

Get real in Network Meta-analysis: a review of the methodology

Orestis Efthimiou^{a,1}, Thomas P. A. Debray^{b,c}, Gert van Valkenhoef^{cd}, Sven Trelle^{ef}, Klea Panayidou^e, Karel G. M. Moons^{b,c}, Johannes B. Reitsma^{b,c}, Aijing Shang^g and Georgia Salanti^a on behalf of GetReal methods review group²

^a *Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece*

^b *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands*

^c *The Dutch Cochrane Centre, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands*

^d *Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands*

^e *Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland*

^f *CTU Bern, Department of Clinical Research, University of Bern, Bern, Switzerland*

^g *F. Hoffmann-La Roche AG*

Abstract

Pairwise meta-analysis is an established statistical tool for synthesizing evidence from multiple trials but it is informative only about the relative efficacy of two specific interventions. The usefulness of pairwise meta-analysis is thus limited in real-life medical practice, where many competing interventions may be available for a certain condition and studies informing some of the pairwise comparisons may be lacking. This commonly encountered scenario has led to the development of network meta-analysis (NMA). In the last decade several applications, methodological developments and empirical studies in NMA have been published and the area is thriving as its relevance to public health is increasingly recognized. This article presents a review of the relevant literature on NMA methodology aiming to pinpoint the developments that have appeared in the field.

Keywords: mixed-treatment comparison; multiple treatment meta-analysis; indirect treatment comparison; comparing multiple interventions;

¹ Corresponding author. Email address: oremiou@gmail.com

² A list of the members of the GetReal methods' review group is given in the Supporting Information.

1 Introduction

When developing a new treatment for a particular disease, a series of clinical trials are conducted where its efficacy is compared to a placebo and/or to a competing treatment that is already implemented in regular care. As a consequence, when reviewing estimates of treatment efficacy researchers often encounter clinical trials comparing different subsets of all the available treatments for the disease, forming a ‘network’ of evidence (Salanti et al., 2008a). In this situation a standard meta-analysis restricted to head-to-head comparisons cannot give a definite answer as to which treatment works best for a specific condition. Network meta-analysis can then be used to synthesize the totality of the available information and provide clinically relevant estimates to better support decision-making (Caldwell et al., 2005; Coleman et al., 2012; Higgins and Whitehead, 1996; Lu and Ades, 2004; Lumley, 2002).

In the past few years NMA has become increasingly popular (Coleman et al., 2012; D’Ascenzo and Biondi-Zoccai, 2014; Donegan et al., 2010; Glenny et al., 2005; Nikolakopoulou et al., 2014; Sobieraj et al., 2013) and several published articles have discussed its advantages and limitations (Cipriani et al., 2012; Cooper et al., 2011; Cucherat and Izard, 2009; EUnetHTA – European network for Health Technology Assessment (last), 2013; Hoaglin et al., 2011; Ioannidis, 2009; Jansen et al., 2011a, 2008; Malone, 2007; Mills et al., 2013b, 2011a; Naci and Fleurence, 2011; Salanti, 2012; Sutton et al., 2008, 2009; Wells et al., 2009; Welton et al., 2012), which have also been explored in empirical assessments (Jonas et al., 2013; Madan et al., 2011; O’Regan et al., 2009; Song et al., 2008, 2003) and simulation studies (Glenny et al., 2005; Jonas et al., 2013; Song et al., 2012). In particular, NMA allows an increase in the precision of estimates of relative effects (as compared to estimates based on direct evidence alone) and it can be used to compare interventions that have never been compared directly, head-to-head. This is especially valuable when active agents are only compared to placebo or standard care for regulatory purposes but not to each other (Falissard et al., 2009). In addition, NMA can support (policy-related) decision-making by explicitly providing a ranking of all competing interventions (Salanti et al., 2011) and to reduce the uncertainty in input parameters for subsequent formal cost-effectiveness models (Thorlund et al., 2014).

Despite the aforementioned advantages, the implementation of NMA may be hindered due to several reasons. First, the methodology of NMA rests on the assumption that different sources of evidence (direct and indirect evidence for the same treatment comparison) are in agreement. This assumption is often viewed as an important limitation of the method because it may be difficult to justify in practice. Moreover, the field of NMA is swiftly evolving; during the last few years there has

been an abundance of published articles presenting alternative approaches to deal with issues related to NMA. Hence, an updated review of the methodology is deemed timely to ensure that interested researchers use state-of-the-art methods for practical applications and also when conducting further methodological research. The present article aims to give a comprehensive account of the currently available methods for NMA, discuss in depth conceptual and statistical ways of evaluating the underlying assumptions of the model while providing guidance for researchers that set out to perform an NMA.

This review is part of work undertaken within the auspices of *GetReal: Incorporating real-life data into drug development*, a project funded by the Innovative Medicines Initiative (IMI) – a European public-private initiative aiming to speed up the development of better and safer medicines for patients. The aim of the GetReal project is to explore how drug development can become more efficient by better incorporating evidence of relative effectiveness in the process, and to propose ways for enriching the decision-making by regulatory authorities and HTA. This article is part of a series of three literature reviews (Debray et al., 2015; Panayidou et al., 2015).

2 Literature review methods and results

Search strategy

For the purposes of this review we searched for published articles that presented new methodology for NMA or articles that evaluated existing methods. We based our search on a previous review of the literature in NMA performed by the ‘*Comparing Multiple Interventions Methods Group*’ of the Cochrane Collaboration, which resulted in a publicly available database (www.zotero.org/groups/network_meta-analysis_methods/items/). This database has been shared with experts in the field to identify missing relevant articles. We also used the results from a recent literature review performed by Donegan et al. (Donegan et al., 2013b). In addition, we searched the MEDLINE database and also hand-searched key journals. Details on the search strategy we employed can be found in Appendix 2.

Inclusion criteria

We included articles that contributed to the methodology of network meta-analysis by introducing new methods and models; articles that reviewed the existing methodology and articles that provided recommendations or gave guidance on how to perform an NMA. We also included papers that discussed the conceptual issues and the assumptions behind NMA and articles that provided some sort of empirical assessment for the conduct of network meta-analyses in general.

Exclusion Criteria

We excluded publications for which one of the following criteria was met:

- full text of the publication was not available
- published in a language other than English
- conference posters
- applications of NMA without a methodological focus

Results

A total of 179 articles were included in our database. The identified articles were organized according to their context and are discussed in the relevant sections of this review. The included papers were classified using a number of tags. Each article was assigned one or more tags according to the type of research presented, one or more tags according to the methodological topics addressed and one or more tags according to the software it used to implement the methods it presented. An online database of all included articles tagged by topic can be found at https://www.zotero.org/groups/wp4_network_meta-analysis/items. More details on the results and the categorization we used are discussed in Appendix 2.

In the next sections we present a review of the methods for NMA as identified in our literature search. In Section 3 we present key concepts and the basic methodology for NMA. We present the underlying assumptions of the model in Section 3.1 and we discuss popular approaches for fitting NMA in Section 3.2. We present methods for detecting inconsistency in the network in Section 4. In Section 5 we give a list of the software options available for fitting NMA models and evaluating inconsistency. In Section 6 we present special issues of the NMA methodology. We discuss the use of different effect sizes (Section 6.1) and how to model time-to-event data (Section 6.2). In Section 6.3 we give methods to allow the inclusion of covariates in the analysis. In Section 6.4 we explore ways for adjusting for possible sources of bias and in Section 6.5 we discuss the reporting of the results of an NMA. We review available methods for analyzing multiple outcomes and repeated measures in Section 6.6 and we discuss the definition of nodes in the network of treatments in Section 6.7. In Section 6.8 we review the use of Individual Participant Data (IPD) in an NMA and in Section 6.9 we discuss the use of data from non-randomized studies in the analysis. Finally in Section 6.10 we discuss how NMA can inform the planning of future studies.

3 Basic methodology in NMA

3.1 Conceptual issues and underlying assumptions

The key feature of NMA is that it allows the synthesis of direct and indirect estimates for the relative effects of many competing treatments for the same health condition. Two treatments A and B may have been *directly* compared in head-to-head studies (in ‘AB’ studies). An *indirect* estimate may also be obtained from studies comparing these two treatments with a common comparator treatment C, i.e. AC and BC studies (Bucher et al., 1997). If both direct and indirect estimates are available, they can be combined in order to estimate a *mixed* treatment effect, as shown in the left panel of Figure 1. In practice, for most health conditions there is a plethora of interventions being compared across various randomized trials, forming a network of evidence. For a given treatment comparison within such a network there may be direct and many different indirect estimates, obtained via many different comparators as shown in the example of the right panel of Figure 1. Using NMA one can synthesize all these different pieces of information to produce an internally consistent overall estimate of all treatments’ relative effects.

FIGURE 1 HERE

Despite the benefits of NMA discussed in the Introduction, there is still controversy among researchers about the validity of using indirect treatment comparisons (*indirect evidence*) for decision making. The use of such evidence is particularly challenged when direct treatment comparisons (*direct evidence*) are also available (Edwards et al., 2009; Gartlehner and Moore, 2008; Ioannidis, 2006). One focal point of criticism is the nature of evidence NMA provides. Even though patients within an RCT are randomized to each of the treatments being compared, the treatments are not randomized across the included trials. Therefore, indirect comparisons are non-randomized comparisons and in fact provide observational, rather than randomized, evidence. As a consequence indirect treatment comparisons may be more subject to biased treatment effect estimates, due, for example, to confounding (e.g. when randomized AB and AC studies are systematically different than BC (Caldwell et al., 2005)) and selection bias (e.g. when the choice of comparator in a study is dependent on the relative treatment effect (Salanti et al., 2008b)). Such considerations are also closely related to the underlying assumptions of NMA; in what follows we discuss these assumptions in detail.

3.1.1 Transitivity

The objective of an NMA is to enhance the decision-making process regarding the choice among alternative treatments for a certain disorder and a target population. Thus, the estimands one aims to estimate in an NMA are the average relative treatment effect sizes between the competing treatments,

as they are expected to exist in the targeted population. The estimates produced by an NMA model for these parameters will be unbiased and consistent provided that the studies included in the dataset yield unbiased estimates and form a representative sample of the population of interest, and when the assumptions employed by the NMA model are valid. NMA adopts the same set of assumptions as a pairwise meta-analysis (Dias et al., 2010b) but it also employs an additional assumption which can be hard to assess (Song et al., 2009) termed transitivity (Baker and Kramer, 2002) (also called similarity (Donegan et al., 2010; Song et al., 2003) or exchangeability (Dias et al., 2013d)). Transitivity implies that information for the comparison between treatment A and B can be obtained via another treatment C, using the comparisons A vs. C and B vs. C. This assumption cannot be tested statistically, but its validity can be evaluated in a conceptual and epidemiological manner (Salanti, 2012).

The transitivity assumption implies that we can combine the direct evidence from AC and BC studies to learn (indirectly) about the comparison AB. This, however, will be questionable if there are important differences in the distribution of the effect modifiers (variables or characteristics which modify the observed relative effects, e.g. mean age of the participants, treatment dosage etc.) across the AC and BC trials which inform the indirect comparison (Baker and Kramer, 2002; Jansen and Naci, 2013). An effect modifier might differ across studies of the same comparison (e.g. mean participant age might be different across the AC trials) but if it has a similar distribution across comparisons (AC and BC) the transitivity assumption may still be valid (Salanti, 2012). Consequently, the plausibility of the transitivity assumption can be evaluated by putting the collection of studies under scrutiny for important differences in the distribution of effect modifiers. If the studies are deemed to be similar then the transitivity assumption might be realistic, provided that there are no unknown modifiers of the relative treatment effect (Donegan et al., 2013b). Obviously, such an evaluation of transitivity may be impossible when the effect modifiers are not reported or when there are few studies per treatment comparison (Cipriani et al., 2013). If important differences are identified and there are enough data available, a network meta-regression can be used to improve the transitivity of the network (see also Section 6.3).

The transitivity assumption also implies that the common comparator treatment C must be similar in the AC and in the BC studies in terms of dose, modes of administration, duration etc. In an NMA of studies comparing fluoride treatments for the prevention of dental carries, the definition of placebo was different between studies of fluoride toothpaste and studies of fluoride rinse (Salanti et al., 2009), casting doubt about the plausibility of the transitivity assumption and thereby challenging the reliability of NMA results. In another example, Julious and Wang (Julious and Wang, 2008) discussed how the use of placebo as an intermediate comparator might bias the results of indirect comparisons due to

changes in the placebo response of the population over the years; for example, when studies comparing treatment A to placebo are older than studies comparing B to placebo the indirect estimate for A vs. B via placebo may be biased.

Other ways of formulating the transitivity assumption is to assume that regardless of the treatments being compared in each study the true relative effect of A vs. B is the same in a fixed effects model or exchangeable across studies in a random effects model (Dias et al., 2013d; Lu and Ades, 2009), that the ‘missing’ treatments in each trial are missing at random (Lu and Ades, 2006) or, equivalently, that the choice of treatment comparisons in the trials is not associated either directly or indirectly with the relative effectiveness of the interventions (Salanti, 2012). Finally, an alternative way of postulating this assumption is to state that the included patients could in principle be randomized to any of the treatments included in the network (Salanti, 2012).

3.1.2 Consistency

The statistical manifestation of transitivity is called consistency (Cipriani et al., 2013). Checking the network for consistency constitutes an additional method of inferring indirectly about the plausibility of the transitivity assumption. Consistency refers to the statistical agreement between the observed direct and the (possibly many) indirect sources of evidence. A simple network may only include three treatments A, B and C. The transitivity assumption then implies that $\mu_{AB} = \mu_{AC} - \mu_{BC}$ (also termed consistency equation), where μ_{AB} denotes the true relative effect of treatment B over C; likewise for μ_{AC} , μ_{BC} . When this equation does not hold for the (direct) estimates, the network is said to be inconsistent (Lu and Ades, 2006) or incoherent (Lumley, 2002). If this is the case, results from an NMA will be more difficult to interpret and become less reliable. In a following section we review various statistical methods and models that have been suggested for identifying inconsistency and thus assessing the transitivity assumption in NMA.

Statistical inconsistency can be thought of as another form of heterogeneity: heterogeneity results from the variation of effect modifiers within a treatment comparison, while inconsistency results from the variation of effect modifiers across treatment comparisons (Jansen and Naci, 2013). Researchers should keep in mind, though, that the consistency of a network can only be assessed statistically when there is both direct and indirect evidence for one or more treatment comparisons. This situation only occurs when there are *closed loops* in the network (i.e. when three or more interventions are connected by a polygon, the edges of which represent head-to-head comparisons between the corresponding treatments). When there are no closed loops present in the network, a statistical assessment of inconsistency will not be possible. In these situations there cannot be inconsistency by definition. This,

however, does not imply that the transitivity assumption will necessarily hold. It should also be noted that the absence of statistical inconsistency (in the presence of loops) does not provide proof for the validity of the transitivity assumption, which, as discussed in the previous section is essentially an untestable assumption. Thus, next to statistical tests for inconsistency, a conceptual/theoretical assessment of the transitivity assumption should always take place before an NMA is conducted (Cipriani et al., 2013) and the studies included in an NMA should always be scanned for important differences in terms of patients, interventions, outcomes, study design, methodological characteristics and reporting biases (Bucher et al., 1997; Caldwell et al., 2005; Coleman et al., 2012; Coory and Jordan, 2010; Cucherat and Izard, 2009; Dewilde and Hawkins, 2012; Donegan et al., 2013b; Glenny et al., 2005; Griffin et al., 2006; Hoaglin et al., 2011; Jansen et al., 2008; Madan et al., 2011; van Valkenhoef et al., 2012c; Wells et al., 2009; Xiong et al., 2013).

3.2 *Statistical methods for NMA*

A simple network may include three treatments of interest, A, B and C. An estimate of the indirect treatment effect of A vs. B can then be obtained by utilizing the direct observations A vs. C and B vs. C as $\hat{\mu}_{AB}^{Ind} = \hat{\mu}_{AC}^{Dir} - \hat{\mu}_{BC}^{Dir}$ (Bucher et al., 1997). This result is sometimes also referred to as “adjusted indirect comparison”. The variance of the indirect estimate is the sum of the variances of the two direct ones. When direct evidence is also available for the A vs. B comparison it can be combined with the indirect estimate using the inverse variance method to produce a mixed estimate. Note that this approach for obtaining indirect estimates is only valid for ‘triangular networks’, where three treatments have been compared in a number of two-arm trials (and it can be readily extended for ‘polygonal networks’ with no diagonals); it can also be used for ‘star-shaped’ networks, where all treatments are compared to a common comparator (e.g. placebo) but not to each other. For complex networks, e.g. like the one depicted in the right panel of Figure 1, there will be multiple sources of indirect information and thus more advanced models need to be used.

Popular implementations of NMA models adopt meta-regression (Section 3.2.1), hierarchical modelling (Section 3.2.2) or a multivariate meta-analysis approach (Section 3.2.3). A common feature of all of these approaches is that the use of the consistency equations minimizes the number of parameters that need to be estimated. The minimum set of parameters needed to model the relative treatment effects is usually termed as the set of “basic parameters” or “basic contrasts”; these parameters are (assuming that the network is connected) in number equal to the number of treatments minus one and can be used to generate estimates for all possible treatment comparisons via the consistency equations (i.e. $\mu_{XY} = \mu_{XZ} - \mu_{YZ}$ for treatments X, Y and Z). The basic parameters can be

chosen arbitrarily as long as they form a “spanning tree” of the evidence, i.e. they form a connected subgraph of the network containing all vertices but no cycles (Lu and Ades, 2006); if this condition is satisfied the actual choice of basic parameters does not affect the NMA results. These parameters are commonly taken to be the relative effects of each treatment versus a reference (e.g. the placebo, if present in the network). For example, for a network of four treatments A, B, C and D three basic parameters are needed. These can be chosen to correspond to the relative treatment effects of all other treatments versus A, i.e. AB, AC and AD. All other treatment effects can be generated from these 3 parameters, e.g. the relative treatment effect for BD can be estimated using the AB and AD parameters. Choosing instead BA, BC and BD as the basic parameters would have no impact on the NMA results.

In what follows we describe some of the most common approaches for fitting an NMA model. These approaches should be considered equivalent; the choice between them should be primarily dictated by the availability of software packages for implementing them and by the technical expertise of the researchers. Different methods may give slightly different estimates, e.g. fitting an NMA in a Bayesian framework as a hierarchical model may lead to slightly different results than fitting it in a frequentist setting using multivariate meta-analysis techniques. These differences, however, should not be seen as a limitation of NMA models in general; they are due to the different approaches employed for making statistical inferences, in a similar way as it applies to a pairwise meta-analysis. We discuss the currently available software options for fitting all the models and implementing all the methods in Section 5.

3.2.1 NMA as a meta-regression

In the meta-regression approach, first proposed by Lumley (Lumley, 2002) the various treatment comparisons are treated as covariates in a meta-regression model (Salanti et al., 2008a). The usual NMA meta-regression model can be summarized in the following equation: $\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon} + \boldsymbol{\delta}$, with \mathbf{y} being the vector of observed relative treatment effects, $\boldsymbol{\mu}$ the vector of basic parameters, $\boldsymbol{\varepsilon}$ the vector of random errors, and $\boldsymbol{\delta}$ the vector of random effects. Note that for a study i comparing T_i different treatments, only $T_i - 1$ observations on treatment comparisons need to enter the model. For a parallel randomized three-arm ABC trial, for example, we only need to include two of the three comparisons, e.g. AB and AC; the BC comparison is just a linear function of the other two. This means that \mathbf{y} , $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ have a length equal to $\sum(T_i - 1)$. Random errors are assumed to follow a multivariate normal distribution, $\boldsymbol{\varepsilon} \sim N(0, \mathbf{S})$, with \mathbf{S} being the (block-diagonal), within-study variance covariance matrix. A study i with T_i treatments arms will contribute a $(T_i - 1) \times (T_i - 1)$ matrix to \mathbf{S} ; a two-arm AB study, for example, will only contribute to \mathbf{S} the variance of the relative treatment effect of A vs. B. A three-

arm trial ABC will contribute to \mathbf{S} a 2×2 matrix with the variances and the covariance of the 2 relative treatment effects chosen to be included in \mathbf{y} , e.g. AB and AC. Similarly, $\boldsymbol{\delta} \sim N(0, \boldsymbol{\Delta})$ for the random effects, with $\boldsymbol{\Delta}$ being the heterogeneity variance-covariance matrix. Matrix \mathbf{X} , the *design matrix*, has as elements 1, -1 and 0 and describes the structure of the network, providing information on which comparison is being performed in each study (Salanti et al., 2008a). If for example the network is built by an AB study (study 1), an AC study (study 2) and a BC study (study 3), the model would be written as:

$$\begin{pmatrix} y_{1AB} \\ y_{2AC} \\ y_{3BC} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_{AB} \\ \mu_{AC} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1AB} \\ \varepsilon_{2AC} \\ \varepsilon_{3BC} \end{pmatrix} + \begin{pmatrix} \delta_{1AB} \\ \delta_{2AC} \\ \delta_{3BC} \end{pmatrix}.$$

The basic parameters can be estimated as $\hat{\boldsymbol{\mu}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}$, with variance $var(\hat{\boldsymbol{\mu}}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$, where \mathbf{W} is the weight matrix, $\mathbf{W} = (\mathbf{S} + \boldsymbol{\Delta})^{-1}$. The within-study variance-covariance matrix \mathbf{S} can be estimated from the observed data (Franchini et al., 2012; Higgins and Whitehead, 1996), while for the between-study variance-covariance matrix $\boldsymbol{\Delta}$ one can use various ways of estimation including likelihood methods or the methods of moments (Berkey et al., 1998; Jackson et al., 2010; van Houwelingen et al., 2002). Estimating $\boldsymbol{\Delta}$ may be difficult especially when the data are sparse or in the presence of multi-arm studies. For this reason, it is common to introduce additional assumptions to reduce the number of parameters in $\boldsymbol{\Delta}$ and simplify the estimation. The most common approach is to assume equal heterogeneity variances across comparisons, i.e. the between-study heterogeneity of the relative treatment effects is the same for all treatment comparisons (Higgins and Whitehead, 1996; Lumley, 2002). This assumption is, however, quite strong and may often be unrealistic. Lu & Ades (Lu and Ades, 2009) discussed how the consistency equations impose restrictions on the heterogeneity of each comparison, based on the (different) heterogeneity variances of each of the basic parameters. Thorlund et al. (Thorlund et al., 2013) presented models for exchangeable heterogeneity variances and also discussed the use of informative prior distributions in the context of a Bayesian analysis.

In a different approach, Lu et al. proposed a two-stage method for performing an NMA as a meta-regression (Lu et al., 2011). At the first stage a meta-analysis is performed in each group of trials comparing the same treatments, e.g. all two-arm trials that compare A vs. B, all three-arm trials that compare A vs. B vs. C, etc. This provides the direct estimates on treatment comparisons. At the second stage of the meta-analysis, a weighted linear regression is performed with the direct estimates as dependent variables. This two-stage method can be used to investigate how the first-stage (direct)

evidence influences the network estimates and may therefore help to assess the consistency of the network (see Section 4).

3.2.2 NMA as a hierarchical model

Hierarchical NMA models (Lu and Ades, 2004; Salanti et al., 2008a) appear to be implemented most often (Coleman et al., 2012; Nikolakopoulou et al., 2014). An important advantage of this approach is that if arm-level data are available, their exact likelihood can be used (Dias et al., 2013a).

The likelihood of the arm-level data is defined in terms of a set of unknown parameters γ and a link function, $g(\gamma)$ which is used to map these parameters in the $(-\infty, \infty)$ range. For a study i comparing treatments A and B we set:

$$\begin{aligned} g(\gamma_{iA}) &= u_i, \\ g(\gamma_{iB}) &= u_i + \theta_{iAB}. \end{aligned}$$

For the case of binary data, for example, we can choose g to be the *logit* function and γ the probability of observing an event. We set:

$$\begin{aligned} \text{logit}(p_{iA}) &= u_i, \\ \text{logit}(p_{iB}) &= u_i + \theta_{iAB}. \end{aligned}$$

Here u represents the log-odds of the outcome for treatment A and θ_{iAB} the log-odds ratio of A versus B; the event probabilities for each arm parameterize the binomial likelihood, $r_{iT} \sim \text{Bin}(p_{iT}, n_{iT})$, with r_{iT} denoting the events and n_{iT} the total number of randomized patients in each treatment arm ($T = A, B$). We then allow $\theta_{iAB} \sim N(\mu_{AB}, \tau_{AB}^2)$ for a random-effects meta-analysis. If two non-reference treatments are compared in a study, e.g. treatments B and C, we utilize the consistency equations by setting $\theta_{iBC} \sim N(\mu_{AC} - \mu_{AB}, \tau_{BC}^2)$. In the presence of multi-arm studies multivariate normal distributions should be used instead, where the within- and between-study variances are replaced by the corresponding variance-covariance matrices \mathbf{S}_i and $\mathbf{\Delta}_i$. Details on how to model other types of data can be found in (Dias et al., 2013a). Note that the issues discussed in the previous section regarding the estimation of the between-trial heterogeneity hold for the hierarchical models as well.

NMA can be fitted as a hierarchical model also if only contrast-level data are available from the studies (i.e. when the reported data is on the relative treatment effects of the treatments being compared, but not on the specific arms). For a two-arm study i comparing A (reference treatment) and B the model is written as $y_{iAB} \sim N(\theta_{iAB}, s_i^2)$. Note here that the normality assumption can be justified even if the underlying patient-level distributions are skewed, due to the central limit theorem (Dias et al., 2013a).

Hierarchical models can also be fitted when a combination of arm-level and contrast-level data is available, using the exact likelihood for the arm-level data and the normal approximation for the contrast-level data in a so-called shared parameter model (Dias et al., 2013a).

3.2.3 NMA as a multivariate meta-analysis model

White et al. (White et al., 2012) suggested a method for performing NMA as a multivariate meta-analysis by treating the basic comparisons as different outcomes and by employing standard multiple-outcome meta-analytical techniques (Mavridis and Salanti, 2013). For this model to work all studies need to report on the reference treatment; if this is not the case for some studies, a data-augmentation technique is required to impute a minimally informative reference treatment arm. The model is written as $\mathbf{y} = \mathbf{X}^* \boldsymbol{\mu} + \boldsymbol{\varepsilon} + \boldsymbol{\delta}$, with \mathbf{X}^* being a matrix with all elements either 0 or 1, depending on which ‘outcomes’ are reported in each study.

Assume for example that treatments A, B and C are compared in a number of studies, and also assume that treatment A is taken to be the reference treatment. In this approach $\boldsymbol{\mu}$ will be a 2×1 vector of the basic parameters, AB and AC. A study comparing A vs. B will contribute an element 1 in the first column of \mathbf{X}^* and 0 in the second, since in this study only the first ‘outcome’ is reported. An A vs. C study will report the second outcome only, thus the relevant elements in \mathbf{X}^* will be 0 and 1 respectively. For a B vs. C study, however, an A arm must be imputed; this study becomes three-arm, and reports on both ‘outcomes’. The model for these three studies is as follows:

$$\begin{pmatrix} y_{1AB} \\ y_{2AC} \\ y_{3AB} \\ y_{3AC} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{AB} \\ \mu_{AC} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1AB} \\ \varepsilon_{2AC} \\ \varepsilon_{3AB} \\ \varepsilon_{3AC} \end{pmatrix} + \begin{pmatrix} \delta_{1AB} \\ \delta_{2AC} \\ \delta_{3AB} \\ \delta_{3AC} \end{pmatrix}$$

Note that the two random errors and also the two random effects that were included for the third study will be correlated. Also note that in this approach the vector of observations, \mathbf{y} has been modified to account for the imputed arms. Standard methods for multiple-outcome meta-analysis can now be used to fit the model.

Alternative models for performing an NMA have also been recently proposed in the literature. Rucker (Rucker, 2012) described the analogy between network meta-analysis and electrical networks and applied graph-theoretical methods to perform a fixed-effects NMA. This method was extended to also accommodate for random-effects and was shown to be equivalent to the meta-regression method described in Section 3.2.1 (Rucker and Schwarzer, 2014). Also, Yang et al. introduced a confidence distribution approach for performing an NMA (Yang et al., 2014). In this approach, instead of combining point estimates from each study, the authors combine confidence distributions.

For more in-depth reviews of the methodology that include the statistical details of the models we presented we refer the reader to (Cucherat and Izard, 2009; Greco et al., 2013; Ohlssen et al., 2013).

4 Statistical methods for evaluating inconsistency in NMA

As we have previously discussed (Section 3.1), transitivity is a central assumption of NMA. A statistical assessment of the plausibility of this assumption can be made by checking whether the various sources of evidence fit together in a coherent way. This assessment is vital for ensuring that the NMA results are valid and interpretable for clinical decision making (Donegan et al., 2012), but may be difficult to do in practice, especially in the case of complex networks or when multi-arm studies are included in the network.

Statistical consistency can be assessed only in closed loops of evidence, and there are two general approaches to do this: either locally (by focusing on the inconsistency of a specific treatment comparison) or globally (by checking for inconsistency in the entire network of evidence). In what follows we discuss methods and models that have been proposed for both of these approaches.

4.1 Local methods to detect inconsistency

Loop-specific approach

A straightforward approach for evaluating the presence of inconsistency in a network is to apply a loop-specific approach (Bucher et al., 1997); in this approach we examine each loop of the network in isolation from the rest of the network. For an ABC loop in the network, for example, we choose one of the comparisons (e.g. B vs. C) and compute the direct ($\hat{\mu}_{BC}^{Dir}$) and indirect estimates ($\hat{\mu}_{BC}^{Ind}$). Their difference measures inconsistency and is usually termed inconsistency factor (Lu and Ades, 2006): $\hat{w}_{ABC} = \hat{\mu}_{BC}^{Dir} - \hat{\mu}_{BC}^{Ind}$, with variance $var(\hat{w}_{ABC}) = var(\hat{\mu}_{BC}^{Ind}) + var(\hat{\mu}_{BC}^{Dir})$. A 95% confidence interval can be obtained as $\hat{w}_{ABC} \pm 1.96\sqrt{var(\hat{w}_{ABC})}$ and a Z-statistic for the null hypothesis of consistency, i.e. $\hat{w}_{ABC} = 0$, can be constructed as $z_{ABC} = \hat{w}_{ABC}/\sqrt{var(\hat{w}_{ABC})}$; this can be compared with the standard normal distribution to obtain a p-value (Bucher et al., 1997). Inconsistency is a property of the loop, in the sense that choosing a different treatment comparison of the loop and repeating the computations would give the exact same results (Dias et al., 2013d; Lu and Ades, 2006); thus we denote the inconsistency factor with an ABC subscript.

The loop-specific approach can be applied for each loop in a network to point out hotspots for inconsistency. The major advantage of this approach is that it is easy to implement, it suffers however from important limitations: when a treatment comparison is part of more than one loop this method does not compare direct evidence for this comparison to all available indirect information, but to

evidence from only one loop at a time; also in this case the tests for different loops sharing this comparison will not be independent. In addition, for networks with many loops there are multiple-testing issues.

Composite test

It is possible to extend the loop-specific approach by accounting for more than one indirect estimates for a treatment comparison using a ‘composite test’ for inconsistency (Caldwell et al., 2010; Dias et al., 2013d). Suppose that there are L loops that provide independent indirect information for the A vs. B comparison; these can be combined with the direct information using the usual inverse variance method to obtain a pooled, overall estimate of the relative treatment effect of A vs. B. Under the null hypothesis that the $L + 1$ different estimates are in agreement a test statistic following a chi-squared distribution with L degrees of freedom can be constructed to check for inconsistency. One should keep in mind, however, that the presence of multi-arm studies induces correlations among the estimates for the treatment effects which this method, as well as the loop-specific approach, fail to account for.

Back-calculation

Dias et al. also proposed two additional methods for locally checking inconsistency (Dias et al., 2010b). The first method (“back-calculation”) can be applied when the only available data are the pooled summaries of the pairwise meta-analyses. In the first step, the data is used to obtain a network estimate for each pairwise comparison in the network. It is then assumed that this estimate is a weighted average of the direct and the indirect evidence, coming from the rest of the network. This allows a back-calculation of the indirect estimate and its variance, which in turn can be used to construct a Z-test for the difference of direct and indirect evidence. Note that this method is problematic for a random-effects meta-analysis, as the posterior distribution of the heterogeneity variance will in general be different between the NMA model and the model for the pairwise meta-analysis.

Node-splitting

The second method proposed by Dias et al. (Dias et al., 2010b), the node-splitting approach, can be used when trial-level data are available. In this method the direct evidence for a specific treatment comparison is excluded from the rest of the network and is used to obtain a direct estimate. The remaining information in the network is used to obtain an indirect estimate for this comparison, after fitting an NMA model. The two estimates, direct and indirect, are then used to evaluate inconsistency with a Z-test. The main drawbacks of this approach (as well as the back-calculation approach) are that

they might be computationally intensive, especially for large networks with many treatment comparisons, and that they cannot properly handle multi-arm studies.

4.2 Global methods to detect inconsistency

The Lu and Ades model

Lu and Ades introduced a model (Lu and Ades, 2006), in which the consistency equations are ‘bent’ by including extra terms, the inconsistency factors. For an ABC loop, for example, the consistency equation is written as $\mu_{BC} = \mu_{AC} - \mu_{AB} + w_{ABC}$, where the w parameter measures the discrepancy of direct and indirect evidence. For networks comprising many loops a different inconsistency factor needs to be included in each loop. When the network only includes two-arm studies the number of independent inconsistency factors (the ‘inconsistency degrees of freedom’, *ICDF*) is $ICDF = C - T - 1$, with C being the number of available pairwise comparisons in the data and T the number of different treatments. The inconsistency factors can be assumed to follow a common distribution in order to increase the power in their estimation. A χ^2 test can be used to assess the inconsistency of the whole network, under the null assumption that all inconsistency factors are zero. In the presence of multi-arm studies, however, this model is problematic (van Valkenhoef et al., 2012b). Higgins et al. (Higgins et al., 2012) showed that different parameterizations of the model (choosing different reference treatment) may lead to different results. Thus, when multi-arm studies are present, the use of the Lu and Ades model should be avoided.

The ‘Design-by-treatment’ inconsistency model

Higgins et al. (Higgins et al., 2012) and White et al. (White et al., 2012) introduced an alternative inconsistency model, the ‘design-by-treatment’ interaction model, which encompasses both loop and design inconsistencies. The latter corresponds to the possible discrepancies in the treatment effects across designs, where ‘design’ refers to the treatments being compared in a study. For example, a study comparing treatments A and B is considered to be an AB design. The A vs. B estimate coming from such a study may be different than the A vs. B estimate coming from a three-arm study comparing treatments A, B and C (ABC design); this difference is referred to as design inconsistency. In the absence of multi-arm studies the Lu and Ades model is equivalent to the design-by-treatment model. Similarly to the Lu and Ades model, when multi-arm studies are present, the estimates of the inconsistency factors depend on the parameterization. Unlike the Lu and Ades model, however, the global statistic for inconsistency in the design-by-treatment interaction model is invariant under re-parameterization. One limitation of the model is the increased complexity and the accompanying loss of power for the statistical test (Higgins et al., 2012). Another point of criticism for this approach is that

it adopts a somewhat arbitrary definition of design inconsistency, as it is mainly dictated by methodological rather than clinical considerations.

Jackson et al. (Jackson et al., 2014) presented an implementation of the design-by-treatment inconsistency model where they assumed random effects on the inconsistency factors. In the same paper the authors also derived R and I^2 statistics to quantify heterogeneity and inconsistency. In addition, an approach based on the notion of design-by-treatment inconsistency was also proposed by Piepho et al. (Piepho et al., 2012; Piepho, 2014; Piepho et al., 2015).

Unrelated mean effects model

An alternative method was introduced by Dias et al. (Dias et al., 2013d). In this model the consistency equations are completely removed and the network meta-analysis model is equivalent to a series of separate, independent meta-analyses for each pairwise contrast, sharing, however, a common heterogeneity variance (Caldwell et al., 2012). The fit of the model is then compared to the standard consistency NMA model using the posterior deviance and the deviance information criterion (DIC) (Spiegelhalter et al., 2002). In addition, estimating the contribution to posterior mean deviance for each data point can help identify possibly ‘problematic’ studies, i.e. studies not fitting well with the rest of the evidence. Each data point is expected to have a contribution of about 1 to the posterior mean deviance. A larger value will suggest a poor fit to the model, pointing out possibly inconsistent pieces of evidence. The use of leverage plots has been also suggested as a diagnostic tool for identifying studies with poor fit (Dias et al., 2010b). A limitation of the unrelated mean effects model is that in the presence of multi-arm studies a re-parameterization will affect the results of a random-effects meta-analysis.

Q statistic for inconsistency

A different approach for globally assessing inconsistency is by using the Q statistic for inconsistency (König et al., 2013; Krahn et al., 2013), which is analogous to the Q statistic for heterogeneity in simple meta-analysis. This approach is based on the two-stage method for fitting NMA (Lu et al., 2011). On the first stage we perform a pairwise meta-analysis for the studies of each design available in the dataset, and obtain the direct relative treatment effect. From the AB studies, for example, we estimate $\hat{\mu}_{AB}$; this can be used to compute the Q_{AB}^{het} statistic for heterogeneity as $Q_{AB}^{het} = \sum_i \frac{1}{s_i^2} (y_i - \hat{\mu}_{AB})^2$, where i runs through all AB studies. Similarly, on the second stage of the analysis we obtain network estimates for all pairwise comparisons; using these estimates and the direct estimates of the first stage, a Q statistic for the inconsistency of the whole network can be obtained and the null hypothesis of consistency in the network can be tested using a χ^2 distribution with $C - T - 1$ degrees

of freedom. This approach can be used in the presence of multi-arm studies. The authors also propose the use of a graphical tool, the *net-heat* plot, which provides a display of the flow of evidence in the network and can help identify potential hotspots for inconsistency.

Also, Rücker and Schwarzer suggested a Q statistic for inconsistency using a graph-theoretical approach and showed that it is identical to the Q statistic described above (Rücker, 2012; Rücker and Schwarzer, 2014).

Multidimensional scaling

In a another graphical approach proposed by Chung and Lumley (Chung and Lumley, 2008) the multidimensional scaling method is used to infer about inconsistency in a network. For each pairwise comparison a usual inverse variance meta-analysis is performed; the magnitude of the relative treatment effects is considered to be a measure of the observed pairwise ‘dissimilarity’ of the treatments. The pairwise estimates are summarized in a dissimilarity matrix, to which a weighted multidimensional scaling is applied in order to obtain the ‘fitted dissimilarities’. Important differences between observed and fitted dissimilarities are an indicator of possible inconsistencies. Note that this method cannot properly handle multi-arm studies.

4.3 Empirical studies and simulations on inconsistency

Empirical studies show that the prevalence of inconsistency in published networks is non-trivial. Song et al. (Song et al., 2011) performed a meta-epidemiological study that included 112 published triangular networks, 16 of which were found to be statistically inconsistent. Veroniki et al. (Veroniki et al., 2013) evaluated inconsistency in 40 published networks including a total of 303 loops. They found that 2-9% of the loops were inconsistent, depending on the effect measure used and the assumptions for heterogeneity; also, approximately one eighth of the networks were found to be inconsistent using the design-by-treatment method.

The various methods for assessing inconsistency have been rarely and poorly applied in published NMAs (Donegan et al., 2010). In a meta-epidemiological study by Nikolakopoulou et al. (Nikolakopoulou et al., 2014) it was found that in 24% of the published NMAs the authors did not use appropriate methods to evaluate inconsistency, while in 44% the authors did not report using any method at all.

Song et al. (Song et al., 2012) performed simulations to evaluate the statistical properties of various methods for inferring about network inconsistency. They explored the use of the loop-specific approach, the node-splitting technique and the Lu & Ades model. They found that even though all methods are unbiased, they have little power in detecting inconsistency. It is also important to note that inferences

on inconsistency heavily depend on the extent of heterogeneity and the method used to evaluate it (Veroniki et al., 2013). Thus, analysts should keep in mind that a statistically non-significant estimate for inconsistency should not be interpreted as proof of consistency. In addition, even when statistically significant inconsistency is found, its magnitude should be interpreted in terms of clinical relevance; thus, a statistically-significant inconsistency in a certain loop might be clinically unimportant.

4.4 Choosing between the methods for evaluating inconsistency

If the network structure allows it, i.e. if there are closed loops in the network, a statistical assessment of inconsistency should always take place after fitting the NMA model. In the previous paragraphs we presented a variety of methods and models currently available for statistically checking the network for consistency and we discussed the advantages and limitations of each approach. In Table 1 we provide an overview of these approaches, including a brief summary of the limitations of each one.

An assessment of inconsistency may start with the loop-specific approach which, despite its shortcomings, is the easiest one to implement and can pinpoint possibly problematic loops. Afterwards, if all studies in the network are two-armed, all presented strategies are valid choices for checking for inconsistencies. We generally recommend the application of both local and global methods to gain a better understanding of the source of possible discrepancies between direct and indirect evidence and the plausibility of the consistency assumption in the network as a whole. If the network includes multi-arm studies only the design-by-treatment model and the Q statistic approach will lead to results that are independent of the parameterization of the model (i.e. the choice of the basic parameters). Researchers may still choose to implement some of the other methods as well, as exploratory analyses; they should bear in mind, however, that their results might not be robust.

Approaches for evaluating inconsistency can also be selected based on the available technical expertise and/or software packages. In Table 2 we provide a summary of the currently available software solutions for implementing the various approaches.

If statistically significant inconsistency is detected, researchers are advised to explore potential sources of it and try to explain it. Local methods for assessing inconsistency can indicate outlying studies, which should be checked for data extraction errors, important differences in the distribution of effect modifiers or other possible biases. In Section 6.3 we present various models for adding covariates and adjusting for suspected biases in the analysis. If sufficient studies are available, such models can be applied to explain and possibly eliminate inconsistencies, while, if inconsistency persists, researchers can consider splitting up the network (see discussion in Section 6.7). Finally, in the case of unexplained inconsistency, researchers may choose not to synthesize the evidence in an NMA

at all, or to present the results from the appropriate inconsistency model (Lu & Ades model when all studies are two-armed; design-by-treatment model when multi-arm trials are present) along with the direct evidence and a warning to the readers of the limitations of the analysis.

TABLE 1 HERE

5 Software options for fitting an NMA model and evaluating inconsistency

Our literature search showed that BUGS software is a popular choice for implementing new methods in NMA, the majority of the articles included in our database reported using WinBUGS, OpenBUGS (Lunn et al., 2009, 2000) or JAGS (Plummer, 2003): (Achana et al., 2013; Ades, 2003; Ades et al., 2010; Caldwell et al., 2012, 2010, 2005; Carlin et al., 2013; Chaimani et al., 2013b; Chaimani and Salanti, 2012; Cooper et al., 2011, 2009; Cope and Jansen, 2013; Dakin et al., 2011; Del Giovane et al., 2013; Dewilde and Hawkins, 2012; Dias et al., 2013a, 2013b, 2013c, 2013d, 2010a, 2010b, 2010c; Ding and Fu, 2012; Donegan et al., 2013a, 2013b; Efthimiou et al., 2014a, 2014b; Franchini et al., 2012; Govan et al., 2010; Griffin et al., 2006; Higgins and Whitehead, 1996; Hoaglin et al., 2011; Hong et al., 2013a, 2013b; Jansen, 2011; Jansen et al., 2008; Jansen and Cope, 2012; Jonas et al., 2013; Jones et al., 2011; Lu et al., 2007; Lu and Ades, 2009, 2006, 2004; Madan et al., 2014, 2011; Mavridis et al., 2013; Mills et al., 2013a; Nixon et al., 2007; Ohlssen et al., 2013; O'Regan et al., 2009; Ouwens et al., 2010; Price et al., 2011a, 2011b; Salanti et al., 2011, 2010, 2009; Saramago et al., 2012; Schmid et al., 2013; Schmidli et al., 2013; Schmitz et al., 2013, 2012; Soares et al., 2013; Song et al., 2012; Spineli et al., 2013; Sturtz and Bender, 2012; Thorlund et al., 2014, 2013; Thorlund and Mills, 2012a; Trinquart et al., 2012a, 2012b; Tu and Faggion, 2012; van Valkenhoef et al., 2012a; Warren et al., 2014; Wells et al., 2009; Welton et al., 2012, 2010, 2009, 2008; White et al., 2012; Woods et al., 2010; Zhang et al., 2013). An alternative option for implementing Bayesian statistical inference is Stan, a recently developed programming language (Stan Development Team, 2014). However, we did not identify any articles using Stan.

Also, there were many articles that reported the use of R (R Core Team, 2014): (Chung and Lumley, 2008; Dias et al., 2010b; Donegan et al., 2013a, 2012; Jansen, 2012; Krahn et al., 2013; Lu et al., 2011; Lu and Ades, 2009; Lumley, 2002; Mills et al., 2013a, 2011b; Naudet et al., 2013; Rucker, 2012; Rucker and Schwarzer, 2014; Salanti et al., 2009; Schmid et al., 2013; Song et al., 2012; Sturtz and Bender, 2012; Tan et al., 2014; Thorlund et al., 2014; Thorlund and Mills, 2012a; Trinquart et al., 2012b; van Valkenhoef et al., 2012a, 2012c; Veroniki et al., 2013; Zhang et al., 2013); some papers used STATA (StataCorp., 2013): (Ballesteros, 2005; Chaimani et al., 2013a; Franchini et al., 2012;

Higgins et al., 2012; Jonas et al., 2013; Madan et al., 2011; Norton et al., 2012; Tu and Faggion, 2012; Veroniki et al., 2013; Warren et al., 2014; White et al., 2012) and a few papers reported using SAS software (SAS Institute, 2011): (Carlin et al., 2013; Hong et al., 2013b; Jonas et al., 2013; Jones et al., 2011; Piepho et al., 2012; Piepho, 2014; Piepho et al., 2015; Signorovitch et al., 2012). Finally, Van Valkenhoef et al. (van Valkenhoef et al., 2012a) presented GeMTC, a freely-available, open-source program for performing NMA.

Neupane et al. performed a review of the available automated packages for performing an NMA in R aiming to summarize the key features and functionality of each package (Neupane et al., 2014). In Table 2 we describe possible software solutions for some of the models presented in this review.

TABLE 2 HERE

6 Other issues in NMA

6.1 *The use of different measures of effect size*

There is a wide choice of summary effect measures that can be used for the meta-analysis of evidence on a binary outcome. The most common choices are the odds ratio (OR), risk ratio for harmful or beneficial outcomes (RR_H and RR_B), risk difference (RD) and hazard ratio (HR) for time-to-event data. Donegan et al. (Donegan et al., 2010) reported that the majority of the published NMAs of dichotomous data used OR and RR (50% and 40% respectively), with RD being used in only 10% of the analyses. Veroniki et al. (Veroniki et al., 2013) analyzed 40 published networks and showed that the choice of effect measure may have an impact on the inferences about the statistical inconsistency. This was also discussed by Coory and Jordan (Coory and Jordan, 2010); using information from published networks they concluded that the use of OR and RR is preferable over RD. In addition, it has been demonstrated that the choice of the scale may have an impact on the results of an NMA (Gleser and Olkin, 2009). In particular, Eckermann et al. (Eckermann et al., 2009) showed that the use of RR may lead to inferential fallacies and advocate the use of OR. Norton et al. (Norton et al., 2012) discussed that different choices of scale may lead to differences in the ranking of the treatments in an NMA. They propose that researchers should explore how sensitive the NMA results are in the choice of effect measure. Van Valkenhoef and Ades (van Valkenhoef and Ades, 2013) on the other hand discuss that a rank reversal is unlikely to take place unless the assumptions underlying NMA do not hold or the data is very sparse; thus, if a rank reversal is observed when using a certain effect measure then the transitivity assumption is unlikely to hold irrespectively of the choice of effect measure.

In principle, the choice of the summary statistic in any meta-analysis should be dictated (if possible) by biological justifications regarding the disease (Walter, 2000) and also by empirical evidence (Deeks,

2002). Alternatively, it is possible to fit models using alternative effect measures and make a choice by comparing model fit. Caldwell et al. proposed the use of the posterior mean deviance and the deviance information criterion (DIC) to evaluate the model fit of the different effect measures in an NMA (Caldwell et al., 2012). The choice of the effect measure can also be guided by comparing the estimates of between-study heterogeneity, with lower values being preferable. This, however, may be impractical when there are few data available, in which case the choice of scale may be driven by the ease of interpretation and understanding of the disease process (Caldwell et al., 2012; Dias et al., 2013a). Analysts should also keep in mind that the choice of the effect measure for modeling may be different than the scale used for decision-making (Egger et al., 1997); e.g. OR may be preferred for data analysis, while RR or RD are easier to interpret and can be used to better communicate results with patients or clinicians (Walter, 2000).

Moreover, HR should always be considered as a suitable choice of scale for the case there is an underlying time-to-event process and the proportional hazards assumption is deemed plausible (Caldwell et al., 2012; Soares et al., 2013) (see Section 6.3 for additional details on the methodology). Also note that the discussion of this paragraph pertains to the analysis of a binary outcome. When continuous data is available the analysts should avoid dichotomization since it translates into a loss of power and also because the choice of cut-off point may impact on the inferences of NMA (Schmitz et al., 2012).

6.2 *Modelling time-to-event data*

In many RCTs the outcome measured is the time to the occurrence of an event (e.g. death, disease progression etc.). Welton et al. (Welton et al., 2010) described a method for simultaneously synthesizing survival and disease progression outcomes in a single NMA analysis; also, Woods et al. (Woods et al., 2010) provided guidance on how to perform an NMA on the log-hazard scale when studies report different survival statistics.

Analysts should keep in mind that the synthesis of time-to-event data in terms of hazard ratios relies on the proportional hazards assumption. Treatment effects, however, may vary over time and this might threaten the validity of the meta-analysis results. For NMA this might have an extra impact on the consistency of the results. Ouwens et al. (Ouwens et al., 2010) and Jansen (Jansen, 2011) modeled the hazard functions using parametric survival curves and fractional polynomials of time respectively; in these models the hazard ratio is allowed to vary over time. Jansen and Cope (Jansen and Cope, 2012) discussed methods for extending these models by including study-level covariates (such as study date) that can reduce possible inconsistencies and bias. In another paper by the same authors (Cope and

Jansen, 2013), various alternative summaries were presented for summarizing the estimates of the relative treatment effects obtained from an NMA of survival data.

6.3 Extension of NMA models to account for effect modifiers

In a pairwise meta-analysis a meta-regression on important effect modifiers can explain the presence of between-study heterogeneity, which may hinder the interpretation of the results and may have important implications in decision making (Dias et al., 2013b). In NMA interpreting results will be even more problematic in the presence of evidence inconsistency; meta-regression techniques in NMA adjust the treatment effects for possible effect-modifiers and can reduce heterogeneity and inconsistency in the results that may be present when these covariates are distributed unevenly among studies (Achana et al., 2013; Cooper et al., 2009; Dias et al., 2013b; Greco et al., 2013; Hoaglin et al., 2011; Jonas et al., 2013; Nixon et al., 2007; Salanti et al., 2009). The effect modifiers can be either categorical or continuous variables, and may represent either patient-level or trial-level characteristics.

6.3.1 General model for including effect modifiers in the analysis

Nixon et al. (Nixon et al., 2007) first combined NMA and meta-regression techniques to develop models that allow the simultaneous comparison of multiple competing treatments while adjusting for study-level covariates in an attempt to investigate and explain heterogeneity. Salanti et al. (Salanti et al., 2009) and Cooper et al. (Cooper et al., 2009) proposed the use of meta-regression as a tool for eliminating inconsistency as well as heterogeneity in NMA. As an example of adding covariates in NMA Salanti et al. (Salanti et al., 2009) considered the year of randomization in each trial as a covariate in an NMA for topical fluoride treatments for the prevention of dental carries. The covariate adjusted for possible time trends in the placebo-controlled comparisons and relative treatment effects were estimated for a pre-defined year of randomization (the year of the most recent study).

In general, there are three main approaches in the meta-regression of study-level covariates for NMA (Cooper et al., 2009; Dias et al., 2013b): using different and unrelated interaction terms (coefficients), using exchangeable interaction terms and using a common interaction term.

Unrelated interaction terms

In this approach there are a number of interaction terms for each covariate equal to the number of the basic parameters of the model. Let us assume for simplicity that we are only interested in one study-level covariate x_i . We can augment the hierarchical random-effects model previously presented as follows: for a study i , comparing treatments B versus C, we allow $\theta_{iBC} \sim N(\mu_{BC} + x_i\beta_{BC}, \tau^2)$, assuming a common heterogeneity variance τ^2 for the treatment effects. If treatment A is chosen to be the reference treatment, we can utilize the consistency equations to write $\theta_{iBC} \sim N(\mu_{AC} -$

$\mu_{AB} + x_i(\beta_{AB} - \beta_{AC}), \tau^2$); in a Bayesian analysis the β_{AT} ‘basic’ coefficients (where $T \neq A$) can be assigned unrelated vague prior distributions.

Exchangeable interaction terms

The model has the same structure as the model for unrelated interaction terms, but now the basic coefficients are drawn from a common distribution, $\beta_{AT} \sim N(b, \tau_b^2)$ where index T runs through all treatments except reference treatment A . The mean b and the variance τ_b^2 of the common distributions can be assigned vague priors.

Common interaction term

The common interaction term model is the same as the exchangeable interaction model described in the previous paragraph, but now all basic interaction terms are assumed equal, $\beta_{AT} = \beta$, for all treatments $T \neq A$. A vague prior is then assigned to β . This model implies that the relative treatment effects between the non-reference treatments are independent of the covariate, since the interaction terms cancel out. In this case the choice of the reference treatment becomes important, as it might change the meta-regression results (Dias et al., 2013b).

6.3.2 NMA meta-regression for baseline risk

The underlying risk of the disease, usually termed as ‘baseline risk’, is a proxy for important patient characteristics that may be possible modifiers of the treatment effects and it can be included as a covariate in an NMA; however, care should be taken to account for its correlation with the treatment effects (Achana et al., 2013; Dias et al., 2013b). Achana et al. (Achana et al., 2013) proposed a random-effects meta-regression model in which the effect of the reference treatment was used as a measure of the baseline risk. In order to include studies not reporting the reference treatment the authors proposed three alternative distributional assumptions for the ‘true’ unobserved baseline risk. Following Cooper et al. (Cooper et al., 2009), the interaction terms were taken to be independent, exchangeable or common. The authors recommended that the goodness of fit of the various alternative configurations can be based on residual deviance.

6.3.3 Limitations of NMA meta-regression models

Analysts should keep in mind that network meta-regression inherits all the interpretation difficulties of meta-regression for the case of pairwise meta-analysis, most importantly the inability to directly infer on causal relationships (Dias et al., 2013b), and the risk of ecological bias if study-level covariates are used to infer about individuals. Another drawback of meta-regression models for decision-making in general is that in order to assess the magnitude of the relative treatment effects the analyst must choose a value of the covariate at which to make the comparison (Hoaglin et al., 2011).

Regarding the choice between the three meta-regression models we presented, the unrelated interaction terms approach requires a sufficient number of studies to inform each pairwise comparison. In the common case where some of the comparisons in the network are performed in only few studies (or even worse in only one study) the coefficients cannot be properly estimated. Moreover, Dias et al. (Dias et al., 2013b) advocate that even though the use of models with exchangeable coefficients seems attractive, they are likely to lead to situations where decision making on the optimal treatment will depend on the value of the covariate. In addition, when there are comparisons in the network informed by a low number of studies the ranking of the treatments might be based in non-robust results. Therefore, even though the exchangeable coefficient model – or even more complex models – can be fitted, the authors suggest that their use should be limited to exploratory analyses. For these reasons the common interaction approach might be the safest choice for most networks, but analysts should keep in mind the rather strong assumption it employs (i.e. all interaction terms are equal) and that it heavily depends on the choice of reference treatment.

6.4 Investigating potential sources of bias in NMA

When combining results from different studies in a (pairwise or network) meta-analysis researchers always run the risk of obtaining biased estimates of the relative treatment effects for the target population. This may be the case when some of the studies provide a biased estimate of the treatment effect for their population of interest (which may or may not be the same as the target population of the meta-analysis). For example treatment effects may be overestimated in studies of low methodological quality. This relates to the so-called ‘internal validity’ of the included studies. The pooled result of a (network) meta-analysis may, however, still be biased even if the included studies have no internal biases. This might be the case, for example, when the included studies refer to populations different than the one targeted by the meta-analysis or when studies without ‘statistically significant’ results have not been published (Turner et al., 2009). This rather relates to the generalizability (sometimes referred to as ‘external validity’) of the included studies with respect to the targeted question or population.

Dias et al. (Dias et al., 2013b) discussed that when confronted with studies of mixed quality, researchers have three options: they can choose to analyze only the high-quality studies, thus ignoring a possibly important amount of information, they can choose to analyze all evidence, thus risking a bias in the pooled estimates, or they can include all studies after taking into account and adjusting for possible biases in the studies. In what follows we focus on the latter, presenting various available

approaches for adjusting for suspected internal biases in the included studies, and also for adjusting for various forms of external bias.

6.4.1 Accounting for study limitations in NMA

Several approaches can be used to account for the potential presence of biases in meta-analysis. Spiegelhalter and Best proposed an additive bias model to adjust for both external and internal biases (Spiegelhalter and Best, 2003). More advanced methods account for multiple sources of bias, where the contribution of each source is modeled separately; see for example (Eddy and Shachter, 1992; Greenland, 2005, 2003; Wolpert and Mengersen, 2004). A straightforward but more resource-intensive approach to adjust for bias is by eliciting bias distributions (Turner et al., 2009). This involves assigning a number of independent experts to evaluate each study in terms of some predefined criteria that may indicate the presence of bias. This information can be used to construct an overall bias distribution. The parameters of this distribution are combined with the estimates of the studies in order to produce a bias-adjusted estimate of the treatment effect in each study. These estimates can then be synthesized using standard NMA techniques.

Models that account for bias usually assume that biased studies yield inaccurate treatment effects of size $\mu + \beta$, where μ is the unbiased treatment effect and β is a bias parameter. For the case of network meta-analysis, if the study-specific bias parameters are assumed to be exchangeable across studies the unbiased treatment effects and the mean bias can be estimated from the network (Dias et al., 2013b, 2010c). Dias et al. (Dias et al., 2010c) presented a model where exchangeable bias parameter with non-zero mean were included in studies that compared an active versus an inactive treatment and were considered to be of a high risk of bias (according to some predefined measure such as allocation concealment, blinding or other trial characteristics). They also explored the use of two different bias parameters, one for active versus inactive, and one for active versus active comparisons; note that in this approach some assumption on the direction of bias in the active versus active trials is necessary to be made. Salanti et al. (Salanti et al., 2010) considered a similar model in which the newest treatments were favored, thus adjusting the treatment effects for possible ‘optimism’ or ‘novelty’ bias (Dias et al., 2010a). Study size can be a proxy for the study’s risk of bias and Chaimani and Salanti (Chaimani and Salanti, 2012) presented a method for adjusting for the ‘small study effects’, the exaggeration of treatment effects in smaller trials. This exaggeration might be due to methodological differences between smaller and larger trials that affect treatment effectiveness, due to publication bias or due to reporting bias. They proposed a network meta-regression model, where the bias parameter is multiplied with the observed variance of the treatment effects in each study; the standard error or the precision

(inverse variance) can be used alternatively. Their model can also adjust for suspected ‘sponsorship’ bias, for the case when interventions are sponsored in some of the studies. A similar model was also presented by Trinquart et al. (Trinquart et al., 2012b).

6.4.2 *Selection model to account for publication bias*

Mavridis et al. (Mavridis et al., 2013) proposed a Bayesian implementation of the Copas selection model (Copas, 1999) for addressing for possible publication bias in NMA. The idea behind selection models is that the observed set of published studies is considered to be a ‘biased’ sample, due to the nature of the publication process. This is addressed by introducing a latent variable for each study, the ‘propensity of publication’, which is assumed to be correlated with the study’s effect size. Mavridis et al. (Mavridis et al., 2013) modeled propensity via a regression model, where it was assumed to be inversely proportional to the standard error of the effect size. They considered alternative scenarios of how the selection model parameters depend on the treatments being compared in each study. Trinquart et al. also presented a selection model (Trinquart et al., 2012b) which modeled the propensity score of a trial as a linear function of the standard error. The effect sizes of the studies were weighted according to their propensity. Their model was shown to yield similar results to the model by Mavridis et al. (Mavridis et al., 2013).

6.4.3 *Accounting for ecological bias*

The meta-analysis of aggregated data can lead to *ecological bias*. This refers to a bias that may arise when using aggregated data in order to make inferences about patient-level interactions. Govan et al. (Govan et al., 2010) proposed an NMA model to control for ecological bias by specifically modeling the effects of the covariates. Their model allows the inclusion of studies that provide information on all covariates, studies that report marginal data on some of the covariates and also studies not providing any covariate information at all. The model allows the joint estimation of both the treatment and the covariate effects.

6.4.4 *Graphical approaches to bias*

In a different approach Salanti et al. (Salanti et al., 2008b) discussed how the geometry of the network can offer insight on the presence of a ‘comparator preference’ bias, i.e. when head-to-head comparisons between effective treatments are deliberately avoided, which in turn would imply that the treatment effects versus the reference treatments might be exaggerated. The authors utilized two indices from ecological literature: diversity, which is a measure of the number of treatments present in the network and how often they were tested, and co-occurrence, which measures whether specific treatment comparisons were preferred in the network while others were avoided. Limited diversity and

significant co-occurrence in a network is an indicator of possible preference bias in the network (Ioannidis, 2009).

Jansen et al. (Jansen et al., 2012) discussed the use of directed acyclic graphs (DAGs) as a graphical tool for conceptually evaluating the consistency assumption and also identifying possible sources of bias in indirect and mixed treatment estimates. By means of DAGs they showed that NMA estimates can be biased when relative treatment effect modifiers vary across comparisons and are not adjusted for in the analysis. They also showed that adjusting for covariates that are not effect modifiers is not only unnecessary, but that it can introduce bias.

6.4.5 Empirical assessments of the impact of bias in NMA

Chaimani et al. (Chaimani et al., 2013b) performed a network meta-epidemiological study to explore the effect of trial characteristics and study precision in NMA. They analyzed 32 networks and found evidence that imprecise studies (studies reporting broader confidence intervals for their estimates) tend to report larger effects compared to more precise studies, thus altering the results of the NMA. However, they found no evidence of association between effect size and previously identified indicators of bias, including blinding, allocation concealment and random sequence generation. Trinquart et al. (Trinquart et al., 2012a) used data from 74 FDA-registered placebo-controlled studies on 12 antidepressants along with 51 corresponding publications in order to assess the impact of publication bias. They found that the effect sizes derived from published studies differed from the ones derived from the FDA data by at least 100% for almost half of the pairwise comparisons. They concluded that reporting bias alters NMA estimates and changes the treatments' ranking. They also noted that the impact of reporting bias may be more important in NMA compared to classical meta-analyses, in the sense that reporting bias in one treatment comparison may have an effect in the ranking of all treatments in NMA.

6.5 Reporting results from NMA

Although the implementation of NMA is increasingly gaining popularity, the quality of reporting has been rather low. Various meta-epidemiological studies of published NMAs showed that the methods used and the assumptions made were not routinely reported (Bafeta et al., 2013; Donegan et al., 2010; Lee, 2013; Nikolakopoulou et al., 2014). Ohlssen et al. (Ohlssen et al., 2013) presented a checklist of items that should be reported in a drug-safety Bayesian NMA while Ades et al. (Ades et al., 2013) and Mills et al. (Mills et al., 2012a) give guidelines for those reviewing an NMA for the purposes of decision making.

One possible hurdle in the reporting of an NMA is that presenting all results can be a challenging

task, especially for networks with many treatments and multiple outcomes. The literature offers a plethora of graphical and tabular methods for visualizing the evidence base (Chaimani et al., 2013a; Chung and Lumley, 2008; Senn et al., 2013), the assumptions made (Chaimani et al., 2013a; Krahn et al., 2013) and the results obtained from an NMA (Chaimani et al., 2013a; Chung and Lumley, 2008; Fadda et al., 2011; König et al., 2013; Salanti et al., 2011; Senn et al., 2013; Tan et al., 2014). In a meta-epidemiological study on the presentational approaches used, Tan et al. (Tan et al., 2013) examined NMAs published in the UK and found that there is no standardized presentational approach for reporting the results of NMA. The authors concluded that a standardization of reporting is required.

6.6 *Modelling multiple outcomes and repeated measures*

RCTs commonly report on more than one outcome. These outcomes may be correlated within a study (due to the fact that observations come from the same set of patients) and in addition the true treatment effects on the outcomes can be correlated across studies (reflecting the way outcomes are related when measured in different settings). The usual meta-analytical approach on multiple outcomes is to analyze each one separately, ignoring all possible correlations. On the other hand, a joint meta-analysis of all outcomes which incorporates possible correlations can increase precision by ‘borrowing of strength’ across outcomes and may reduce the impact of outcome reporting bias (Jackson et al., 2011; Kirkham et al., 2012).

Welton et al. (Welton et al., 2008) described a method for performing an NMA of two correlated outcomes for the case when all studies are two-armed. Hong et al. (Hong et al., 2013a) presented a model for multiple outcomes that does not take into account within-study correlations. Schmid et al. (Schmid et al., 2013) proposed a model for unordered categorical data that also allows the inclusion of studies with partially observed data. Efthimiou et al. (Efthimiou et al., 2014a) developed a method for multiple binary outcomes, where information on the correlations is elicited from expert clinicians. Also, Efthimiou et al. (Efthimiou et al., 2014b) presented two models that can be used to perform a joint NMA for correlated dichotomous, continuous, time-to-event or mixed outcomes: the first of these models can be used when within-study correlations are available, the second when these correlations are unavailable.

Competing risks is a special case of multiple-outcome structure where the outcomes are mutually exclusive; Ades et al. (Ades et al., 2010) presented methods for performing a competing-risks NMA. Price et al. (Price et al., 2011a) discussed methods for an NMA in multi-state Markov models; a model averaging technique was also proposed (Price et al., 2011b) for combining estimates from alternative multi-state models.

In some cases, studies may report on a single outcome for multiple time points, which leads to a series of correlated observations. Lu et al. (Lu et al., 2007) proposed a hierarchical NMA model for synthesizing repeated measures of a discrete outcome. Dakin et al. (Dakin et al., 2011) suggested a model that can handle a continuous outcome, but did not include in the analysis the correlations between the observations. Ding and Fu (Ding and Fu, 2012) also presented a model for a continuous outcome that automatically modeled the correlations between the observations at different time points. Madan et al. (Madan et al., 2014) presented methods for analyzing two dichotomous outcomes reported on multiple time points, for studies comparing complex interventions.

Using NMA results to decide which of all available treatments is optimal for a specific condition might be a non-trivial issue, when the treatments are compared for more than one outcome. In order to facilitate decision making in the presence of multiple outcomes, Van Valkenhoef et al. (van Valkenhoef et al., 2012c) proposed a method for multiple criteria benefit-risk assessment of all competing treatments in an NMA; also, Hong et al. (Hong et al., 2013b) described a similar method for producing an overall ranking of the treatments in the network using a scoring system for combining efficacy and safety outcomes.

6.7 Definition of nodes in the treatment network

One important decision that analysts must make at the onset of an NMA regards the number of nodes (treatments) to be included in the network, with a simple choice being to include all relevant treatments. As discussed in Section 3.1.1 the parameters of interest in an NMA are the relative treatment effects between the various competing treatments for a target disorder and a target population; however, a decision-maker may not be interested in all pairwise comparisons. The relative treatment effect between two active treatments might be more important to decision making than the relative effects of active treatments versus placebo. When this is the case, researchers may simplify the network by focusing on just a subset of the available treatments, i.e. the ones that are deemed to be clinically relevant (e.g. newer/more effective treatments). This, however, poses a dilemma: including in the evidence base studies that compare treatments that are not clinically interesting may provide additional indirect evidence for the clinically interesting ones, which in turn may increase the precision of the estimates (Li et al., 2011; Salanti, 2012); on the other hand, obtaining all relevant evidence in order to include clinically uninteresting treatments in the network may be very time consuming and it may also lead to additional inconsistencies (Sturtz and Bender, 2012).

Hawkins et al. (Hawkins et al., 2009b) performed an empirical study that supported the use of all potentially relevant data; in another empirical study Mills et al. (Mills et al., 2013a) concluded that the

exclusion of treatments in an NMA might have an important effect on the results and might limit its usefulness, if important comparisons are excluded. To address this issue Hawkins et al. (Hawkins et al., 2009a) presented two alternative iterative search strategies for identifying an efficient set of evidence, where the comparators included in each search is determined by the results of the previous iteration. In addition, Cooper et al. (Cooper et al., 2011) showed that extending the network to include more treatments might lead to increased heterogeneity, which in turn will increase the uncertainty in the results despite the inclusion of additional information.

Another challenging issue of NMA relates to the definition of nodes in the network. It is not uncommon for treatments – although defined similarly – to be administered in different ways across the trials; for instance trials that involve the same drug may apply different doses. Alternatively, it might be the case that some of the drugs belong to the same drug class. In such scenarios it is unclear whether and how similar treatments should be grouped into nodes. Also, as the included studies adopt increasingly different definitions of the nodes transitivity becomes a less plausible assumption. This, in turn may lead to increased inconsistency and/or heterogeneity in the results (Dewilde and Hawkins, 2012; Salanti, 2012). Decisions about whether or not to group similar treatments in a node should be therefore guided by clinical considerations regarding their similarity and the research question at hand. In addition, Del Giovane et al. (Del Giovane et al., 2013) and also Warren et al. (Warren et al., 2014) discussed several alternative approaches to account for variability in treatment definition. The choice of the network structure can then be guided by statistical considerations, after comparing goodness-of-fit measures between alternative plausible models.

In another frequently encountered scenario in which the definition of nodes can be of importance is when interventions are administered as a combination of more than one treatment; the simplest approach would be to analyze each combination as a different node in the NMA. Welton et al. (Welton et al., 2009) and Mills et al. (Mills et al., 2012b) proposed possible scenarios for modeling how interventions interact with each other when combined into a complex intervention, with one of the approaches being the assumption of additive treatment effects. Thorlund et al. (Thorlund and Mills, 2012a) performed a simulation study which showed that when the treatment effects are truly additive, the conventional NMA model performs poorly in comparison to the additive effects model.

In summary, even though there is no exact recipe available for setting up the network and defining the nodes, the choice should be guided by considerations on the transitivity assumption, the presence of statistical inconsistency, the possibility of bias and also practical constraints on the resources available for setting up the database. Ideally, whenever possible, decisions about the network structure should be described a priori in the protocol in order to avoid selective use of data (Salanti, 2012).

6.8 *Incorporating Individual Patient Data (IPD) in NMA*

The NMA models we have discussed so far can be used only for the analysis of aggregated data (AD) while the meta-regression approaches presented in Section 6.3.1 allow the exploration of the effects of only study-level covariates to the relative treatment effects. On the other hand the use of IPD in an NMA (either exclusively or in combination with AD) is expected to increase precision and also allows the distinction between within-study from across-study associations to be made, so as to avoid possible ecological bias (Jansen, 2012). We extensively discuss the statistical methodology and the potential advantages of an IPD-MA when pooling head-to-head trials in a companion paper (Debray et al., 2015). These advantages also apply to NMA, and access to IPD is particularly relevant when the number of included studies is small and the validity of using meta-regression of study-level covariates becomes increasingly questionable. The use of patient-level covariates will allow a better evaluation of the heterogeneity and inconsistency in the network (Greco et al., 2013; Hoaglin et al., 2011; Jansen et al., 2011b, 2008).

A few models for including IPD in an NMA have been recently proposed. Saramago et al. (Saramago et al., 2012) developed a series of NMA models set in a Bayesian background that can be used for the simultaneous synthesis of IPD and AD while incorporating both study and individual-level covariates. Their models also allow the inclusion of studies with different designs (cluster and individual allocation). The authors found that the incorporation of IPD in the network resulted in an increase in the precision compared to an AD-only analysis, even when IPD are available only for a fraction of the studies. Donegan et al. (Donegan et al., 2013a, 2012) presented a model for combining IPD and AD in a single analysis with three alternative specifications (unrelated, exchangeable and common interactions; see also Section 6.3.1). The inclusion of both IPD and AD in the analysis was shown to lead to an increased precision of the estimates of the regression coefficients and a better assessment of the consistency assumption. A similar model was proposed by Jansen (Jansen, 2012). In the same paper a second, alternative model was also suggested for the case of a binary covariate. The author performed a simulation study indicating that the second model is less affected by bias at the cost of larger uncertainty in the results. Finally Ali et al. (Ali et al., 2013) discussed the use of IPD in order to identify possible interactions between treatment effects and potential effect modifiers; when such modifiers are found to be unevenly distributed among studies, the authors suggest that NMA models need to account for these differences.

6.9 Utilizing data from non-randomized and observational studies

Ades and Sutton (Ades and Sutton, 2006) discuss that results obtained from RCTs may not be necessarily generalizable to a wide population and that randomized studies' results could be combined with information from observational studies or patient registries, after adjusting for potential biases. Randomized and non-randomized evidence can be regarded as being complimentary, in the sense that observational studies can be considered to be reliable sources of information regarding the population baseline, while RCTs regarding intervention effects data. Dias et al. (Dias et al., 2013c) describe how non-randomized studies can be used to inform a 'baseline natural history model'. Evidence from such studies can be used to estimate the absolute effect for a reference treatment. This can in turn be combined with NMA results for the relative effects of active treatments, in order to obtain an estimate of the absolute treatment effects.

Schmitz et al. (Schmitz et al., 2013) proposed three alternative models for jointly synthesizing information from RCTs as well as non-randomized studies: the simplest approach presented was to perform a naïve pooling, disregarding differences in study design; the second approach was to utilize non-randomized studies as prior information, while adjusting for bias due to study design; the third was a three-level hierarchical model which accounts for bias and for heterogeneity between trial designs. The first of the models (naïve pooling) should only be used as the first step of the analysis, since it disregards potential biases in non-randomized trials. The second model (using observational evidence as prior information) allows adjusting for biases, but between-trial design heterogeneity is not taken into account, and it is not possible to include more than two different trial designs. The third model (three-level hierarchical model) addresses these issues and should be preferred.

Finally, Soares et al. (Soares et al., 2013) discussed the use of observational data for the case that there are sparse and few data in an NMA. In their approach such data were used to inform the baseline effects, but did not directly contribute to the relative treatment effects.

6.10 Planning future studies

The issue of planning future studies based on the results of an existing NMA has received little attention in the literature. Thorlund and Mills (Thorlund and Mills, 2012b) and also Snapinn and Jiang (Snapinn and Jiang, 2011) provided sample size considerations for determining the statistical power of indirect evidence and Mills et al. (Mills et al., 2011b) performed a simulation study to estimate the power of indirect comparisons; however, there is currently little guidance on the design (i.e. treatments to be compared) and the sample size needed for updating an existing NMA in an optimal manner.

Naci and O' Connor (Naci and O'Connor, 2013) alternatively suggest the design and conduct of prospective NMAs; this would go against the current practice of retrospective NMA, where each individual study is planned in isolation from others. They also suggest that the regulatory agencies should have an active role in the design of future trials, especially in the selection of comparators and in ensuring that the patient populations are comparable in terms of treatment effect modifiers.

7 Concluding remarks

In this article we have presented the currently available methodology on NMA. We have described major developments in the field, discussed all known relevant issues, highlighted key assumptions while providing guidance for the interested researchers.

Our review has several limitations. Pragmatic decisions needed to be made given the lack of a widely accepted terminology referring to network meta-analysis, the abundance of recently published articles and the complexity of new methods in order to ensure a timely publication of this review. Thus, there may have been articles that presented methodological advances for NMA which we failed to identify by not including in our search more online databases and by not hand-searching additional journals. We believe, however, that even if the identified set of articles might not be complete, it is representative of the currently available methods for NMA and that the most important methodological aspects, challenges and solutions of NMA are covered. Moreover, although we present some of the mathematical features of the various models and methods, we do it in a descriptive manner and we do not provide all relevant details. Hence, this review serves as a roadmap for researchers: the keen reader should refer to the original articles for details, keeping also in mind that NMA is still an active, rapidly developing research field.

Acknowledgements

The research leading to these results was conducted as part of the GetReal consortium. For further information please refer to www.imi-getreal.eu

Funding

This work was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement No 115546, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007- 2013) and EFPIA companies' in kind contribution

References

- Achana, F.A., Cooper, N.J., Dias, S., Lu, G., Rice, S.J., Kendrick, D., Sutton, A.J., 2013. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. *Stat.Med.* 32, 752–771. doi:10.1002/sim.5539
- Ades, A.E., 2003. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat.Med.* 22, 2995–3016. doi:10.1002/sim.1566
- Ades, A.E., Caldwell, D.M., Reken, S., Welton, N.J., Sutton, A.J., Dias, S., 2013. Evidence synthesis for decision making 7: a reviewer’s checklist. *Med.Decis.Making* 33, 679–691. doi:10.1177/0272989X13485156
- Ades, A.E., Mavranouzouli, I., Dias, S., Welton, N.J., Whittington, C., Kendall, T., 2010. Network meta-analysis with competing risk outcomes. *Value.Health* 13, 976–983. doi:10.1111/j.1524-4733.2010.00784.x
- Ades, A.E., Sutton, A.J., 2006. Multiparameter evidence synthesis in epidemiology and medical-decision making: current approaches. *J.Roy.Statist.Soc.A* 169, 5–35.
- Ali, S., Mealing, S., Hawkins, N., Lescrauwaet, B., Bjork, S., Mantovani, L., Lampertico, P., 2013. The use of individual patient-level data (IPD) to quantify the impact of pretreatment predictors of response to treatment in chronic hepatitis B patients. *BMJ Open* 3. doi:10.1136/bmjopen-2012-001309
- Bafeta, A., Trinquart, L., Seror, R., Ravaud, P., 2013. Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review. *BMJ* 347, f3675.
- Baker, S.G., Kramer, B.S., 2002. 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.* 2, 13.
- Ballesteros, J., 2005. Orphan comparisons and indirect meta-analysis: a case study on antidepressant efficacy in dysthymia comparing tricyclic antidepressants, selective serotonin reuptake inhibitors, and monoamine oxidase inhibitors by using general linear models. *J.Clin.Psychopharmacol.* 25, 127–131.
- Berkey, C.S., Hoaglin, D.C., Antczak-Bouckoms, A., Mosteller, F., Colditz, G.A., 1998. Meta-analysis of multiple outcomes by regression with random effects. *Stat. Med.* 17, 2537–2550. doi:10.1002/(SICI)1097-0258(19981130)17:22<2537::AID-SIM953>3.0.CO;2-C
- Bucher, H.C., Guyatt, G.H., Griffith, L.E., Walter, S.D., 1997. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J.Clin.Epidemiol.* 50, 683–691.
- Caldwell, D.M., Ades, A.E., Higgins, J.P.T., 2005. Simultaneous comparison of multiple treatments:

- combining direct and indirect evidence. *BMJ* 331, 897–900. doi:10.1136/bmj.331.7521.897
- Caldwell, D.M., Welton, N.J., Ades, A.E., 2010. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *J.Clin.Epidemiol.* 63, 875–882. doi:10.1016/j.jclinepi.2009.08.025
- Caldwell, D.M., Welton, N.J., Dias, S., Ades, A.E., 2012. Selecting the best scale for measuring treatment effect in a network meta-analysis: a case study in childhood nocturnal enuresis. *Res.Synth.Meth.* 3, 126–141.
- Carlin, B.P., Hong, H., Shamliyan, T.A., Sainfort, F., Kane, R.L., 2013. Case study comparing bayesian and frequentist approaches for multiple treatment comparisons. (Prepared by the Minnesota Evidence-Based Practice Center under Contract No. 290-2007-10064-I2). AHRQ Publication No. 12(13)-EHC103-EF. Rockville, MD: Agency for Healthcare Research and Quality.
- Chaimani, A., Higgins, J.P.T., Mavridis, D., Spyridonos, P., Salanti, G., 2013a. Graphical Tools for Network Meta-Analysis in STATA. *PloS One* 8, e76654. doi:10.1371/journal.pone.0076654
- Chaimani, A., Salanti, G., 2012. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Res.Synth.Meth.* 3, 161–176.
- Chaimani, A., Vasiliadis, H.S., Pandis, N., Schmid, C.H., Welton, N.J., Salanti, G., 2013b. Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. *Int. J. Epidemiol.* 42, 1120–1131. doi:10.1093/ije/dyt074
- Chung, H., Lumley, T., 2008. Graphical exploration of network meta-analysis data: the use of multidimensional scaling. *Clin.Trials* 5, 301–307. doi:10.1177/1740774508093614
- Cipriani, A., Barbui, C., Rizzo, C., Salanti, G., 2012. What is a multiple treatments meta-analysis? *Epidemiol.Psychiatr.Sci.* 21, 151–153. doi:10.1017/S2045796011000837
- Cipriani, A., Higgins, J.P.T., Geddes, J.R., Salanti, G., 2013. Conceptual and technical challenges in network meta-analysis. *Ann. Intern. Med.* 159, 130–137. doi:10.7326/0003-4819-159-2-201307160-00008
- Coleman, C.I., Phung, O.J., Cappelleri, J.C., Baker, W.L., Kluger, J., White, C.M., Sobieraj, D.M., 2012. Use of mixed treatment comparison in systematic reviews. (Prepared by the university of Connecticut/Harford Hospital Evidence-Based Practice Center under Contract No. 290-2007-10067-I). AHRQ Publication No. 12-EHC119-EF. Rockville , MD: Agency for Healthcare Research and Quality.
- Cooper, N.J., Peters, J., Lai, M.C., Juni, P., Wandel, S., Palmer, S., Paulden, M., Conti, S., Welton, N.J., Abrams, K.R., Bujkiewicz, S., Spiegelhalter, D., Sutton, A.J., 2011. How valuable are

- multiple treatment comparison methods in evidence-based health-care evaluation? *Value.Health* 14, 371–380. doi:10.1016/j.jval.2010.09.001
- Cooper, N.J., Sutton, A.J., Morris, D., Ades, A.E., Welton, N.J., 2009. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat.Med.* 28, 1861–1881. doi:10.1002/sim.3594
- Coory, M., Jordan, S., 2010. Frequency of treatment-effect modification affecting indirect comparisons: a systematic review. *Pharmacoeconomics*. 28, 723–732. doi:10.2165/11535670-000000000-00000
- Copas, J., 1999. What works?: selectivity models and meta-analysis. *J. R. Stat. Soc. Ser. A Stat. Soc.* 162, 95–109. doi:10.1111/1467-985X.00123
- Cope, S., Jansen, J.P., 2013. Quantitative summaries of treatment effect estimates obtained with network meta-analysis of survival curves to inform decision-making. *BMC Med. Res. Methodol.* 13, 147. doi:10.1186/1471-2288-13-147
- Cucherat, M., Izard, V., 2009. Summary report: Indirect comparisons: Methods and validity. *HAS*.
- Dakin, H.A., Welton, N.J., Ades, A.E., Collins, S., Orme, M., Kelly, S., 2011. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Stat.Med.* doi:10.1002/sim.4284
- D’Ascenzo, F., Biondi-Zoccai, G., 2014. Network Meta-Analyses: The “White Whale” for Cardiovascular Specialists. *J. Cardiothorac. Vasc. Anesth.* 28, 169–173. doi:10.1053/j.jvca.2013.01.004
- Debray, T.P.A., Moons, K.G.M., van Valkenhoef, G., Efthimiou, O., Hummel, N., Groenwold, R.H.H., Reitsma, J.B., on behalf of GetReal methods review group, 2015. Methods for Individual Participant Data meta-analysis of relative treatment effects: a systematic review. *Res. Synth. Methods* Submitted.
- Deeks, J.J., 2002. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat. Med.* 21, 1575–1600. doi:10.1002/sim.1188
- Del Giovane, C., Vacchi, L., Mavridis, D., Filippini, G., Salanti, G., 2013. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. *Stat.Med.* 32, 25–39. doi:10.1002/sim.5512
- Dewilde, S., Hawkins, N., 2012. Investigating incoherence gives insight: clopidogrel is equivalent to extended-release dipyridamole plus aspirin in secondary stroke prevention. *J.Clin.Epidemiol.*

65, 835–845. doi:10.1016/j.jclinepi.2012.01.019

- Dias, S., Sutton, A.J., Ades, A.E., Welton, N.J., 2013a. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.* 33, 607–617. doi:10.1177/0272989X12458724
- Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E., 2013b. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. *Med.Decis.Making* 33, 618–640. doi:10.1177/0272989X13485157
- Dias, S., Welton, N.J., Ades, A.E., 2010a. Study designs to detect sponsorship and other biases in systematic reviews. *J.Clin.Epidemiol.* 63, 587–588. doi:10.1016/j.jclinepi.2010.01.005
- Dias, S., Welton, N.J., Caldwell, D.M., Ades, A.E., 2010b. Checking consistency in mixed treatment comparison meta-analysis. *Stat.Med.* 29, 932–944. doi:10.1002/sim.3767
- Dias, S., Welton, N.J., Marinho, V.C.C., Salanti, G., Higgins, J.P.T., Ades, A.E., 2010c. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *J.Roy.Statist.Soc.A* 173, 613–629.
- Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E., 2013c. Evidence synthesis for decision making 5: the baseline natural history model. *Med.Decis.Making* 33, 657–670. doi:10.1177/0272989X13485155
- Dias, S., Welton, N.J., Sutton, A.J., Caldwell, D.M., Lu, G., Ades, A.E., 2013d. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med.Decis.Making* 33, 641–656. doi:10.1177/0272989X12455847
- Ding, Y., Fu, H., 2012. Bayesian indirect and mixed treatment comparisons across longitudinal time points. *Stat.Med.* doi:10.1002/sim.5688
- Donegan, S., Williamson, P., D'Alessandro, U., Garner, P., Tudur Smith, C., 2013a. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. *Stat.Med.* 32, 914–930. doi:10.1002/sim.5584
- Donegan, S., Williamson, P., D'Alessandro, U., Tudur Smith, C., 2013b. Assessing key assumptions of network meta-analysis: a review of methods. *Res. Synth. Methods* 4, 291–323. doi:10.1002/jrsm.1085
- Donegan, S., Williamson, P., D'Alessandro, U., Tudur Smith, C., 2012. Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates.

Stat.Med. 31, 3840–3857. doi:10.1002/sim.5470

- Donegan, S., Williamson, P., Gamble, C., Tudur Smith, C., 2010. Indirect comparisons: a review of reporting and methodological quality. *PLoS.One*. 5, e11054. doi:10.1371/journal.pone.0011054
- Eckermann, S., Coory, M., Willan, A.R., 2009. Indirect comparison: relative risk fallacies and odds solution. *J.Clin.Epidemiol.* 62, 1031–1036. doi:10.1016/j.jclinepi.2008.10.013
- Eddy, D.M., Shachter, R., 1992. *Meta-Analysis by the Confidence Profile Method*, Har/Dskt edition. ed. Academic Press, Boston.
- Edwards, S.J., Clarke, M.J., Wordsworth, S., Borrill, J., 2009. Indirect comparisons of treatments based on systematic reviews of randomised controlled trials. *Int. J. Clin. Pract.* 63, 841–854. doi:10.1111/j.1742-1241.2009.02072.x
- Efthimiou, O., Mavridis, D., Cipriani, A., Leucht, S., Bagos, P., Salanti, G., 2014a. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Stat. Med.* 33, 2275–2287. doi:10.1002/sim.6117
- Efthimiou, O., Mavridis, D., Riley, R.D., Cipriani, A., Salanti, G., 2014b. Joint synthesis of multiple correlated outcomes in networks of interventions. *Biostat. Oxf. Engl.* doi:10.1093/biostatistics/kxu030
- Egger, M., Smith, G.D., Phillips, A.N., 1997. Meta-analysis: Principles and procedures. *BMJ* 315, 1533–1537. doi:10.1136/bmj.315.7121.1533
- EUnetHTA – European network for Health Technology Assessment (last), 2013. Comparators and comparisons: Direct and indirect comparisons. Available [Httpwww.eunethta.eu/sites5026fedimbelgiumbefiles/Direct20and20indirect20comparisons.pdf](http://www.eunethta.eu/sites5026fedimbelgiumbefiles/Direct20and20indirect20comparisons.pdf).
- Fadda, V., Maratea, D., Trippoli, S., Messori, A., 2011. Results can be summarised in a simple figure. *BMJ* 342, d1555–d1555. doi:10.1136/bmj.d1555
- Falissard, B., Zylberman, M., Cucherat, M., Iazard, V., Meyer, F., 2009. Real medical benefit assessed by indirect comparison. *Therapie* 64, 225–232. doi:10.2515/therapie/2009032
- Franchini, A.J., Dias, S., Ades, A.E., Jansen, J.P., Welton, N.J., 2012. Accounting for correlation in network meta-analysis with multi-arm trials. *Res.Synth.Meth.* 3, 142–160.
- Gartlehner, G., Moore, C.G., 2008. Direct versus indirect comparisons: a summary of the evidence. *IntJTechnolAssessHealth Care* 24, 170–177. doi:10.1017/S0266462308080240
- Gasparrini, A., Armstrong, B., Kenward, M.G., 2012. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat. Med.* 31, 3821–3839. doi:10.1002/sim.5471
- Glenny, A.M., Altman, D.G., Song, F., Sakarovitch, C., Deeks, J.J., D’Amico, R., Bradburn, M., Eastwood, A.J., 2005. Indirect comparisons of competing interventions. *Health TechnolAssess*

9, 1–iv.

- Gleser, L., Olkin, L.V., 2009. Stochastically dependent effect sizes. In *The Handbook of Research Synthesis and Meta-Analysis*, (Cooper H, Hedges LV (eds). Russell Sage Foundation.
- Govan, L., Ades, A.E., Weir, C.J., Welton, N.J., Langhorne, P., 2010. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. *Stat.Med.* 29, 1340–1356. doi:10.1002/sim.3869
- Greco, T., Landoni, G., Biondi-Zoccai, G., D'Ascenzo, F., Zangrillo, A., 2013. A Bayesian network meta-analysis for binary outcome: how to do it. *StatMethods MedRes*. doi:10.1177/0962280213500185
- Greenland, S., 2005. Multiple-bias modelling for analysis of observational data. *J. R. Stat. Soc. Ser. A Stat. Soc.* 168, 267–306. doi:10.1111/j.1467-985X.2004.00349.x
- Greenland, S., 2003. The Impact of Prior Distributions for Uncontrolled Confounding and Response Bias: A Case Study of the Relation of Wire Codes and Magnetic Fields to Childhood Leukemia. *J. Am. Stat. Assoc.* 98, 47–54.
- Griffin, S., Bojke, L., Main, C., Palmer, S., 2006. Incorporating direct and indirect evidence using bayesian methods: an applied case study in ovarian cancer. *Value.Health* 9, 123–131. doi:10.1111/j.1524-4733.2006.00090.x
- Hawkins, N., Scott, D.A., Woods, B., 2009a. How far do you go? Efficient searching for indirect evidence. *Med.Decis.Making* 29, 273–281. doi:10.1177/0272989X08330120
- Hawkins, N., Scott, D.A., Woods, B.S., Thatcher, N., 2009b. No study left behind: a network meta-analysis in non-small-cell lung cancer demonstrating the importance of considering all relevant data. *Value.Health* 12, 996–1003. doi:10.1111/j.1524-4733.2009.00541.x
- Higgins, J.P.T., Jackson, D., Barrett, J.K., Lu, G., Ades, A.E., White, I.R., 2012. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res.Synth.Meth.* 3, 98–110.
- Higgins, J.P.T., Whitehead, A., 1996. Borrowing Strength from External Trials in a Meta-Analysis. *Stat. Med.* 15, 2733–2749. doi:10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0
- Hoaglin, D.C., Hawkins, N., Jansen, J.P., Scott, D.A., Itzler, R., Cappelleri, J.C., Boersma, C., Thompson, D., Larholt, K.M., Diaz, M., Barrett, A., 2011. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value.Health* 14, 429–437. doi:10.1016/j.jval.2011.01.011

- Hong, H., Carlin, B.P., Chu, H., Shamliyan, T.A., Wang, S., Kane, R.L., 2013a. A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons, *AHRQ Methods for Effective Health Care*. Agency for Healthcare Research and Quality (US), Rockville (MD).
- Hong, H., Carlin, B.P., Shamliyan, T.A., Wyman, J.F., Ramakrishnan, R., Sainfort, F., Kane, R.L., 2013b. Comparing Bayesian and Frequentist Approaches for Multiple Outcome Mixed Treatment Comparisons. *Med.Decis.Making*. doi:10.1177/0272989X13481110
- Ioannidis, J.P., 2009. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ*. 181, 488–493. doi:10.1503/cmaj.081086
- Ioannidis, J.P., 2006. Indirect comparisons: the mesh and mess of clinical trials. *Lancet* 368, 1470–1472. doi:10.1016/S0140-6736(06)69615-3
- Jackson, D., Barrett, J.K., Rice, S., White, I.R., Higgins, J.P.T., 2014. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Stat. Med.* 33, 3639–3654. doi:10.1002/sim.6188
- Jackson, D., Riley, R., White, I.R., 2011. Multivariate meta-analysis: Potential and promise. *Stat. Med.* 30, 2481–2498. doi:10.1002/sim.4172
- Jackson, D., White, I.R., Thompson, S.G., 2010. Extending DerSimonian and Laird’s methodology to perform multivariate random effects meta-analyses. *Stat. Med.* 29, 1282–1297. doi:10.1002/sim.3602
- Jansen, J.P., 2012. Network meta-analysis of individual and aggregate level data. *Res.Synth.Meth.* 3, 177–190.
- Jansen, J.P., 2011. Network meta-analysis of survival data with fractional polynomials. *BMC.Med.Res.Methodol.* 11, 61. doi:10.1186/1471-2288-11-61
- Jansen, J.P., Cope, S., 2012. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. *BMC.Med.Res.Methodol.* 12, 152. doi:10.1186/1471-2288-12-152
- Jansen, J.P., Crawford, B., Bergman, G., Stam, W., 2008. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. *Value.Health* 11, 956–964. doi:10.1111/j.1524-4733.2008.00347.x
- Jansen, J.P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., Lee, K., Boersma, C., Annemans, L., Cappelleri, J.C., 2011a. 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 J. Int. Soc. Pharmacoeconomics Outcomes Res.* 14, 417–428. doi:10.1016/j.jval.2011.04.002
- Jansen, J.P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., Lee, K., Boersma, C., Annemans, L., Cappelleri, J.C., 2011b. 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* 14, 417–428. doi:10.1016/j.jval.2011.04.002
- Jansen, J.P., Naci, H., 2013. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med.* 11, 159. doi:10.1186/1741-7015-11-159
- Jansen, J.P., Schmid, C.H., Salanti, G., 2012. Directed acyclic graphs can help understand bias in indirect and mixed treatment comparisons. *J. Clin. Epidemiol.* 65, 798–807. doi:10.1016/j.jclinepi.2012.01.002
- Jonas, D.E., Wilkins, T.M., Bangdiwala, S., Bann, C.M., Morgan, L.C., Thaler, K.J., Amick, H.R., Gartlehner, G., 2013. Findings of Bayesian mixed treatment comparison meta-analysis: comparison and exploration using real-world trial data and simulation. (Prepared by RTI-UNC Evidence-Based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality.
- Jones, B., Roger, J., Lane, P.W., Lawton, A., Fletcher, C., Cappelleri, J.C., Tate, H., Moneuse, P., 2011. Statistical approaches for conducting network meta-analysis in drug development. *Pharm. Stat.* 10, 523–531. doi:10.1002/pst.533
- Julious, S.A., Wang, S.J., 2008. How biased are indirect comparisons, particularly when comparisons are made over time in controlled trials. *Drug Inform. J.* 42, 625.
- Kirkham, J.J., Riley, R.D., Williamson, P.R., 2012. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Stat. Med.* 31, 2179–2195. doi:10.1002/sim.5356
- König, J., Krahn, U., Binder, H., 2013. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Stat. Med.* n/a–n/a. doi:10.1002/sim.6001
- Krahn, U., Binder, H., König, J., 2013. A graphical tool for locating inconsistency in network meta-analyses. *BMC Med. Res. Methodol.* 13, 35. doi:10.1186/1471-2288-13-35
- Lee, A.W., 2013. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J. Clin. Epidemiol.* doi:10.1016/j.jclinepi.2013.07.014
- Lin, L., Zhang, J., Chu, and H., 2015. pcnetmeta: Patient-Centered Network Meta-Analysis. Available

at: <http://www.r-project.org/>.

- Li, T., Puhan, M.A., Vedula, S.S., Singh, S., Dickersin, K., The Ad Hoc Network Meta-analysis Methods Meeting Working Group, 2011. Network meta-analysis-highly attractive but more methodological research is needed. *BMC Med.* 9, 79. doi:10.1186/1741-7015-9-79
- Lu, G., Ades, A., 2009. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics.* 10, 792–805. doi:10.1093/biostatistics/kxp032
- Lu, G., Ades, A.E., 2006. Assessing evidence inconsistency in mixed treatment comparisons. *J.Amer.Statist.Assoc.* 101, 447–459.
- Lu, G., Ades, A.E., 2004. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat. Med.* 23, 3105–3124. doi:10.1002/sim.1875
- Lu, G., Ades, A.E., Sutton, A.J., Cooper, N.J., Briggs, A.H., Caldwell, D.M., 2007. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat.Med.* 26, 3681–3699. doi:10.1002/sim.2831
- Lu, G., Welton, N.J., Higgins, J.P.T., White, I.R., Ades, A.E., 2011. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Res.Synth.Meth.* 2, 43–60.
- Lumley, T., 2002. Network meta-analysis for indirect treatment comparisons. *Stat. Med.* 21, 2313–2324. doi:10.1002/sim.1201
- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - A Bayesian Modelling Framework: Concepts, Structure, and Extensibility. *Stat. Comput.* 10, 325–337. doi:10.1023/A:1008929526011
- Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: Evolution, critique and future directions. *Stat. Med.* 28, 3049–3067. doi:10.1002/sim.3680
- Madan, J., Chen, Y.-F., Aveyard, P., Wang, D., Yahaya, I., Munafo, M., Bauld, L., Welton, N., 2014. Synthesis of evidence on heterogeneous interventions with multiple outcomes recorded over multiple follow-up times reported inconsistently: a smoking cessation case-study. *J. R. Stat. Soc. Ser. A Stat. Soc.* 177, 295–314. doi:10.1111/rssa.12018
- Madan, J., Stevenson, M.D., Cooper, K.L., Ades, A.E., Whyte, S., Akehurst, R., 2011. Consistency between direct and indirect trial evidence: is direct evidence always more reliable? *Value.Health* 14, 953–960. doi:10.1016/j.jval.2011.05.042
- Malone, D.C., 2007. Using indirect comparisons in pharmacoeconomic studies--time for implementation. *Clin.Ther.* 29, 2454–2455. doi:10.1016/j.clinthera.2007.11.011
- Mavridis, D., Salanti, G., 2013. A practical introduction to multivariate meta-analysis. *Stat. Methods Med. Res.* 22, 133–158. doi:10.1177/0962280211432219

- Mavridis, D., Sutton, A., Cipriani, A., Salanti, G., 2013. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. *Stat.Med.* 32, 51–66. doi:10.1002/sim.5494
- Mills, E.J., Bansback, N., Ghement, I., Thorlund, K., Kelly, S., Puhan, M.A., Wright, J., 2011a. Multiple treatment comparison meta-analyses: a step forward into complexity. *Clin.Epidemiol.* 3, 193–202. doi:10.2147/CLEP.S16526
- Mills, E.J., Ghement, I., O'Regan, C., Thorlund, K., 2011b. Estimating the power of indirect comparisons: a simulation study. *PLoS.One.* 6, e16237. doi:10.1371/journal.pone.0016237
- Mills, E.J., Ioannidis, J.P., Thorlund, K., Schunemann, H.J., Puhan, M.A., Guyatt, G.H., 2012a. How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA* 308, 1246–1253. doi:10.1001/2012.jama.11228
- Mills, E.J., Kanters, S., Thorlund, K., Chaimani, A., Veroniki, A.A., Ioannidis, J.P., 2013a. The effects of excluding treatments from network meta-analyses: survey. *BMJ* 347, f5195.
- Mills, E.J., Thorlund, K., Ioannidis, J.P., 2013b. Demystifying trial networks and network meta-analysis. *BMJ* 346, f2914.
- Mills, E.J., Thorlund, K., Ioannidis, J.P., 2012b. Calculating additive treatment effects from multiple randomized trials provides useful estimates of combination therapies. *J.Clin.Epidemiol.* 65, 1282–1288. doi:10.1016/j.jclinepi.2012.07.012
- Naci, H., Fleurence, R., 2011. Using indirect evidence to determine the comparative effectiveness of prescription drugs: Do benefits outweigh risks? *Health.Outc.Res.Med.* 2, 241–249.
- Naci, H., O'Connor, A.B., 2013. Assessing comparative effectiveness of new drugs before approval using prospective network meta-analyses. *J. Clin. Epidemiol.* 66, 812–816. doi:10.1016/j.jclinepi.2013.04.002
- Naudet, F., Millet, B., Charlier, P., Reymann, J.M., Maria, A.S., Falissard, B., 2013. Which placebo to cure depression? A thought-provoking network meta-analysis. *BMC.Med.* 11, 230. doi:10.1186/1741-7015-11-230
- Neupane, B., Richer, D., Bonner, A.J., Kibret, T., Beyene, J., 2014. Network Meta-Analysis Using R: A Review of Currently Available Automated Packages. *PLoS ONE* 9, e115065. doi:10.1371/journal.pone.0115065
- Nikolakopoulou, A., Chaimani, A., Veroniki, A.A., Vasiliadis, H.S., Schmid, C.H., Salanti, G., 2014. Characteristics of networks of interventions: a description of a database of 186 published networks. *PloS One* 9, e86754. doi:10.1371/journal.pone.0086754
- Nixon, R.M., Bansback, N., Brennan, A., 2007. Using mixed treatment comparisons and meta-

- regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat.Med.* 26, 1237–1254. doi:10.1002/sim.2624
- Norton, E.C., Miller, M.M., Wang, J.J., Coyne, K., Kleinman, L.C., 2012. Rank reversal in indirect comparisons. *Value.Health* 15, 1137–1140. doi:10.1016/j.jval.2012.06.001
- Ohlssen, D., Price, K.L., Amy, X.H., Hong, H., Kerman, J., Fu, H., Quartey, G., Heilmann, C.R., Ma, H., Carlin, B.P., 2013. Guidance on the implementation and reporting of a drug safety Bayesian network meta-analysis. *Pharm.Stat.* doi:10.1002/pst.1592
- O'Regan, C., Ghement, I., Eyawo, O., Guyatt, G.H., Mills, E.J., 2009. Incorporating multiple interventions in meta-analysis: an evaluation of the mixed treatment comparison with the adjusted indirect comparison. *Trials* 10, 86. doi:10.1186/1745-6215-10-86
- Ouwens, M.J.N.M., Philips, Z., Jansen, J.P., 2010. Network meta-analysis of parametric survival curves. *Res.Synth.Meth.* 1, 258–271.
- Panayidou, K., Gsteiger, S., Kilcher, G., Egger, M., Carreras, M., Efthimiou, O., Debray, T.P.A., Trelle, S., Hummel, N., on behalf of GetReal methods review group, 2015. Mathematical modelling for predicting real world effectiveness from RCT efficacy data: a systematic review. *Res. Synth. Methods* Submitted.
- Piepho, H.-P., 2014. Network-meta analysis made easy: Detection of inconsistency using factorial analysis-of-variance models. *BMC Med. Res. Methodol.* 14, 61. doi:10.1186/1471-2288-14-61
- Piepho, H.-P., Madden, L.V., Williams, E.R., 2015. Multiplicative interaction in network meta-analysis. *Stat. Med.* 34, 582–594. doi:10.1002/sim.6372
- Piepho, H.P., Williams, E.R., Madden, L.V., 2012. The use of two-way linear mixed models in multitreatment meta-analysis. *Biometrics* 68, 1269–1277.
- Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.
- Price, M.J., Welton, N.J., Ades, A.E., 2011a. Parameterization of treatment effects for meta-analysis in multi-state Markov models. *Stat.Med.* 30, 140–151. doi:10.1002/sim.4059
- Price, M.J., Welton, N.J., Briggs, A.H., Ades, A.E., 2011b. Model averaging in the presence of structural uncertainty about treatment effects: influence on treatment decision and expected value of information. *Value.Health* 14, 205–218. doi:10.1016/j.jval.2010.08.001
- R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Rücker, G., 2012. Network meta-analysis, electrical networks and graph theory. *Res.Synth.Meth.* 3, 312–324.

- Rücker, G., Schwarzer, G., 2014. Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. *Stat. Med.* 33, 4353–4369. doi:10.1002/sim.6236
- Rücker, G., Schwarzer, G., Krahn, U., König, J., 2015. netmeta: Network Meta-Analysis using Frequentist Methods. Available at: <http://www.r-project.org/>.
- Salanti, G., 2012. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res. Synth. Methods* 3, 80–97. doi:10.1002/jrsm.1037
- Salanti, G., Ades, A.E., Ioannidis, J.P.A., 2011. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J. Clin. Epidemiol.* 64, 163–171. doi:10.1016/j.jclinepi.2010.03.016
- Salanti, G., Dias, S., Welton, N.J., Ades, A.E., Golfinopoulos, V., Kyrgiou, M., Mauri, D., Ioannidis, J.P., 2010. Evaluating novel agent effects in multiple-treatments meta-regression. *Stat.Med.* 29, 2369–2383. doi:10.1002/sim.4001
- Salanti, G., Higgins, J.P., Ades, A.E., Ioannidis, J.P., 2008a. Evaluation of networks of randomized trials. *StatMethods MedRes* 17, 279–301. doi:10.1177/0962280207080643
- Salanti, G., Kavvoura, F.K., Ioannidis, J.P., 2008b. Exploring the geometry of treatment networks. *Ann.Intern.Med.* 148, 544–553.
- Salanti, G., Marinho, V., Higgins, J.P., 2009. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J.Clin.Epidemiol.* 62, 857–864. doi:10.1016/j.jclinepi.2008.10.001
- Saramago, P., Sutton, A.J., Cooper, N.J., Manca, A., 2012. Mixed treatment comparisons using aggregate and individual participant level data. *Stat.Med.* 31, 3516–3536. doi:10.1002/sim.5442
- SAS Institute, 2011. The SAS system for Windows. Release 9.2. SAS Inst., Cary, NC.
- Schmid, C.H., Trikalinos, T.A., Olkin, I., 2013. Bayesian network meta-analysis for unordered categorical outcomes with incomplete data. *Res. Synth. Methods* n/a–n/a. doi:10.1002/jrsm.1103
- Schmidli, H., Wandel, S., Neuenschwander, B., 2013. The network meta-analytic-predictive approach to non-inferiority trials. *Stat. Methods Med. Res.* 22, 219–240. doi:10.1177/0962280211432512
- Schmitz, S., Adams, R., Walsh, C., 2013. Incorporating data from various trial designs into a mixed treatment comparison model. *Stat.Med.* doi:10.1002/sim.5764
- Schmitz, S., Adams, R., Walsh, C., 2012. The use of continuous data versus binary data in MTC models: a case study in rheumatoid arthritis. *BMC.Med.Res.Methodol.* 12, 167. doi:10.1186/1471-2288-12-167

- Senn, S., Gavini, F., Magrez, D., Scheen, A., 2013. Issues in performing a network meta-analysis. *StatMethods MedRes* 22, 169–189. doi:10.1177/0962280211432220
- Signorovitch, J., Erder, M.H., Xie, J., Sikirica, V., Lu, M., Hodgkins, P.S., Wu, E.Q., 2012. Comparative effectiveness research using matching-adjusted indirect comparison: an application to treatment with guanfacine extended release or atomoxetine in children with attention-deficit/hyperactivity disorder and comorbid oppositional defiant disorder. *Pharmacoepidemiol. Saf* 21 Suppl 2, 130–137. doi:10.1002/pds.3246
- Snapinn, S., Jiang, Q., 2011. Indirect comparisons in the comparative efficacy and non-inferiority settings. *Pharm.Stat.* 10, 420–426. doi:10.1002/pst.509
- Soares, M.O., Dumville, J.C., Ades, A.E., Welton, N.J., 2013. Treatment comparisons for decision making: facing the problems of sparse and few data. *J.Roy.Statist.Soc.A.*
- Sobieraj, D.M., Cappelleri, J.C., Baker, W.L., Phung, O.J., White, C.M., Coleman, C.I., 2013. Methods used to conduct and report Bayesian mixed treatment comparisons published in the medical literature: a systematic review. *BMJ Open* 3. doi:10.1136/bmjopen-2013-003111
- Song, F., Altman, D.G., Glenny, A.M., Deeks, J.J., 2003. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 326, 472. doi:10.1136/bmj.326.7387.472
- Song, F., Clark, A., Bachmann, M.O., Maas, J., 2012. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC Med. Res. Methodol.* 12, 138. doi:10.1186/1471-2288-12-138
- Song, F., Harvey, I., Lilford, R., 2008. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *J. Clin. Epidemiol.* 61, 455–463. doi:10.1016/j.jclinepi.2007.06.006
- Song, F., Loke, Y.K., Walsh, T., Glenny, A.M., Eastwood, A.J., Altman, D.G., 2009. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 338, b1147.
- Song, F., Xiong, T., Parekh-Bhurke, S., Loke, Y.K., Sutton, A.J., Eastwood, A.J., Holland, R., Chen, Y.F., Glenny, A.M., Deeks, J.J., Altman, D.G., 2011. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ* 343, d4909.
- Spiegelhalter, D.J., Best, N.G., 2003. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Stat. Med.* 22, 3687–3709. doi:10.1002/sim.1586
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model

- complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64, 583–639. doi:10.1111/1467-9868.00353
- Spineli, L.M., Higgins, J.P., Cipriani, A., Leucht, S., Salanti, G., 2013. Evaluating the impact of imputations for missing participant outcome data in a network meta-analysis. *Clin. Trials Lond. Engl.* 10, 378–388. doi:10.1177/1740774512470317
- Stan Development Team, 2014. Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0.
- StataCorp., 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.
- Sturtz, S., Bender, R., 2012. Unsolved issues of mixed treatment comparison meta-analysis: network size and inconsistency. *Res.Synth.Meth.* 3, 300–311.
- Sutton, A., Ades, A.E., Cooper, N., Abrams, K., 2008. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics.* 26, 753–767.
- Sutton, A.J., Cooper, N.J., Jones, D.R., 2009. Evidence synthesis as the key to more coherent and efficient research. *BMC.Med.Res.Methodol.* 9, 29. doi:10.1186/1471-2288-9-29
- Tan, S.H., Bujkiewicz, S., Sutton, A., Dequen, P., Cooper, N., 2013. Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons. *JHealth ServResPolicy* 18, 224–232. doi:10.1177/1355819613498379
- Tan, S.H., Cooper, N.J., Bujkiewicz, S., Welton, N.J., Caldwell, D.M., Sutton, A.J., 2014. Novel presentational approaches were developed for reporting network meta-analysis. *J. Clin. Epidemiol.* doi:10.1016/j.jclinepi.2013.11.006
- Thorlund, K., Mills, E., 2012a. Stability of additive treatment effects in multiple treatment comparison meta-analysis: a simulation study. *Clin.Epidemiol.* 4, 75–85. doi:10.2147/CLEP.S29470
- Thorlund, K., Mills, E.J., 2012b. Sample size and power considerations in network meta-analysis. *Syst.Rev.* 1, 41. doi:10.1186/2046-4053-1-41
- Thorlund, K., Thabane, L., Mills, E.J., 2013. Modelling heterogeneity variances in multiple treatment comparison meta-analysis--are informative priors the better solution? *BMC.Med.Res.Methodol.* 13, 2. doi:10.1186/1471-2288-13-2
- Thorlund, K., Zafari, Z., Druyts, E., Mills, E.J., Sadatsafavi, M., 2014. The impact of incorporating Bayesian network meta-analysis in cost-effectiveness analysis - a case study of pharmacotherapies for moderate to severe COPD. *Cost Eff. Resour. Alloc. CE* 12, 8. doi:10.1186/1478-7547-12-8
- Trinquart, L., Abbe, A., Ravaud, P., 2012a. Impact of reporting bias in network meta-analysis of antidepressant placebo-controlled trials. *PLoS.One.* 7, e35219.

doi:10.1371/journal.pone.0035219

- Trinquart, L., Chatellier, G., Ravaud, P., 2012b. Adjustment for reporting bias in network meta-analysis of antidepressant trials. *BMC.Med.Res.Methodol.* 12, 150. doi:10.1186/1471-2288-12-150
- Turner, R.M., Spiegelhalter, D.J., Smith, G.C.S., Thompson, S.G., 2009. Bias modelling in evidence synthesis. *J. R. Stat. Soc. Ser. A Stat. Soc.* 172, 21–47. doi:10.1111/j.1467-985X.2008.00547.x
- Tu, Y.K., Faggion, C.M., 2012. A primer on network meta-analysis for dental research. *ISRN.Dent.* 2012, 276520. doi:10.5402/2012/276520
- van Houwelingen, H.C., Arends, L.R., Stijnen, T., 2002. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat. Med.* 21, 589–624.
- van Valkenhoef, G., Ades, A.E., 2013. Evidence synthesis assumes additivity on the scale of measurement: response to “Rank reversal in indirect comparisons” by Norton et al. *Value.Health* 16, 449–451. doi:10.1016/j.jval.2012.11.012
- van Valkenhoef, G., Kuiper, J., 2015. *gemtc: Network Meta-Analysis Using Bayesian*. Available at: <http://www.r-project.org/>.
- van Valkenhoef, G., Lu, G., de Brock, B., Hillege, H., Ades, A.E., Welton, N.J., 2012a. Automating network meta-analysis. *Res.Synth.Meth.* 3, 285–299.
- van Valkenhoef, G., Tervonen, T., de Brock, B., Hillege, H., 2012b. Algorithm parameterization of mixed treatment comparisons. *Stat.Comput.* 22, 1099–1111.
- van Valkenhoef, G., Tervonen, T., Zhao, J., de, B.B., Hillege, H.L., Postmus, D., 2012c. Multicriteria benefit-risk assessment using network meta-analysis. *J.Clin.Epidemiol.* 65, 394–403. doi:10.1016/j.jclinepi.2011.09.005
- Veroniki, A.A., Vasiliadis, H.S., Higgins, J.P., Salanti, G., 2013. Evaluation of inconsistency in networks of interventions. *Int.J.Epidemiol.* 42, 332–345. doi:10.1093/ije/dys222
- Viechtbauer, W., 2010. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* 36, 1–48. doi:10.18637/jss.v036.i03
- Walter, S.D., 2000. Choice of effect measure for epidemiological data. *J. Clin. Epidemiol.* 53, 931–939. doi:10.1016/S0895-4356(00)00210-9
- Warren, F.C., Abrams, K.R., Sutton, A.J., 2014. Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. *Stat. Med.* doi:10.1002/sim.6131
- Wells, G.A., Sultan, S.A., Chen, L., Khan, M., Coyle, D., 2009. Indirect evidence: Indirect treatment comparisons in meta-analysis. *Ott. Can. Agency Drugs Technol. Health.*

- Welton, N.J., Caldwell, D.M., Adamopoulos, E., Vedhara, K., 2009. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *Am.J.Epidemiol.* 169, 1158–1165. doi:10.1093/aje/kwp014
- Welton, N.J., Cooper, N.J., Ades, A.E., Lu, G., Sutton, A.J., 2008. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. *Stat.Med.* 27, 5620–5639. doi:10.1002/sim.3377
- Welton, N.J., Sutton, A.J., Cooper, N., Abrams, K.R., Ades, A.E., 2012. Mixed and indirect treatment comparisons, in: *Evidence Synthesis for Decision Making in Healthcare*. John Wiley & Sons, New York.
- Welton, N.J., Willis, S.R., Ades, A.E., 2010. Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. *Res.Synth.Meth.* 1, 239–257.
- White, I.R., Barrett, J.K., Jackson, D., Higgins, J.P.T., 2012. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res.Synth.Meth.* 3, 111–125.
- Wolpert, R.L., Mengersen, K.L., 2004. Adjusted Likelihoods for Synthesizing Empirical Evidence from Studies That Differ in Quality and Design: Effects of Environmental Tobacco Smoke. *Stat. Sci.* 19, 450–471.
- Woods, B.S., Hawkins, N., Scott, D.A., 2010. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: a tutorial. *BMC.Med.Res.Methodol.* 10, 54. doi:10.1186/1471-2288-10-54
- Xiong, T., Parekh-Bhurke, S., Loke, Y.K., Abdelhamid, A., Sutton, A.J., Eastwood, A.J., Holland, R., Chen, Y.F., Walsh, T., Glenny, A.M., Song, F., 2013. Overall similarity and consistency assessment scores are not sufficiently accurate for predicting discrepancy between direct and indirect comparison estimates. *J.Clin.Epidemiol.* 66, 184–191. doi:10.1016/j.jclinepi.2012.06.022
- Yang, G., Liu, D., Liu, R.Y., Xie, M., Hoaglin, D.C., 2014. Efficient network meta-analysis: A confidence distribution approach. *Stat. Methodol.* doi:10.1016/j.stamet.2014.01.003
- Zhang, J., Carlin, B.P., Neaton, J.D., Soon, G.G., Nie, L., Kane, R., Virnig, B.A., Chu, H., 2013. Network meta-analysis of randomized clinical trials: Reporting the proper summaries. *Clin.Trials.* doi:10.1177/1740774513498322

Figure 1: Each circle represents an intervention, lines represent direct comparisons. Left panel: Three interventions A, B, and C form a simple network. The indirect AB comparison is estimated via C, i.e. using the direct AC and BC comparisons. The mixed relative treatment effect for AB is estimated by combining the direct comparison and the indirect comparison. Right panel: A more complex network with five interventions and eight direct comparisons. Overall, one direct comparison and four indirect comparisons contribute evidence to A versus B (indirect comparisons are: via C; via E; via C and D; via E and D).

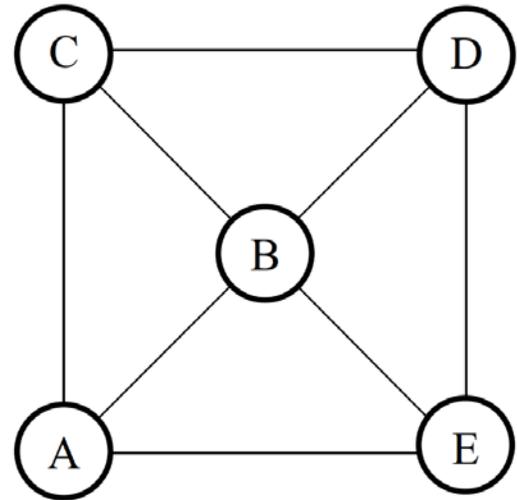
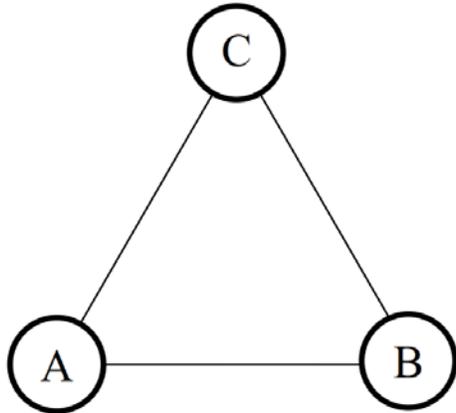


Table 1: An overview of the methods for assessing statistical inconsistency along with the limitations of each method

Approach to inconsistency	Method	Limitations
Local methods	Loop-specific	<ul style="list-style-type: none"> • Does not compare direct to all indirect evidence for each comparison • Different loops sharing a comparison are not independent • Multiple-testing issues
	Composite test	Fails to account for correlations induce by multi-arm studies
	Back-calculation	<ul style="list-style-type: none"> • Problematic for a random-effects meta-analysis • Cannot properly handle multi-arm studies
	Node splitting	<ul style="list-style-type: none"> • Computationally intensive • Cannot properly handle multi-arm studies
Global methods	Lu and Ades model	Depends on parameterization when multi-arm studies are included
	Design-by-treatment model	Non-intuitive definition of inconsistency.
	Unrelated mean effects model	Depends on parameterization when multi-arm studies are included
	Q statistic for inconsistency	Based on the notion of design-by-treatment inconsistency model (non-intuitive definition of inconsistency)
	Multidimensional scaling	Cannot properly handle multi-arm studies

Table 2: Available software solutions for NMA

Method	Software
Bucher method	<ul style="list-style-type: none"> • Stata (indirect command) • ITC (Microsoft Windows application, available at https://www.cadth.ca/resources/itc-user-guide/introduction)
NMA as meta-regression	<ul style="list-style-type: none"> • Stata (metareg command: when only two-armed trials are included in the network; mvmeta and network commands: in the presence of multi-armed studies in the network; lincom command can be used to estimate the functional parameters in both cases.) • R (rma command from metafor package: when only two-armed studies are included) (Viechtbauer, 2010)
Hierarchical NMA model	<ul style="list-style-type: none"> • BUGS codes available at: http://www.mtm.uoi.gr and http://www.bris.ac.uk/social-community-medicine/projects/mpes/ • R (gemtc package (van Valkenhoef and Kuiper, 2015), information available at https://drugis.org/software/r-packages/gemtc; pcnetmeta package (Lin et al., 2015) available at http://cran.r-project.org/web/packages/pcnetmeta/index.html)
NMA as a multivariate meta-analysis	<ul style="list-style-type: none"> • Stata (mvmeta; network) • R (mvmeta) (Gasparrini et al., 2012)
Graph-theoretical approach to NMA	<ul style="list-style-type: none"> • R (netmeta package: accounts for multi-armed studies available at http://cran.r-project.org/web/packages/netmeta/index.html) (Rücker et al., 2015)
Loop-specific approach for inconsistency	<ul style="list-style-type: none"> • R (routines available at http://mtm.uoi.gr/index.php/how-to-do-an-mtm) • Stata (network graphs, available at http://mtm.uoi.gr/index.php/stata-routines-for-network-meta-analysis)
Node splitting approach	<ul style="list-style-type: none"> • BUGS (codes available at http://www.bristol.ac.uk/cobm/research/mpes) • GeMTC software • Stata (network package)
Lu & Ades model	<ul style="list-style-type: none"> • BUGS (codes available at http://www.bristol.ac.uk/social-community-medicine/projects/mpes) • Stata (mvmeta) (Gasparrini et al., 2012) • GeMTC software
Design-by-treatment model	<ul style="list-style-type: none"> • Stata (mvmeta; network)
Q-statistics in NMA	<ul style="list-style-type: none"> • R (netmeta) • Stata (mvmeta; network)
Graphical presentation	<ul style="list-style-type: none"> • Stata (network graphs, available at http://mtm.uoi.gr/index.php/stata-routines-for-network-meta-analysis) • R (netmeta)