
Jeroen P Jansen PhD 1, Rachael Fleurence PhD 2, Beth Devine PharmD MBA PhD 3, Robbin Itzler PhD 4, Annabel Barrett BSc 5, Neil Hawkins PhD 6, Karen Lee MA 7, Cornelis Boersma PhD MSc 8, Joseph C Cappelleri PhD MPH 9

1 Mapi Values, Boston, MA, USA
2 Center for Health Economics and Science Policy, United BioSource Corporation, Bethesda, MD, USA
3 Pharmaceutical Outcomes Research & Policy Program, School of Pharmacy, School of Medicine, University of Washington, USA
4 Merck Research Laboratories, North Wales, PA, USA
5 Eli Lilly and Company Ltd, Windlesham, Surrey, UK
6 Oxford Outcomes Ltd, Oxford, UK
7 Canadian Agency for Drugs and Technologies in Health (CADTH), Ottawa, ON, Canada
8 University of Groningen / HECTA, Groningen, Netherlands
9 Pfizer Inc, New London, CT, USA

Author responsible for correspondence:

Jeroen Jansen
Mapi Values
133 Portland Street
Boston MA, 02114
+1 617 720 0001
jeroen.jansen@mapivalues.com

Running title: Interpreting indirect comparisons and network meta-analysis
Abstract
Evidence-based health care decision-making requires comparisons of all relevant competing interventions. In the absence of randomized controlled trials involving a direct comparison of all treatments of interest, indirect treatment comparisons and network meta-analysis provide useful evidence for judiciously selecting the best choice(s) of treatment. Mixed treatment comparisons, a special case of network meta-analysis, combine direct with indirect evidence for particular pairwise comparisons thereby synthesizing a greater share of the available evidence than traditional meta-analysis. This report from the ISPOR Task Force on good research practices provides guidance on the interpretation of indirect treatment comparisons and network meta-analysis to assist policy-makers & health care professionals in using its findings for decision-making. We start with an overview how networks of randomized controlled trials allow for multiple treatment comparisons of competing interventions. Next, an introduction to the synthesis of the available evidence with a focus on terminology, assumptions, validity and statistical methods is provided, followed by advice on critically reviewing and interpreting an indirect treatment comparison or network meta-analysis to inform decision-making. We finish with a discussion what to do if there are no direct or indirect treatment comparisons of randomized controlled trials possible, and a health-care decision still needs to be made.

Keywords: comparative effectiveness, indirect treatment comparison, mixed treatment comparison, network meta-analysis, Bayesian, decision-making,
Introduction

Systematic reviews of randomized controlled trials (RCTs) are considered the standard basis for evidence-based health care decision-making for clinical treatment guidelines and reimbursement policies. Many systematic reviews use meta-analysis to combine quantitative results of several similar and comparable studies and summarize the available evidence (Higgins & Green 2009). Sound decision-making requires comparisons of all relevant competing interventions. Ideally, robustly designed RCTs would simultaneously compare all interventions of interest. Unfortunately, such studies are almost never available, thereby complicating decision-making (Caldwell et al., 2005; Ioannidis 2006; Sutton et al., 2008; Wells et al., 2009). New drugs are often compared with placebo or standard care, but not against each other, in trials aimed to contribute (as expeditiously as possible) toward obtaining approval for drug licensing; there may be no commercial incentive to compare the new treatment with an active control treatment (Sutton et al. 2008; Wells, et al. 2009). Even if there was an incentive to incorporate competing interventions in a RCT, the interventions of interest may vary by country or have changed over time due to new evidence and treatment insights. Therefore, for some indications the number of competing interventions makes a trial incorporating all of them impractical.

In the absence of trials involving a direct comparison of treatments of interest, an indirect comparison can provide useful evidence (which otherwise would be lacking) for judiciously selecting the best choice(s) of treatment. For example, if two particular treatments have never been compared against each other, head-to-head, but these two treatments have been compared against a common comparator, then an indirect treatment comparison (ITC) can use the relative effects of the two treatments versus the common comparator (Bucher et al. 1997, Song et al. 2003, Lu and Ades 2004; Song et al., 2009).

Although it is often argued that indirect comparisons are needed when direct comparisons are not available, it is important to realize that both direct and indirect evidence contribute to the total body of evidence. The results from indirect evidence in combination with the direct evidence may strengthen the assessment between treatments directly evaluated (Caldwell et al., 2005). Even when the results of the direct evidence are conclusive, combining them with the results of indirect estimates in a mixed treatment comparison (MTC) may yield a more refined and precise estimate for the interventions directly compared and broaden inference to the population sampled, as it links and maximizes existing information within the network of treatment comparisons (Lu and Ades 2004).

If the available evidence consists of a network of multiple RCTs involving treatments compared directly or indirectly or both, it can be synthesized by means of so-called network meta-analysis (Lumley 2002). In traditional meta-analysis all included studies compare the same intervention with the same comparator. Network meta-analysis extends this concept by including multiple pairwise comparisons.
across a range of interventions and provides comparative effectiveness information on multiple treatment comparisons.

Given the great value of ITC and network meta-analysis for health care decision-making, this paper provides practical guidance for policy-makers and other health-care practitioners, to enrich their understanding of these evidence synthesis methods. We start with an overview of how RCTs of competing interventions form networks of evidence, which allow multiple treatment comparisons. We then discuss the synthesis of the available evidence with a focus on terminology, assumptions, validity and statistical methods, followed by some advice on critically reviewing and interpreting an ITC or network meta-analysis. The last section discusses what to do if there are no direct or indirect treatment comparisons of RCTs possible and a health-care decision still needs to be made.

Multiple treatment comparisons and evidence networks

Figure 1 shows networks of increasing complexity in which multiple treatments have been compared. Each node reflects an intervention, and a line connecting two nodes reflects one or more RCTs. For every intervention in a connected network, a relative treatment effect can be estimated versus another intervention. Suppose that the main comparison of interest is between intervention B and intervention C, but no direct assessment has compared them. In the first network on the left in Figure 1, intervention B has been compared against intervention A in an AB trial, and C has been compared against A in an AC trial, so an indirect comparison can estimate the relative efficacy of B and C. The ITC of C versus B is 'anchored' on A. (We favor this more-descriptive term, rather than “adjusted,” which appears in the literature as well.) A could represent an active treatment comparator or placebo. Of key importance in an ITC is not to ‘break randomization’ (Glenny et al., 2005; Sutton et al. 2008; Song et al., 2009): For example, if A, B, and C are interventions for rheumatoid arthritis patients, it is incorrect to simply compare the observed fraction of responders on drug B in the AB trials with the observed fraction of responders on drug C in the AC trials. Using the data in this way fails to separate the efficacy of the drugs from possible placebo effects. (RCTs are designed to separate drug effects from other effects.) Another reason to avoid 'breaking randomization' is that differences in response may reflect different baseline risks, even where the relative risk is consistent between trials (Sutton et al. 2008; Jansen et al., 2008). Using data only from the treatment arms of interest to draw comparisons, omitting the data from the control or placebo arms, is called a “naïve indirect comparison,” results in bias, and should be avoided. In order to preserve the randomization within each trial, one must compare the relative treatment effects (e.g., compare the odds ratio for B versus A from the AB trials with the odds ratio for C versus A from the AC trials).
The second network in Figure 1 would permit an ITC of interventions B, C, D, and E, anchored on the common comparator A. Because these interventions are all connected in the network (i.e., each pair has a path from one to the other), indirect comparisons can be performed for B vs. C, B vs. D, B vs. E, C vs. D, C vs. E, and D vs. E. As an example of such a ‘star-shaped’ network, in a recent ITC of bisphosphonate therapy for osteoporosis the direct comparisons of four interventions were all against placebo (Jansen et al. 2009). For some of the interventions multiple placebo-controlled trials were available, and the analysis can be labeled a network meta-analysis. Another example is the ITC of intracoronary drug-eluting stents by Bondi-Zoccai et al. (2005).

In the third network not all trials have a common comparator, but all interventions are still connected. The additional interventions, F and G, are connected with A, B, C, D, and E by the EF trials and the FG trials, and an indirect comparison of each intervention with any other is possible (though comparisons with longer paths will have lower precision) (Lu and Ades 2004).

The fourth network structure consists of interventions A, B, and C (as in the first network), but now head-to-head RCT data are available for every comparison; the network of evidence consists of AB trials, AC trials, and BC trials. An important characteristic of this network is the ‘closed loop’: each comparison has both direct evidence and indirect evidence. For example, the BC comparison has direct evidence from the BC trials, and indirect evidence from the AB and AC trials (and similarly for the AB and AC comparisons). A network in which some of the pairwise comparisons have both direct and indirect evidence is called a MTC (Lu and Ades, 2004; Caldwell et al., 2005).

The fifth network also involves a MTC for interventions A, B, and C, but interventions A, C, E, and F form another, longer loop. For networks that contain loops it is important that the indirect comparisons be consistent with the direct comparisons, as discussed in the next section (Ades et al., 2003; Lu and Ades, 2004; Cooper et al., 2009). Recent examples of network meta-analysis with ‘loops’ include the network meta-analysis of first line antihypertensive therapies by Psaty et al. (2003), the study of stroke prevention among patients with atrial fibrillation by Cooper et al. (2006), and network meta-analysis of opioids for breakthrough cancer pain by Vissers et al. (2010). Salanti et al. (2008) provide an overview of different network structures of some of the recently published studies.

Whatever the structure of the network, pairwise comparisons, either direct or indirect or both, can be made between interventions that are connected. The terms ITC, MTC, and network meta-analysis are sometimes used interchangeably. We propose to use network meta-analysis when the evidence base consists of more than 2 RCTs connecting more than 2 interventions. If the network consist of at least one closed loop, labeling the analysis a MTC is appropriate. Any analysis of an open loop network can be called an ITC.
**Synthesis of the evidence**

**Assumptions**

Given a network of interventions and RCTs comparing them, the objective of the analysis is to synthesize the results from the individual RCTs, thereby obtaining (pooled) estimates of relative treatment effects for pairwise comparisons. Although the comparators shared by the RCTs form the basis of the network, the key question is whether the trials in the network are sufficiently similar to yield meaningful results for the ITC and MTC.

A traditional meta-analysis combines the results of several RCTs that compared the same interventions, say A and B, to get an overall estimate of relative effect (e.g., odds ratio, relative risk, or difference in change from baseline) and a corresponding estimate of uncertainty. It is important to realize that randomization holds within each RCT of A and B, but not across the RCTs. Thus, the trials may differ
on study and patient characteristics. If these characteristics are modifiers of the relative treatment effect of B versus A, then the studies are said to be heterogeneous. A meta-analysis can use fixed-effects or random-effects approaches to combine the results of the AB trials. If there is no heterogeneity, and one assumes that the observed differences among study results are due solely to chance, a fixed-effects approach can be used. However, if there is heterogeneity, a random-effects approach is needed (Borenstein et al., 2009) A random-effects approach typically assumes that true individual study results are exchangeable (i.e. the prior position of expecting underlying effects to be similar, but non-identical) and can be described as a sample from a normal distribution whose mean is the pooled relative effect and whose standard deviation reflects the heterogeneity. (Skene and Wakefield, 1990; Gelman et al, 1995; Higgins & Green 2009; Borenstein et al., 2009, Cappelleri et al., 2010).

Similarly, in a network meta-analysis of RCTs involving multiple treatment comparisons, the randomization holds only within the individual trials. Relative treatment effects for a particular pairwise comparison may exhibit heterogeneity. Also, if the trials differ among the direct comparisons (e.g., AB trials differ from AC trials) and these differences are modifiers of the relative treatment effects, then the estimate of the indirect comparison is biased (Song et al., 2003; Jansen et al., 2008, Cooper et al., 2009). In other words, if the distribution of interactions between relative treatment effects and covariates is not balanced across trials that are comparing different sets of interventions, the similarity assumption of an ITC is violated, and confounding biases the analysis (Jansen et al., 2008; Cooper et al., 2009). Figure 2 depicts the comparisons involved in the similarity assumption of an ITC. If the AB trials and the AC trials are comparable on effect modifiers, then an indirect estimate for the relative effect of C versus B ($d_{BC}$, which can be a difference in normally distributed data, or a log odds ratio, or log hazards ratio, etc) can be obtained from the estimates of the effect of B versus A ($d_{AB}$) and the effect of C versus A ($d_{AC}$): $d_{BC} = d_{AC} - d_{AB}$. In essence, this implies that the same true $d_{BC}$ is obtained as would have been estimated in a three arm ABC trial (Lu and Ades, 2004).

When direct and indirect evidence are combined for a particular pairwise comparison, it is important that the indirect estimate is not biased, and there is no discrepancy between the direct and indirect comparisons (Lu and Ades 2006; Salanti et al., 2008a; Cooper et al. 2009; Dias et al. 2010). Therefore, consistency between these direct and indirect comparisons should be accounted for. Figure 3 depicts the components involved in the consistency assumption. The network has both direct and indirect evidence for every pairwise comparison of interventions A, B, and C. (For example, $d_{BC}$ can be obtained from the BC trials, but also indirectly from the AC trials and the AB trials.) For consistency, the following equation needs to be satisfied: $d_{BC} = d_{AC} - d_{AB}$ (Lu and Ades 2006; Cooper et al., 2009). If there is an imbalance in modifiers of the relative treatment effects across studies for one or more of the comparisons,
the consistency assumption may not be justifiable. Consistency only applies to the loops of evidence. It is not meaningful to say, for example, that the AB comparison is consistent with the AC comparison. We can only say that the AB, AC, and BC comparisons are consistent. As a simple example of inconsistency in an ABC network with an AB trial, an AC trial, and a BC trial (assuming no sampling error), let’s assume that the odds ratio (OR) of response with treatment B relative to A is 0.4 (OR\textsubscript{AB}=0.4) and the OR with C versus A is 0.5 (OR\textsubscript{AC}=0.5) then we would expect the OR of C versus B to be OR\textsubscript{BC}=OR\textsubscript{AC}/OR\textsubscript{AB}=0.5/0.4=1.25. There is inconsistency if the BC trial shows OR\textsubscript{BC} \neq 1.25. (Of course, in actual MTC analysis there is always sampling error, and this kind of strict evaluation of consistency based on the point estimates is not appropriate. Here it aims to illustrate the inconsistency concept.)

In summary, heterogeneity pertains to variation in the same treatment effect among studies, whereas evidence inconsistency is the discrepancy between direct and indirect comparisons.

**Figure 2:** Similarity assumption in an indirect treatment comparison. AB trials and the AC trials are comparable on effect modifiers, and an unbiased indirect estimate for the relative effect of C versus B can be obtained from the estimates of the effect of B versus A and the effect of C versus A.

**Figure 3:** Consistency assumption in a mixed treatment comparison. AB trials, AC trials, and BC trials are comparable on effect modifiers, and for each pairwise comparison the direct and indirect estimates are consistent.
Interpreting Indirect Treatment Comparison for Health Care Decision-making – Part1: FOR COMMENT

174

**Analysis**

In order to synthesize the results of the different RCTs in the network and obtain relative efficacy estimates for all possible pairwise comparisons, an analysis method needs to be used that preserves randomization within trials and minimize bias due to lack of randomization across trials.

In Figure 1 an overview of the analysis methods are presented and discussed in more detail here below. Whatever the method of analysis, the pooling of individual study results and indirect comparisons should be based on relative effect measures (e.g. odds ratio, difference in change from baseline, hazards ratio, etc) in order to preserve randomization. If the network does not consist of loops, the results of the RCTs available for each of the direct comparisons can be combined by means of multiple traditional meta-analyses (e.g. a meta-analysis of the AB trials and a meta-analysis of the AC trials) followed by an indirect comparison of the pooled results of each of these meta-analyses (Bucher et al., 1997; Lumley et al., 2002).

If the network of interventions consists of a loop then the analysis method needs to combine estimates of the direct comparisons with estimates of the indirect comparisons. In the ABC network where for each of the pairwise comparisons we have RCTs (network 4 in Figure 1), the pooled relative efficacy of the BC comparison from the BC trials needs to be combined with the indirect estimate based on the AB trials and the AC trials. The same applies to the AB and AC comparisons. It is clear that the more complex the network is, the more burdensome and potentially confusing such a stepwise approach is.

As an alternative to multiple sequential meta-analyses and indirect comparisons, a statistical model can be defined that reflects the mathematical relations between the relative effect estimates of the direct and indirect comparisons in the complete network (Lu and Ades 2004). Given a network of A, B, and C comparisons the relative effect estimates can be expressed as follows: \( d_{BC} = d_{AC} - d_{AB} \) (assuming similarity/consistency assumptions hold). When this expression is generalized to any network with multiple different interventions the following is obtained: \( d_{bk} = d_{ak} - d_{Ab} \) with \( k \) the intervention and \( b \) the comparator. Depending on the network \( k \) can be intervention B, C, D, E, etc. Comparator \( b \) can be A, B, C, D, etc, as long as \( k \) is alphabetically after \( b \).

This expression implies that any estimate for a particular pairwise comparison can be expressed by the relative efficacy estimates of the intervention and comparator relative to an overall reference treatment A as long as all interventions are connected in one network. \( d_{AB}, d_{AC}, d_{AD}, \ldots, d_{Ak} \) are called *basic parameters* of the model that are estimated based on the available studies. \( d_{BC}, d_{BD}, d_{CD}, \ldots \) are *functional parameters* and can be calculated based on the (pooled) estimates for the basic parameters (Lu and Ades 2006). For a network involving \( K \) treatments and \( T \) types of comparisons, there are \( K-1 \) basic parameters and \( T-K+1 \) functional parameters.
To summarize, a network meta-analysis model is an extension of a traditional meta-analysis model consisting of not one, but K-1 parameters that need to be estimated to allow for multiple pairwise comparisons across a range of K interventions. Such a network meta-analysis model applies to networks with and without loops, that is, both ITCs and MTCs.

Network meta-analysis can be performed with fixed or random-effects models. With a fixed effects model, it is assumed that there is no variation in relative treatment effects across studies for a particular pair-wise comparison (Jansen et al., 2008; Borenstein et al. 2009). For any given treatment comparison, in a fixed-effects model, the following question arises: “What is the true treatment effect?” However, if there is heterogeneity - variation in true (or underlying) relative treatment effects for a particular pair wise comparison - random effects models must be used. With a random-effects approach it is assumed that the true relative effects across studies are not identical, but follow some distribution; they are considered exchangeable. With a random-effects model the heterogeneity is explicitly modeled, and in network meta-analysis this variance is often assumed to be constant for all pairwise comparisons. One could argue that with a random-effects model, the question asked is “What is the average of the true treatment effects, and how much do these effects vary across trials?” (Higgins & Green 2009).

Although a random effects model explicitly models heterogeneity, it does not explain heterogeneity. Extending network meta-analysis models with treatment*covariate interactions attempts to explain heterogeneity in relative treatment effects and estimates relative treatment effects for different levels of the covariate. As outlined above, network meta-analysis will be biased if there are differences in study level covariates across those studies that are indirectly compared and act as modifiers of the relative treatment effect (Jansen et al, 2008; Cooper et al., 2009). This implies that by taking into account these covariates with treatment by covariate interactions in a meta-regression model (i.e. a model that includes study level covariates) the impact of bias due to similarity and/or consistency violations can be reduced (Cooper, 2009). (Covariates that vary across studies but are not effect modifiers do not need to be taken into consideration in a meta-regression model.)

Unfortunately the number of studies in a network is often limited, and adjustment by incorporating study level covariates with meta-regression approaches may sometimes be questionable (Glenny et al., 2005; Berlin et al. 2002; Jansen et al. 2008). In addition, aggregate-level covariate adjustment might produce ecological bias limiting the interpretation of estimated results for subgroups (Greenland and Morgenstern 1989; Berlin et al., 2002; Lambert et al., 2002). In contrast, patient-level network meta-analyses usually have sufficient power to estimate meta-regression models thereby reducing inconsistency and providing the opportunity to explore differences in effect among subgroups. However, obtaining patient-level data for all RCTs in the network may be considered unrealistic.
Since, with a random-effect model the study specific treatment effects are explicitly modeled, a random-effects model ‘fits’ the data better than a fixed-effects model. Similarly, extending a fixed or random-effects model by incorporating treatment-by-covariate interaction terms can also improve model fit. However, for any given data set, the more parameters that need to be estimated, the more uncertain the estimates of these parameters will be. Hence, the objective is to use a model that sufficiently fits the data (and reduces confounding bias) but that provides stable parameter estimates. The choice of a fixed or random effect meta-analysis model, with or without covariate interactions can be made by comparing different competing models regarding their goodness of fit to the data. The goodness of fit can be estimated by calculating the difference between the deviance for the fitted model and the deviance for the saturated model (which fits the data perfectly). Within a frequentist framework the Akaike information criterion (AIC) which uses the likelihood function can be used for model selection. In a Bayesian framework the Bayesian information criterion (BIC) or deviance information criterion (DIC) can easily be used for the same purpose. (McCullagh and Nelder, 1989; Dempster, 1997; Spiegelhalter et al., 2002)

Network meta-analysis can be performed within a frequentist or Bayesian framework. In a frequentist approach, the result of the analysis is a point estimate with a 95% confidence interval. In a frequentist approach the 95% confidence interval, under repeated sampling, would contain the true population parameter 95% of the time. It must be noted that confidence intervals obtained with a frequentist approach cannot be interpreted in terms of probabilities. (The 95% confidence interval does not mean there is 95% probability that ‘true’ or population value is between the boundaries of the interval.) (Goodman, 1999).

Bayesian methods involve a formal combination of a prior probability distribution (that reflects a prior belief of the possible values of the pooled effect) with a (likelihood) distribution of the pooled effect based on the observed data, to obtain a posterior probability distribution of the pooled effect (Sutton et al. 2000) (The likelihood informs us about the extent to which different values for the parameter of interest are supported by the data.).

‘Frequentists’ use the sampling distribution as the basis of statistical inference which is proportional to the likelihood function (Spiegelhalter et al, 2004). The posterior distribution (as obtained with the Bayesian approach) can be interpreted in terms of probabilities, which allows for a more intuitive interpretation (e.g. “There is an x% probability that treatment A results in a greater response than treatment B”). This is different from the interpretation of the findings within a conventional frequentist approach.

A major advantage of the Bayesian approach is that the method naturally leads to a decision framework that supports decision-making (Luce et al., 1999; Spiegelhalter et al., 2004; Sutton et al., 2000). Other advantages of a Bayesian meta-analysis include the straightforward way to make predictions, and the
possibility to incorporate different sources of uncertainty (Spiegelhalter et al., 2004; Sutton et al., 2000). In order not to influence the observed results by the prior distribution, an often heard critique of the Bayesian approach, a non-informative prior distribution can be used for the treatment effect parameter(s). With such a ‘flat’ prior it is assumed that before seeing the data any parameter value is ‘equally’ likely. As a consequence, posterior results are not influenced by the prior distribution but are driven by the data as in a conventional frequentist meta-analysis. For a network meta-analysis, an additional advantage of a Bayesian approach is that it allows calculating the probability of which of the competing interventions is best, and other probability statements (Goodman, 1999). This aspect of a Bayesian analysis is providing information that is directly relevant to health care decision-makers (e.g. policy-makers & health care professionals/clinicians). However, as discussed below, there is a risk of over-interpreting this probability.

Critically reviewing and interpreting an indirect treatment comparison or network meta-analysis

To assist health care decision makers in using the findings of network meta-analyses, we describe in this section how to critically review and interpret such studies. The importance of correctly assessing results of network meta-analyses cannot be understated as these are intended to inform comparative effectiveness choices and are likely to have coverage implications. Understanding the validity of these studies is therefore critical. In the following section we briefly review issues related to internal and external validity of network meta-analyses. We provide a list of items that we recommend be reported for a network meta-analysis to allow proper evaluation and interpretation of findings to inform decision-making.

Internal and external validity

Decision-makers making use of results of ITC or network meta-analyses will need to assess whether the differences between treatments are most likely true, or whether they can be explained by bias in the analysis. The internal validity of the network meta-analyses is contingent upon three factors: 1) the appropriate identification of the studies that make up the evidence network; 2) the quality of the individual RCTs; and 3) the extent of confounding bias due to similarity and consistency violations.

Appropriate search and selection methods of all relevant RCTs must be conducted, although the delimitation of what constitutes the evidence network is a matter of current research (Hawkins et al., 2009a; Hawkins et al., 2009b). Nevertheless, even with rigorous and extensive literature search methods, the extent of publication bias must be assessed. It is well known that negative or small trials are less likely to be published so the evidence network may be limited accordingly (Salanti et al., 2008b). Furthermore, in a network of RCTs, specific comparisons can heavily outweigh less compared interventions, resulting in asymmetric networks (Salanti et al., 2008a). The validity of a network meta-analysis will also be contingent upon the internal validity of the single RCTs included in the evidence network. The inclusion of poor
quality trials may be an issue. Randomization does not guarantee that an RCT is unbiased. (Cappelleri et al., 1996; Abel & Koch, 1999; Song et al., 2003) There may be lack of adequate allocation concealment; patients may be excluded after randomization, which may result in an imbalance between groups; or lack of blinding of the outcome may overestimate the treatment effect. (Schulz et al., 1995). Thus, each RCT included in a network meta-analysis should be critically evaluated for bias.

After addressing the threats to internal validity associated with the development of the evidence network, the similarity between the trials included in the network will also be a determinant of the internal validity of the analyses. Studies may differ with respect to the characteristics of the patients, the way in which the outcomes were measured or defined, the protocol requirements including the concomitant interventions allowed, the length of follow-up as well as differential loss to follow-up, and the time frame during which the studies were conducted (Glenny et al., 2005).

As outlined earlier, an ITC or MTC is affected by confounding bias if there are differences across trials that are indirectly compared regarding relative treatment effect modifiers. This bias may be reduced by adjusting for these differences by incorporating treatment*covariate interactions in the statistical models used. However, one can only judge the similarity of trials and potentially adjust for bias regarding study level covariates that are measured. Hence, differences in baseline risks and placebo responses across trials should be assessed as these can reflect additional important differences in study or patient characteristics across studies.

The external validity of the network meta-analysis will naturally be limited by the external validity of the RCTs included in the evidence network, and health care decision-makers will need to review whether results can be extrapolated to the population of interest. It is important to remember that registration trials for regulatory purposes are more likely to include selective homogeneous populations, which compromises external validity (Schwartz and Lellouch, 1967; Rothwell, 2006). From a decision-making perspective, a certain degree of variation in the patient populations may be welcome for comparative and cost effectiveness evaluations, if it reflects real world practice. Hence, some heterogeneity across trials in the evidence network may arguably increase external validity, as long as the heterogeneity within direct comparisons is greater than the variation of effect modifiers across studies that are indirectly compared in order to avoid similarity violations as much as possible. Although we are not aware of any network meta-analysis that evaluated this explicitly, a possible approach is by means of an analysis of variance of the within versus between direct comparisons relative treatment effects.
**Box 1:** Simplified checklist to assist decision-makers in evaluating a reported network meta-analysis

<table>
<thead>
<tr>
<th>Report section</th>
<th>Checklist Item</th>
<th>What to look for in the paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>Are the rationale for the study and the study objectives stated clearly?</td>
<td>A clear rationale for the review</td>
</tr>
<tr>
<td></td>
<td>Are the outcome measures described?</td>
<td>Justification of outcome measures selected for analysis</td>
</tr>
<tr>
<td>Methods</td>
<td>Does the methods section include the following?</td>
<td>A systematic review of the literature in accordance with Centre for Reviews and Dissemination (CRD) guidelines and PRISMA</td>
</tr>
<tr>
<td></td>
<td>- Description of eligibility criteria</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Information sources</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Search strategy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Study selection process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Data extraction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- (validity assessment of individual studies)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is there a description of methods for analysis/synthesis of evidence? Do the methods described include the following?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Description of analyses methods/models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Handling of potential bias/inconsistency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Analysis framework</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Description and justification of statistical model(s) used: Multiple meta-analysis of pairwise comparisons vs. network meta-analysis models; Fixed vs. random effects models; Models without or with covariate (interactions)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Description of whether analyses were performed in frequentist vs. Bayesian approach.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Description of how possible bias / inconsistency was evaluated (either qualitative or quantitative, e.g. comparison of direct evidence with the indirect evidence). If meta-regression models are used rational for selection of covariates in models.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Description of comparative effectiveness estimates used for presentation of findings (e.g. odds ratio, relative risk, hazard ratio, difference in change from baseline).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Description of whether relative effect measures were transformed into expected (absolute) outcomes (e.g. proportion of</td>
<td></td>
</tr>
<tr>
<td>Report section</td>
<td>Checklist Item</td>
<td>What to look for in the paper</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
|                | Are sensitivity analyses presented?                                            | Rationale for and description of sensitivity analyses  
• Studies included  
• Prior distributions for model parameters in Bayesian framework |
| Results        | Do the results include a summary of the studies included in the network of evidence? | • Description of results of study identification and selection process  
• Table/list of studies with information regarding study design and patient characteristics (that might act as effect modifiers). These are important to judge potential similarity/consistency issues.  
• Figure of network of studies  
• Table with raw data by study and treatment as used for the analysis/model. (Optionally present relative effects of available direct comparisons of each study) |
|                | Does the study describe an assessment of model fit? Are competing models being compared? | Justification of model results |
|                | Are the results of the evidence synthesis (ITC/MTC) presented clearly?        | • Table/Figure with results for the pairwise comparisons as obtained with analyses; Point estimates and measure of uncertainty (95% intervals)  
• In Bayesian framework probability to reflect decision uncertainty (i.e. probability which treatment is best if multiple treatments are being compared, and probability that one treatment is better than the comparator) |
| Sensitivity/scenario analyses |                                                                                   | Description of (different) findings with sensitivity/scenario analysis. |
| Discussion     | Does the discussion include the following?                                     | • Summary of findings  
• Internal validity (individual trials; publication bias; differences across trials |
In Box 1, we present a simplified checklist of items that should be included in a report of a network meta-analysis in order to enable health care decision-makers to interpret the findings on comparative health outcomes. The checklist is not exhaustive but is intended as a general guide. Some caution should be exercised when using this list to judge the quality of published network meta-analyses, as this list focuses on reporting quality, and does not capture explicit items to judge or score the internal and external validity of a network meta-analysis.

A clear statement of the objectives should clarify what the decision problem is, with a specific focus on the patient population and the competing interventions of interest. The treatments that will be compared with the network meta-analysis may be limited to all drugs in a class, but can also include competing drugs of different classes, and in some cases, other medical interventions. Whatever the scope of interventions, a clear rationale for the choice should be described in the introduction section of the report.

In the methods section, the development of the evidence network should be described and should follow systematic review procedures that include an explicit search strategy in a variety of databases, and pre-specified inclusion and exclusion criteria for the study selection process. A protocol is recommended to list these elements as well as pre-specify the outcomes to be analyzed to avoid outcome selection bias.

Rigorous data extraction methods should be used, including whether double data extraction was performed (and how disagreements were resolved) and how missing data were handled. These methods have been described in detail elsewhere (Centre for Reviews and Dissemination Handbook) and should be reported following the PRISMA statement (Liberati et al., 2009; Moher et al., 2009).

The data analysis section should provide a comprehensive overview of the statistical methods used, including the justification of outcomes and endpoints, relative effect estimates, choice of fixed or random effects models and whether the models were extended with study level covariates in an attempt to improve similarity and reduce inconsistency. If the analyses are performed within a Bayesian framework the choice
of prior distributions for the model parameters that are going to be estimated should be defined. A description of different sensitivity analyses pertaining to studies included in the networks and prior distributions (if analyses are performed within a Bayesian framework) should be reported.

It is not the mandate of the Task Force to be prescriptive in recommending elements to be reported in the results section. Nevertheless, we recommend that, at a minimum, the elements in the following section be reported for users of a network meta-analysis to be able to judge the credibility of the analyses.

A list of the specific studies identified by the systematic review and those included in the network meta-analysis (in some instances, these will differ if there were insufficient data reported to include in the actual analysis) should be provided.

A flow diagram that illustrates the way in which trials were selected can be helpful. A list of key patient and study characteristics of each study should be provided in table format and preferably discussed. This is essential in order to judge whether there are differences across trials that might act as effect modifiers thereby causing bias in the analysis. The reader is referred to the PRISMA statement for specific recommendations on how to report the results of a systematic review. A graphical representation of the network with labels of the different RCTs is very helpful and will improve transparency of the analyses.

The source of data for the outcomes of interest from individual trials should be reported, including point estimates as well as measures of uncertainty. Although the network meta-analysis will be performed using the relative effect measures of the different trials, either relative effect measures or outcomes by treatment arm can be presented in tabular format for the individual studies. The advantage of presenting the results by treatment arm in a format like Table 1 is that it facilitates the understanding of the network. It also facilitates the comparison of placebo responses across trials, as outlined above. However, presenting relative treatment effects for each of the RCTs in a table (See Table 2) or figure (e.g. forest plot), helps relate the pooled results of the network meta-analysis to the individual study results.

In the section of the report where the results of the network meta-analysis are presented, competing models have to be compared in terms of their goodness of fit to the data, and residual deviance calculations may be provided to justify the study’s choice of the base case model. As a minimum, the estimates of relative treatment effects (e.g., odds ratios, hazard ratios, differences in means) along with 95% confidence or credible intervals (depending on the framework of analysis) compared to a common reference treatment or anchor should be reported for clarity reasons (see Table 3). However, in order to appreciate the value of a network meta-analysis, results of all (relevant) pairwise comparisons (as a reflection of the functional parameters) are recommended to be presented (Table 4). As an alternative to tables, forest plots can be very informative to present pairwise comparisons, as illustrated by Mills et al. (2009) and Vissers et al. (2010). The companion paper provides an example how to plot results graphically as well.
It should also be noted that it may be useful to decision-makers to report estimates of relative treatment on a different scale than that used for the model analysis. For example, it may be useful to report results from a network meta-analysis conducted on an odds ratio as relative risks, absolute risk differences, and numbers needed to treat. These estimates will depend on the estimated probability of response for the reference treatment. Analyses using the (Bayesian) statistical software package WinBUGS (Spiegelhalter et al., 2003) facilitate the derivation of estimates of relative treatment effects on difference scales.

If the analyses are performed within a Bayesian framework, the uncertainty in the comparative effectiveness estimates can be translated into probabilities of decision uncertainty. For example, the odds ratio along with the 95% credible intervals for each of the interventions relative to a common anchor enables the calculation of the probability that each treatment is the most efficacious regarding the outcome of interest out of all treatments compared. For example in Table 3, there is a 38.5% probability that anidulafungin is the most efficacious in preventing mortality out of the eight interventions compared. (Before considering the available evidence, each treatment would have an a priori change of 12.5%.) Although this illustrates a great advantage of the Bayesian approach for multiple treatment comparisons, caution should be applied where only the probability of a treatment being ranked first is provided as this does not provide information of the ‘spread’ of rankings for a treatment. For example, a treatment for which there is little trial data and consequently a wide confidence interval, may have a probability approaching 50% of being the best treatment, but may also have probability of 50% of being the worst treatment (See Table 3 itraconazole). Hence, in addition, it is useful to calculate the expected ranking of efficacy for all treatments based on the probabilities of all treatment rankings (i.e., probability of being the best, probability of second best, and so on)

In addition to estimates of relative treatment effects it may be useful to report estimates of the absolute probability of outcome, at least for binary outcomes. This will require an estimate of the baseline probability for the anchor treatment. This may be derived from the trial data, other sources, or may be subject to sensitivity analyses. The method used to estimate the baseline probability should be clearly stated. For example, Mills et al. (2009) reported the expected mortality based on the results of the network meta-analysis. (See Table 3)

The discussion section of a report should present a critical assessment of the results with respect to internal and external validity as discussed above. Authors should provide a thoughtful discussion of the assumptions of similarity and consistency and whether these can be assumed to hold for the analysis at hand. The discussion should also address whether the network meta-analysis results are in line with expectations based on previous meta-analyses and other (observational) evidence available (GRADE
Furthermore, an explanation for observed differences between compared interventions from both biological and clinical perspectives is recommended. Apart from the appropriateness of the results, the relevance of the findings for real world clinical (and reimbursement) decision-making should be addressed.

**Interpretation of findings**

After assessing the validity and results of a network meta-analysis, decision-makers will want to carefully consider whether the findings can be applied to their decision problems. Is one treatment better than another and do these results apply to the population of interest to the decision-maker?

Frequently a small number of studies in an ITC or network meta-analysis limit the possibility to adjust for possible bias due to similarity issues by means of statistical techniques. Rather than immediately ignoring the results of the analysis by claiming that trials are not comparable, the decision-maker should make an attempt to hypothesize the possible direction of bias in the indirect estimates. An important question to ask is how different a non-biased indirect comparison would be, and whether this would lead to a different conclusion and decision.

An issue to consider is whether a treatment can be considered more efficacious than another when only a limited number of outcomes have been analyzed. Selection of outcomes analyzed must be clearly justified at the outset of the analysis (for example, do these reflect primary outcomes used in clinical trials?). Synthesizing the results of multiple ITCs or MTCs must be considered. How do we interpret situations where Drug A is better on a number of clinical outcomes but not all of these outcomes? How is the decision made in these cases? A possible approach is to weigh the different endpoints based on the relative importance according to the decision maker and calculate probabilities of which treatment is best, taking into account these weights (Jansen et al., 2010). These are not issues specific to network-meta-analysis. Indeed, the development of measures such as the quality adjusted life year (QALY) has been fuelled by the need to compare disparate health outcomes using a common metric by means of decision models. Nevertheless, it is an issue to consider when interpreting network meta-analysis, as well.

Furthermore identification of the ‘best’ or most appropriate treatment cannot be made on the basis of efficacy endpoints alone. To inform health care decision making for clinical treatment guidelines and reimbursement policies, the efficacy findings of a network meta-analysis must be interpreted in light of other available (observational) evidence and other characteristics of the competing interventions, such as safety and convenience.

The general development of evidence synthesis methods to provide systematic exhaustive evidence on which to base decisions has a natural place in comprehensive decision models that include both costs and effects and are used to determine the cost-effectiveness of interventions by such bodies as the National...
Institute for Health and Clinical Effectiveness (NICE) (Petitti, 2000; Cooper et al., 2004; Ades et al., 2006; Sutton et al. 2008). Network meta-analysis represents a valuable set of analytical tools to inform clinical evidence in cost-effectiveness analysis.

**Decision-making in the absence of direct and indirect treatment comparisons of RCTs**

Pragmatic randomized naturalistic head-to-head trials are arguably the gold standard to obtain comparative effectiveness estimates given their high internal and external validity. (Tunis et al., 2003). However, these trials take a long time to complete and can never provide relative efficacy information for all competing interventions, especially when new treatments are developed continuously. Hence, an ITC or network meta-analysis can be considered a useful and realistic alternative. To minimize bias, a network meta-analysis requires RCT evidence. However, evidence from RCTs may not be available in a significant proportion of situations that decision-makers face. For example, in oncology Phase II trials often have a single arm. A study by the Institute of Medicine in 2007 estimated that, in the United States, less than half of all treatments delivered were supported by evidence. Furthermore, a review of practice guidelines found that few recommendations were based on high-quality evidence and many were based on expert opinion, individual case studies, and standards of care (Tricoci, 2009). There are often good reasons for the absence of RCTs. Time may be too short to conduct RCTs of rapidly emerging technologies (Chambers et al. 2009). RCTs may be unethical if clinicians believe there is a causal association between the intervention and the outcome, for example, between sleeping position and sudden infant death. Limited resources can also be a factor for the lack of RCT evidence.

Whatever the reasons for the absence of RCTs, and therefore the absence of indirect treatment comparisons based on randomized evidence, health care payers, health care professionals and patients may need to make decisions. It is wrong to assume that these stakeholders can postpone the decision and wait for the ‘appropriate’ evidence. In particular, decisions have to be made on the available set of possible choices. In reality, not covering or prescribing an intervention is a tacit decision to stay with the status quo. This decision has societal implications. It may or may not maximize health benefits for the population (if made by the decision maker) or health benefits for the patient (if made by the health care provider and patient) (Claxton, 1999).

A critical question for decision-makers, then, is whether to use observational comparative studies if RCTs or indirect comparisons of RCTs are not available. In order to answer this question, it is important to remember that in an ITC or MTC of RCTs, the value of randomization does not hold across trials. If study or patient characteristics differ among trials for interventions indirectly compared, and are modifiers of relative treatment effects, the analysis will be biased. Hence, an ITC or network meta-analysis of RCTs is a form of observational evidence, but less prone to (confounding) bias than is a cohort study (or any other
observational design). A cohort study is biased if differences in unmeasured covariates affect both the
intervention and the outcome, whereas an ITC or network meta-analysis of RCTs is biased only if
differences in unmeasured covariates among trials are modifiers of relative treatment effects, which is
arguably much more unlikely. Thus, asking whether comparative observational studies should be used in
the absence of an ITC or network meta-analysis of RCTs is synonymous with asking what level of
observational evidence can be considered to have sufficient internal validity to inform decision-making. Or,
more specifically, what level of observational evidence are decision-makers comfortable with? Is the
minimum acceptable level of observational evidence an ITC or network meta-analysis of RCTs, or is a
cohort study sufficient?

In order to answer such questions, decision-makers must recognize that, the lower the internal
validity, the greater the risk of biased results, and therefore the greater the risk of making an inferior
decision. If the new treatment is chosen over the standard treatment because of biased estimates of
comparative effectiveness, and the true outcomes favor the standard treatment, then health benefits are
foregone.

The debate over the proper use of RCT evidence, indirect comparison of RCTs, and ‘traditional’
observational studies is likely to continue. Fortunately, the needs for comparative effectiveness research are
driving developments in evidence synthesis and its understanding. At this stage, we conclude that, in the
absence of (head-to-head) RCTs, decision-makers can use observational evidence as long as they are aware
of the potential risks in using evidence of lower quality and are comfortable taking these risks. If decision-
makers prefer to wait for head-to-head RCTs or, the next best thing, an ITC of RCTs, they must realize that
they are choosing the ‘old’ treatment over the ‘new’ treatment, with potential societal implications.

In essence, for every appraisal of a new intervention, a decision-maker must make the trade-off
between, on the one hand, the risk of making the wrong decision and therefore losing health benefits by
relying on lower-quality evidence, and, on the other hand, postponing the decision and therefore possibly
also forgoing health benefits. This trade-off is influenced by considerations including the burden of disease
and the number of currently available treatments. Whatever the outcome of the debate, the quality of
decision making will be increased by being transparent and explicit about which type of evidence is being
used and evaluating its limitations and consequences.

Conclusion

This paper, the first part of the report from the Task Force, outlines the key concepts of ITC and
MTC and provides guidance for reviewing and interpreting these studies to inform decision-making. ITC
and network meta-analysis can be considered an extension of traditional meta-analysis by including
multiple different pair-wise comparisons across a range of different interventions to allow for multiple
treatment comparisons in the absence of head-to-head evidence. Furthermore, the methodology can
combine direct with indirect treatment comparisons thereby synthesizing a greater share of the available
evidence than traditional meta-analysis. Although the evidence networks underlying network meta-analysis
typically include RCTs, randomization does not hold across trials and there is a risk of confounding bias,
compromising internal validity. Accordingly, an ITC or MTC must be considered observational evidence,
but is arguably less prone to confounding bias than an observational comparative (prospective) cohort
study. While the methodological issues regarding indirect comparisons and network meta-analysis are
recognized, application of this method is expected to continue and grow, simply because failing to view
accumulation of information as an evolving process would undermine the role played by scientific evidence
in shaping health care decision-making. For that reason, the goal of the Task Force is to help educate policy
makers and health care professionals about these studies and identify areas for future research.
References


Caldwell DM. Ades AE, and Higgins JPT; Simultaneous Comparison of Multiple Treatments: Combining Direct and Indirect Evidence; British Medical Journal 2005;331: 897-900.


Chambers D, Rodgers M, Woolacott N. Not only randomized controlled trials, but also case series should be considered in systematic reviews of rapidly developing technologies. J Clin Epidemiol. 2009;62:1253-1260.


Cooper NJ, Sutton AJ, Lu G. Mixed Comparison of Stroke Prevention Treatments in Individuals with Nonrheumatic Atrial Fibrillation; Archives of Internal Medicine 2006; 166:1269-1275.


Glenny AM; Statistical Methods for Indirect Treatment Comparisons; Health Technology Assessment; 2005;9:17 – 49


Rothwell PM. Factors That Can Affect the External Validity of Randomized Controlled Trials. PLOS Clin Trials 2006;1:e9.

Interpreting Indirect Treatment Comparison for Health Care Decision-making – Part 1: FOR COMMENT


Spiegelhalter D, Abrams K, Myles J. Bayesian Approaches to Clinical Trials and Health Care Evaluation. Chichester: John Wiley & Sons, 2004a


27
Interpreting Indirect Treatment Comparison for Health Care Decision-making – Part 1: FOR COMMENT

716 Systematic Reviews: CRD’s guidance for undertaking reviews in health care
717 Centre for Reviews and Dissemination, 2009


722 Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003;290:1624-32


Table 1: Example table how source data as used in network meta-analysis can be presented - Data by treatment arm

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Fluconazole</th>
<th>Caspofungin</th>
<th>Amphotericin B Deoxycholate</th>
<th>Amphotericin B Liposomal</th>
<th>Voriconazole</th>
<th>Micafungin</th>
<th>Anidulafungin</th>
<th>Itraconazole</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abele-Horn et al. 1996</td>
<td>15/36</td>
<td>14/36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaissie et al. 1996</td>
<td>9/75</td>
<td>9/67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kujath et al. 1993</td>
<td>6/20</td>
<td>5/20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rex et al. 1994</td>
<td>34/103</td>
<td>41/103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phillips et al. 1997</td>
<td>13/50</td>
<td>11/53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mora-Duarte et al. 2002</td>
<td>39/109</td>
<td>38/115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kullberg et al. 2005</td>
<td>51/122</td>
<td></td>
<td>88/248</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>van’t Wout et al. 1991</td>
<td></td>
<td>6/16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4/16</td>
</tr>
<tr>
<td>Pappas et al. 2007</td>
<td>27/188</td>
<td></td>
<td></td>
<td></td>
<td>57/390</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuse et al. 2007</td>
<td></td>
<td>62/267</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61/264</td>
</tr>
<tr>
<td>Reboli et al. 2007</td>
<td>37/118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29/127</td>
</tr>
</tbody>
</table>

Source: based on Mills et al., 2009
Table 2: Example table how source data as used in network meta-analysis can be presented - Relative efficacy data by study

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Odds Ratio</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abele-Horn et al. 1996</td>
<td>0.89</td>
<td>(0.35, 2.27)</td>
</tr>
<tr>
<td>Anaissie et al. 1996</td>
<td>1.14</td>
<td>(0.42, 3.06)</td>
</tr>
<tr>
<td>Kujath et al. 1993</td>
<td>0.78</td>
<td>(0.19, 3.13)</td>
</tr>
<tr>
<td>Rex et al. 1994</td>
<td>1.34</td>
<td>(0.76, 2.37)</td>
</tr>
<tr>
<td>Phillips et al. 1997</td>
<td>0.75</td>
<td>(0.30, 1.86)</td>
</tr>
<tr>
<td>Mora-Duarte et al. 2002</td>
<td>0.89</td>
<td>(0.51, 1.54)</td>
</tr>
<tr>
<td>Kullberg et al. 2005</td>
<td>0.77</td>
<td>(0.49, 1.19)</td>
</tr>
<tr>
<td>van’t Wout et al. 1991</td>
<td>0.56</td>
<td>(0.12, 2.55)</td>
</tr>
<tr>
<td>Pappas et al. 2007</td>
<td>1.02</td>
<td>(0.62, 1.67)</td>
</tr>
<tr>
<td>Kuse et al. 2007</td>
<td>0.99</td>
<td>(0.66, 1.49)</td>
</tr>
<tr>
<td>Reboli et al. 2007</td>
<td>0.65</td>
<td>(0.37, 1.14)</td>
</tr>
</tbody>
</table>

Source: based on Mills et al., 2009
Table 3: Example table how results of a network meta-analysis can be presented - Relative effectiveness of each treatment relative to overall comparator (fluconazole) expressed as odds ratios (with 95% credible intervals); expected mortality (with 95% credible intervals); and probability best as a measure of decision uncertainty

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Odds Ratio</th>
<th>95% Credible Interval</th>
<th>Expected mortality</th>
<th>95% Credible Interval</th>
<th>Probability best out of all treatments compared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluconazole</td>
<td>1</td>
<td>reference</td>
<td>28.4%</td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td>Caspofungin</td>
<td>1.34</td>
<td>(0.65, 2.48)</td>
<td>33.8%</td>
<td>(18.6%, 55.3%)</td>
<td>0.008</td>
</tr>
<tr>
<td>Amphotericin B Deoxycholate</td>
<td>1.14</td>
<td>(0.78, 1.62)</td>
<td>30.9%</td>
<td>(21.8%, 42.4%)</td>
<td>0.001</td>
</tr>
<tr>
<td>Amphotericin B Liposomal</td>
<td>1.81</td>
<td>(0.71, 3.90)</td>
<td>40.0%</td>
<td>(26.9%, 58.5%)</td>
<td>0.004</td>
</tr>
<tr>
<td>Voriconazole</td>
<td>0.9</td>
<td>(0.48, 1.53)</td>
<td>25.8%</td>
<td>(7.8%, 60.1%)</td>
<td>0.09</td>
</tr>
<tr>
<td>Micafungin</td>
<td>1.73</td>
<td>(0.74, 3.49)</td>
<td>39.2%</td>
<td>(27.6%, 55.4%)</td>
<td>0.001</td>
</tr>
<tr>
<td>Anidulafungin</td>
<td>0.67</td>
<td>(0.36, 1.14)</td>
<td>20.8%</td>
<td>(3.5%, 66.4%)</td>
<td>0.385</td>
</tr>
<tr>
<td>Itraconazole</td>
<td>0.83</td>
<td>(0.11, 2.94)</td>
<td>21.8%</td>
<td>(0.4%, 99.2%)</td>
<td>0.504</td>
</tr>
</tbody>
</table>

Source: based on Mills et al., 2009
Table 4: Example table how results of a network meta-analysis can be presented - Relative mortality for pairwise comparisons expressed as odds ratios (with 95% credible intervals)

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Comparator</th>
<th>Fluconazole</th>
<th>Caspofungin</th>
<th>Amphotericin B Deoxycholate</th>
<th>Amphotericin B Liposomal</th>
<th>Voriconazole</th>
<th>Micafungin</th>
<th>Anidulafungin</th>
<th>Itraconazole</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluconazole</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caspofungin</td>
<td>1.34</td>
<td>1</td>
<td></td>
<td>(0.65, 2.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphotericin B Deoxycholate</td>
<td>1.14</td>
<td>0.92</td>
<td>1</td>
<td>(0.78, 1.62)</td>
<td>(0.50, 1.54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphotericin B Liposomal</td>
<td>1.81</td>
<td>1.35</td>
<td>1.58</td>
<td>(0.71, 3.90)</td>
<td>(0.77, 2.20)</td>
<td>(0.68, 3.17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voriconazole</td>
<td>0.9</td>
<td>0.72</td>
<td>0.78</td>
<td>(0.48, 1.53)</td>
<td>(0.33, 1.38)</td>
<td>(0.49, 1.19)</td>
<td>0.57</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Micafungin</td>
<td>1.73</td>
<td>1.28</td>
<td>1.51</td>
<td>(0.74, 3.49)</td>
<td>(0.86, 1.85)</td>
<td>(0.72, 2.81)</td>
<td>0.98</td>
<td>2.02</td>
<td>1</td>
</tr>
<tr>
<td>Anidulafungin</td>
<td>0.67</td>
<td>0.56</td>
<td>0.6</td>
<td>(0.36, 1.14)</td>
<td>(0.21, 1.22)</td>
<td>(0.29, 1.12)</td>
<td>0.44</td>
<td>0.81</td>
<td>0.45</td>
</tr>
<tr>
<td>Itraconazole</td>
<td>0.83</td>
<td>0.67</td>
<td>0.72</td>
<td>(0.11, 2.94)</td>
<td>(0.08, 2.44)</td>
<td>(0.10, 2.50)</td>
<td>0.53</td>
<td>0.97</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Source: based on Mills et al., 2009