Conceptual and Technical Challenges in Network Meta-analysis

Andrea Cipriani, PhD; Julian P.T. Higgins, PhD; John R. Geddes, MD; and Georgia Salanti, PhD

The increase in treatment options creates an urgent need for comparative effectiveness research. Randomized, controlled trials comparing several treatments are usually not feasible, so other methodological approaches are needed. Meta-analyses provide summary estimates of treatment effects by combining data from many studies. However, an important drawback is that standard meta-analyses can compare only 2 interventions at a time. A new meta-analytic technique, called network meta-analysis (or multiple treatments meta-analysis or mixed-treatment comparison), allows assessment of the relative effectiveness of several interventions, synthesizing evidence across a network of randomized trials. Despite the growing prevalence and influence of network meta-analysis in many fields of medicine, several issues need to be addressed when constructing one to avoid conclusions that are inaccurate, invalid, or not clearly justified. This article explores the scope and limitations of network meta-analysis and offers advice on dealing with heterogeneity, inconsistency, and potential sources of bias in the available evidence to increase awareness among physicians about some of the challenges in interpretation.

Ann Intern Med. 2013;159:130-137. www.annals.org
For author affiliations, see end of text.

The increase in alternative medical treatment options has led to the need for comparative effectiveness research (1, 2). Randomized, controlled trials comparing many treatment options are usually not feasible, so other methodological approaches are needed. A meta-analysis embedded in a systematic review is a useful statistical tool that provides a summary estimate of treatment effect by combining data from many studies. However, a key limitation of standard (or pairwise) meta-analyses is that they can compare only 2 interventions at a time. When several treatment options are available, a series of individual meta-analyses provides only partial information because it can only answer questions about pairs of treatments. This does not support optimal clinical decision making because each meta-analysis is only one part of the whole picture.

The need for a method to summarize evidence across many interventions has been increasingly recognized (3). Network meta-analysis (or multiple treatments meta-analysis or mixed-treatment comparison) has been developed to assess the relative effectiveness of several interventions and synthesize evidence across a network of randomized trials (4–6). The method is based on the simultaneous analysis of direct evidence (which comes from studies directly randomizing treatments of interest) and indirect evidence (which comes from studies comparing treatments of interest with a common comparator) (7).

Several applications and methodological articles have outlined the benefits of network analysis (1, 8, 9), which has become increasingly popular (**Appendix Figure 1**, available at www.annals.org). Although most of the assumptions underlying network meta-analysis are shared with those for pairwise meta-analysis, it is far from being fully accepted and has been criticized (10).

This article elucidates the main characteristics of network meta-analysis, focuses on questions about its main conceptual and technical challenges, and offers advice on how they might be addressed in practice. Definitions of the statistical terms used are provided in **Appendix Table 1** (available at www.annals.org).

What Is a Network Meta-analysis?

Figure 1 presents a network of pharmacologic treatments for acute mania. The primary outcome for short-term treatment efficacy is the change score on a rating scale for manic symptoms (11). The lines between treatment nodes indicate which comparisons have been made in randomized trials. If there is no line between 2 nodes, then no studies (that is, no direct evidence) compare the 2 drugs. A network meta-analysis is a simultaneous analysis of the data from all of these randomized trials. With a network meta-analysis, the relative effectiveness of 2 treatments can be estimated even if no studies directly compare them. For example, no single study has compared aripiprazole and risperidone but using a common comparator (placebo) allows for an indirect comparison between them. Denoting aripiprazole, risperidone, and placebo as treatments A, B, and C, respectively, an indirect comparison (AB) can be obtained by subtracting the meta-analytic estimates of all studies of risperidone versus placebo (BC) from the estimate of all studies of aripiprazole versus placebo (AC): $AB_{\text{indirect meta-analysis}} = AC_{\text{direct meta-analysis}} - BC_{\text{direct meta-analysis}}$.

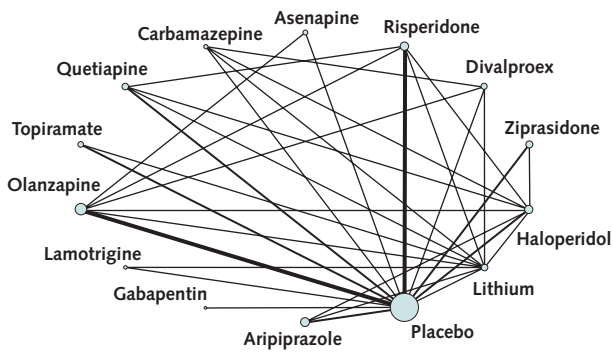
If direct evidence is available (such as haloperidol vs. risperidone in **Figure 1**), the network meta-analysis combines direct and indirect estimates and calculates a mixed effect size as the weighted average of the direct evidence (studies comparing haloperidol and risperidone directly) and the indirect evidence (for example, studies comparing haloperidol and risperidone via placebo). The network formed by studies of haloperidol versus risperidone, haloperidol versus placebo, and risperidone versus placebo is often called a loop of evidence. In more complex networks (with ≥4 competing treatments), indirect estimates can be derived through several loops with many different intermediate comparators (**Figure 2**). The use of indirect estimates can provide information on comparisons for which no trials exist. It can also improve the precision of the direct estimate by reducing the width of the CIs compared with the direct evidence alone (8).

Is Indirect Evidence Randomized or Observational Evidence?

Current hierarchies of evidence tend to place indirect and mixed comparisons below direct evidence because indirect comparisons may have biases similar to those in observational studies. For example, the Canadian Agency for Drugs and Technologies in Health allows indirect and mixed comparisons only as a sensitivity or supportive analysis to supplement the direct evidence (12).

Indirect comparisons can be viewed as a special case of subgroup meta-analysis; the studies form 2 groups of comparison pairs (that is, AC and BC), and the difference between the 2 subgroup summary estimates is the indirect estimate of A versus B (**Figure 1**). An indirect estimate is calculated using estimates from direct meta-analyses, so the within-trial randomization is preserved. This process is often called adjusted indirect comparison (8); thus, network meta-analyses are based on randomized comparisons. Nevertheless, indirect comparison and consequently network meta-analysis provide observational evidence in that the treatments being compared have not been randomized directly within the individual trials. Standard meta-analysis has been described as providing evidence of an observational nature (13) because competing treatments have been randomized within but not across trials.

Figure 1. **Network of eligible comparisons for the multiple treatments meta-analysis for efficacy in acute mania.**



The lines between treatment nodes indicate the comparisons made within randomized trials. The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomly assigned participants (sample size). If there is no line between 2 nodes, then no studies (that is, no direct evidence) compare the 2 drugs. Stata (StataCorp, College Station, Texas) and R routines (R Foundation for Statistical Computing, Vienna, Austria) are used to plot a network, and their results, as well as the routines and their help files, are available at www.mtm.uoi.gr. Data from reference 11.

When Are Indirect and Mixed Comparisons Valid?

The main assumption underpinning the validity of indirect and mixed comparison is that there are no important differences between the trials making different comparisons other than the treatments being compared. This assumption has been described alongside several terms in the literature, including similarity (14, 15), transitivity (16, 17), consistency (4), and coherence (18). We will discuss our interpretation of these terms and different manifestations of the main assumption, highlighting the important methodological issues and their implications for drawing clinically meaningful conclusions.

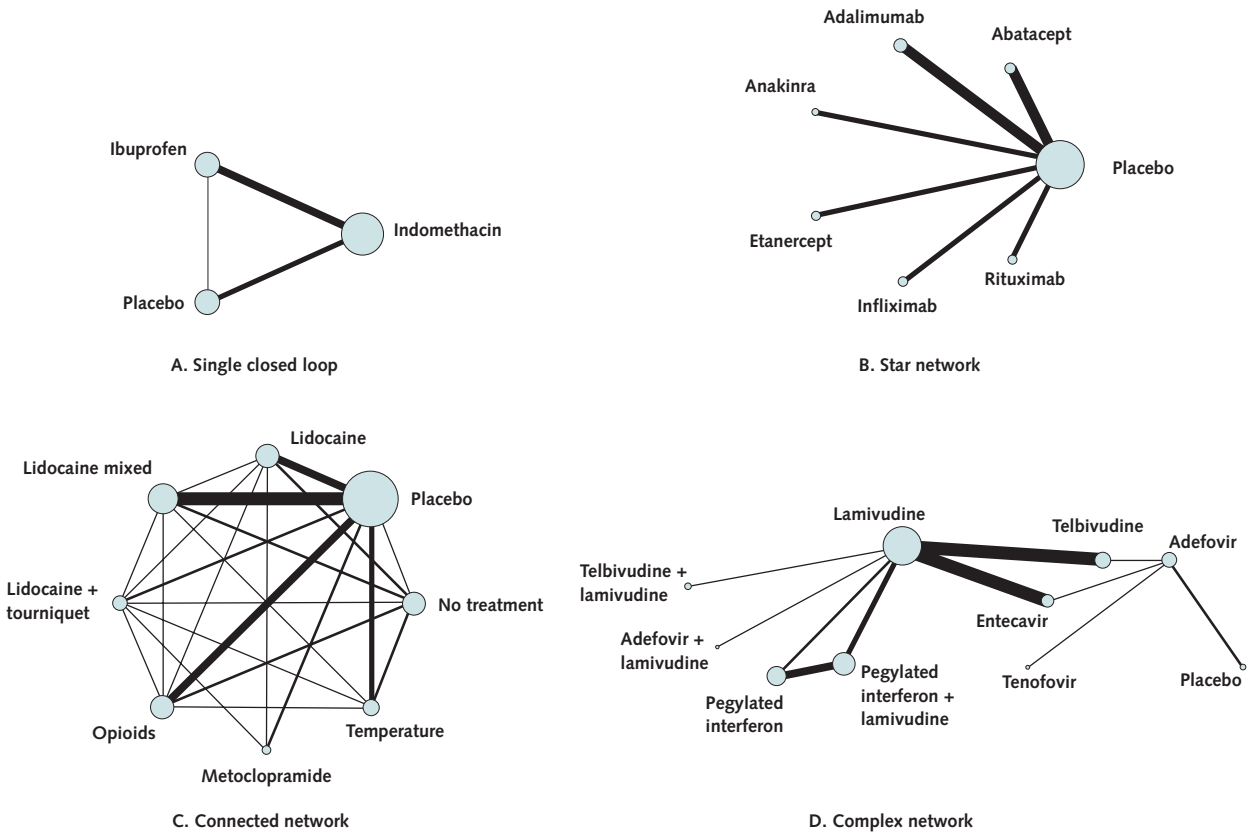
Transitivity and Similarity

The synthesis of studies making a direct comparison of 2 treatments makes sense only when the studies are sufficiently similar in important clinical and methodological characteristics (effect modifiers). These effect modifiers need not be, and usually are not, identical (leading to heterogeneity of effects across studies). A valid indirect comparison (such as AB) requires that the sets of AC and BC studies are similar in their distributions of effect modifiers (for example, severity of illness at baseline, treatment dose, sample size, and study quality). Only when this is the case can we assume that the intervention effects are transitive (that is, the previously mentioned subtraction equation holds). Transitivity can be viewed as the extension of clinical and methodological homogeneity to comparisons across groups of studies that compare treatments. In complex network structures, the transitivity assumption should hold for all cases where indirect or mixed estimates are derived. Note that similarity within each comparison in the network is not sufficient to justify the transitivity assumption. Suppose that all AC studies include patients with severe illness and all BC studies include patients with moderate illness. Each study set is similar within itself (at least according to this particular characteristic), but the 2 sets deal with clinically different populations of patients. So, if severity is an effect modifier, the transitivity assumption would not hold, and synthesis of these 2 meta-analyses would not give a valid AB estimate (17).

A special case of an effect modifier that can vary across comparisons violating the transitivity assumption is the nature of the common comparator. If comparator C is systematically different in AC and BC studies (for example, treatment C is administered as an oral tablet in AC studies but as an intramuscular injection in BC studies), then the transitivity assumption probably will not hold, and the indirect comparison between treatments A and B might not be valid.

The plausibility of the transitivity assumption requires judgment to decide whether differences in the distributions of the effect modifiers across studies are large enough to make network meta-analysis invalid. However, practical challenges are that effect modifiers are often not reported in studies and that too few studies are available per comparison to make a reasonable judgment.

Figure 2. **Examples of network structures.**



Nodes represent a treatment or an intervention; lines show where direct comparisons exist from 1 or more RCTs. RCT = randomized, controlled trial. **A.** A single closed loop involves 3 interventions and can provide data to calculate direct comparisons and indirect comparisons (mixed evidence). **B.** A "star network" in which all interventions have a single mutual comparator. **C.** A well-connected network in which all interventions have been compared with each other in several trials. **D.** A complex network with many loops and groups that may have sparse connections.

If an imbalanced distribution of effect modifiers is identified, adjustment can be used to improve transitivity through network meta-regression (19, 20). Note that this use of meta-regression for adjustment is for a subtly different purpose from its traditional use to examine how treatment effects depend on study-level characteristics. Adjustment should take place only for study or patient characteristics that are effect modifiers (such as severity of illness at baseline, number of previous episodes, age, or sex) (21).

Consistency

Consistency (or coherence) is the statistical manifestation of transitivity and occurs when the subtraction equation is supported by the data. It can be evaluated only when a loop in the evidence network exists, that is, when there is direct and indirect evidence for a particular comparison of interventions. The distinction between transitivity and consistency is analogous to the distinction between clinical or methodological heterogeneity and statistical heterogeneity seen in standard meta-analysis. Heterogeneity refers to the degree of disagreement between study-specific treatment effects and is measured by differences in estimates of study treatment effect beyond what chance can explain. Inconsistency similarly refers to the degree of disagreement between source-specific (and not study-specific) treatment effects and is measured by differences between direct and indirect estimates beyond what chance can explain (16). Heterogeneity is usually evaluated by the Cochran Q test or the I^2 statistic (22). Consistency in a network meta-analysis can be evaluated statistically by comparing the direct and indirect summary effects in specific loops (8, 18) or across a network by fitting models that allow and do not allow for inconsistency (4, 23, 24). A "leave-one-comparison-out" approach, often called "node splitting," can also be applied, with each direct comparison being excluded from the network and then estimating the difference between this direct evidence and the indirect evidence from the network (25). In an empirical study of 40 published networks, inconsistency was seen in 9% of evidence loops and in 1 in 8 networks (26). The power of tests for

inconsistency is expected to be generally low (26) and might depend on the method used to estimate heterogeneity (27).

How Should Inconsistent Networks Be Addressed?

Despite the best efforts of investigators to construct a consistent network, statistically significant inconsistency may arise, which should be investigated when found. Systematic review protocols list potential sources of heterogeneity, possibly using them to form more homogeneous subgroups of studies and generate hypotheses for effect modifiers. Similarly, network meta-analysis should describe in the protocol a clear strategy to deal with inconsistency. Some strategies used to address heterogeneity in standard meta-analysis can be adapted and employed in network meta-analysis to tackle inconsistency, which is described in the **Table**. Representing all triangular closed loops in a network, a graph of the differences between direct and indirect estimates can show which loops have statistically significant inconsistency. This graph is shown for the main efficacy outcome for the acute mania example in **Figure 3**. Only 4 of a total of 31 loops showed inconsistency, yielding 95% CIs that exclude 0 (that is, the direct estimate of the summary effect differs statistically from the indirect estimate).

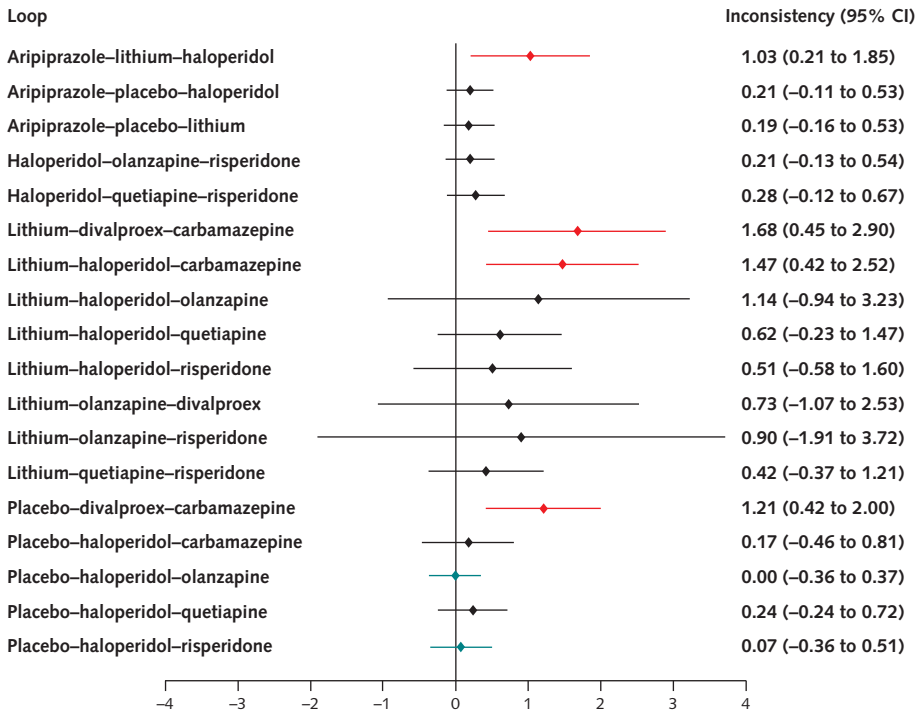
Meta-analyses (or, as is often the case, single studies) contributing to inconsistent loops should be scrutinized for outlying studies, and primary data should be checked for errors. If data are shown to be correct after double checking and inconsistency is still evident, researchers should explore other sources of inconsistency. Doing so (through network meta-regression or subgroup analysis) may enhance understanding of how interventions seem to work under different conditions and why certain sources are more reliable than others. In the acute mania example, data extraction and data entry were found to be correct. No important variables could be identified that differed across comparisons in the 4 inconsistent loops, although the numbers of included studies were very small in these loops. Models with and without imposition of the consistency assumption fitted similarly well with the data, which suggested that the more parsimonious consistency model might be preferable (12).

When inconsistency cannot be explained after considering effect modifiers, researchers may synthesize the data to reflect this additional uncertainty. Models can be applied to relax the consistency assumption by adding an

Table. **Strategies to Respond to Heterogeneity and Inconsistency**

Action	Heterogeneity	Inconsistency	Details and Warnings
Check the data	Studies that stand out in the forest plot are checked for data extraction errors (28, 29).	Inconsistency in a loop may suggest data errors in extraction or transitivity violations (26).	Look for data extraction errors in loops that present inconsistency and in comparisons with large heterogeneity. Amend the data and redo the analysis with the correct data set.
Try to bypass	Empirical evidence shows that some effect measures are associated with larger heterogeneity than others (30, 31).	Empirical evidence shows that the 3 measures for dichotomous outcomes are similarly prone to inconsistency (26).	Statistical heterogeneity and inconsistency might change if the measure of treatment effect is changed. Relative and standardized measures (odds ratio, risk ratio, standardized mean difference, and ratio of means) may be more homogeneous than difference measures (mean differences and risk difference).
Explore it	Study-level variables (ideally prespecified) can be investigated in a subgroup analysis or meta-regression (32).	The network may be split into subgroups, or network meta-regression can be used to account for differences across studies and comparisons. The variables, including bias-related characteristics, should ideally be specified in the protocol (18, 19).	Adjustment for factors that can vary across comparisons may improve the plausibility of the transitivity assumption and reduce heterogeneity. As with simple meta-regression, network meta-regression can have low statistical power and high rates of false-positive results if heterogeneity is ignored and may be prone to ecological bias when aggregated data are used as covariates. Statistical expertise and careful interpretation of the results are required.
Encompass it	Random-effects meta-analysis can be applied (33, 34).	Models can be applied that relax the consistency assumption (4). These models account for the extra variation in the summary network estimates due to unexplained inconsistency.	Just as random effects are not a remedy for large heterogeneity and should be applied only for moderate amounts of unexplained heterogeneity, inconsistency models should be used to allow for only moderate amounts of unexplained inconsistency. Fitting inconsistency models needs statistical expertise, and their results need to be interpreted with caution.
Resign to it	Investigators may decide not to undertake meta-analysis in the presence of important unexplained heterogeneity.	Investigators may decide not to synthesize the network in the presence of important unexplained inconsistency.	If important statistical heterogeneity or inconsistency remains even after adjustment for possible explanatory variables, researchers might choose not to synthesize the data. This decision should be guided by clinical understanding of the field and careful examination of the consequences for policymakers and patients.

Figure 3. **An inconsistency plot reporting the first 18 triangular closed loops in the network meta-analysis comparing antimanic drugs.**



Each diamond represents the difference between direct and indirect estimates in terms of standardized mean difference for efficacy; the corresponding horizontal line represents its 95% CI. For loops with no inconsistency (in green), this difference is 0 and the CI has 1 negative and 1 positive limit (across 0). If the limits of the CI are greater or less than 0, there is statistically significant inconsistency (in red). Data from reference 11.

extra random effect to each loop (4). Such models have a conceptual analogy to the random-effects model in standard pairwise meta-analysis, which postulates that the true study-specific relative treatment effects are not identical but are similar in that they form a common distribution around the average treatment effect. Random effects in pairwise meta-analysis incorporate the additional uncertainty associated with heterogeneity. Similarly, a model with random inconsistency effects assumes that the different source-specific effects (direct and various indirect effects via different routes) are not identical but are similar. However, incorporation is not the same as adjustment. Because random effects should be applied only for a moderate amount of unexplained heterogeneity, inconsistency models would be most appropriately used to encompass a moderate amount of unexplained inconsistency and not viewed as an attempt to adjust for it.

When Can We Rely on Direct, Indirect, or Mixed Evidence?

The issue of choosing among direct, indirect, and mixed evidence may arise, particularly when important inconsistency is identified. Arguments exist for giving priority to direct evidence because it does not rely on the transitivity assumption. For example, if the transitivity assumption does not hold, then conclusions from direct evidence may relate to different populations or contexts, but each can be valid for the contexts in which the trials were done. If transitivity seems plausible, then mixed evidence may be preferable to direct evidence because of the ability to compare treatments that have not been directly compared in any trial and the improvement in precision for the estimated effect sizes. In some situations, indirect evidence may be more reliable than direct or mixed evidence. This may be the case if all of the direct evidence (for example, novel treatments compared with placebo) is subject to similar types of bias (for example, in favor of the novel treatment), such that these can theoretically be cancelled out when indirect comparisons are made (35). More important, if there are inconsistencies between direct and indirect evidence, one should not choose between sources of evidence but should instead investigate the sources of inconsistency to explain the differences. If a large amount of inconsistency remains unexplained after considering effect modifiers, researchers may decide not to synthesize the available evidence in a network meta-analysis. This decision should be acknowledged and discussed in the report,

and other possible reasons for inconsistency should be thoroughly investigated.

How Might Inconsistency in Network Meta-analysis Be Explained?

Meta-regression can be used to explore inconsistency due to differences across studies and comparisons (19, 20). It can be used to examine causes of inconsistency (typically by using covariates at the comparison level) or to adjust for different distributions of effect modifiers across studies making different comparisons. For example, interventions that are given to more severely ill patients probably will be favored over placebo. Adjustment for factors that vary across comparisons may improve the plausibility of the transitivity assumption while reducing heterogeneity. Through network meta-regression, uncovering the "true" relative effectiveness of each drug is theoretically possible.

As with simple meta-regression, network meta-regression can have low statistical power and high rates of false-positive results if heterogeneity is ignored. Specific statistical expertise and careful interpretation of the results are required (36). Individual-participant data will often be needed to explore effect modification when participant-level covariates are of interest.

Inconsistency may be due to bias. For example, it might be difficult to blind participants in trials of interventions with important side effects but straightforward to do so in those without such side effects. Such trial characteristics can be accounted for in a subgroup analysis or through network meta-regression. By combining studies that compare treatments with various comparators, network meta-analysis enables researchers to explore biases that are difficult to assess in head-to-head meta-analysis. For example, if treatment A is always industry-funded when compared with placebo (C), the potential effect of industry funding (sponsorship bias) in AC trials alone is impossible to account for. However, in a network, we might have studies that compare A with C (all sponsored by the manufacturer of treatment A) and those that compare A with other active treatments in which A is not always the sponsored group (for example, if some AB studies are sponsored by the manufacturer of treatment A and others by the manufacturer of treatment B). We can adjust for sponsoring in all comparisons, including the AC comparisons, as long as we can assume that any such bias is similar wherever it exists across all sponsored interventions (37–39).

How Can Results From Network Meta-analysis Be Presented in a Clinically Useful Way?

Graphical presentation of the evidence network, as seen in **Figure 1**, provides an accessible and understandable format for describing the evidence base contributing to a network meta-analysis (16, 40). By plotting nodes and edges proportional to the amount of information they carry, the graphical presentation conveys how much evidence is available for each direct comparison and which contributes most to the network meta-analysis.

It is useful to report direct estimates and estimates from the network meta-analysis (which may be mixed or based on indirect evidence alone), all with their corresponding uncertainty intervals. This provides a complete and transparent picture of the available evidence and helps to highlight where possible disagreements may occur. Most published network meta-analyses use a table or graph to report this information. Effect sizes are often presented as a league table or in a forest plot against a common comparator (11).

Network meta-analyses also enable estimation of the probability that each intervention is the best for each outcome. Probabilities for each treatment taking each possible rank can be plotted in absolute and cumulative rankograms (37, 41–43) or presented in a table (44) (**Appendix Figures 2** and **3**, available at www.annals.org). It is important to look at the probabilities rather than the naive rankings before drawing conclusions because the latter may be misleading. Ranking measures and probabilities are straightforward to understand and a convenient way to present network results because clinicians usually want to know the preferential order of treatments that could be prescribed to an average patient. However, clinicians should always be interested in the effect sizes and the rankings because a good rank does not necessarily imply a large or clinically important effect size.

How Many Studies Are Needed to Conduct a Network Meta-analysis, and Which Is the Best Statistical Approach to Use?

In general, the more treatments and studies included in a network, the more clinically informative the network meta-analysis. Large data sets might, however, increase the variability across comparisons and studies and render the transitivity assumption difficult to defend. Technically, each treatment of interest must be represented by at least 1 study and the network needs to be connected. A large network informed by few studies often yields imprecise estimates.

Network meta-analysis can be done using frequentist or Bayesian statistical techniques, although to date most have used the latter. Bayesian applications are typically done in WinBUGS (Medical Research Council's Biostatistics Unit, Cambridge, United Kingdom) (45), and codes for various types of data are available online (for example, nicedsu.org.uk, www.mtm.uoi.gr, and www.bristol.ac.uk/social-community-medicine/projects/mpes/).

Conclusion

We have outlined many conceptual and technical issues in network meta-analysis to help clinicians familiarize themselves with this increasingly important method (**Appendix Table 2**). As for all systematic reviews, network meta-analysis should be based on detailed protocols, ideally registered prospectively (46, 47). A detailed protocol protects against data-driven decisions, such as selective use of indirect evidence. It should describe the strategy that the researchers plan to follow to evaluate transitivity and consistency, list factors that could introduce heterogeneity and inconsistency, detail the conditions under which a network meta-analysis will be used to synthesize the results, and specify how inconsistency will be addressed if found. Online registries or databases of systematic reviews (such as PROSPERO [www.crd.york.ac.uk/prospero] or The Cochrane Library [www.thecochranelibrary.com]) provide natural destinations for publication of these details.

A comprehensive search of the literature, careful assessment of the body of evidence with respect to the plausibility of the transitivity assumption, and thoughtful discussion of the potential impact of trial-specific biases on the effect estimates can maximize transparency of a network meta-analysis and avoid errors in its interpretation (15). The methods for network meta-analysis will no doubt continue to develop and will be informed by the experience gained from increased use. Such applications, with accompanying critical evaluation of the methods by collaborative groups of epidemiologists, statisticians, clinicians, and policymakers, will establish and maintain standards for reporting and interpreting network meta-analysis.

From University of Verona, Verona, Italy; University of Oxford, Oxford, United Kingdom; University of Bristol, Bristol, United Kingdom; University of York, North Yorkshire, United Kingdom; and University of Ioannina School of Medicine, Ioannina, Greece.

Acknowledgment: The authors thank Anna Chaimani for her help and support in drafting Figure 3 and Appendix Figure 1.

Grant Support: By the European Research Council (grant 260559 IMMA; Dr. Salanti).

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M12-3072.

Requests for Single Reprints: Andrea Cipriani, PhD, Department of Public Health and Community Medicine, University of Verona, Policlinico G.B. Rossi, Piazzale L.A. Scuro 10, 37134 Verona, Italy; e-mail, andrea.cipriani@psych.ox.ac.uk.

Current author addresses and author contributions are available at www.annals.org.

References

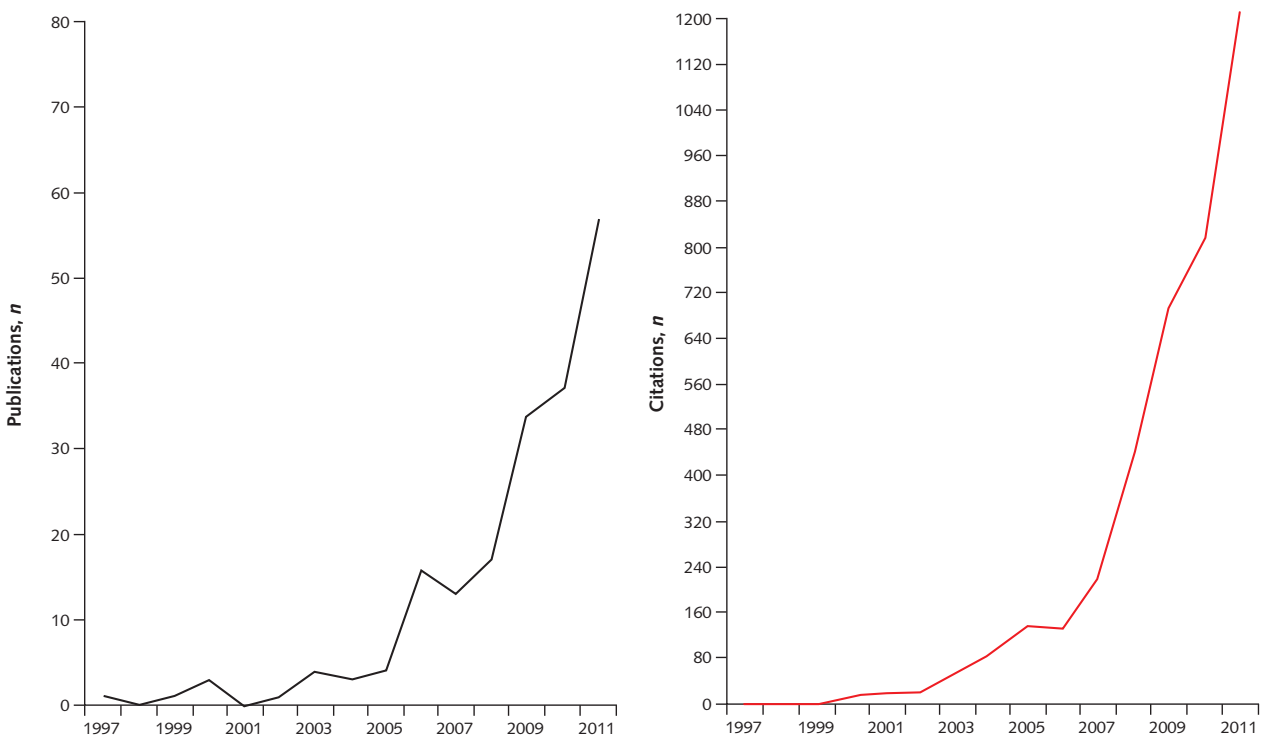
1. **Caldwell DM, Ades AE, Higgins JP.** Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005;331:897-900. [PMID: 16223826]
2. **Li T, Vedula SS, Scherer R, Dickersin K.** What comparative effectiveness research is needed? A framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. Ann Intern Med. 2012;156:367-77. [PMID: 22393132]
3. **Mitka M.** US government kicks off program for comparative effectiveness research. JAMA. 2010;304:2230-1. [PMID: 21098764]
4. **Lu G, Ades AE.** Assessing evidence inconsistency in mixed treatment comparisons. J Am Stat Assoc. 2006;101:447-59.
5. **Salanti G, Higgins JP, Ades AE, Ioannidis JP.** Evaluation of networks of randomized trials. Stat Methods Med Res. 2008;17:279-301. [PMID: 17925316]
6. **Higgins JP, Whitehead A.** Borrowing strength from external trials in a meta-analysis. Stat Med. 1996;15:2733-49. [PMID: 8981683]
7. **Mills EJ, Ioannidis JP, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH.** How to use an article reporting a multiple treatment comparison meta-analysis. JAMA. 2012;308:1246-53. [PMID: 23011714]
8. **Bucher HC, Guyatt GH, Griffith LE, Walter SD.** The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50:683-91. [PMID: 9250266]
9. **Mills EJ, Ghement I, O'Regan C, Thorlund K.** Estimating the power of indirect comparisons: a simulation study. PLoS One. 2011;6:e16237. [PMID: 21283698]
10. **Ioannidis JP.** Indirect comparisons: the mesh and mess of clinical trials. Lancet. 2006;368:1470-2. [PMID: 17071265]
11. **Cipriani A, Barbui C, Salanti G, Rendell J, Brown R, Stockton S, et al.** Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. Lancet. 2011;378:1306-15. [PMID: 21851976]
12. **Wells GA, Sultan SA, Chen L, Khan M, Coyle D.** Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.
13. **Egger M, Smith GD, Phillips AN.** Meta-analysis: principles and procedures. BMJ. 1997;315:1533-7. [PMID: 9432252]
14. **Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG.** Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ. 2009;338:b1147. [PMID: 19346285]
15. **Donegan S, Williamson P, Gamble C, Tudur-Smith C.** Indirect comparisons: a review of reporting and methodological quality. PLoS One. 2010;5:e11054. [PMID: 21085712]
16. **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. Research Synthesis Methods. 2012;3:80-97.
17. **Baker SG, Kramer BS.** The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? BMC Med Res Methodol. 2002;2:13. [PMID: 12429069]
18. **Lumley T.** Network meta-analysis for indirect treatment comparisons. Stat Med. 2002;21:2313-24. [PMID: 12210616]
19. **Salanti G, Marinho V, Higgins JP.** A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. J Clin Epidemiol. 2009;62:857-64. [PMID: 19157778]
20. **Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ.** Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med. 2009;28:1861-81. [PMID: 19399825]
21. **Jansen JP, Schmid CH, Salanti G.** Directed acyclic graphs can help understand bias in indirect and mixed treatment comparisons. J Clin Epidemiol. 2012;65:798-807. [PMID: 22521579]
22. **Higgins JP, Thompson SG, Deeks JJ, Altman DG.** Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-60. [PMID: 12958120]
23. **Dias S, Welton NJ, Ades AE.** Study designs to detect sponsorship and other biases in systematic reviews. J Clin Epidemiol. 2010;63:587-8. [PMID: 20434021]
24. **White IR, Barrett JK, Jackson D, Higgins JPT.** Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Research Synthesis Methods. 2012;3:111-25.

25. **Dias S, Welton NJ, Caldwell DM, Ades AE.** Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29:932-44. [PMID: 20213715]
26. **Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G.** Evaluation of inconsistency in networks of interventions. Int J Epidemiol. 2013;42:332-45. [PMID: 23508418]
27. **Song F, Clark A, Bachmann MO, Maas J.** Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. BMC Med Res Methodol. 2012;12:138. [PMID: 22970794]
28. **Gøtzsche PC, Hróbjartsson A, Maric K, Tendal B.** Data extraction errors in meta-analyses that use standardized mean differences. JAMA. 2007;298:430-7. [PMID: 17652297]
29. **Horton J, Vandermeer B, Hartling L, Tjosvold L, Klassen TP, Buscemi N.** Systematic review data extraction: cross-sectional study showed that experience did not increase accuracy. J Clin Epidemiol. 2010;63:289-98. [PMID: 19683413]
30. **Deeks JJ.** Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med. 2002;21:1575-600. [PMID: 12111921]
31. **Friedrich JO, Adhikari NK, Beyene J.** Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. J Clin Epidemiol. 2011;64:556-64. [PMID: 21447428]
32. **Thompson SG, Higgins JP.** How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002;21:1559-73. [PMID: 12111920]
33. **DerSimonian R, Laird N.** Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177-88. [PMID: 3802833]
34. **Higgins JP, Thompson SG, Spiegelhalter DJ.** A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009;172:137-159. [PMID: 19381330]
35. **Song F, Harvey I, Lilford R.** Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. J Clin Epidemiol. 2008;61:455-63. [PMID: 18394538]
36. **Higgins JP, Thompson SG.** Controlling the risk of spurious findings from meta-regression. Stat Med. 2004;23:1663-82. [PMID: 15160401]
37. **Dias S, Welton NJ, Marinho VCC, Salanti G, Ades AE.** Estimation and adjustment of bias in randomised evidence using mixed treatment comparison meta-analysis. J R Stat Soc Ser A Stat Soc. 2010;173:613–29.
38. **Salanti G, Dias S, Welton NJ, Ades AE, Golfinopoulos V, Kyrgiou M, et al.** Evaluating novel agent effects in multiple-treatments meta-regression. Stat Med. 2010;29:2369-83. [PMID: 20687172]
39. **Chaimani A, Salanti G.** Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Research Synthesis Methods. 2012;3:161-76.
40. **Li T, Puhan MA, Vedula SS, Singh S, Dickersin K; Ad Hoc Network Meta-analysis Methods Meeting Working Group.** Network meta-analysis-highly attractive but more methodological research is needed. BMC Med. 2011; 9:79. [PMID: 21707969]
41. **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;64:163-71. [PMID: 20688472]
42. **Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, et al.** Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet. 2009;373:746-58. [PMID: 19185342]
43. **Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al.** Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health. 2011;14:417-28. [PMID: 21669366]
44. **Hawkins N, Scott DA, Woods BS, Thatcher N.** No study left behind: a network meta-analysis in non-small-cell lung cancer demonstrating the importance of considering all relevant data. Value Health. 2009;12:996-1003. [PMID: 19402854]
45. **Lunn DJ, Best N, Thomas A, Wakefield J, Spiegelhalter D.** Bayesian analysis of population PK/PD models: general concepts and software. J Pharmacokinet Pharmacodyn. 2002;29:271-307. [PMID: 12449499]
46. **Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al.** The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev. 2012;1:2. [PMID: 22587842]
47. **Van der Wees P, Qaseem A, Kaila M, Ollenschlaeger G, Rosenfeld R; Board of Trustees of the Guidelines International Network (G-I-N).** Prospective systematic review registration: perspective from the Guidelines International Network (G-I-N). Syst Rev. 2012;1:3. [PMID: 22587933]

Current Author Addresses: Dr. Cipriani: Department of Public Health and Community Medicine, University of Verona, Policlinico G.B. Rossi, Piazzale L.A. Scuro 10, 37134 Verona, Italy.
Dr. Higgins: School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, United Kingdom.
Dr. Geddes: Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, United Kingdom.
Dr. Salanti: Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, University Campus Ioannina, 45110 Ioannina, Greece.

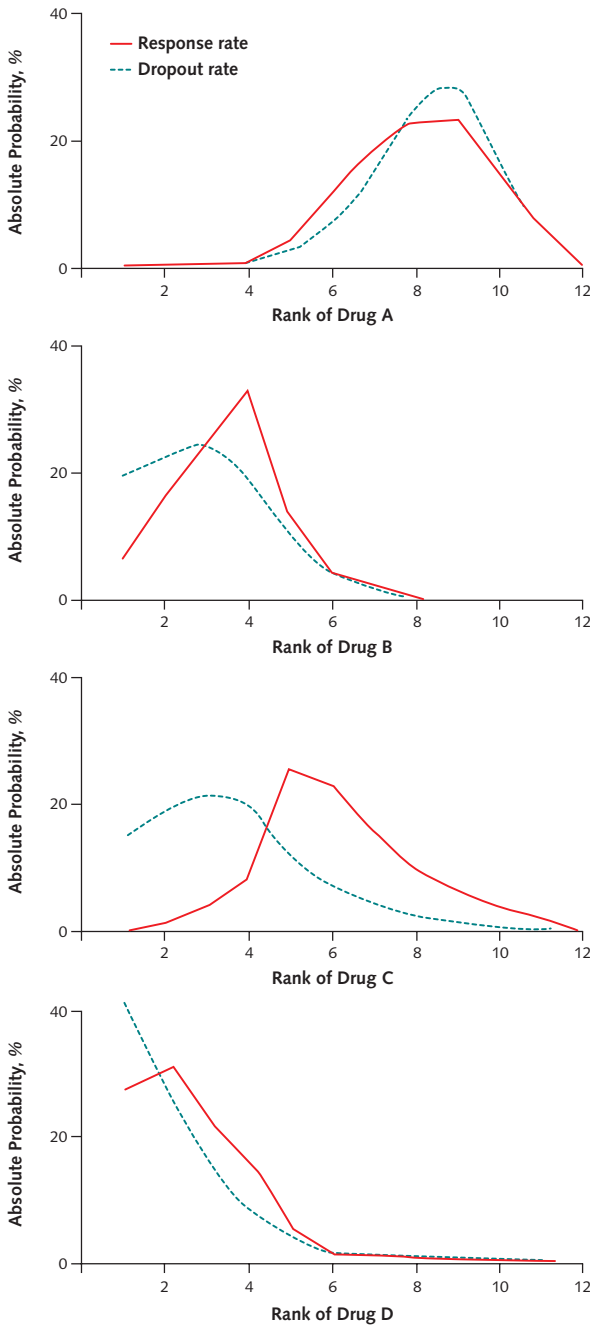
Author Contributions: Conception and design: A. Cipriani, J.R. Geddes, G. Salanti.
Analysis and interpretation of the data: A. Cipriani, J.R. Geddes, G. Salanti.
Drafting of the article: A. Cipriani, J.P.T. Higgins, J.R. Geddes, G. Salanti.
Critical revision of the article for important intellectual content: J.P.T. Higgins, J.R. Geddes, G. Salanti.
Final approval of the article: J.P.T. Higgins, J.R. Geddes, G. Salanti.
Statistical expertise: J.P.T. Higgins, G. Salanti.
Collection and assembly of data: A. Cipriani, G. Salanti.

Appendix Figure 1. **Number of network meta-analyses published in the scientific literature and their citations since 1997.**



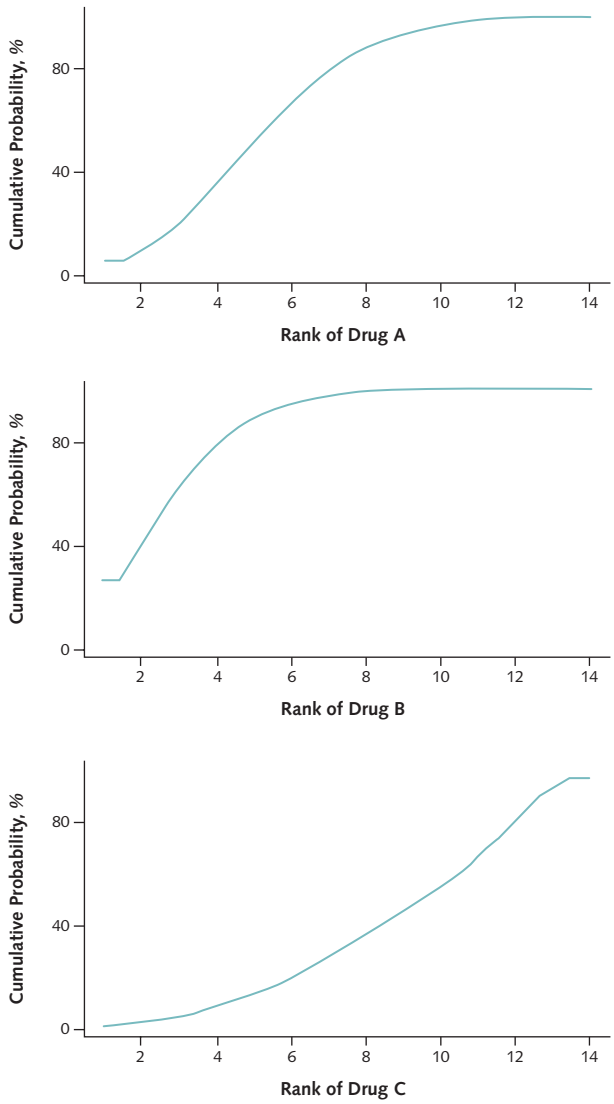
We defined a network meta-analysis as any meta-analysis that used a form of valid indirect relative treatment estimates.

Appendix Figure 2. **Absolute rankograms for presenting probabilities and rankings in network meta-analysis.**



Network meta-analyses enable estimation of the probability that each intervention is the best for each outcome. Probabilities for each treatment taking each possible rank can be plotted in absolute rankograms or cumulative rankograms. These absolute rankograms (modified from reference 41) rank for response rate (efficacy is the solid red line) and withdrawal rate (acceptability is the dotted blue line). Ranking indicates the probability to be the best treatment, the second best, the third best, and so on, among the group of 12 antidepressants under investigation. For example, drug A has a higher probability to be among the worst antidepressants in terms of efficacy and acceptability, although drug B is good.

Appendix Figure 3. **Cumulative rankograms for presenting probabilities and rankings in network meta-analysis.**



Network meta-analyses enable estimation of the probability that each intervention is the best for each outcome. Probabilities for each treatment taking each possible rank can be plotted in absolute rankograms or cumulative rankograms. These cumulative rankograms (modified from reference 41) show the distribution of the probabilities of each treatment to be ranked at each of the possible 14 positions within the group of antimanic drugs under investigation. The larger the surface below the cumulative ranking curve (usually called SUCRA), the more probable the drug will be among the lowest ranks (that is, the more effective or acceptable the treatment). For example, drugs A and B have a higher probability to be among the best antimanic drugs (drug B is better than drug A), although drug C is likely among the worst. The SUCRA can be quantified and reported in tables to show its mean values together with the 95% CIs.

Appendix Table 1. Glossary of Terms

Indirect evidence: Information about a comparison of 2 treatments (AB) that comes entirely from combining studies by means of 1 or more comparator treatments. It may be available through different routes and more than 1 intermediate comparator in a network; we might consider these to be compound indirect evidence. Networks that have compound indirect evidence are most conveniently analyzed using network meta-analysis.

Mixed evidence: Information about a comparison of 2 treatments (AB) that comes from direct evidence (AB studies) and the synthesis of studies by means of 1 or more comparator treatments (indirect evidence, such as the combination of AC and BC studies). If the indirect evidence comes from only 1 route (i.e., simple indirect evidence), then a mixed estimate can be obtained as a weighted average of the direct and indirect estimates of treatment effect. When many indirect routes between A and B are in the network, this approach can be extended, but the most convenient approach to combine them all is a network meta-analysis.

Network meta-analysis:* Synthesis of information over a network of comparisons to assess the comparative effectiveness of more than 2 alternative treatment options for the same condition. The method relies on mixed comparison and synthesizes direct and indirect evidence over the entire network to obtain the relative treatment effects for all comparisons and a ranking of the treatments.

Transitivity and similarity: The assumption that an indirect comparison is a valid method to compare 2 treatments. In its strong form, transitivity is a well-defined mathematical concept that says $A - B = A - C - (B - C)$. It requires that the sets of randomized, controlled trials used to obtain the indirect comparison are sufficiently similar about moderators of the treatment effect. Moderators (13) include clinical similarity, which refers to similarity in patients' characteristics, interventions, settings, length of follow-up, and outcomes measured; and methodological similarity, which refers to aspects of trials associated with the risk of bias. The validity of the results from indirect comparisons, mixed comparisons, and network meta-analysis depends on the plausibility of the transitivity assumption.

Consistency or coherence: The manifestation of transitivity in the data from a network of interventions. It exists when treatment effects from direct and indirect evidence are in agreement (subject to the usual variation due to heterogeneity in the direct evidence). Unlike transitivity, consistency can be evaluated statistically.

Closed loop: A network of 3 comparisons, each of which has been compared directly with the others.

Inconsistent loop: A closed loop in which treatment effect estimates from direct and indirect evidence are in statistically significant disagreement.

Star network: A network where all treatments have been compared with a common comparator treatment but not between themselves. Statistical evaluation of consistency is impossible in a star network.

* Synonyms include multiple treatments meta-analysis or mixed treatment comparison.

Appendix Table 2. Conceptual and Technical Challenges in Interpreting Results From a Network Meta-analysis

Challenges	Explanation and Proposed Plan of Action
Indirect evidence from randomized trials can be considered to be observational evidence.	Although within-trial randomization is preserved, indirect comparisons and network meta-analysis provide observational evidence in that the treatments being compared have not been randomized directly in the individual trials. We believe that indirect evidence should always be considered when available unless there are concerns about the plausibility of the transitivity assumption (see next row). If there is disagreement with direct evidence, investigators should explore possible reasons for it.
Indirect and mixed comparisons may not be valid.	We can obtain a valid, indirect AB estimate (i.e., the transitivity assumption will hold) only when the sets of AC and BC studies are similar with respect to their distributions of effect modifiers. Evaluation of the transitivity assumption requires clinical and methodological understanding of the research field. As in a pairwise meta-analysis, possible effect modifiers should be identified a priori when planning a network meta-analysis and their distributions across comparisons should be compared. Adjustment can be used to account for differences in the distribution of effect modifiers and improve transitivity via network meta-regression. In practice, this might be challenging because of a small number of studies per comparison.
An imbalance in the distribution of effect modifiers might occur even in a set of carefully selected studies.	Judgment should be used to infer about the plausibility of transitivity in a network of trials and decide whether differences in the distributions of the effect modifiers across studies are large enough to make network meta-analysis invalid. Unfortunately, effect modifiers are often not reported in studies and supplemental unpublished material should be retrieved by asking study authors and drug companies. If too few studies are available per comparison to judge transitivity, the review group should carefully consider whether and how to proceed.
Despite the best efforts of investigators to construct a network where the transitivity assumption would hold, significant statistical inconsistency may arise.	Network meta-analysis should describe in the protocol a clear strategy to deal with inconsistency (**Appendix Table 1**). Meta-analyses in inconsistent loops should be scrutinized for outlying studies, and primary data should be checked for errors. If data are correct and inconsistency is still evident, meta-regression can be used to explore inconsistency due to differences across studies and comparisons and why certain sources of evidence are more reliable than others.
Sometimes inconsistency cannot be explained after considering effect modifiers.	Researchers may decide to synthesize the data in a way that reflects the additional (residual) uncertainty due to inconsistency by adding an extra random effect to each loop. Such models have a conceptual analogy to the random-effects model in standard, pairwise meta-analyses. Accordingly, an inconsistency model in network meta-analyses assumes that the different source-specific effects (direct and various indirect effects via different routes) are not identical but are similar. Application and interpretation of inconsistency models require statistical expertise and caution (**Table**).
Study-level and review-level biases may be present in the network of evidence.	As in pairwise meta-analysis, the risk of bias introduced by limitations of individual studies must be considered, but network meta-analysis offers some methodological advantages. For example, by combining studies that compare treatments with various comparators, network meta-analysis enables researchers to explore biases that are difficult to assess in head-to-head meta-analysis. Compared with sensitivity analyses in traditional meta-analyses, adjustment for bias in a network of interventions can increase power if biases can be assumed to operate in similar and consistent ways across comparisons included in the same network.