervention 2 vs. 3) and one indirect comparison (intervention 1 vs. 3) in a network meta-analysis. Panel C demonstrates 3 direct comparisons (intervention 1 vs. 2, intervention 2 vs. 3, and intervention 1 vs. 3) that form a closed loop.

Figure 2. Examples of Network Plots

Connected network plots (Jabbar-Lopez et al., 2017). Nodes represent individual interventions and nodes connected by lines indicate that these two interventions have previously been directly compared in a study. In these examples, the nodes are weighted by the number of studies evaluating this treatment and the lines are weighted by the number of studies evaluating this treatment comparison. Each panel is a network plot of interventions reporting the outcome of interest: (a) clear/nearly clear (minimal residual activity/PASI > 90/0 or 1 on PGA), (b) mean change in the dermatology life quality index, and (c) withdrawal due to adverse events. Abbreviations: adalimumab (ADA), etanercept (ETA), infliximab (INF), ixekizumab (IXE), methotrexate (MTX), psoriasis area and severity index (PASI), placebo (PBO), physician’s global assessment (PGA), secukinumab (SEC), ustekinumab (UST).

Figure 3. Examples of Surface under the Cumulative Ranking (SUCRA) Curves

Surface under the cumulative ranking (SUCRA) curves of treatments evaluating the psoriasis area and severity index (PASI) 75 at 12/16 weeks (Jabbar-Lopez et al., 2017). The cumulative probability that each treatment is ranked among the top \( n \) (e.g. 1, 2, ..., 8) treatments (y-axis) is plotted against each possible rank (x-axis) for treatments in the network. Predictive probabilities incorporate the uncertainty in our network estimates from heterogeneity. Ixekizumab (IXE) has the highest SUCRA value (96.4%) and placebo (PBO) has the lowest SUCRA value (0%). Abbreviations: adalimumab (ADA), etanercept (ETA), infliximab (INF), ixekizumab (IXE), methotrexate (MTX), placebo (PBO), secukinumab (SEC), ustekinumab (UST).

Figure 4. Example of an Inconsistency Plot

This is an example of an inconsistency plot with closed triangular loops of treatment comparisons evaluating the psoriasis area and severity index (PASI) 75 at 12/16 weeks (Jabbar-Lopez et al., 2017). The x-axis represents the scale for the inconsistency factors (IFs). The PBO-INF-MTX loop shows evidence of inconsistency between direct and indirect evidence because the 95% CI for the IF does not include zero. There is no significant inconsistency identified in any of the other loops. Abbreviations: adalimumab (ADA), etanercept (ETA), infliximab (INF), ixekizumab (IXE), methotrexate (MTX), placebo (PBO), secukinumab (SEC), ustekinumab (UST), inconsistency factor (IF), confidence interval (CI).
Figure 5. Example of a Comparison-Adjusted Funnel Plot

This is an example of a comparison-adjusted funnel plot of treatment comparisons evaluating the psoriasis area and severity index (PASI) 75 at 12/16 weeks (Jabbar-Lopez et al., 2017). Comparisons are color-coded as per the legend at the bottom of the figure. The y-axis represents the standard error of each study-specific effect estimate. The x-axis represents the difference between the ln(odds ratio) for each study-specific effect estimate and the pooled effect estimate for each comparison (e.g. all of the study-specific estimates reporting on the PBO vs ADA comparison). The blue diagonal line represents a linear regression of the x-axis variable on the y-axis variable. The paucity of studies in the bottom left of the plot indicates there may be small studies missing that would have favored established treatments. Abbreviations: adalimumab (ADA), etanercept (ETA), infliximab (INF), ixekizumab (IXE), methotrexate (MTX), placebo (PBO), secukinumab (SEC), ustekinumab (UST).

Figure 6. Example of a Rank-Heat Plot

This is an example of a rank-heat plot of outcomes associated with insulin use in patients with type 1 diabetes mellitus. Each ring represents a different outcome. Outcomes are also specified in the legend. Each “slice” represents a different treatment. Treatments are ranked according to their surface under the cumulative ranking curve (SUCRA) values. Higher SUCRA values (in green) indicate more efficacious and safer treatments. Uncolored areas indicate that the treatment was not included in the NMA of that outcome. Abbreviations: hemoglobin A1c (A1c), twice daily (bid), once daily (OD), four times per day (qid).
Multiple Choice Questions

1. Which of the following are advantages of conducting a network meta-analysis as compared to a pairwise meta-analysis?
   a. make indirect comparisons between interventions that have not been previously compared in RCTs
   b. rank interventions in terms of their relative efficacy or safety
   c. increase the precision of our summary effect estimates by including both direct and indirect evidence
   d. all of the above

2. You read an article reporting the results of a systematic review and network meta-analysis. The authors report there was no inconsistency detected in their network meta-analysis models. You should:
   a. accept the network meta-analysis results as robust because there was no inconsistency identified
   b. read further in the study methods and results section to see if the authors evaluated the transitivity assumption prior to conducting the network meta-analysis
   c. consider the similarities and differences between the studies included in the network meta-analysis to evaluate the transitivity assumption
   d. b and c

3. Which of the following model outputs are common to both pairwise and network meta-analysis?
   a. summary effect estimate (e.g. odds ratio, mean difference)
   b. mean rank
   c. surface under the cumulative ranking curve (SUCRA) value
   d. inconsistency plot

4. Which of the following scenarios best describes a homogeneous comparison?
   a. the mean age of patients enrolled in studies evaluating comparison AB is 65; whereas, the mean age of patients enrolled in studies evaluating comparison AC is 70
   b. among three studies evaluating comparison AB, the mean age of patients enrolled in study #1 is 65, the mean age of patients enrolled in study #2 is 66, and the mean age of patients enrolled in study #3 is 63
   c. the mean age of patients enrolled in studies evaluating comparison AB is 65; whereas, the mean age of patients enrolled in studies evaluating comparison AC is 66
   d. among three studies evaluating comparison AB, the mean age of patients enrolled in study #1 is 65, the mean age of patients enrolled in study #2 is 45, and the mean age of patients enrolled in study #3 is 80

5. You conduct a network meta-analysis on the comparative risk of death from new drugs used to treat atopic dermatitis. The mean ranks for four of the new drugs are as follows:
   Drug A     6.2
   Drug B     3.4
Drug C 8.1
Drug D 1.5

Which of the following is true?

a. Drug A is associated with a greater risk of death compared to Drug B
b. Drug D is associated with a lower risk of death compared to Drug C
c. Drug A is associated with a lower risk of death compared to Drug B
d. Drug D is associated with a lower risk of death compared to Drug A

Answers:

1. The correct answer is d. All of the listed choices are advantages of network meta-analysis over pairwise meta-analysis.
2. The correct answer is d. You should never rely solely on tests of inconsistency to detect inconsistency in a network meta-analysis. You must first conduct an assessment of transitivity across comparisons in the network to ensure effect modifiers are balanced. Authors should conduct an assessment of transitivity and they should provide a way for readers of their study to assess the transitivity assumption as well.
3. The correct answer is a. Summary effect estimates are reported in both pairwise and network meta-analyses.
4. The correct answer is b. The three studies that have compared treatments A and B have a similar distribution of patient ages, which indicates there is homogeneity with regards to patient age within this treatment comparison. Choice c is an example of the transitivity assumption. Patients enrolled in studies comparing treatments A and B and treatments A and C are similar in age, which confirms the transitivity assumption to be valid with regards to the potential effect modifier of patient age.
5. The correct answer is a. Drugs with a lower mean rank are associated with a higher risk of death.
References


Mills EJ, Ioannidis JPA, Thorlund K, Schunemann HJ, Puhan MA, Guyatt GH. How to Use an Article Reporting a Multiple Treatment Comparison Meta-analysis. JAMA 2012;308(12):1246-53.


Figure 1. Illustration of Intervention Comparisons in Pairwise and Network Meta-Analysis. Panel A demonstrates a pairwise comparison between interventions 1 and 2. Panel B demonstrates two direct comparisons (intervention 1 vs. 2 and intervention 2 vs. 3) and one indirect comparison (intervention 1 vs. 3) in a network meta-analysis. Panel C demonstrates 3 direct comparisons (intervention 1 vs. 2, intervention 2 vs. 3, and intervention 1 vs. 3) that form a closed loop.
Figure 2. Examples of Network Plots

Connected network plots (Jabbar-Lopez et al., 2017). Nodes represent individual interventions and nodes connected by lines indicate that these two interventions have previously been directly compared in a study. In these examples, the nodes are weighted by the number of studies evaluating this treatment and the lines are weighted by the number of studies evaluating this treatment comparison. Each panel is a network plot of interventions reporting the outcome of interest: (a) clear/nearly clear (minimal residual activity/PASI > 90/0 or 1 on PGA), (b) mean change in the dermatology life quality index, and (c) withdrawal due to adverse events. Abbreviations: adalimumab (ADA), etanercept (ETA), infliximab (INF), ixekizumab (IXE), methotrexate (MTX), psoriasis area and severity index (PASI), placebo (PBO), physician’s global assessment (PGA), secukinumab (SEC), ustekinumab (UST).
Surface under the cumulative ranking (SUCRA) curves of treatments evaluating the psoriasis area and severity index (PASI) 75 at 12/16 weeks (Jabbar-Lopez et al., 2017). The cumulative probability that each treatment is ranked among the top n (e.g. 1, 2, ..., 8) treatments (y-axis) is plotted against each possible rank (x-axis) for treatments in the network. Predictive probabilities incorporate the uncertainty in our network estimates from heterogeneity. Ixekizumab (IXE) has the highest SUCRA value (96.4%) and placebo (PBO) has the lowest SUCRA value (0%). Abbreviations: adalimumab (ADA), etanercept (ETA), infliximab (INF), ixekizumab (IXE), methotrexate (MTX), placebo (PBO), secukinumab (SEC), ustekinumab (UST).
This is an example of an inconsistency plot with closed triangular loops of treatment comparisons evaluating the psoriasis area and severity index (PASI) 75 at 12/16 weeks (Jabbar-Lopez et al., 2017). The x-axis represents the scale for the inconsistency factors (IFs). The PBO-INF-MTX loop shows evidence of inconsistency between direct and indirect evidence because the 95% CI for the IF does not include zero. There is no significant inconsistency identified in any of the other loops. Abbreviations: adalimumab (ADA), etanercept (ETA), infliximab (INF), ixekizumab (IXE), methotrexate (MTX), placebo (PBO), secukinumab (SEC), ustekinumab (UST), inconsistency factor (IF), confidence interval (CI).

Figure 4. Example of an Inconsistency Plot
This is an example of a comparison-adjusted funnel plot of treatment comparisons evaluating the psoriasis area and severity index (PASI) 75 at 12/16 weeks (Jabbar-Lopez et al., 2017). Comparisons are color-coded as per the legend at the bottom of the figure. The y-axis represents the standard error of each study-specific effect estimate. The x-axis represents the difference between the ln(odds ratio) for each study-specific effect estimate and the pooled effect estimate for each comparison (e.g. all of the study-specific estimates reporting on the PBO vs ADA comparison). The blue diagonal line represents a linear regression of the x-axis variable on the y-axis variable. The paucity of studies in the bottom left of the plot indicates there may be small studies missing that would have favored established treatments. Abbreviations: adalimumab (ADA), etanercept (ETA), infliximab (INF), ixekizumab (IXE), methotrexate (MTX), placebo (PBO), secukinumab (SEC), ustekinumab (UST).
Figure 6. Example of a Rank-Heat Plot

This is an example of a rank-heat plot of outcomes associated with insulin use in patients with type 1 diabetes mellitus. Each ring represents a different outcome. Outcomes are also specified in the legend. Each "slice" represents a different treatment. Treatments are ranked according to their surface under the cumulative ranking curve (SUCRA) values. Higher SUCRA values (in green) indicate more efficacious and safer treatments. Uncolored areas indicate that the treatment was not included in the NMA of that outcome. Abbreviations: hemoglobin A1c (A1c), twice daily (bid), once daily (OD), four times per day (qid).
Research Techniques Made Simple:
Understanding Network Meta-Analysis

Dr. Jennifer Watt, MD
Dr. Andrea Tricco, PhD
Dr. Sharon Straus, MD MSc
Dr. Areti Angeliki Veroniki, PhD
Dr. Gary Naglie, MD
Dr. Aaron Drucker, MD ScM
Outline of Presentation

• Comparing Pairwise and Network Meta-Analysis
• Assumptions of Network Meta-Analysis
• Interpreting a Network Meta-Analysis
• Summary
• References
Comparing Pairwise and Network Meta-Analyses
Comparing Pairwise and Network Meta-Analyses

<table>
<thead>
<tr>
<th></th>
<th>Pairwise Meta-Analysis</th>
<th>Network Meta-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Comparators</td>
<td>2</td>
<td>&gt; 2</td>
</tr>
<tr>
<td>Questions Answered by</td>
<td>(1) What is the</td>
<td>(1) Is the intervention efficacious and/or safe?</td>
</tr>
<tr>
<td>Analysis Method</td>
<td>efficacy or risk of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>harm associated with</td>
<td></td>
</tr>
<tr>
<td></td>
<td>one intervention</td>
<td></td>
</tr>
<tr>
<td></td>
<td>compared to another?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) What intervention</td>
<td>(2) What intervention is the most efficacious and/or safe?</td>
</tr>
<tr>
<td></td>
<td>is the most efficacious</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and/or safe?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) What is the</td>
<td>(3) What is the</td>
</tr>
<tr>
<td></td>
<td>comparative efficacy</td>
<td>comparative efficacy</td>
</tr>
<tr>
<td></td>
<td>and/or safety</td>
<td>and/or safety</td>
</tr>
<tr>
<td></td>
<td>between two</td>
<td>between two</td>
</tr>
<tr>
<td></td>
<td>interventions that</td>
<td>interventions that</td>
</tr>
<tr>
<td></td>
<td>haven't been</td>
<td>haven't been</td>
</tr>
<tr>
<td></td>
<td>directly compared?</td>
<td>directly compared?</td>
</tr>
<tr>
<td>Assumptions</td>
<td>(1) Homogeneity</td>
<td>(1) Summary effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>estimate (e.g. OR, MD,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMD)</td>
</tr>
<tr>
<td></td>
<td>(2) Network connectivity</td>
<td>(2) Mean rank</td>
</tr>
<tr>
<td></td>
<td>(3) Transitivity</td>
<td>(3) SUCRA value or P-score</td>
</tr>
<tr>
<td></td>
<td>(4) Consistency</td>
<td>(4) Inconsistency plot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5) Comparison-adjusted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>funnel plot</td>
</tr>
<tr>
<td>Model Outputs</td>
<td>(1) Summary effect</td>
<td></td>
</tr>
<tr>
<td></td>
<td>estimate (e.g. OR, MD, SMD)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Funnel plot</td>
<td></td>
</tr>
<tr>
<td>Reporting Guidelines</td>
<td>PRISMA</td>
<td>PRISMA-NMA</td>
</tr>
</tbody>
</table>
## Assumptions of Network Meta-Analysis

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Questions to Consider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogeneity</td>
<td>(1) Is there any clinical, methodological, or statistical heterogeneity between studies that compare the same interventions? &lt;br&gt; (2) Are there effect modifiers (e.g. age, gender, illness severity) between studies making the same treatment comparison that could influence the summary effect estimate?</td>
</tr>
<tr>
<td>Network Connectivity</td>
<td>(1) Do all of the interventions form a connected network (as in Figure 2)?</td>
</tr>
<tr>
<td>Transitivity</td>
<td>(1) Is there an imbalance in effect modifiers among studies included in the network? &lt;br&gt; (2) In theory, could any patient randomized in one study within a network have been randomized to any of the other studies in this same network?</td>
</tr>
<tr>
<td>Consistency</td>
<td>(1) Where possible to assess, are the direct and indirect effect estimates from closed loops in the network in agreement?</td>
</tr>
</tbody>
</table>
## Common Network Meta-Analysis Outputs

<table>
<thead>
<tr>
<th>Network Meta-Analysis Output</th>
<th>Description</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network Plot</strong></td>
<td>a diagram depicting how interventions (nodes) are connected to one another through direct comparisons (lines) (see Figure 2)</td>
<td>provides an overview of the available evidence; a network estimate of an intervention’s relative efficacy or safety compared to other interventions in the network can only be calculated if it is connected to the network</td>
</tr>
<tr>
<td><strong>Summary Effect Estimate</strong></td>
<td>estimate of the relative efficacy of interventions in the network (e.g. OR, MD, SMD, HR) compared to other network interventions, reported with a measure of uncertainty (e.g. confidence/credible intervals or predictive intervals)</td>
<td>same interpretation as a summary effect estimate in a pairwise meta-analysis</td>
</tr>
<tr>
<td><strong>Ranking Statistics</strong></td>
<td>frequently presented as a mean/median rank, SUCRA value (or P-Score) or probability of being the best treatment</td>
<td>an intervention with a higher ranking, SUCRA value, or probability of being the best is more efficacious or more likely to cause harm</td>
</tr>
<tr>
<td><strong>Inconsistency Plot</strong></td>
<td>a plot reporting the inconsistency factors (absolute difference between direct and indirect effect estimates) for each comparison in a closed network loop (see Figure 4)</td>
<td>an inconsistency factor with a confidence interval that does not include zero indicates that there is significant inconsistency between direct and indirect effect estimates</td>
</tr>
<tr>
<td><strong>Comparison-Adjusted Funnel Plot</strong></td>
<td>similar to a funnel plot in pairwise MAs; however, (1) the x-axis is the difference between each study-specific effect estimate and pooled effect estimate for each comparison and (2) comparisons have been ordered in a meaningful way (e.g. chronological treatment order) (see Figure 5)</td>
<td>asymmetry in the plot indicates publication bias/small-study effects</td>
</tr>
</tbody>
</table>
Examples of Network Plots

(from Jabbar-Lopez et al., 2017)
Example of a SUCRA Plot
Example of an Inconsistency Plot

(from Jabbar-Lopez et al., 2017)
Example of a Comparison-Adjusted Funnel Plot

(from Jabbar-Lopez et al., 2017)
Summary

• Comparing Pairwise and Network Meta-Analyses
  – Pairwise meta-analyses synthesize evidence comparing 2 interventions; network meta-analyses (NMA) compare >2 interventions
  – NMAs can rank interventions in terms of their relative safety or efficacy

• Limitations
  – Must consider important assumptions underlying NMA, such as transitivity and consistency
  – RCTs included in an NMA are subject to the same biases as those included in pairwise meta-analyses
References


Question 1:

Which of the following are advantages of conducting a network meta-analysis as compared to a pairwise meta-analysis?

❍ make indirect comparisons between interventions that have not been previously compared in RCTs

❍ rank interventions in terms of their relative efficacy or safety

❍ increase the precision of our summary effect estimates by including both direct and indirect evidence

● all of the above

*Explanation:*

All of the listed choices are advantages of network meta-analysis over pairwise meta-analysis.

Question 2:

You read an article reporting the results of a systematic review and network meta-analysis. The authors report there was no inconsistency detected in their network meta-analysis models. You should:

❍ accept the network meta-analysis results as robust because there was no inconsistency identified

❍ read further in the study methods and results section to see if the authors evaluated the transitivity assumption prior to conducting the network meta-analysis

❍ consider the similarities and differences between the studies included in the network meta-analysis to evaluate the transitivity assumption

● b and c
You should never rely solely on tests of inconsistency to detect inconsistency in a network meta-analysis. You must first conduct an assessment of transitivity across comparisons in the network to ensure effect modifiers are balanced. Authors should conduct an assessment of transitivity and they should provide a way for readers of their study to assess the transitivity assumption as well.

**Question 3:**

Which of the following model outputs are common to both pairwise and network meta-analysis?

- summary effect estimate (e.g. odds ratio, mean difference)

**Explanation:**

Summary effect estimates are reported in both pairwise and network meta-analyses.

- mean rank
- surface under the cumulative ranking curve (SUCRA) value
- inconsistency plot

**Question 4:**

Which of the following scenarios best describes a homogeneous comparison?

- the mean age of patients enrolled in studies evaluating comparison AB is 65; whereas, the mean age of patients enrolled in studies evaluating comparison AC is 70

- among three studies evaluating comparison AB, the mean age of patients enrolled in study #1 is 65, the mean age of patients enrolled in study #2 is 66, and the mean age of patients enrolled in study #3 is 63

**Explanation:**
The three studies that have compared treatments A and B have a similar distribution of patient ages, which indicates there is homogeneity with regards to patient age within this treatment comparison. Choice c is an example of the transitivity assumption. Patients enrolled in studies comparing treatments A and B and treatments A and C are similar in age, which confirms the transitivity assumption to be valid with regards to the potential effect modifier of patient age.

- the mean age of patients enrolled in studies evaluating comparison AB is 65; whereas, the mean age of patients enrolled in studies evaluating comparison AC is 66

- among three studies evaluating comparison AB, the mean age of patients enrolled in study #1 is 65, the mean age of patients enrolled in study #2 is 45, and the mean age of patients enrolled in study #3 is 80

**Question 5:**

You conduct a network meta-analysis on the comparative risk of death from new drugs used to treat atopic dermatitis. The mean ranks for 4 new drugs are as follows: Drug A 6.2 Drug B 3.4 Drug C 8.1 Drug D 1.5 Which of the following is true?

- Drug A is associated with a greater risk of death compared to Drug B

  *Explanation:*

  Drugs with a lower mean rank are associated with a higher risk of death.

- Drug D is associated with a lower risk of death compared to Drug C

- Drug A is associated with a lower risk of death compared to Drug B

- Drug D is associated with a lower risk of death compared to Drug A