UNIVERSITY OF IOANNINA
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

SECTION OF SOCIAL MEDICINE & MENTAL HEALTH
DEPARTMENT OF HYGIENE & EPIDEMIOLOGY

Investigating bias in network meta-analysis

Anna Chaimani

DOCTORAL THESIS

IOANNINA 2014, GREECE

UNIVERSITY OF IOANNINA
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

SECTION OF SOCIAL MEDICINE & MENTAL HEALTH
DEPARTMENT OF HYGIENE & EPIDEMIOLOGY

Investigating bias in network meta-analysis

Anna Chaimani

DOCTORAL THESIS

IOANNINA 2014, GREECE

“Η έγκριση της διδακτορικής διατριβής από το τμήμα Ιατρικής του Πανεπιστημίου Ιωαννίνων δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.”

(Ν. 5343/32, άρθρο 202, παράγραφος 2)

Ημερομηνία αίτησης της κ. Χαϊμάνη Άννας: 18-1-2011

Ημερομηνία ορισμού Τριμελούς Συμβουλευτικής Επιτροπής: 708ª/8-3-2011

Μέλη Τριμελούς Συμβουλευτικής Επιτροπής:

Επιβλέπουσα

Ντζάνη Ευαγγελία, Επίκουρη Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία

Μέλη

Σαλαντή Γεωργία, Λέκτορας Επιδημιολογίας

Christopher Henry Schmid, Καθηγητής Ιατρικής Σχολής του Πανεπιστημίου Tufts της Βοστώνης

Ανασύσταση Τριμελούς Συμβουλευτικής Επιτροπής: Αριθμ. Συνεδρ. 714ª/7-6-2011

Επιβλέπουσα

Σαλαντή Γεωργία, Επίκουρη Καθηγήτρια Επιδημιολογίας

Μέλη

Ντζάνη Ευαγγελία, Επίκουρη Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία

Christopher Henry Schmid, Καθηγητής Ιατρικής Σχολής του Πανεπιστημίου Tufts της Βοστώνης

Ημερομηνία ορισμού Θέματος: 4-5-2011

«Μελέτη των συστηματικών σφαλμάτων στη μετα-ανάλυση πολλαπλών παρεμβάσεων»

ΔΙΟΡΙΣΜΟΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ: 763ª/10-4-2014

1. Higgins Julian Piers Thomas, Professor of Evidence Synthesis, University of Bristol and University of York, H. B.
2. Li Tianjing, Assistant Professor in Epidemiology, John Hopkins Bloomberg School of Public Health
3. Christopher Henry Schmid, Professor of Biostatistics, Brown University Rhode Island (USA)
4. Ντζάνη Ευαγγελία, Επίκουρη Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία του τμήματος Ιατρικής του Παν/μιου Ιωαννίνων
5. Σαλαντή Γεωργία, Επίκουρη Καθηγήτρια Επιδημιολογίας του τμήματος Ιατρικής του Παν/μιου Ιωαννίνων
6. Τσιλίδης Κωνσταντίνος, Λέκτορας Υγιεινής με έμφαση στην Επιδημιολογία του τμήματος Ιατρικής του Παν/μιου Ιωαννίνων
7. Μαυρίδης Δημήτριος, Λέκτορας Μαθηματικών και Στατιστικής στην Εκπαίδευση

Έγκριση Διδακτορικής Διατριβής με βαθμό **«αριστα» στις 22-5-2014**

ΠΡΟΕΔΡΟΣ ΤΟΥ ΤΜΗΜΑΤΟΣ ΙΑΤΡΙΚΗΣ

Φωτόπουλος Ανδρέας

Καθηγητής Πυρηνικής Ιατρικής

Acknowledgements

The work presented in this Thesis was funded by a grant from the European Research Council (IMMA 260559 project), which supported my research at the Department of Hygiene and Epidemiology of the University of Ioannina during my PhD studies.

Foremost, I need to thank my supervisor Dr. Georgia Salanti for her best effort to convey to me her professional expertise and knowledge and for the continuous support and guidance she provided to me. I also appreciate a lot that she trusted me and involved me in important research projects and she gave me the opportunity to collaborate with top experts in my research field. I also thank my advisor Prof. Christopher Schmid as well as Prof. Julian Higgins for their extremely valuable input in my PhD research.

I wouldn't be able to finish this Thesis without the support of my parents and my brother, who were always giving priority to my problems, and I am more than grateful to them for this.

Many thanks to my collegue Dimitris for the important help and suggestions he provided to me during my PhD research as well as for making the working hours funnier. Thanks also to my collegue Adriani for her support at several difficulties.

Finally, I would like to thank all my friends in Athens for offering me nice breaks of entairtenment, which kept me eager for work.

Contents

1 Introduction

Systematic reviews and meta-analyses have been established as an integral part of comparative effectiveness research and are used worldwide by health-care agencies (e.g. National Institute for Health and Clinical Excellence (NICE) or Agency for Healthcare Research and Quality (AHRQ)) to inform policy-making. A major advantage of meta-analysis is that usually it outperforms any single study in terms of statistical power. Thus, it has become a highly valuable tool particularly when individual studies are inconclusive (1).

The usefulness of meta-analysis goes beyond the synthesis of relative effects for comparing two interventions for a specific clinical condition and outcome; it can be also employed for the investigation of characteristics that may affect the treatment effects (2,3). More specifically, substantial clinical and methodological discrepancies across the studies of a meta-analysis may cause differences between the study-specific *true underlying effects*. Such differences, known as *heterogeneity*, can be accounted for in a meta-analysis as additional variation across studies on the top of random error. This variation is reflected in the estimated *summary relative effect* as the between-study variance, which refers to the deviation of the study-specific underlying treatment effects from their common mean (4).

Heterogeneity can sometimes be explained using *meta-regression* models, which are extensions of the meta-analysis model that incorporate covariates (5,6). Nevertheless, associations between such study-specific characteristics and relative effects cannot always be explored adequately within a meta-analysis due to the small number of included studies. Therefore, *meta-epidemiological* research uses collections of many meta-analyses to assess the possible effect of several factors using a wide range of studies typically collected as part of individual meta-analyses (7,8).

Meta-regression and meta-epidemiology approaches in the context of *pairwise meta-analysis* (comparing only two interventions) are well-known methodologies. However, the development of a more complex evidence synthesis tool, which enables the simultaneous comparison of three or more competing interventions, requires the adaptation of these methods into its framework. This tool, known as *network meta-analysis* (NMA) (also called mixed-treatment comparison or multiple-treatment meta-analysis), synthesizes data from a set of studies, which may compare the same or different sets of treatments for a common clinical outcome. NMA integrates *direct* (from individual studies directly comparing interventions) with *indirect* evidence (information on a treatment effect via a connected indirect root) to infer about the relative effect of any pair of treatments included in the network (9–11). This allows us to compare treatments that have not been directly compared in any of the individual studies and also *network estimates* usually have increased precision compared to the respective direct estimates. Furthermore, the inclusion of all competing treatments in the same meta-analytic model allows the estimation of their relative ranking (12), which provides a concise summary of the findings and can be used to inform decision-making.

This introductory Chapter offers a brief overview of the available meta-analytic methods for pairwise meta-analysis. It also presents meta-regression models, which allow the relative effects to differ according to the values of a covariate. Next, it describes meta-epidemiological models that synthesize parameters across different meta-analyses. The last Section outlines the objectives of this Thesis.

1.1 Models for pairwise meta-analysis

1.1.1 Basic concepts of pairwise meta-analysis

Consider that a systematic review includes S eligible studies and each study i ($i = 1, \dots, S$) has estimated a relative effect y_i, called also the *observed effect* of study

i, between two competing interventions X and Y with (observed) variance s_i^2. The true underlying effects θ_i of the studies can be modeled either under the *fixed effect* or the *random effects* assumption (4,13,14). The former assumption implies that the study-specific underlying effects θ_i are identical, hence

$$\theta_i = \mu$$

where μ is the true summary relative effect of treatment X vs. Y. On the other hand, the random effects assumption allows the underlying effects to differ across studies so that

$$\theta_i = \mu + \delta_i \tag{1.1}$$

The random effects δ_i are assumed to share a common (usually normal) distribution with zero mean and variance the heterogeneity parameter τ^2, namely

$$\delta_i \sim N(0, \tau^2) \tag{1.2}$$

and reflect the deviation of each θ_i from their common mean μ. Equations (1.1) and (1.2) can be written equivalently as

$$\theta_i \sim N(\mu, \tau^2) \tag{1.3}$$

Setting $\tau^2 = 0$ or $\delta_i = 0$ renders the random effects model as a fixed effect model.

Pairwise meta-analysis can be fitted either as a linear or a hierarchical model. Both approaches as described in the following Section.

Note that throughout this Thesis bold letters represent matrices and regular letters represent scalars. Also, there is distinction between lowercase and uppercase letters.

1.1.2 Modeling approach for pairwise meta-analysis

Pairwise meta-analysis as a linear model

In terms of linear regression and using matrix notation the meta-analysis model can be described as

$$\boldsymbol{y} = \boldsymbol{X\mu} + \boldsymbol{\varepsilon} + \boldsymbol{\delta} \tag{1.4}$$

where, $\boldsymbol{y} = (y_1, \dots, y_S)^T$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_S)^T$**.** The vector of the random error terms $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_S)^T$ is assumed normally distributed with

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{s}^2)$$

($\boldsymbol{s}^2 = diag(s_i^2, \ldots, s_S^2)$). The matrix $\boldsymbol{X} = (1, \ldots, 1)$ is the *design matrix* and $\boldsymbol{\mu} \equiv \mu$ is the summary effect. The model of Equation (1.4) assumes that variances are known and equal to their sample counterparts. Also, the random effects δ_i are assumed independent from the random errors ε_i.

Pairwise meta-analysis as a hierarchical model

A hierarchical model for pairwise meta-analysis can be constructed assuming that

$$y_i \sim N(\theta_i, s_i^2) \tag{1.5}$$

and then employing Equation (1.3). Note that hierarchical models for meta-analysis are often fitted in a Bayesian framework; in this case prior distributions are necessary for μ and τ^2.

Alternatively, Equation (1.5) can be replaced by arm-level likelihoods. These likelihoods are presented below for the two most common types of outcome data:

- For a dichotomous outcome the number of events in each study arm r_{iX} and r_{iY} (for treatments X and Y respectively) of study i is assumed to follow a binomial distribution

$$r_{iX} \sim Bin(\pi_{iX}, n_{iX})$$
$$r_{iY} \sim Bin(\pi_{iY}, n_{iY})$$

  with probability of experiencing an event π_{iX} and π_{iY} respectively. The n_{iX} and n_{iY} are the total number of participants in each arm.

- Continuous outcomes are usually modelled by assuming a normal distribution for each arm-specific mean score m_{iX}, m_{iY}

$$m_{iX} \sim N(\varphi_{iX}, \sigma_{iX}^2)$$
$$m_{iY} \sim N(\varphi_{iY}, \sigma_{iY}^2)$$

  with means φ_{iX}, φ_{iY} and variances $\sigma_{iX}^2 = \frac{sd_{iX}^2}{\sqrt{n_{iX}}}$, $\sigma_{iY}^2 = \frac{sd_{iY}^2}{\sqrt{n_{iY}}}$ (sd_{ik} is the standard deviation of the observations in arm k of study i).

Then, the true probabilities π_{iX}, π_{iY} and the true means φ_{iX}, φ_{iY} are parameterized using a link function to derive the underlying relative effect sizes θ_i. For example,

the log-odds ratio or the mean difference of X vs. Y can be estimated using the $logit(\,)$ or identity function respectively

$$\theta_i = logit(\pi_{iX}) - logit(\pi_{iY})$$
$$\theta_i = \varphi_{iX} - \varphi_{iY} \tag{1.6}$$

An overview of all possible likelihoods and link functions for any type of data can be found in Dias et al. (15).

1.2 Models incorporating covariates for pairwise meta-analysis

In the presence of important heterogeneity, conclusions formed on the basis of the summary effect might be misleading. However, heterogeneity often can be explained by differences in several characteristics across the studies of a systematic review. If information on such study-specific characteristics is available, it can be incorporated in the analysis via a meta-regression model as described below (6).

Consider p different characteristics $z_{i1}, \dots, z_{ip}$ that differ in every study i and are possible explanatory variables for heterogeneity. These variables can be binary, categorical or continuous and are added as extra linear terms in Equation (1.1). The meta-regression model including p covariates would be

$$\theta_i^* = \mu^* + \delta_i^* + \sum_{j=1}^{p} \beta_j z_{ij} \tag{1.7}$$

where the superscript $(*)$ denotes that the parameters from this model are different from the parameters of the model without covariates.

Each coefficient β_j $(j = 1, \dots, p)$ shows the change, on average, in the summary effect μ^*, if z_j increases one unit and $z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_p$ remain constant. The summary effect μ^* estimated from this model corresponds to studies for which all covariates are equal to zero. We assume

$$\delta_i^* \sim N(0, \tau^{*2})$$

Then, Equations (1.4) and (1.5), depending on the fitting approach (linear or hierarchical model), are modified into

$$\boldsymbol{y} = \boldsymbol{X}^* \begin{pmatrix} \boldsymbol{\mu}^* \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\varepsilon}^* + \boldsymbol{\delta}^* \quad (1.8)$$

(with $\boldsymbol{X}^* = \begin{pmatrix} 1 & z_{11} & \cdots & z_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{S1} & \cdots & z_{Sp} \end{pmatrix}$, $\boldsymbol{\mu}^* \equiv \mu^*$, $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_p\right)^T$) and

$$y_i \sim N(\theta_i^*, s_i^2)$$

(with $\theta_i^* \sim N(\mu^*, \tau^{*2})$) respectively.

If the selected explanatory variables are indeed factors that cause heterogeneity, the estimated between-study variance $\hat{\tau}^{*2}$ is expected to be smaller than that estimated from the model without covariates (i.e. $\hat{\tau}^{*2} < \hat{\tau}^2$). Note that the use of fixed effect meta-regression is not recommended, since it is not reasonable to assume that the total amount of statistical heterogeneity is explained (16).

Meta-regression models aim to identify factors acting as effect modifiers; that is study-level characteristics that affect the results of the individual studies. However, meta-regression can be subject to false-positive findings, particularly when a large number of explanatory variables has been selected. For this reason the potential effect modifiers should be specified *a priori* based on the clinical experience and understanding of the researchers.

Despite its usefulness, meta-regression is not free of limitations. A major drawback is that, in the absence of many studies or in the presence of extreme heterogeneity, the regression coefficients cannot be estimated adequately due to low power (17). Existing guidance suggests that performing meta-regression requires at least ten studies for each included covariate to ensure that coefficients could be estimated with sufficient power (13). In addition, meta-regression is observational evidence, since the benefits of randomization are not preserved when associations are investigated across the studies. Thus, any meta-regression analysis suffers from the potential biases of other observational studies, such as confounding. An extended discussion on all stengths and limitations of meta-rergession is available in Thompson et al. (16).

1.3 Models synthesizing parameters across pairwise meta-analyses

The low power of meta-regression implies that, when only few studies are available, it may fail to detect associations between study-level characteristics and treatment effects (13). However, the impact of such characteristics on relative effects is very often not specific to a single meta-analysis, but affects any meta-analysis irrespective of the interventions or the outcomes of interest across many research areas. Meta-epidemiology exploits this fact to overcome the lack of power of meta-regression and explores the impact of possible effect modifiers using collections of meta-analyses (7,8).

For example, the appropriate blinding of patients is supposed an important trial characteristic that strengthens the validity of findings. This implies that studies, in which blinding has not been conducted adequately might give different results from studies, in which patients have been blinded properly. If there are no reasons to believe that blinding impacts in a different way (in magnitude and direction) across a range of clinical outcomes (e.g. when active treatments are compared to a control), information on blinding can be synthesized across all meta-analyses evaluating such outcomes. This approach would give a summary estimate of the impact of blinding on treatment effects pertaining to the entire collection of meta-analyses. This estimate would be more precise compared to any coefficient estimated from a single meta-regression.

Quantitative synthesis of results assumes that the parameters (e.g. coefficients from meta-regression models) showing the effect of the characteristic of interest are comparable across meta-analyses and aims to estimate a common summary parameter.

Consider again the $z_1, \dots, z_p$ explanatory variables that may act as effect modifiers in a collection of M meta-analyses. Either a one-stage or a two-stage approach can be employed. The latter approach initially applies a meta-regression model (see Section 1.2) within each meta-analysis m ($m = 1, \dots, M$) separately to estimate the

set of p meta-analysis-specific coefficients $\beta_{m,1}, \dots, \beta_{m,p}$. Then, the estimated coefficients $\hat{\beta}_{m,j}$ $(j = 1, \dots, p)$ are synthesized across the M meta-analyses using standard meta-analytic methods (see Section 1.1). A random effects model can be employed to allow for between-meta-analysis variability in the estimated coefficients. On the other hand, the one-stage approach estimates simultaneously in the same model both the meta-analysis-specific coefficients and their summary coefficient (18). The advantage of using this approach is that meta-analyses with few studies can borrow strength from larger meta-analyses and in this way the summary parameters might be estimated with increased precision.

The one-stage meta-epidemiological model can be constructed by extending Equation (1.1) into

$$\theta'_{m,i} = \mu'_m + \delta'_{m,i} + \sum_{j=1}^{p} \left[\left(B_j^{overall} + \beta'_{m,j} \right) z_{m,ij} \right] \tag{1.9}$$

with

$$\beta'_{m,j} \sim N(0, \omega_j^2)$$

The parameters B_j are the mean summary coefficient for all M meta-analyses that shows how the underlying treatment effects change for one unit increase in each covariate z_j. The superscript ($'$) shows that the parameters in the above equation are different from the respective parameters in Equations (1.1) or (1.7). The $\beta'_{m,j}$ are assumed random effects to allow for variability in the regression coefficients between the different meta-analyses. This variability is reflected by the parameter ω_j^2 that represents the between-meta-analysis variance of their common mean for each characteristic j. A fixed effect model for the regression coefficients can be derived by setting every $\beta'_{m,j} = 0$. A less restrictive meta-epidemiological model has been suggested by Welton et al. (18), which allows for between-study variance in the regression coefficietns as well as for uncertainty in the estimation of the overall coefficients across meta-anlayses. Following Welton et al., Equation (1.9) would be

$$\theta'_{m,i} = \mu'_m + \delta'_{m,i} + \sum_{j=1}^{p}\left[\left(B_j^{overall} + \beta'_{m,j} + \gamma_{m,ij}\right)z_{m,ij}\right]$$

with

$$\gamma_{m,ij} \sim N(0, u_j^2) \quad , \quad \beta'_{m,j} \sim N(0, \omega_j^2)$$

and

$$B_j^{overall} \sim N(B_j, V_j)$$

An important constraint of the meta-epidemiological models in the selection of meta-analyses is that they need to be independent (i.e. to include different sets of individual studies) to avoid correlation between the meta-analysis-specific coefficients.

Note that usually meta-epidemiological studies do not investigate simultaneously (in the same model) many characteristics, but only one or two (i.e. $p = 1,2$). Synthesizing jointly many covariates across a collection of meta-analyses would mitigate the benefits of meta-epidemiology, since the increase in power compared to a single meta-regression model might be only minor.

1.4 Objectives & outline of the Thesis

The increasing number of competing interventions in many medical fields and the advantages of analyzing simultaneously all available evidence (as outlined at the beginning of this Chapter) have made NMA a popular statistical tool (19–22). However, the lack of a user-friendly implementation framework has rendered it to be a privilege of researchers with strong statistical and computational skills. Consequently, less advanced software options and easily-understood tools for the evaluation of assumptions and presentation of results are necessary to make NMA accessible to non-statisticians. In addition, the properties of the various different approaches are still under investigation and the appropriate conduct of NMA is very often doubtful. Thus, empirical studies using networks of interventions are

needed to evaluate the validity of conclusions drawn from NMA in the literature as well as to explore different characteristics that may affect NMA results. However, accounting for the potential of biased effect sizes in the context of NMA might be challenging. This is because the inclusion of covariates in NMA models requires assumptions about the direction of bias for the different treatment comparisons. The aim of this Thesis is to address the above described issues and contribute to improved conduct and reporting of future NMA.

The rest of the Thesis is structured as follows. Chapter 2 initially provides an overview of NMA methodology and the existing software options for fitting the various NMA models. Then, it introduces a series of new or modified graphical presentation tools, which I developed and implemented in the STATA software (23), for addressing the different steps of the analysis. Chapter 3 describes a database of published networks of interventions that I compiled for the conduct of two empirical analyses. Chapter 4 describes how meta-regression and meta-epidemiology methods can be extended in the framework of NMA and presents different possible assumptions that can be employed for the regression coefficients. I refer to these methods as *network meta-regression* and *network meta-epidemiology* respectively. Finally, Chapter 5 presents applications of network meta-epidemiology and network meta-regression, through which I investigated the impact of five study characteristics on NMA results. It also describes a case study that I conducted in order to exemplify how different assumptions about the direction of bias can be modeled.

2 Methodology & software for network meta-analysis

2.1 Introduction

NMA can be seen from several perspectives which are, in principal, equivalent, but differ with respect to the ease of implementation in standard software packages. All approaches are based on the idea of indirect and *mixed* comparisons that were first introduced as a valid statistical tool by Bucher et al. (9) for the case of only three competing interventions X, Y, Z. The method implies that the summary relative effect YZ (i.e. Y vs. Z) can be estimated indirectly (I) by subtracting the direct (D) relative effects XZ and XY

$$\hat{\mu}_{YZ}^{I} = \hat{\mu}_{XZ}^{D} - \hat{\mu}_{XY}^{D} \tag{2.1}$$

and its variance is the sum of the respective variances

$$\hat{v}_{YZ}^{I} = \hat{v}_{XZ}^{D} + \hat{v}_{XY}^{D} \tag{2.2}$$

Then, a mixed estimate (M) YZ can be derived as the weighted average of the direct and indirect treatment effects

$$\hat{\mu}_{YZ}^{M} = \frac{\frac{1}{\hat{v}_{YZ}^{D}}\hat{\mu}_{YZ}^{D} + \frac{1}{\hat{v}_{YZ}^{I}}\hat{\mu}_{YZ}^{I}}{\frac{1}{\hat{v}_{YZ}^{D}} + \frac{1}{\hat{v}_{YZ}^{I}}} \tag{2.3}$$

with variance

$$\hat{v}_{YZ}^{M} = \frac{1}{\frac{1}{\hat{v}_{YZ}^{D}} + \frac{1}{\hat{v}_{YZ}^{I}}} \tag{2.4}$$

In all above Equations the superscripts D, I and M denote the type of each relative effect estimate (i.e. direct, indirect and mixed respectively).

Extending this idea to a larger network of interventions (with more than three competing treatments) allows the incorporation of several indirect estimates from different roots into the *network estimates*.

Indirect and mixed comparisons rely on the fundamental assumption of *transitivity*, which suggests that one can learn about the relative effect of Y vs. Z via X. This implies that the common comparator treatment X is similar (e.g. administered the same way and in similar populations) in the XY and XZ studies (11,24). Alternative equivalent interpretations of transitivity are that 'the missing arms are missing at random' from all included studies or that all competing treatments in the network should be 'jointly randomizable'. Missing at random means that the choice of treatment arms in studies should be independent of their effectiveness (25). The notion of jointly randomizable treatments implies that one could imagine a clinically meaningful randomized controlled trial (RCT) comparing simultaneously all treatments (11). Transitivity can be assessed statistically by comparing the distribution of potential effect modifiers across the available direct comparisons in the network, when such information is available (26). If the distributions are not similar between two or more pairwise comparisons, the transitivity assumption would be probably violated leading to invalid indirect and mixed estimates.

At the level of model parameters, transitivity is reflected by the *consistency equations*

$$\mu_{YZ} = \mu_{XZ} - \mu_{XY} \tag{2.5}$$

which imply that direct and indirect evidence in the network are in agreement regarding the true (summary) underlying relative effects. If the *consistency assumption* does not hold in parts or in the entire network, the results of NMA might be questionable (24,27). However, alternative NMA models have been developed that relax the consistency equations and allow for extra variability (on the top of heterogeneity) in the network due to potential *inconsistency* (25,28,29). These models are usually called *inconsistency models*.

The following Section (2.2) presents currently available methods for fitting a NMA, while Section 2.3 describes different approaches for evaluating heterogeneity and inconsistency in the network. Section 2.4 introduces a series of new or modified graphical tools that I developed to aid the presentation and

interpretation of data and findings from NMA. Then, Section 2.5 provides an overview of the available software options for NMA and describes the routines I developed for the implementation of the new graphical tools in STATA (23).

2.2 Models for network meta-analysis

2.2.1 Basic concepts of network meta-analysis

Consider a network that involves T competing interventions and S studies providing information for C^D direct comparisons between pairs XY ($X, Y = \{1, \dots, T\}$ with $X \neq Y$) of treatments. NMA models are extensions of the pairwise meta-analysis models (see Section 1.1) that can estimate more than one single summary effect μ_{XY} assuming consistency. Equation (2.5) suggests that only $T-1$ *basic comparisons* or *basic parameters* are sufficient for the estimation of the relative effects for all $C = \binom{T}{2}$ possible comparisons in the network. All other *functional comparisons* can be estimated as a linear function of the elements of the vector containing the basic parameters, which is denoted with $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{T-1})^T$. In general, the choice of the basic comparisons is arbitrary and usually does not affect the results. The only restriction is that every treatment $X \in \{1, \dots, T\}$ should be expressed via at least one basic parameter (31). However, it has been shown that computationally some parameterizations may not work (30).

The inclusion of studies with more than two treatment arms must account for the correlation between both the observed and the underlying relative effects from such studies. Within a k-arm study i (with $k = 1, \dots, K_i$, $K_i \geq 2$ the number of arms in study i), consistency holds implicitly and consequently only a subset of $K_i - 1$ relative effects need to be included in the NMA model from this study. Thus, the variance-covariance matrix of the vector containing the observed relative effects $\boldsymbol{y_i} = \left(y_{i1}, \dots, y_{i(K_i-1)}\right)^T$ would be

$$\boldsymbol{s_i^2} = \begin{pmatrix} s_{i1}^2 & \cdots & cov(y_{i1}, y_{i(K_i-1)}) \\ \vdots & \ddots & \vdots \\ cov(y_{i1}, y_{i(K_i-1)}) & \cdots & s_{i(K_i-1)}^2 \end{pmatrix}$$

and the respective matrix for the underlying effects $\boldsymbol{\theta_i} = (\theta_{i1}, \dots, \theta_{i(K_i-1)})^T$

$$\boldsymbol{\tau_i^2} = \begin{pmatrix} \tau_1^2 & \cdots & cov(\theta_{i1}, \theta_{i(K_i-1)}) \\ \vdots & \ddots & \vdots \\ cov(\theta_{i1}, \theta_{i(K_i-1)}) & \cdots & \tau_{K_i-1}^2 \end{pmatrix} \quad (2.6)$$

The off-diagonal elements in $\boldsymbol{s_i^2}$ and $\boldsymbol{\tau_i^2}$ contain the covariance terms between observed and true relative effects respectively within study i and the diagonal elements the variances of the respective relative effects in $\boldsymbol{y_i}$ and $\boldsymbol{\theta_i}$. The variances $\tau_1^2, \dots, \tau_{K_i-1}^2$ are the comparison-specific heterogeneity parameters representing the between-study variance for each of the $(K_i - 1) \in C$ comparison. Similarly to pairwise meta-analysis, the random effects $\boldsymbol{\delta_i} = (\delta_{i1}, \dots, \delta_{i(K_i-1)})^T$ are assumed to be normally distributed with $\boldsymbol{\delta_i} \sim N(\mathbf{0}, \boldsymbol{\tau}^2)$ and $\boldsymbol{\tau}^2 = diag(\boldsymbol{\tau_i^2})$. Note that setting $\boldsymbol{\tau}^2 = \mathbf{0}$ or $\boldsymbol{\delta_i} = \mathbf{0}$ gives the fixed effect NMA model.

The several modeling strategies for NMA presented below differ mainly in the way multi-arm studies are treated and in the implementation of the consistency assumption.

2.2.2 Modeling approach for network meta-analysis

Network meta-analysis as a 'multivariate meta-regression' model

The first approach suggests that NMA can be considered as a standard meta-regression model (see Section 1.2) that treats the different treatment comparisons in the network as covariates (32). Specifically, to allow for multiple treatment comparisons and multiple study arms the meta-regression model of Equation (1.4) would be

$$\boldsymbol{y} = \boldsymbol{X\mu} + \boldsymbol{\varepsilon} + \boldsymbol{\delta}$$

with $\boldsymbol{y} = (\boldsymbol{y_1}, \dots, \boldsymbol{y_S})$, $\boldsymbol{\delta} = (\boldsymbol{\delta_1}, \dots, \boldsymbol{\delta_S})$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon_1}, \dots, \boldsymbol{\varepsilon_S})$. Each vector $\boldsymbol{\varepsilon_i} = \left(\varepsilon_{i1}, \dots, \varepsilon_{i(K_i-1)}\right)^T$) contains the random error terms for study i ($i = 1, \dots, S$) and is assumed to follow a multivariate normal distribution $\boldsymbol{\varepsilon_i} \sim N(\boldsymbol{0}, \boldsymbol{s_i^2})$.

Note that $\boldsymbol{y_i}$, $\boldsymbol{\delta_i}$ and $\boldsymbol{\varepsilon_i}$ are vectors reduced to scalars for each two-arm study i.

Each $(K_i - 1) \times (T - 1)$ matrix $\boldsymbol{X_i}$ is the contribution of the study i to the design matrix $\boldsymbol{X} = (\boldsymbol{X_1}, \dots, \boldsymbol{X_S})^T$ (of dimension $[\sum_{i=1}^{S}(K_i - 1)] \times (T - 1)$). This matrix contains values $x_{i(K_i-1)b} = -1,0,1$ (with $b = 1, \dots, T - 1$) and expresses the linear relationships between the $K_i - 1$ study-specific comparisons and the $T - 1$ basic comparisons based on the consistency equations (Equation (2.5)). This meta-regression model has been called also *multivariate meta-regression* model (33), because the inclusion of multi-arm studies requires assuming multivariate distributions for the parameters representing random errors and random effects.

Graph-theoretical methods have been recently applied in the context of NMA (34). This approach is based on the analogy between electrical and treatment networks to perform a fixed effect NMA and is totally equivalent to the above meta-regression model for NMA. An extension to random effects has been developed by adjusting the variances within multi-arm studies as if the elements of $\boldsymbol{y_i}$ were independent (35).

Network meta-analysis as a 'two-stage meta-regression' model

The above meta-regression model can be equivalently fitted using a two-stage approach (29,36,37). At the first stage of this approach a pairwise meta-analysis model (see Section 1.1) is used to derive the available direct relative effects for each pairwise comparison.

Multi-arm studies, if present, should be pooled separately. More specifically, consider that each k-arm study has a specific *study design* g; that is a specific set of treatments being compared. Suppose that a network is consisted of G different study designs and each design g includes S_g ($g = 1, \dots, G$) studies with K_g treatment arms. Then, the vector of the direct (D) summary relative effects

$\boldsymbol{\mu_g^D} = \left(\mu_{1g}^D, \dots, \mu_{(K_g-1)g}^D\right)^T$ derived from all S_g studies with design g can be estimated by replacing Equation (1.1) with

$$\boldsymbol{\theta_{i_g}^D} = \boldsymbol{X_{ig}\mu_g^D} + \boldsymbol{\delta_{i_g}^D} \tag{2.7}$$

where $\boldsymbol{\delta_{i_g}^D} \sim N\left(\mathbf{0}, \left[\boldsymbol{\tau_g^D}\right]^2\right)$, $\left[\boldsymbol{\tau_g^D}\right]^2 = \begin{pmatrix} \left[\tau_{1g}^D\right]^2 & \cdots & cov\left(\theta_{i_g1}, \theta_{i_g(K_g-1)}\right) \\ \vdots & \ddots & \vdots \\ cov\left(\theta_{i_g1}, \theta_{i_g(K_g-1)}\right) & \cdots & \left[\tau_{(K_g-1)g}^D\right]^2 \end{pmatrix}$.

Note that $\boldsymbol{X_{ig}} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}$ is a $\left(K_g - 1\right) \times \left(K_g - 1\right)$ design matrix. Then, Equation (1.4) becomes

$$\boldsymbol{y_g} = \boldsymbol{X_g\mu_g^D} + \boldsymbol{\varepsilon_g^D} + \boldsymbol{\delta_g^D}$$

with $\boldsymbol{y_g} = \left(\boldsymbol{y_{1_g}}, \dots, \boldsymbol{y_{S_g}}\right)^T$, $\boldsymbol{X_g} = \left(\boldsymbol{X_{1g}}, \dots, \boldsymbol{X_{S_gg}}\right)$, $\boldsymbol{\delta_g^D} = \left(\boldsymbol{\delta_{1_g}^D}, \dots, \boldsymbol{\delta_{S_g}^D}\right)$, $\boldsymbol{\varepsilon_g^D} = \left(\boldsymbol{\varepsilon_{1_g}^D}, \dots, \boldsymbol{\varepsilon_{S_g}^D}\right)$ and $\boldsymbol{\varepsilon_{i_g}^D} \sim \boldsymbol{N(0, s_{i_g}^2)}$. Note that the direct relative effects $\mu_{1g}^D, \dots, \mu_{(K_g-1)g}^D$ are correlated within each design g with $K_g > 2$. The design-specific direct estimates $\boldsymbol{\hat{\mu}^D} = \left(\boldsymbol{\hat{\mu}_1^D}, \dots, \boldsymbol{\hat{\mu}_G^D}\right)^T$ with estimated variance-covariance matrix $\boldsymbol{\hat{v}^D} = diag(\boldsymbol{\hat{v}_1^D}, \dots, \boldsymbol{\hat{v}_G^D})$ are used as input data at the second stage of the analysis, so that

$$\boldsymbol{\hat{\mu}^D} = \boldsymbol{X^D\mu} + \boldsymbol{\varepsilon^N} \tag{2.8}$$

with $\boldsymbol{\varepsilon^N} = \left(\boldsymbol{\varepsilon_1^N}, \dots, \boldsymbol{\varepsilon_G^N}\right)$ and $\boldsymbol{\varepsilon_g^N} \sim N(\mathbf{0}, \boldsymbol{\hat{v}_g^D})$ being the random error terms of the direct estimates for design g compared to the true network (N) relative effects $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{T-1})^T$. The design matrix $\boldsymbol{X^D} = \left(\boldsymbol{X_1^D}, \dots, \boldsymbol{X_G^D}\right)$ in this model is constructed similarly to $\boldsymbol{X}$ (defined in the previous approach); it expresses the linear relationship between the direct comparisons of design g and the basic comparisons according to the consistency equations (Equation (2.5)).

In summary, this approach first uses the observed relative effects of the studies $\boldsymbol{y_{1_g}}, \dots, \boldsymbol{y_{S_g}}$ to estimate the direct treatment effects separately for each design g. Then, at the second stage it pools the direct estimates $\boldsymbol{\hat{\mu}_1^D}, \dots, \boldsymbol{\hat{\mu}_G^D}$ across all study designs to obtain the network estimates. Additional details and examples for this approach can be found in Lu et al. (36) and Krahn et al. (29).

Network meta-analysis as a hierarchical model

Moving into a hierarchical model a multivariate normal likelihood would be assumed on the observed relative effects for each study i

$$\boldsymbol{y_i} \sim N(\boldsymbol{\theta_i}, \boldsymbol{s_i^2}) \tag{2.9}$$

The underlying effects are expressed through

$$\boldsymbol{\theta_i} \sim N(\boldsymbol{\mu_i}, \boldsymbol{\tau_i^2}) \tag{2.10}$$

where $\boldsymbol{\mu_i} = \left(\mu_1, \dots, \mu_{K_i - 1}\right)^T$ is the vector including the network (true) relative effects for all XY comparisons modelled for study i (38,39).

For example, suppose a three-arm trial comparing treatments A, B and C. If the AB and AC comparisons are selected to be modelled, Equation (2.10) would be

$$\begin{pmatrix} \theta_{AB} \\ \theta_{Ac} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{AB} \\ \mu_{Ac} \end{pmatrix}, \begin{pmatrix} \tau_{AB}^2 & cov(\theta_{AB}, \theta_{AC}) \\ cov(\theta_{AB}, \theta_{AC}) & \tau_{AC}^2 \end{pmatrix} \right)$$

Similarly to pairwise meta-analysis (see Section 1.1) arm-specific likelihoods can be employed for every study arm $1, \dots, K_i$ instead of the normal likelihood for the observed effects assumed in Equation (2.9) (15).

Network meta-analysis as a multivariate meta-analysis model

An alternative approach treats the $T - 1$ basic comparisons as different outcomes reported in studies and fits NMA using the methodology of pairwise meta-analysis for multiple outcomes (33) (see Jackson et al. (40) and Mavridis et al. (41) for a review on multivariate meta-analysis). This method requires a reference intervention X common to all studies and takes as basic parameters-outcomes all XY comparisons with $Y = 1, \dots, T$ $(Y \neq X)$ (33). An X arm with minimal information needs to be imputed for studies not evaluating the reference intervention X. In this way consistency is imposed in the model by assuming that the reference arm is missing at random, when it is not evaluated in a study. Then, the model is described by Equation (1.4) as in the multivariate meta-regression approach.

The difference between this and the multivariate meta-regression approach is that here the $K_i - 1$ study-specific comparisons modelled for each study i are

necessarily a subset of the $T-1$ basic parameters; elements of $\boldsymbol{y_i}$ corresponding to comparisons not reported in study i are filled with missing values. Also, the elements of each matrix $\boldsymbol{X_i}$ in this model do not express the consistency equations but are equal to 1 for comparisons evaluated in study i and missing values otherwise.

2.3 Approaches for evaluating heterogeneity & inconsistency in network meta-analysis

The distinction between heterogeneity and inconsistency in NMA is not always straightforward; both terms refer to additional variation in the treatment effects that cannot be explained by chance. Inconsistency can be considered as a special form of heterogeneity. More specifically, heterogeneity refers to the between-study disagreement within each comparison in the network and inconsistency the between-source disagreement (direct and indirect evidence) across the different comparisons (11,42).

There are two general types of inconsistency; the *loop inconsistency* and the *design inconsistency*. Loop inconsistency considers the possible differences between direct and indirect (or between the several indirect) estimates. The notion of study design (see Section 2.2.2) implies that inconsistency might be also present because of differences between the estimates from studies with different designs; hence the design inconsistency.

The following Sections briefly describe the most commonly used assumptions and approaches that can be employed to assess the presence and the level of heterogeneity and inconsistency in a network of interventions.

It is important to note that the statistical tests for heterogeneity and inconsistency that follow often have limited power to detect them as statistically significant even when they are present. Therefore, conclusions regarding the presence of heterogeneity and/or inconsistency should not be based solely on the statistical

significance of the related parameters, but also considering their magnitude and the respective confidence interval (CI).

2.3.1 Assumptions and measures for heterogeneity

Assumptions for heterogeneity

Between-study variance is usually assumed common across all comparisons in a network, namely $\tau_{XY}^2 = \tau^2$ for every $X, Y \in \{1, \dots, T\}$ $(X \neq Y)$. In this case the covariance of the underlying effects within a study is $\frac{\tau^2}{2}$ due to the consistency equations (43) and Equation (2.6) becomes

$$\boldsymbol{\tau}_i^2 = \begin{pmatrix} \tau^2 & \cdots & \frac{\tau^2}{2} \\ \vdots & \ddots & \vdots \\ \frac{\tau^2}{2} & \cdots & \tau^2 \end{pmatrix}$$

Models that allow for different heterogeneity parameters across the different comparisons have been developed as well (33,44). This alternative assumption suggests that

$$\left|\tau_{XY}^2 - \tau_{XZ}^2\right| \leq \tau_{YZ}^2 \leq \left|\tau_{XY}^2 + \tau_{XZ}^2\right|$$

which results again from the consistency equations (Equation (2.5)).

Assessing heterogeneity using predictive intervals

The impact of heterogeneity (either common (τ^2) or comparison-specific (τ_{XY}^2)) on the treatment effects can be evaluated by estimating the *predictive intervals* (PI) of the estimated summary effects (45). A PI is the interval within which the estimate of a future study is expected to lie (46,47) and for the relative effect X vs. Y can be derived by the formula

$$\hat{\mu}_{XY} \pm t_{df}^{\alpha} \sqrt{\hat{\tau}_{XY}^2 + \hat{v}_{XY}}$$

where t_{df}^{α} is the $100\left(1-\frac{\alpha}{2}\right)\%$ percentile of the t-distribution. A common arbitrary choice for the degrees of freedom for a single pairwise meta-analysis is $df = S - 2$ (with S the total number of studies). For the case of network meta-analysis, I suggest modifying this into $df = S - C^D - 1$ (with C^D the total number of available direct comparisons in the network) (45,48). In terms of a hierarchical model the PI for the relative effect μ_{XY} can be derived assuming that

$$\mu_{XY}^{new} \sim N(\mu_{XY}, \tau_{XY}^2) \tag{2.11}$$

where μ_{XY}^{new} is the relative effect of X vs. Y expected in a future study.

Pairwise relative effects for which the PI shows highly increased uncertainty compared to the respective confidence/credible interval (CI/CrI) might be affected substantially by the heterogeneity. For these comparisons conclusions should be drawn with greater caution.

Assessing heterogeneity using empirical distributions

The interpretation of the magnitude of the heterogeneity parameter estimate $\hat{\tau}^2$ (or $\hat{\tau}_{XY}^2$) requires consideration of the setting of the studies, such as the outcome and the type of treatment comparison they evaluate. For dichotomous outcomes empirical distributions of the heterogeneity have been developed for different types of meta-analyses (49). Comparing the estimated between-study variance with the corresponding empirical distribution can give insight on whether $\hat{\tau}^2$ reflects unimportant or substantial heterogeneity.

Assessing heterogeneity using Q-statistics

In pairwise meta-analysis the Q-test is one of the most common tools for the assessment of heterogeneity. A Q statistic for heterogeneity (H) in NMA could be derived as the sum of all comparison-specific Q within each study design g (29,37), hence

$$Q_H = \sum_{g=1}^{G} Q_g$$

where

$$Q_g = \sum_{i=1}^{S_g} \left(\boldsymbol{y}_{i_g} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{g}}^{\boldsymbol{D}}\right)^T \left(\widehat{\boldsymbol{v}}_{\boldsymbol{g}}^{\boldsymbol{D}}\right)^{-1} \left(\boldsymbol{y}_{i_g} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{g}}^{\boldsymbol{D}}\right)$$

with $\boldsymbol{y}_{i_g}$ the vector of observed effects for each study in design g and $\widehat{\boldsymbol{v}}_{\boldsymbol{g}}^{\boldsymbol{D}}$ the variance-covariance matrix of the direct summary effects $\widehat{\boldsymbol{\mu}}_{\boldsymbol{g}}^{\boldsymbol{D}}$ (see Section 2.2.2). The statistic Q_H is assumed to follow a χ^2 distribution with $\sum_{g=1}^{G}\left[\left(\sum_{i=1}^{S_g}(K_i - 1)\right) - \left(K_g - 1\right)\right]$ degrees of freedom (where S_g, K_g are the number of studies and treatmetns respectively in design g).

Assessing heterogeneity using the I^2 measure

The I^2 measure can be employed to estimate the proportion of variability in a meta-analysis that cannot be explained by random error (50). The extension of I^2 into the context of multivariate meta-analysis (51) enabled the adaption of this multivariate definition into NMA (52). This definition is based on the multivariate version of R^2 for heterogeneity (51) assuming the $T - 1$ basic comparisons being different outcomes. The R^2 in this case reflects the inflation in the volumes of the normal approximations to the confidence/credible regions for all relative effect parameters between a fixed effect and a random effects model.

2.3.2 Approaches for evaluating inconsistency locally

Local approaches for inconsistency aim to assess whether different pieces of evidence are in agreement for a specific comparison in the network informed either from a single loop or from the entire network. They consider as a measure of inconsistency for any XY comparison the absolute difference between two estimated effects from different sources (i.e. between direct and indirect or between indirect from different roots), commonly called an *inconsistency factor* (IF). Then, they infer about inconsistency based on the magnitude, the statistical

significance and the CI of the estimated inconsistency factors considering also the clinical setting and outcome.

Loop-specific approach for inconsistency

The *loop-specific approach* evaluates inconsistency separately in every closed loop of evidence (triangular, quadratic or higher order loops) (27). In a network with L total number of loops (excluding loops that include nested loops of lower order) the inconsistency factor within each loop l ($l = 1, \dots, L$) is estimated as

$$\widehat{w}_l^{XY} = \left|\hat{\mu}_{XY}^{D} - \hat{\mu}_{XY}^{I_l}\right| \tag{2.12}$$

with variance

$$v(\widehat{w}_l^{XY}) = \hat{v}_{XY}^{D} + \hat{v}_{XY}^{I_l} \tag{2.13}$$

for any XY comparison in the loop. The superscripts D and I_l denote the direct and indirect estimates respectively. Note that for each XY comparison there is one direct estimate common to all loops (i.e. $\hat{\mu}_{XY}^{D_1} \equiv \cdots \equiv \hat{\mu}_{XY}^{D_L}$), whereas different loops may give different indirect estimates (i.e. $\hat{\mu}_{XY}^{I_1} \neq \cdots \neq \hat{\mu}_{XY}^{I_L}$). The indirect effect and its variance $\hat{\mu}_{XY}^{I_l}$, $\hat{v}_{XY}^{I_l}$ are estimated separately for each l loop by Equations (2.1) and (2.2) respectively. The choice of the comparison within each loop does not affect the results; for example if $l \equiv XYZ$, then $\widehat{w}_l^{XY} = \widehat{w}_l^{XZ} = \widehat{w}_l^{YZ} = \widehat{w}_l$.

The null hypothesis $H_0: w_l = 0$ is assessed using the following z-statistic

$$z = \frac{\widehat{w}_l^{XY}}{\sqrt{v(\widehat{w}_l^{XY})}} \sim N(0,1) \tag{2.14}$$

Note that this approach as well as the next two approches for assessing local inconsistency do not account for the correlation induced by multi-arm trials. More specifically, each k-arm study is treated as $\frac{k\times(k-1)}{2}$ independent two-arm studies.

'Composite test' for inconsistency

The *composite test for inconsistency* extends this method and incorporates the entire network to allow also for discrepancies between the several indirect estimates

from different routes (53). More specifically, if a network provides a direct and F indirect estimates from independent indirect roots for a XY comparison, these can be synthesized into their weighted average $\hat{\mu}_{XY}^{M}$ using Equation (2.3). The discrepancy between the $F+1$ sources of evidence can be assessed by the statistic

$$U_{XY} = \frac{1}{\hat{v}_{XY}^{D}}(\hat{\mu}_{XY}^{D} - \hat{\mu}_{XY}^{M})^2 + \sum_{j=1}^{F} \frac{1}{\hat{v}_{XY}^{I_j}} \left(\hat{\mu}_{XY}^{I_j} - \hat{\mu}_{XY}^{M}\right)^2$$

which follows a χ_F^2 distribution under the null hypothesis $H_0: \hat{\mu}_{XY}^{D} = \hat{\mu}_{XY}^{I_1} = \cdots = \hat{\mu}_{XY}^{I_F} = \hat{\mu}_{XY}^{M}$.

Node-splitting & back calculation approaches for inconsistency

The *node-splitting* and *back-calculation* approaches are similar to the previous method but differ in the way the indirect effects are estimated (54). The former excludes one direct comparison at a time. Then, the network estimate for the excluded comparison XY based on the rest of the network is the indirect estimate $\hat{\mu}_{XY}^{I}$ coming from the synthesis of all available indirect evidence.

The back-calculation method is equivalent to the node-splitting but the indirect effect estimate and its variance are derived by solving Equations (2.3) and (2.4) into

$$\hat{\mu}_{XY}^{I} = \left(\frac{\hat{\mu}_{XY}^{M}}{\hat{v}_{XY}^{M}} - \frac{\hat{\mu}_{XY}^{D}}{\hat{v}_{XY}^{D}}\right)\hat{v}_{XY}^{I}$$

and

$$\frac{1}{\hat{v}_{XY}^{I}} = \frac{1}{\hat{v}_{XY}^{M}} - \frac{1}{\hat{v}_{XY}^{D}}$$

For both methods Equations (2.12), (2.13) and (2.14) can be employed to estimate the statistical significance of the inconsistency factor $\widehat{w}^{XY}$ for every comparison XY in the network. In case these methods are fitted in a Bayesian framework the presence of inconsistency can be judged by comparing the posterior distributions of the direct and indirect estimates for every comparison XY.

'Net-heat' approach for inconsistency

Finally, the *net-heat* approach uses the model of Equation (2.8) for the estimation of the design-specific contribution to inconsistency using the following Q statistic for inconsistency (IN)

$$Q_{IN}^{g} = \left(\widehat{\boldsymbol{\mu}}_{\boldsymbol{g}}^{\boldsymbol{D}} - \boldsymbol{X}_{\boldsymbol{g}}^{\boldsymbol{D}}\widehat{\boldsymbol{\mu}}\right)^{T}(\widehat{\boldsymbol{v}})^{-1}\left(\widehat{\boldsymbol{\mu}}_{\boldsymbol{g}}^{\boldsymbol{D}} - \boldsymbol{X}_{\boldsymbol{g}}^{\boldsymbol{D}}\widehat{\boldsymbol{\mu}}\right) \quad (2.15)$$

(with $\widehat{\boldsymbol{v}}$ the variance-covariance matrix of the network estimates $\widehat{\boldsymbol{\mu}}$) (see Section 2.2.2). Under the null hypothesis of consistency Q_{IN}^{g} is assumed to follow a χ^2 distribution with $S_g - K_g + 1$ degrees of freedom. Large values of Q_{IN}^{g} correspond to designs that might be important sources of inconsistency. Then, estimating each Q_{IN}^{g} after taking one design out at a time can further reveal designs that cause inconsistency in other parts of the network (29). This approach differs from the aforementioned approaches in the sense that it does not focus on the inconsistency for a specific comparison in the network but rather looks at all the comparisons in a specific design.

2.3.3 Approaches for evaluating inconsistency globally

Global approaches for inconsistency include inconsistency models (i.e. models that relax the consistency equations) as well as measures for inconsistency, such as Q and I^2 statistics. Three inconsistency models have been suggested so far in the literature; the *Lu & Ades* model (25) and the *design-by-treatment interaction* model (28,29,52) and the *unrelated mean effects* model (55).

The Lu & Ades model

The Lu & Ades inconsistency model (25) can be constructed by adding an additional linear term in the consistency equations (Equation (2.5)) in every loop for which there is potential for inconsistency

$$\mu_{YZ} = \mu_{XZ} - \mu_{XY} + w_l$$

The total number of inconsistency factors required in the model is $C^D - T + 1$. In the presence of multi-arm trials this number might need modification subtracting *"the number of independent inconsistency relations in which the corresponding parameters are supported by no more than two independent sources of evidence"*. For an extended discussion on the parameterization of the model see Lu et al. (25,28).

The rest of the model remains the same as the consistency NMA model (see Section 2.2).

The inconsistency factors can be assumed fixed or random parameters in analogy to the fixed and random effects models. A *random inconsistency* model assumes that

$$w_l \sim N(0, \sigma^2) \tag{2.16}$$

where σ^2 is the inconsistency variance. Large w_l suggest important inconsistency in the l loop. Comparing σ^2 and τ^2 might reveal how much inconsistency exists compared to heterogeneity. The statistical significance of all w_l jointly can be assessed using a χ^2 test.

The design-by-treatment interaction model

This model is an extension of the Lu & Ades model that accounts also for design inconsistency (i.e. disagreement between different study designs). The maximum number of potential inconsistencies (i.e. required inconsistency factors) in a network according to this model is $\sum_{g=1}^{G} C_g^D - T + 1$ with C_g^D the number of independent comparisons within each design g; that is $C_g^D = K_g - 1$. Then, the consistency equations (Equation (2.5)) are modified into

$$\mu_{YZ} = \mu_{XZ} - \mu_{XY} + w_l^{loop} + w_l^{des}$$

where the parameter w_l^{loop} reflects the potential loop inconsistency and w_l^{des} the potential design inconsistency in the l loop ($l \equiv XYZ$). There might be loops requiring the addition of either only w_l^{loop} or only w_l^{des}. However, the distinction between loop and design inconsistency is not always straightforward.

The unrelated mean effects model

The unrelated mean effects model is an inconsistency model that totally omits the consistency equations (Equation (2.5)) and renders the NMA model into a series of pairwise meta-analyses sharing a common heterogeneity parameter. Usually this model is compared with a model assuming consistency for model fit and parsimony as well as the magnitude of heterogeneity (see Section 2.2). If the inconsistency model fits the data better or presents lower heterogeneity, this is an indication that consistency might not be plausible (55).

Assessing inconsistency using Q-statistics

Following the definition of Q_H for heterogeneity in NMA, a Q_{IN} for inconsistency would be the sum of all design-specific contributions to inconsistency, hence

$$Q_{IN} = \sum_{g=1}^{G} Q_{IN}^{g} \sim \chi^2_{df}$$

where Q_{IN}^{g} is given by Equation (2.15) and $df = \sum_{g=1}^{G}(S_g - 1) - T + 1$ (29,36,37).

Assessing inconsistency using the I^2 measure

The I^2 statistic for inconsistency is estimated similarly to that for heterogeneity in NMA, which has been presented in the previous Section. In the case of inconsistency the respective R^2 would express the inflation in the volumes of the normal approximations to the confidence/credible regions for all relative effect parameters between a consistency and an inconsistency effects model.

2.4 New graphical tools for presenting data, assumptions & results

NMA has been criticized for the increased complexity of the evidence base, assumptions and results, which is induced by the presence of multiple treatment comparisons. Various graphical tools have been suggested for presenting the

findings and evaluating assumptions in pairwise meta-analysis (56,57). Adapting these tools into the context of NMA and using them to draw meaningful conclusions requires several modifications or extensions. In this section, I introduce a series of new or modified graphs aiming to make the outputs from NMA well-understandable to researchers that are less familiar with advanced statistical methods. To enhance the use and interpretation of each suggested graphical summary, I use examples of published NMA.

2.4.1 Presenting the evidence base

Network plot/diagram

A graphical illustration of the set of competing interventions along with the available direct pairwise comparisons can be provided by the *network plot*, called also *network diagram*. It consists of nodes representing the eligible interventions and edges linking pairs of nodes if the corresponding treatments are compared in at least one individual study.

I employ weighting schemes for nodes and/or edges to reveal differences between the included treatments and direct comparisons. For example, the least and most studied interventions can be identified by weighting the size of nodes according to the number of studies evaluating them or the number of participants allocated to each treatment arm. Weighting the thickness of edges according to the distribution of potential effect modifiers, such as baseline risk or study duration, can be the first step for the evaluation of transitivity.

Figure 2.1 shows the plot of a network that compares 13 treatments and placebo for acute mania in terms of efficacy (58). The graph shows that Olanzapine is the active treatment evaluated in most studies, while suggests only small differences in the average control group risk (CGR) across the direct comparisons (active treatment vs. placebo).

Similar to pairwise meta-analysis, interpretation of results from NMA should always be in light of the quality of the available evidence (59). The inclusion of individual studies with design limitations can mitigate the validity of findings. I suggest plotting the network with colored edges to identify comparisons that might be more prone to biased summary estimates. Usually, studies are classified as being of low, unclear or high risk of bias (RoB) according to a specific design characteristic (allocation concealment, blinding, etc.) (59). Alternatively, researchers may be willing to use more levels of study classification. The comparison-specific level of overall design limitations can be estimated as a function (e.g. average, weighted average, mode, etc.) of the respective study-specific levels.

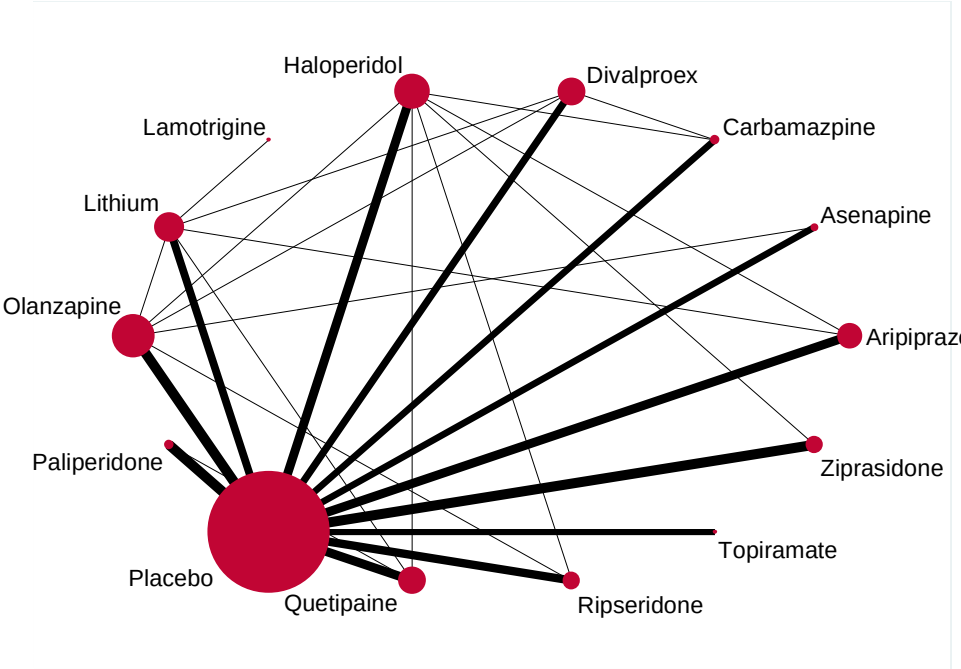

Figure 2.1. Network plot of a network comparing the effectiveness of 13 treatments and placebo for acute mania. Edges connecting active treatments with placebo are weighted according to the mean control group risk of the respective direct comparison, while edges connecting two active treatments have been given minimal weight. Nodes are weighted according to the number of studies evaluating each intervention.

In **Figure 2.2** the acute mania network is presented with color-adjusted edges with respect to the appropriate conduct of allocation concealment. In six comparisons the majority of studies have been assessed as being at low RoB (green lines) and in the rest of comparisons at unclear RoB.

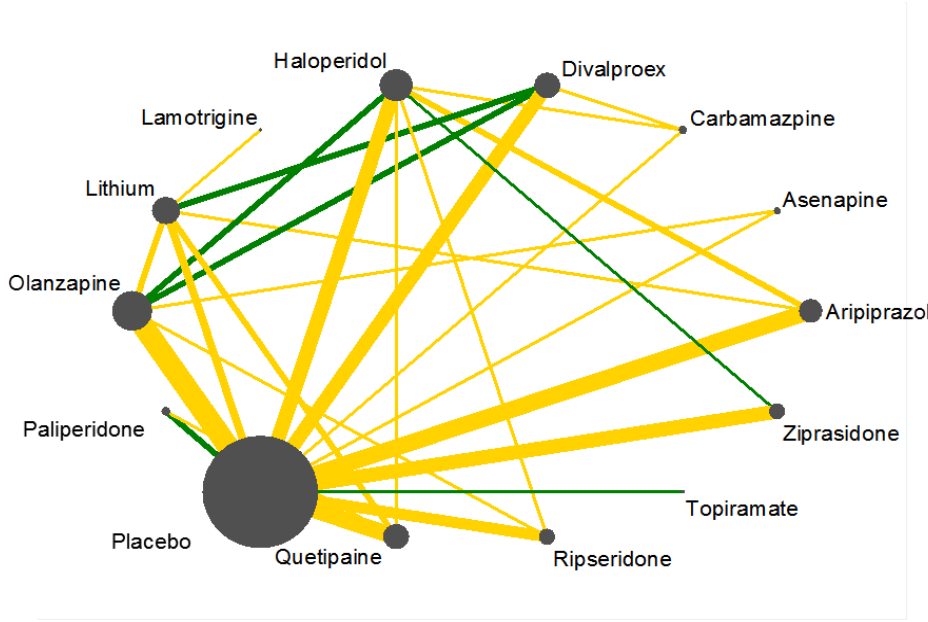

Figure 2.2. Network plot of a network comparing the effectiveness of 13 treatments and placebo for acute mania with color-adjusted edges. Edges are colored according to the adequacy of allocation concealment estimated as the weighted average of the study-specific bias levels with weights the inverse of study variances. Yellow edges represent direct comparisons with unclear average level of bias and green edges represent comparisons with low average level of bias.

Contribution plot

Constructing the evidence base of a network of trials needs careful consideration of the eligibility criteria for interventions and studies. Often researchers include in the analysis interventions that might be not of direct interest but provide useful indirect evidence, such as placebo or standard care. The inclusion of such interventions is valuable only when the risk of introducing intransitivity and inconsistency does not outweigh their contribution in the estimation. In addition, such information can be useful for drawing conclusions regarding the quality of evidence (60). I developed the *contribution plot* or *contribution matrix*, which can help in making these judgments. This matrix plot shows the percentage contribution of each direct piece of evidence in the network estimates and in the entire network. The weight of each available direct comparison in the network estimates can be estimated using the weighted least squares solution of Equation (2.8)

$$\hat{\boldsymbol{\mu}}^{\boldsymbol{N}} = \boldsymbol{X}^{\boldsymbol{N}}\hat{\boldsymbol{\mu}}$$

$$\widehat{\boldsymbol{\mu}} = \left[\left(\boldsymbol{X^D}\right)^T\left(\widehat{\boldsymbol{v}}^{\boldsymbol{D}}\right)^{-1}\boldsymbol{X^D}\right]\left(\boldsymbol{X^D}\right)^T\left(\widehat{\boldsymbol{v}}^{\boldsymbol{D}}\right)^{-1}\widehat{\boldsymbol{\mu}}^{\boldsymbol{D}}$$

where $\widehat{\boldsymbol{\mu}}$ is the vector of network estimates for the basic parameters and $\widehat{\boldsymbol{\mu}}^{\boldsymbol{N}} = (\hat{\mu}_1^N, \dots, \hat{\mu}_C^N)$ the vector of the network estimates for all possible pairwise comparisons C. The design matrix $\boldsymbol{X^N}$ contains the linear relationships between the C comparisons and the basic parameters through the consistency equations (Equation (2.5)). The $\left(C^D \times \sum_{g=1}^{G}[K_g - 1]\right)$ matrix

$$\boldsymbol{H} = \left[\left(\boldsymbol{X^D}\right)^T\left(\widehat{\boldsymbol{v}}^{\boldsymbol{D}}\right)^{-1}\boldsymbol{X^D}\right]\left(\boldsymbol{X^D}\right)^T\left(\widehat{\boldsymbol{v}}^{\boldsymbol{D}}\right)^{-1}$$

known as the *hat matrix* (61), maps the direct estimates to the network estimates for the basic comparisons.

An 'extended' hat matrix, which maps the direct estimates into network estimates for all C pairwise comparisons, so as

$$\widehat{\boldsymbol{\mu}}^{\boldsymbol{N}} = \boldsymbol{H}^*\widehat{\boldsymbol{\mu}}^{\boldsymbol{D}}$$

can be estimated as

$$\boldsymbol{H}^* = \boldsymbol{X^N}\left[\left(\boldsymbol{X^N}\right)^T\left(\widehat{\boldsymbol{v}}^{\boldsymbol{D}*}\right)^{-1}\boldsymbol{X^N}\right]\left(\boldsymbol{X^N}\right)^T\left(\widehat{\boldsymbol{v}}^{\boldsymbol{D}*}\right)^{-1} \quad (2.17)$$

where $\widehat{\boldsymbol{v}}^{\boldsymbol{D}*}$ is the extended $\widehat{\boldsymbol{v}}^{\boldsymbol{D}}$ matrix with imputed large variances (e.g. 10^4) for those $C - C^D$ comparisons that are not reported in any study.

To facilitate the interpretation of the elements of the $\boldsymbol{H}^*$ matrix, I ignore the different study designs; in this way all studies reporting the same direct comparisons are synthesized irrespective of their design. Then, $\boldsymbol{H}^*$ is of dimension $(C \times C^D)$ and each element h_{cd} ($c = 1, \dots, C$ and $d = 1, \dots, C^D$) represents the weight of the column-defining direct comparison in the row-defining network estimate. I express the weights of each row as percentages

$$h_{cd}\% = \frac{|h_{cd}|}{|h_{c1}| + \cdots + |h_{cC^D}|}$$

and then the percentage contribution of the direct comparison d to the entire network is

$$h_{cd}[\%N] = \frac{\sum_{c=1}^{C}|h_{cd}|}{\sum_{d=1}^{C^D}\sum_{c=1}^{C}|h_{cd}|}$$

Figure 2.3 presents the contribution plot of a network evaluating the relative effectiveness of four different percutaneous coronary interventions for non-acute

coronary artery disease (62). The graph shows that the comparison of bare-metal stents vs. percutaneous transluminal balloon coronary angioplasty is the most frequent in the network (33 studies). However, the comparison of bare-metal stents vs. drug-eluting stent has the highest percentage contribution overall in the network (31.2%).

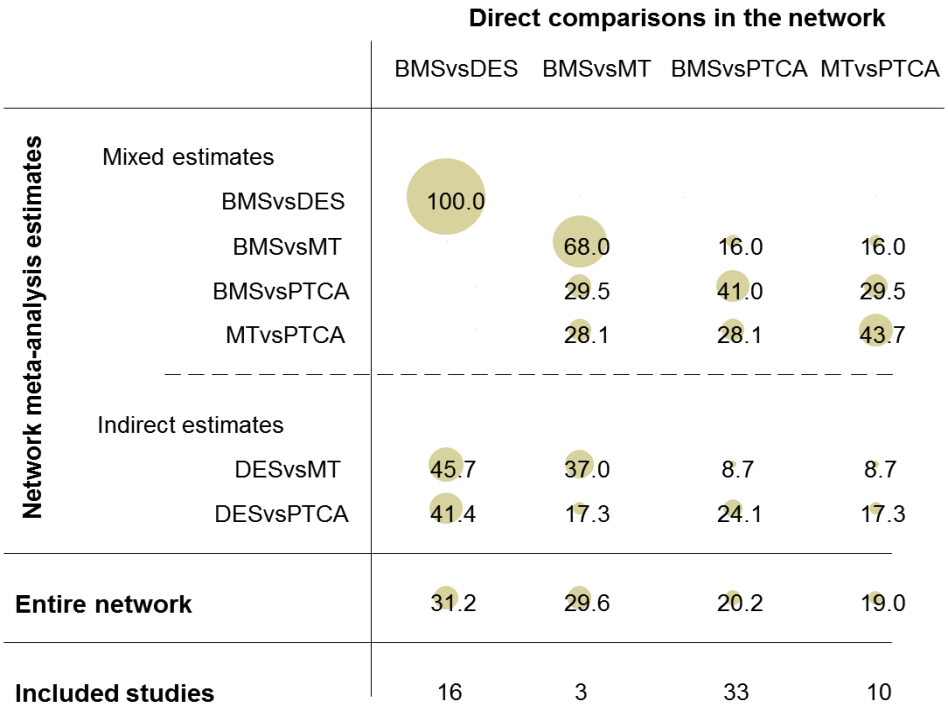

Figure 2.3. Contribution plot of a network comparing the effectiveness of four different percutaneous coronary interventions for non-acute coronary artery disease. The numbers are the percentage contributions of the column-defining direct comparisons to the row-defining network estimates. The size of the circles is proportional to these percentages. (MT = medical therapy, PTCA = percutaneous transluminal balloon coronary angioplasty, BMS = bare-metal stents, DES = drug eluting stents)

2.4.2 Presenting the assumptions

The evaluation of both homogeneity and consistency is crucial when undertaking a NMA. Section 2.3 describes several approaches that aim to assess whether important heterogeneity or inconsistency can be present in a network of interventions. Some of these methods can be summarized in the presentation of a single value (e.g. the p-value of a χ^2 test or the value of the I^2 measure). However, other approaches need to provide more information to derive an overall inference,

if possible, for the network. Summarizing this information in a single forest plot may offer a straightforward way to get a first impression whether the two assumptions are likely to hold (63).

More specifically, in this Section I focus on two approaches; the assessment of heterogeneity based on PI (see Section 2.3.1) and the loop-specific approach for evaluating inconsistency (see Section 2.3.2). In **Figure 2.4** and **Figure 2.5** respectively, I present examples of forest plots including the results from these two methods.

Predictive intervals plot

The graph in **Figure 2.4** corresponds to a star-shaped network including six biologic agents for rheumatoid arthritis compared directly only with placebo (64). The between-study variance (τ^2) was estimated 0.26 and according to the mean summary relative effects (black circles) all six active treatments seem more effective than placebo. I used colored horizontal lines to represent the PIs for all pairwise comparisons by extending the lines that correspond to the CIs. The plot suggests that only for the comparison of infliximab versus placebo the PI does not support the statistically significant effect due to the additional uncertainty anticipated in future studies.

Inconsistency plot

The forest plot in **Figure 2.5** shows the inconsistency results for all loops based on the loop-specific approach (see Section 2.3.2) for the acute mania network (see Section 2.4.1). To facilitate the interpretation of the IFs, I re-expressed them in the effect size scale. This means that the squares in the plot represent the ratio of odds ratios (ROR) between direct and indirect estimates ($ROR_l^{XY} = exp(\widehat{w}_l^{XY})$). Only one (ARI-HAL-LITH-QUE) of the 21 loops in the network seems to be marginally subject to statistically significant inconsistency. However, the limited power of the test for inconsistency implies that non-significant but large RORs in magnitude or

with large CI (depending on the clinical setting and outcome) should be explored as well. I truncated all CI in **Figure 2.5** to the null value, since absolute values of the IFs are estimated and thus RORs smaller than 1 are difficult to interpret.

Note that in this graph I assumed a common heterogeneity for the entire network ($\hat{\tau}^2 = 0.07$), which was derived after performing NMA with the multivariate meta-analysis approach (see Section 2.2.2). Alternative options would be to assume a common heterogeneity within each loop but different between loops or a different heterogeneity for each pairwise comparison. These different assumptions about heterogeneity as well as the use of different methods to estimate the τ^2 might impact on the inconsistency results (65).

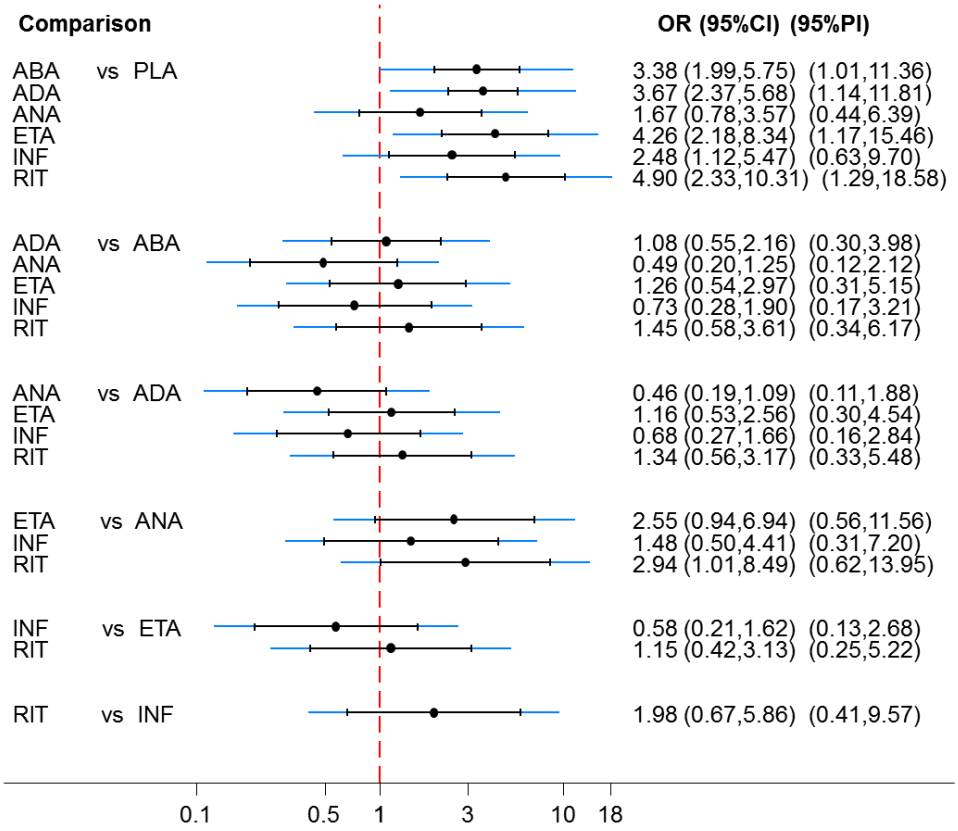

Figure 2.4. Predictive interval plot of a network network comparing the effectiveness of six biologic agents and placebo for rheumatoid arthritis. The circles are the summary odds ratios (OR), the black horizontal lines their confidence intervals (CI) and the blue lines the respective oredictive intervals (PI). The red dashed line is the line of no effect (OR=1). (PLA = placebo, ABA = abatacept, ADA = adalimumab, ANA = anakinra, ETA = etanercept, INF = infliximab, RIT = rituximab)

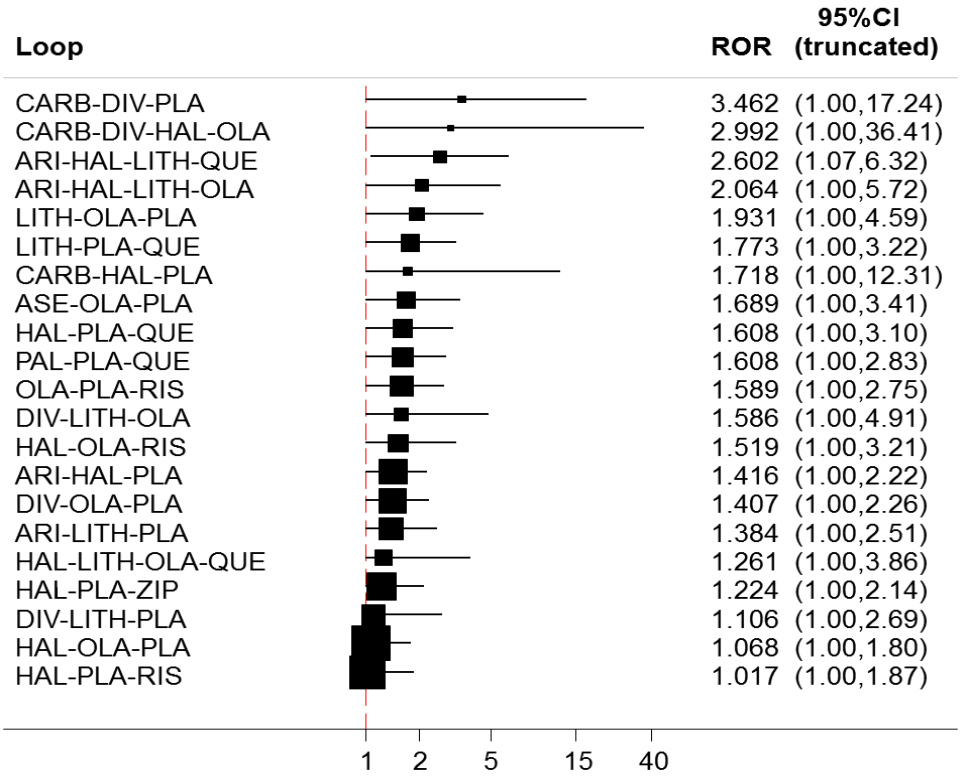

Figure 2.5. Inconsistency plot of a network comparing the effectiveness of 13 treatments and placebo for acute mania. A common heterogeneity $\hat{\tau}^2 = 0.07$ has been assumed for all loops estimated from the restricted maximum likelihood method.

2.4.3 Presenting the results

A major advantage of NMA is that provides estimates of relative effects for all possible comparisons between pairs of treatments. However, in the presence of many competing treatments the presentation of all pairwise estimates becomes cumbersome and not very helpful in drawing conclusions regarding which treatments seem to work best. Thus, the relative ranking of treatments is often used as a supplementary output to facilitate the identification of treatments that perform well enough with respect to the studied outcome.

Ranking plots for a sinlge outcome using probabilities

There are several methods to estimate the ranking of treatments, which may sometimes give different results. One of the most common approaches is to

estimate the probabilities for each treatment being ranked at a specific place (12). Such probabilities can be estimated either within a Bayesian environment using Markov Chain Monte Carlo (MCMC) simulations or in a frequentist framework using resampling methods (e.g. bootstrap methodology (66)). Treatments are ranked according to their relative effects versus a common reference in each cycle and the probability for treatment t $(t = 1, \dots, T)$ being ranked r^{th} $(r = 1, \dots, T)$ would be

$$p_{rt} = \frac{\#\ simulations\ (t = r)}{total\ \#\ simulations}$$

Inference on the treatment ranking should account for the uncertainty in ranking by incorporating the estimated probabilities for all possible ranks. These are often presented with a probability curve for each treatment, called also *rankograms* (**Figure 2.6**). To summarize all p_{rt} and get the relative ranking of treatments *cumulative ranking probabilities* can be employed. These are given by the formula

$$cp_{rt} = p_{1t} + \cdots + p_{rt}$$

and express the probability for treatment t being within the first r places. Then, the *surface under the cumulative ranking curves* (SUCRA) is used as a relative ranking measure; larger areas under the curve corresponding to better treatments. **Figure 2.7** shows an example of these graphs for the rheumatoid arthritis network where the relative ranking of two different NMA models is compared; the standard hierarchical NMA model (Section 2.2.2) and a model controlling for differences in precision across studies (see Section 5.5 for a description of the model). The figure suggests that accounting for the impact of small-study effects minimizes the differences in effectiveness between the treatments and assigns to them less distinct ranks. The surface under these curves can be expressed also as a percentage

$$SUCRA_t = \frac{\sum_{r=1}^{T-1} cp_{rt}}{T - 1}$$

which is interpreted as the percentage of effectiveness/safety of a treatment ranked always first without any uncertainty.

An alternative ranking measure is the *mean rank*, which is the weighted average of all possible ranks with weights the ranking probabilities

$$mr_t = \sum_{r=1}^{T}(p_{rt} \times r)$$

The mathematical relationship between SUCRAs and mean ranks is expressed by the equation

$$SUCRA_t = \frac{(2T-1) \times mr_t - T}{T-1}$$

This implies that mean rank is equivalent to SUCRA and give the same relative ranking results. Note that I provide the proof of the above equation in the Appendix.

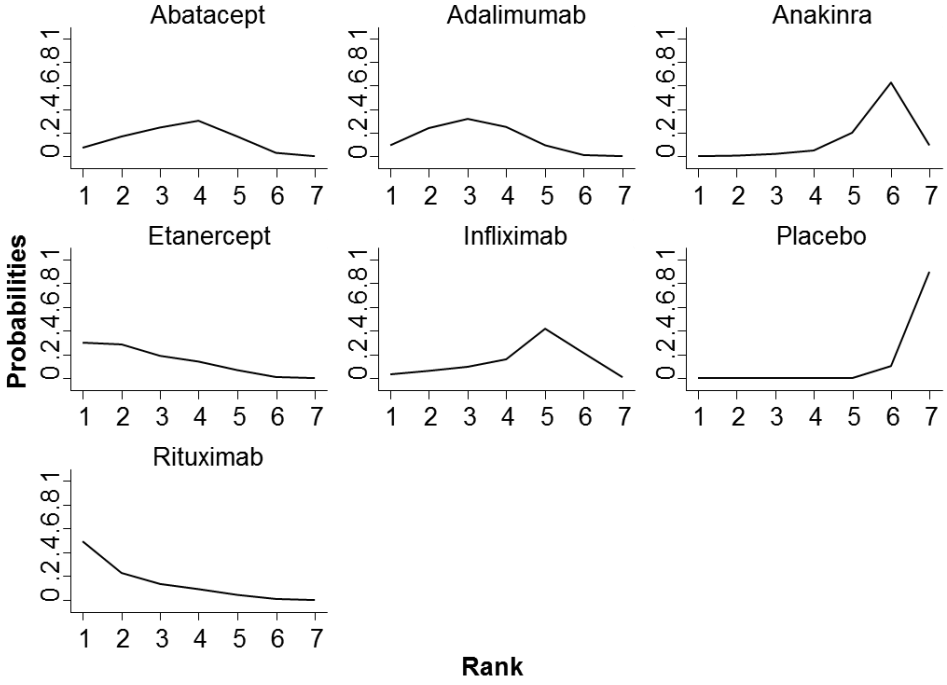

Figure 2.6. Probability ranking curves (rankograms) of a network network comparing the effectiveness of six biologic agents and placebo for rheumatoid arthritis. The horizontal axis contains the possible ranks in the network and the vertical axis the ranking probabilities.

Ranking plot for a single outcome using multidimensional scaling

I also suggest a different approach to estimate the treatment ranking using Multidimensional Scaling (MDS) techniques (67). To apply this method, I treat the NMA summary estimates between all pairs of interventions as proximity data with aim to reveal their latent structure.

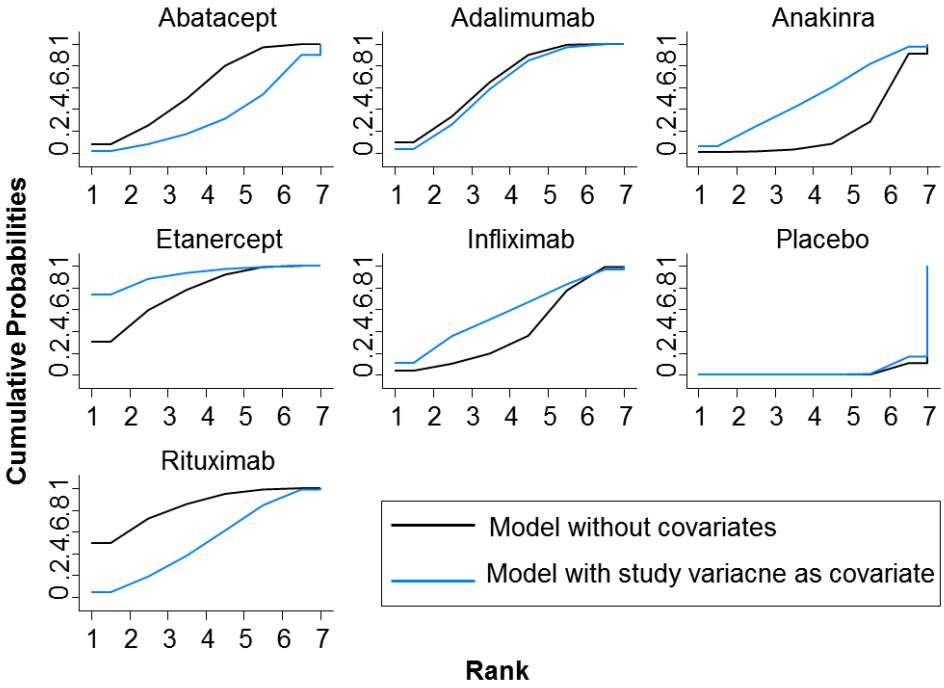

Figure 2.7. Cumulative probability ranking curves of a network comparing the effectiveness of six biologic agents and placebo for rheumatoid arthritis. The horizontal axis contains the possible ranks in the network and the vertical axis the cumulative ranking probabilities. Larger areas under the curve correspond to more effective treatments.

In this way the absolute value of the network estimate $|\hat{\mu}_{XY}|$ $(X, Y = 1, \dots, T)$ defines the dissimilarity between the two treatments (X, Y) with $|\hat{\mu}_{XX}| = 0$. I weight the absolute effects sizes by their inverse standard errors or variances to ensure that the assumption of a common distribution between the elements of the matrix is plausible. Assuming that the rank of the treatments is the only dimension underlying the outcome the purpose of the MDS would be to reduce the $T \times T$ matrix into a $T \times 1$ vector. This vector involves the set of distances being as close as possible to the observed dissimilarities (i.e. relative effects) and would represent the treatment relative ranking. The scatterplot in **Figure 2.8** presents the relative ranking of treatments for the rheumatoid arthritis network using the estimated MDS dimension. There are small differences in the ranking of treatments between the SUCRA (or mean ranks) and the MDS in the four best treatments.

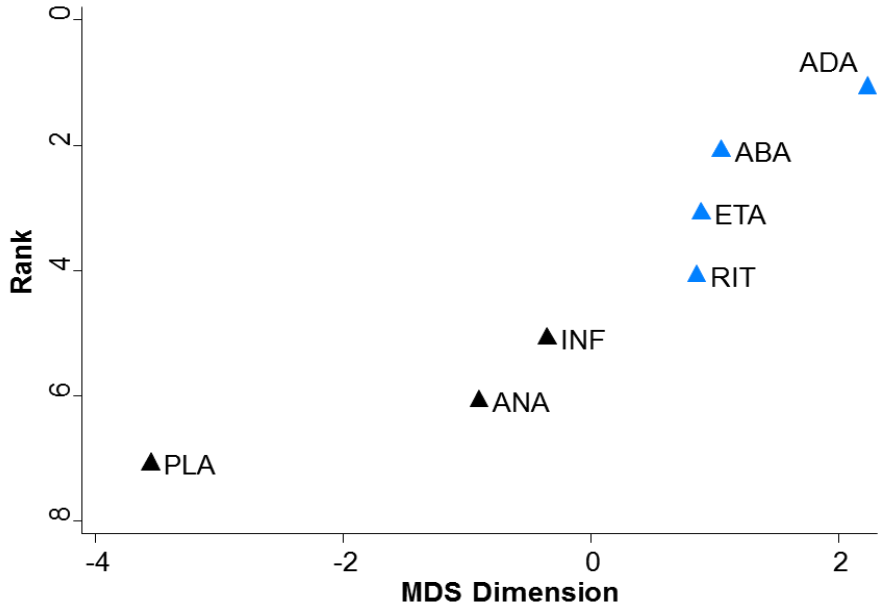

Figure 2.8. Ranking scatterplot of a network network comparing the effectiveness of six biologic agents and placebo for rheumatoid arthritis. Treatments have been ranked using multidimensional scaling (MDS) methods. Blue triangles represent treatments for which MDS and ranking probabilities give different order. More effective treatments lie in the right upper corner.

Cluster ranking plot for two outcomes

In a decision-making context, recommending a specific interventon for a clinical condition should consider more than one outcome. For instance, highly effective treatments sometimes have serious adverse events. Methods for analysing jointly multiple outcomes are well-established for pairwise meta-analyses (40,41) and have been extended into the context of NMA (68,69). However, to date there is no available methodology to estimate a single treatment relative ranking incorporating information for more than one outcomes.

I suggest the use of two-dimensional plots as a possible tool suitable to draw inference for the relative ranking of treatments based on two (competing) outcomes. More specifically, I plot the values of a relative ranking measure (e.g. SUCRA, mean ranks or MDS dimension) for two outcomes jointly in a system of co-ordinate axes. I additionally employ clustering methods to form meaningful groups of treatments with respect to their performance on both outcomes. Cluster analysis aims to group different objects based on their characteristics in a way to result in high association within each cluster but in low association between

clusters (70). The two-dimensional plot for the acute mania network is presented in **Figure 2.9**, where the ranking results for efficacy and acceptability have been put together. According to the graph the best group of studies considering both outcomes includes 7 treatments.

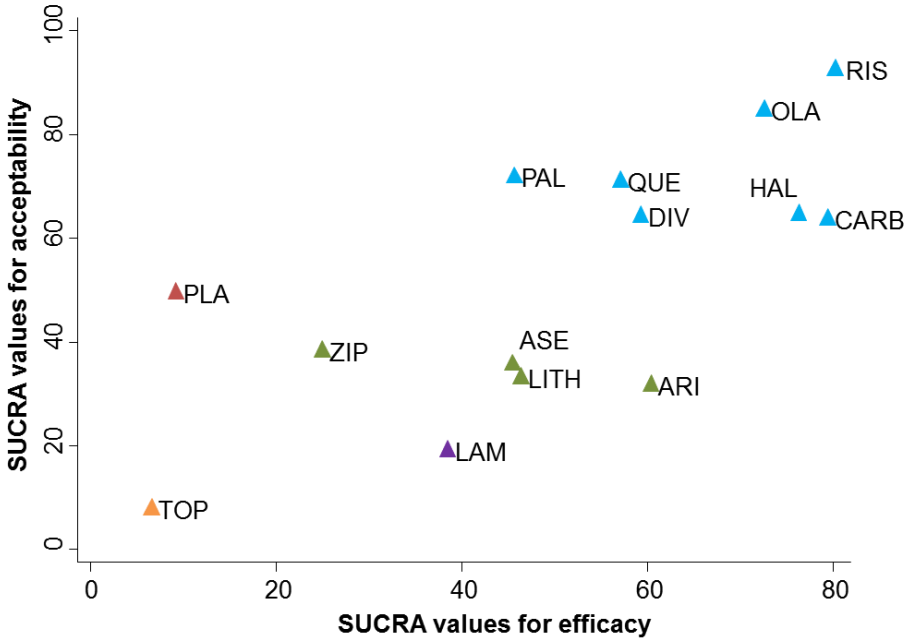

Figure 2.9. Two-dimensional ranking plot of a network comparing 13 treatments and placebo for acute mania in terms of efficacy and acceptability. Treatments that perform well on both outcomes lie in the right upper corner. Treatments have been grouped using clustering methods and different colors represent the different treatment groups.

2.5 Software options for network meta-analysis

The rapid development of NMA methodology underlines the need for flexible and user-friendly software options that would facilitate the appropriate conduct and comprehensive reporting of the analysis. In recent years, the hierarchical model fitted within a Bayesian framework has been the most popular approach for NMA (20,22); the reason is that until recently software to fit non-Bayesian approaches could not properly handle the inclusion of multi-arm trials or fit inconsistency models. Consequently, researchers not familiar with Bayesian methods and

software environment avoided applications of NMA. The development of new graphical presentation tools (see Section 2.4) can aid the ease of presentation and interpretation of NMA, particularly when they are implemented in conventional software.

The two following Sections (2.5.1 and 2.5.2) provide an overview of the available software options for fitting the NMA models and evaluating the required assumptions. Then, Section 2.5.3 describes a series of eight STATA routines that I developed suitable to produce the graphs described in Section 2.4 (45).

2.5.1 Software for performing network meta-analysis

All approaches to NMA presented in Section 2.2.2 can be performed, in theory, in many statistical packages. However, the variety of software options is, in practice, limited due to lack of readily available routines suitable for NMA. The most popular packages in research related to NMA are BUGS (i.e. WinBUGS, OpenBUGS)/JAGS (71), STATA (23) and R (72); SAS (73) is less frequently used and thus is not considered in this section. Another software package developed specifically for NMA is the GeMTC (30). All software options for fitting each modeling approach for NMA (that require no self-programming by the user) are described below.

Performing indirect comparisons

Indirect comparisons can be performed either using the `indirect` command in STATA (74) or the *ITC software* (ITC=indirect treatment comparison) available from the Canadian Agency for Drugs and Technologies in Health (CADTH) (75). **Figure 2.10** shows the dialog box of the STATA command.

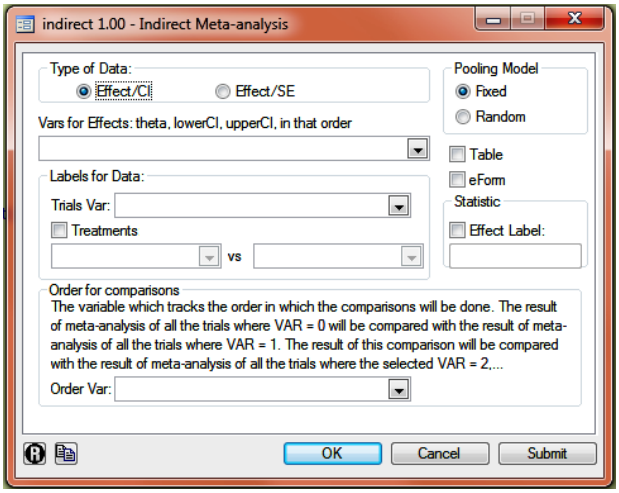

Figure 2.10. Dialog box of the `indirect` command in STATA.

Fitting network meta-analysis as a multivariate meta-regression model

This approach for NMA can be fitted using any software routine able to perform meta-regression. However, conventional routines developed for standard meta-regression in pairwise meta-analysis fail to model properly multi-arm studies and can be employed only for a subset of networks (with only two-arm studies). Such routines are the `metareg` command in STATA (76), the `metafor` (77) and `meta` (78) packages in R or codes in BUGS provided in Dias et al. (79).

The inherent correlation in multi-arm trials can be incorporated in the estimation by routines developed for multivariate meta-analysis that allow for covariates. Both the `mvmeta` command in STATA (33,80) and the `mvmeta` package in R (81,82) can be used but only STATA can impose the assumption of a common heterogeneity across comparisons. Note that the STATA `mvmeta` command can be used either as a standalone command or via the package `network` in STATA (available from http://www.mrc-bsu.cam.ac.uk/IW_Stata/meta).

The use of graph-theoretical methods in NMA has been implemented in the `netmeta` package in R (35).

Fitting network meta-analysis as a two-stage meta-regression model

R functions are also available for fitting the meta-regression model for NMA in two stages as described in Section 2.2.2. and can be found in Lu et al. (36) and Krahn et al. (29).

Fitting network meta-analysis as a hierarchical model

The hierarchical model can be fitted in Bayesian environment using BUGS codes available online from the Integrating Multiple Meta-Analysis project (IMMA, University of Ioannina) and the Multi-Parameter Evidence Synthesis program (MPES, University of Bristol) at www.mtm.uoi.gr and www.bris.ac.uk/social-community-medicine/projects/mpes. Several BUGS codes are also provided in Dias et al. (15). Alternatively, the hierarchical model can be fitted by the GeMTC software.

Fitting network meta-analysis as a multivariate meta-analysis model

The last approach that treats NMA as a multivariate meta-analysis model can be performed using the `mvmeta` command in STATA (33,80) and the `mvmeta` package in R (81,82) with only STATA allowing for a common heterogeneity across comparisons.

2.5.2 Software for evaluating assumptions

This section focuses on the methods presented in Section 2.3 for the evaluation of heterogeneity and inconsistency in a network of interventions.

Assessing heterogeneity using predictive intervals

To enable the estimation of PI for every pairwise relative effect of a NMA in STATA, I developed the `intervalplot` command (45). The command should be

run after performing NMA with the `mvmeta` command (as a multivariate meta-analysis model, see Section 2.2.2). The command can be used also via a dialog box (**Figure 2.11**). Alternatively, PI can be estimated by fitting the hierarchical model in BUGS (Equation (2.11)) using the codes provided at www.mtm.uoi.gr.

Assessing heterogeneity using Q-statistics

The Q-statistics for heterogeneity in NMA have been implemented in the R function given by Krahn et al. (29) and in the `netmeta` package in R (35).

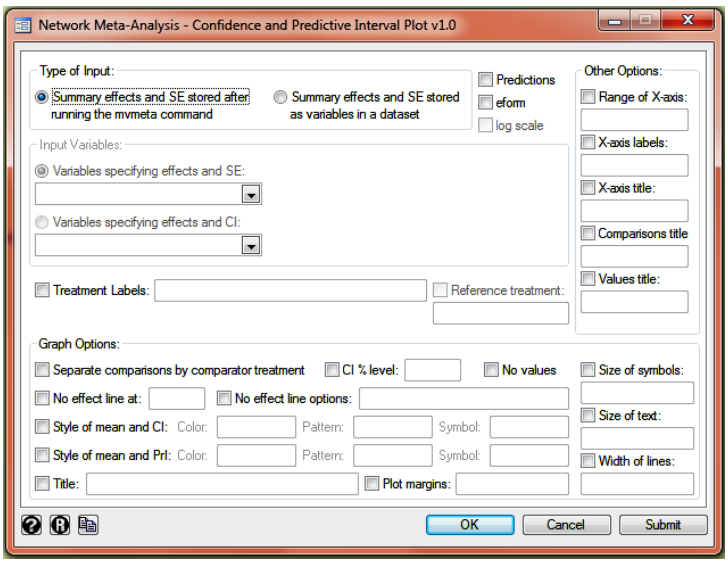

Figure 2.11. Dialog box of the `intervalplot` command in STATA.

Performing the loop-specific approach for inconsistency

I implemented the loop-specific approach for inconsistency in STATA by programming the `ifplot` command (45). An equivalent `ifplot` function in R is available from www.mtm.uoi.gr. The two routines can incorporate different assumptions about the heterogeneity; different heterogeneities across the direct comparisons, a common heterogeneity within each loop, a common heterogeneity for the entire network (65). The STATA command is available also in a dialog box (**Figure 2.12**). **Box 2.1** describes how the command identifies the existing triangular and quadtratic loops in a network.

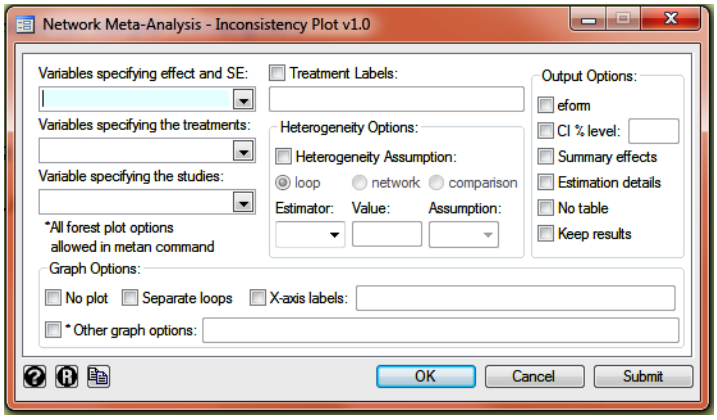

Figure 2.12. Dialog box of the `ifplot` command in STATA.

Box 2.1. Algorithm used in the `ifplot` STATA command to identify the available triangular and quadratic loops.

1. Identify the total number of treatments (T)
2. Identify the available direct comparisons (C^D) between pairs of treatments (t_{1d}, t_{2d}) with $d = 1, \dots, C^D$ and $t_{1d}, t_{2d} = 1, \dots, T$, $t_{1d} \neq t_{2d}$
3. For every (t_{1d}, t_{2d}), create a 1×2 matrix $\boldsymbol{T}_{t_{1d},t_{2d}} = (t_{1d}, t_{2d})$
4. For $T > k > j > i > 1$, if the matrices $\boldsymbol{T}_{i,j}, \boldsymbol{T}_{j,k}, \boldsymbol{T}_{i,k}$ exist, create a 1×6 matrix $\boldsymbol{T}_{i,j,k} = (\boldsymbol{T}_{i,j}, \boldsymbol{T}_{j,k}, \boldsymbol{T}_{i,k})$
5. For $T > z > k > j > i > 1$,
   if the matrices $\boldsymbol{T}_{i,j}, \boldsymbol{T}_{j,k}, \boldsymbol{T}_{k,z}, \boldsymbol{T}_{i,z}$ exist, create a 1×8 matrix $\boldsymbol{T}_{i,j,k,z} = (\boldsymbol{T}_{i,j}, \boldsymbol{T}_{j,k}, \boldsymbol{T}_{k,z}, \boldsymbol{T}_{i,z})$
   if the matrices $\boldsymbol{T}_{i,j}, \boldsymbol{T}_{j,z}, \boldsymbol{T}_{k,z}, \boldsymbol{T}_{i,k}$ exist, create a 1×8 matrix $\boldsymbol{T}_{i,j,k,z} = (\boldsymbol{T}_{i,j}, \boldsymbol{T}_{j,z}, \boldsymbol{T}_{k,z}, \boldsymbol{T}_{i,k})$
   if the matrices $\boldsymbol{T}_{i,k}, \boldsymbol{T}_{j,k}, \boldsymbol{T}_{j,z}, \boldsymbol{T}_{i,z}$ exist, create a 1×8 matrix $\boldsymbol{T}_{i,j,k,z} = (\boldsymbol{T}_{i,k}, \boldsymbol{T}_{j,k}, \boldsymbol{T}_{j,z}, \boldsymbol{T}_{i,z})$
   if the matrices $\boldsymbol{T}_{i,k}, \boldsymbol{T}_{k,z}, \boldsymbol{T}_{j,z}, \boldsymbol{T}_{i,j}$ exist, create a 1×8 matrix $\boldsymbol{T}_{i,j,k,z} = (\boldsymbol{T}_{i,k}, \boldsymbol{T}_{k,z}, \boldsymbol{T}_{j,z}, \boldsymbol{T}_{i,j})$
6. For $T > k > j > i > 1$, if the matrix $\boldsymbol{T}_{i,j,k}$ exists, create a dummy variable representing the loop $i - j - k$
7. For $T > z > k > j > i > 1$, if the matrix $\boldsymbol{T}_{i,j,k,z}$ exists, create a dummy variable representing the loop $i - j - k - z$

Performing the 'node-splitting' approach for inconsistency

The node-splitting method has been implemented into the `network` package in STATA and the GeMTC software. Relevant BUGS codes can be found also online at www.bris.ac.uk/social-community-medicine/projects/mpes.

Performing the 'net-heat' approach for inconsistency

The net-heat approach can be conducted through the R function in Krahn et al. (29) and the `netmeta` package in R (35).

Fitting the Lu & Ades inconsistency model

The parameterization of the Lu & Ades model can be done automatically only in the GEMTC software.

Fitting the design-by-treatment inconsistency model

The design-by-treatment model can be performed in STATA (`mvmeta` command (80)) and the parameterization can be done automatically via the `network` package (available from http://www.mrc-bsu.cam.ac.uk/IW_Stata/meta).

Assessing inconsistency using Q-statistics

The Q-statistics for inconsistency can be estimated by the `netmeta` package in R (35) and the `mvmeta` command in STATA (commonly known as the χ^2-test for the IF) (80).

2.5.3 Implementation of the new graphical presentation tools in STATA

I implemented all graphical tools described in Section 2.4 in STATA with the development of the package `network_graphs`, available from www.mtm.uoi.gr. The most important features of the commands are presented below. Additional details can be found in Chaimani et al. (45) as well as in the help files. Also, the codes of the STATA commands are available online and can be obtained by downloading the .ado files from mtm.uoi.gr/images/network_graphs.rar.

The `networkplot` command can be used to create network diagrams. I allow several weighting schemes for nodes and edges (**Figure 2.1**) via the options `nodeweight()` and `edgeweight()` respectively. In the default graph, both nodes and edges are weighted according to the number of studies including each intervention and direct comparison respectively. I incorporated the use of colored edges with the option `edgecolor()` (**Figure 2.2**). To plot all the treatmetns in a

circle, I used the algorithm provided in **Box 2.2**. The command can be used also via a dialog box (**Figure 2.13**).

Box 2.2. Algorithm used in the `networkplot` command to plot all competing treatments of a network in a circle.

1. Identify the total number of treatments (T)
2. Assume that all treatments lie on a circle with centre $(x, y) = (0,0)$ and radius equal to 1.
3. Divide the circle in T equal radians and calculate the central angle formed by each pair of treatments as $a^o = \frac{2\pi}{T}$
4. Specify the coordinates (x, y) for every treatment $t = 1, \dots, T$ using the following rule:
   - For $t = 1$, $(x[1], y[1]) = (1,0)$
   - For $t = 2, \dots, T$, $(x[t], y[t]) = \left((t-1)\cos[a^o], (t-1)\sin[a^o]\right)$

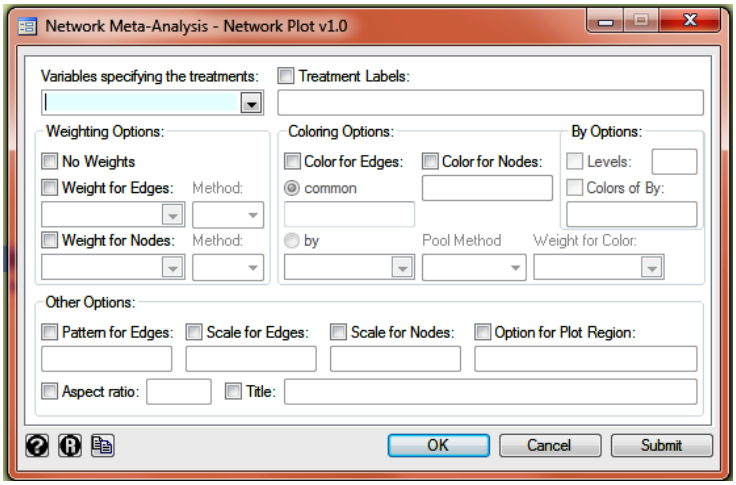

Figure 2.13. Dialog box of the `networkplot` command in STATA.

The `netweight` command estimates the weight of each direct comparison in the network estimates and creates the contribution plot (**Figure 2.3**). In **Box 2.3**, I provide the algorithm I used to define the basic comparisons and the design matrix. This is more challenging when there is not a reference treatment compared to any other treatment in the network. I have incorporated several assumptions for the comparison-specific heterogeneities. More specifically, the options `fixed`, `random` or `tau2(#)` can be added to specify that direct summary effects would be estimated assuming a fixed effect model, a comparison-specific random effects [default] model or a common heterogeneity for all comparisons, if this is already known (see also Sections 2.2.2 and 2.4.1). I added the option `bargraph()`, which

presents the comparison-specific contributions for every network estimate in a bar graph; the bars are colored according to a specific characteristic of each direct comparison (e.g. the average risk of bias of the studies (60)). The dialog box of the command is presented in **Figure 2.14**.

Box 2.3. Algorithm used in the `netweight` command to choose the basic comparisons properly and create the design matrix for any network.

1. Identify the total number of treatments (T)
2. Identify the available direct comparisons (C^D) between pairs of treatments (t_{1d}, t_{2d}) with $d = 1, \dots, C^D$ and $t_{1d}, t_{2d} = 1, \dots, T$, $t_{1d} \neq t_{2d}$
3. Choose the basic comparisons ($b = 1, \dots, T-1$) using the following steps:
   - Create the $2C^D \times 2$ matrix $\boldsymbol{BD} = \begin{pmatrix} t_{11} & 1 \\ t_{21} & 1 \\ \vdots & \vdots \\ t_{1C^D} & C^D \\ t_{2C^D} & C^D \end{pmatrix}$
   - For every $\boldsymbol{BD}[i,1] = \boldsymbol{BD}[i-1,1]$ ($i = 1, \dots, 2C^D$), delete the i^{th} row.
   - For every $\boldsymbol{BD}[i,2] = \boldsymbol{BD}[i-1,2]$ ($i = 1, \dots, 2C^D$), delete the i^{th} row.
   - Set the remaining comparisons (B) in the second column of the matrix $\boldsymbol{BD}$ as basic comparisons.
   - If $B < T-1$, add (any) $T-1-B$ comparisons in the set of basic comparisons (B).
4. Define the entries in the design matrix using the following steps:
   - Create the $T-1$ dummy variables $c_{t_{11},t_{21}}, \dots, c_{t_{1B},t_{2B}}$ representing the $b = 1, \dots, B$ basic comparisons with t_{1b}, t_{2b} the two treatments compared in each comparison b and set them 0.
   - For every possible comparison in the network (observed or unobserved) (C), replace the elements of the respective row in the design matrix ($\boldsymbol{X}$) based on the consistency equations using the procedure:
     i. If $(t_{1d}, t_{2d}) = (t_{1b}, t_{2b})$, set $c_{t_{1b},t_{2b}}$ in row d
     ii. For $T > k > j > i > 1$,
         if $(t_{1d}, t_{2d}) = (i,k)$ & $c_{j,k} = 0$, replace $c_{i,j} = 1$
         if $(t_{1d}, t_{2d}) = (i,k)$ & $c_{i,j} = 1$, replace $c_{j,k} = 1$
     iii. For $T > k > j > i > 1$,
         if $(t_{1d}, t_{2d}) = (i,j)$ & $c_{j,k} = 0$, replace $c_{i,k} = 1$
         if $(t_{1d}, t_{2d}) = (i,j)$ & $c_{i,k} = 1$, replace $c_{j,k} = -1$
     iv. For $T > k > j > i > 1$,
         if $(t_{1d}, t_{2d}) = (j,k)$ & $c_{i,k} = 0$, replace $c_{i,j} = -1$
         if $(t_{1d}, t_{2d}) = (j,k)$ & $c_{i,j} = -1$, replace $c_{i,k} = 1$
     v. Identify the comparisons that need more than three non-zero entries in the design matrix
     vi. Create dummy variables that represent functions of two basic comparisons
     vii. Use the steps (i-iv) to define the functional dummy variables
     viii. Express the variables in (vi) via the basic comparisons
     ix. Repeat the steps (vi-viii) until all comparisons are expressed through the basic comparions

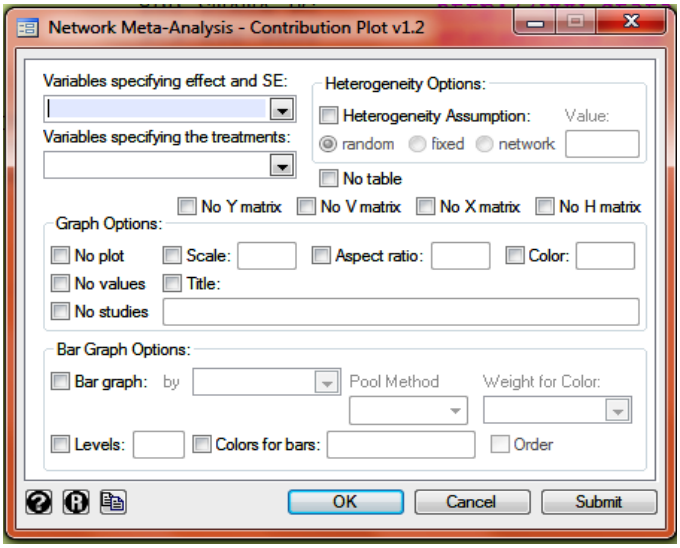

Figure 2.14. Dialog box of the `netweight` command in STATA.

The `intervalplot` command (**Figure 2.11**) is suitable to produce a forest plot showing the estimated relative effects with their CI. I allowed plotting the PI simultaneously with the CI via the option `predictions` (**Figure 2.4**), while the option `eform` can be used to produce the plot in logarithmic scale. Note that the command should be used directly after using the `mvmeta` command specifying the option `mvmetaresults`.

I enabled drawing a forest plot including the estimated inconsistency factors from the loop-specific approach in STATA (**Figure 2.5**) with the `ifplot` command (**Figure 2.12**), which identifies all triangular and quadratic loops in the network (**Box 2.1**). The three different assumptions for heterogeneity (comparison-specific, loop-specific [default], network-specific) can be specified in the option `tau2()`. To produce the graph in logarithmic scale plotting the $exp(w_l^{XY})$, the `eform` option can be employed again. Different estimators for heterogeneity are possible when loop-specific heterogeneity is assumed.

For drawing rankograms and cumulative ranking probability curves, I developed the `sucra` command. The option `mvmetaresults` is necessary to specify that the ranking probabilities have been derived from the `mvmeta` command (other input formats are also possible). The relative ranking results from two different analyses can be plotted jointly using the option `compare()`. Note that the output of the

command provides also the mean ranks on the top of the SUCRA percentages. **Figure 2.15** shows also the dialog box of the command.

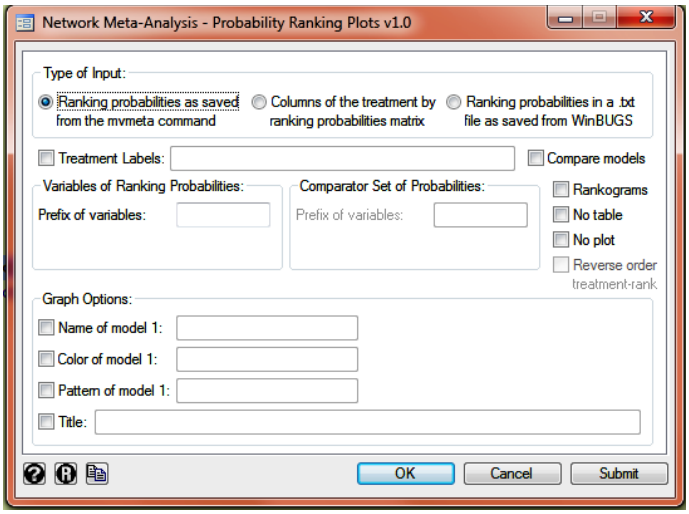

Figure 2.15. Dialog box of the `sucra` command in STATA.

I implemented the use of the MDS approach for treatment ranking in the `mdsrank` command, which can produce the scatterplot of **Figure 2.8**. The option `best(max|max)` specifies whether larger [default] or smaller values of the MDS dimension correspond to better treatments with respect to the studied outcome. The dialog box of the command is shown in **Figure 2.16**.

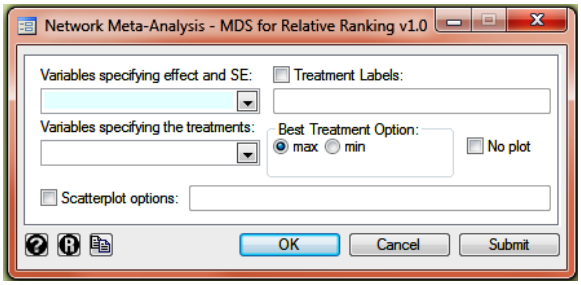

Figure 2.16. Dialog box of the `mdsrank` command in STATA.

The last command in my package is the `clusterank` command, which is suitable to create a two-dimensional ranking plot for drawing conclusions based on two outcomes allowing for grouping the treatments with clustering methods (**Figure 2.9**). These clustering methods are described in **Box 2.4**. Again the option

`best(max|max)` can be added to show whether larger [default] of smaller values in both dimensions represent more beneficial treatments. The command can be used also via the dialog of **Figure 2.17**.

Box 2.4. Measures used in the `clusterank` command to choose the metric and linkage method as well as the optimal number of clusters.

The competing treatments of a network are grouped using a hierarchical agglomerative clustering method. Different metrics (Euclidean, squared Euclidean, absolute-value distance, etc.) and linkage methods (single, average, weighted, complete, ward, centroid, median) are evaluated (1). The choice of the appropriate metric and linkage criterion is driven from the cophenetic correlation coefficient, which measures how faithfully the output dendrogram represents the dissimilarities among observations (2). The optimal level of dendrogram and the optimal number of resulting partitions are chosen using an internal cluster validation measure based on a value of 'clustering gain'. This measure has been designed to have a maximum value when intra-cluster similarity is maximized and inter-cluster similarity is eliminated (3). More details on these methods are available in the papers:

1. Kaufman L, Rousseeuw PJ. Finding groups in data: An introduction to cluster analysis. New York: Wiley, 1990.
2. Handl J, Knowles J, Kell DB. Computational cluster validation in post-genomic data analysis. Bioinformatics 2005; 21: 3201-3212.
3. Jung Y, Park H, Du D. A decision criterion for the optimal number of clusters in hierarchical clustering. J Global Optimazation 2003; 25: 91-111.

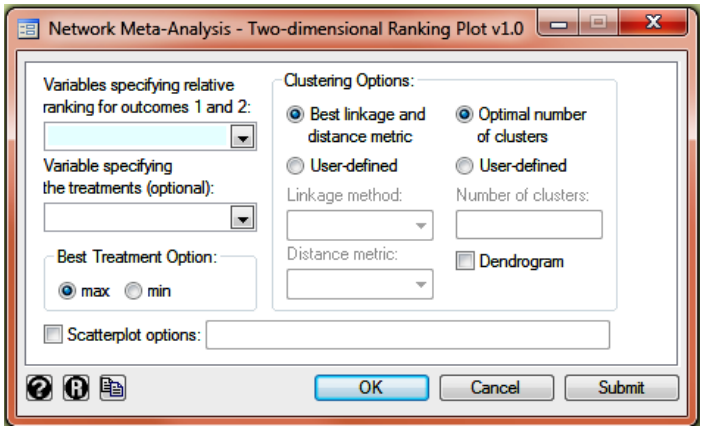

Figure 2.17. Dialog box of the `clusterank` command in STATA.

2.6 Discussion

The continuing development of new methodologies for NMA has resolved several issues of the early approaches and established it as a useful evidence synthesis tool

in comparative effectiveness research. However, the debate for the potential of NMA to yield valid findings and inform decision-making is still on-going (83). This skepticism around NMA may be partly explained by the lack of user-friendly software environment that would simplify the procedure of the analysis and would make it more comprehensible to a wider audience.

The availability of new software options implemented in conventional statistical packages (e.g. STATA or R), which are used routinely in the conduct of pairwise meta-analysis, will possibly bring researchers a step further to demystify this complex statistical tool. In addition, the recent publication of some well-described application (84,85) and tutorial papers (10,11,86) might familiarize researchers with the pitfalls of each approach and the foundamental issues related to the validity and interpretation of NMA results.

3 Characteristics of published networks of interventions

3.1 Introduction

The various approaches to fit a NMA are essentially equivalent; however certain network characteristics sometimes may restrict the choice of the method of analysis. For example, networks that include multi-arm studies require the use of multivariate methods and until recently could be analyzed only as a hierarchical model fitted in Bayesian environment. Also, inconsistency can be assessed statistically only in networks with closed loops; in star-shaped networks (where all treatments are compared directly only with a common comparator intervention) researchers might be able only to assess intransitivity relying on clinical and epidemiological criteria (e.g. comparability of populations across comparisons). Hence, it is interesting to explore how the new methodologies for performing NMA and evaluating assumptions (see Chapter 2) have been employed by researchers undertaking NMA and indirect comparisons.

The fact that NMA is a relatively new statistical tool implies the need for several empirical and simulation studies to investigate the properties of each approach. Such studies should be ideally planned based on the characteristics of real networks of interventions. An overview of the characteristics of published networks of interventions is a useful resource of information for methodologists that aim to update the current knowledge on appraising NMA methods.

This Chapter focuses on a description of a database I compiled, which consists of 88 published networks of interventions. Sections 3.2 and 3.3 present the selection process of the networks including search strategy and eligibility criteria as well as the descriptive analyses. Then, Section 3.4 shows the results and describes the

most important size, clinical and methodological characteristics of the identified networks.

3.2 Selection process of published networks

I searched for networks of randomized controlled trials (RCTs) published until March 2011 in Medline via PubMed using the following code: "(network OR mixed treatment* OR multiple treatment* OR mixed comparison* OR indirect comparison* OR umbrella OR simultaneous comparison*) AND (meta-analysis)". Networks needed to include at least three competing interventions and been analyzed using a valid indirect comparison method or NMA; hence I excluded networks analyzed with the 'naive approach' (i.e. comparison of single arms pooled across studies) (31,87). Networks that did not report the method of analysis were considered eligible if the presented indirect estimates were in agreement with that derived from the Bucher method (9). Meta-analyses including observational or diagnostic test accuracy studies were excluded. Finally, I excluded networks in which the number of studies did not exceed the number of competing treatments to ensure that enough information would be available for each comparison (i.e. relative effects would be estimable).

3.3 Data extraction & analyses

For every included network, information was extracted on the name of first author, journal, year of publication, primary outcome, number of included studies, competing interventions and control intervention (i.e. an inactive intervention or standard care). Arm-level outcome data were extracted, when they were available, or study-level otherwise for the primary outcome of each network. For networks that did not define clearly the primary outcome, the outcome reported first in the

manuscript was taken as such. I requested the required outcome data from the authors of the networks, when they were not available in the publication.

I classified the networks into star-shaped and full (i.e. with at least one closed loop) networks. The network (primary) outcomes were considered as being harmful or beneficial as well as objective, semi-objective or subjective according to the criteria described in Turner et al. (49). The included networks were further classified into three categories depending on the type of treatment comparison. Following Turner et al., these categories of comparisons were defined using the scheme described below:

- *Pharmacological vs. control*. Networks including only pharmacological treatments as well as a control intervention.
- *Pharmacological vs. pharmacological*. Networks including only pharmacological treatments and no 'obvious' control intervention.
- *Non-pharmacological vs. any intervention*. Networks with at least one non-pharmacological intervention.

The modeling approach to NMA and the method to evaluate inconsistency were also recorded. All available methods are described in detail in Sections 2.2.2, 2.3.2 and 2.3.3. Finally, data extraction included information on the type of outcome data (dichotomous, continuous, time-to-event, rates) and the effect size that was used to synthesize them (odds ratio (OR), risk ratio (RR), risk difference, standardized mean difference (SMD), mean difference (MD), ratio of means, hazard ratio (HR)).

Descriptive and frequency statistics were estimated for publication, size, clinical and methodological characteristics. Analyses considered star and full networks both separately and jointly, when possible. The prevalence of each modeling approach for indirect comparisons or NMA and evaluation of inconsistency was estimated exploring its possible association with size characteristics (number of included studies, treatments, participants, etc.). Continuous characteristics are reported in terms of median numbers with the respective interquartile range (IQR). Subsequently, I conducted two empirical analyses that used this network

collection; the first investigated the impact of four RoB items (random sequence generation, allocation concealment, blinding of participants, blinding of outcome assessors) on treatment effects and the second the impact of study precision. These studies are described explicitly in Sections 5.4 and 5.5 respectively. Note that three-treatment networks were not included in the descriptive analyses since the characteristics of such networks have been explored in previous studies (27,88). However, they were considered eligible for the empirical analyses.

3.4 Results

3.4.1 Identified networks

The search identified 890 relevant abstracts; 276 of them were assessed as potentially eligible and their full articles were screened. After excluding duplicated publications, 145 networks met all the inclusion criteria; 102 (70%) were full and 43 (30%) star networks. Finally, 20 star and 68 full networks, for which outcome data were obtained, were included in the database. The full selection process of the networks is summarized in the flow chart of **Figure 3.1**. **Figure 3.2** and **Figure 3.3** show the increase in network publications over time and the seven most popular journals in the area of NMA respectively.

3.4.2 Size characteristics of networks

In a total of 88 networks, the median number of studies and treatments per network were 21.5 (IQR:13-37.5) and 6 (IQR:4-9) respectively. The identified full networks included, on average, more studies and competing interventions than the star networks (**Table 3.1**). In a subset of 82 networks (that reported the number of participants in each study) the median sample size per network was 7729

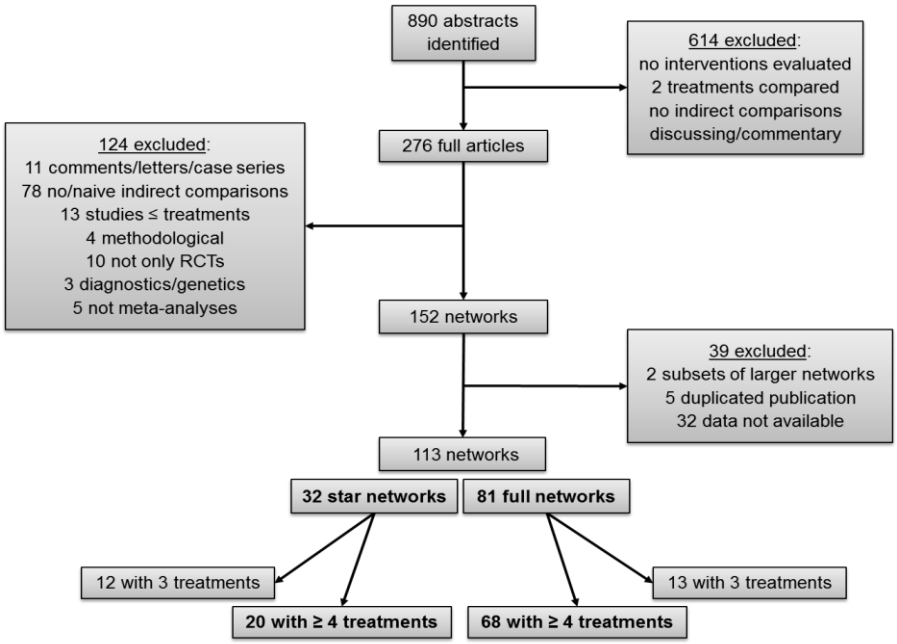

Figure 3.1. Selection process of published networks of interventions.

(IQR:3,043-24,987), while in 80 networks (that reported the number of participants in each arm) the median sample size per comparison was 577 (IQR:208-1,707). Star networks appeared overall being more compact than full networks including fewer direct comparisons, but with larger number of studies and participants per comparison.

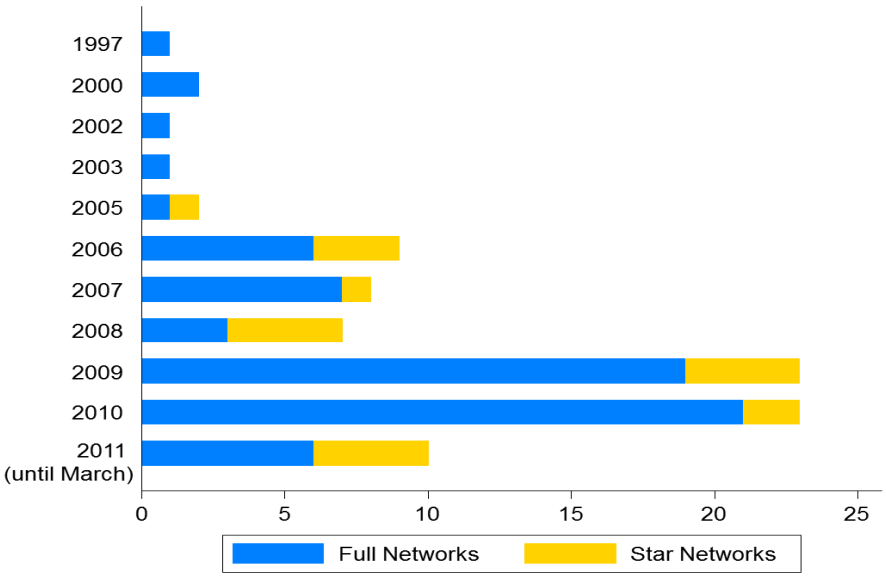

Figure 3.2. Number of eligible published networks by year.

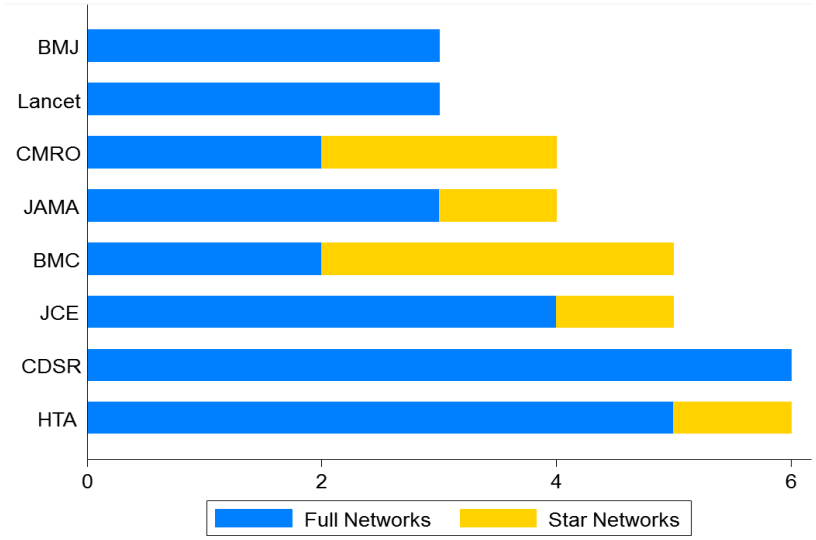

Figure 3.3. Number of eligible published networks by journal. (BMJ=British Medical Journal, CMRO=Current Medical Research and Opinion, JAMA= Journal of the American Medical Association, BMC= BioMed Central, JCE= Journal of Clinical Epidemiology, CDSR= Cochrane Database of Systematic Reviews, HTA= Health Technology Assessment)

The presence of multi-arm studies was also examined in the 68 full networks. At least one three-arm study was included in 56 (82%) of them and at least one four-arm trial in 18 (26%). The total number of closed loops in these networks was 426 and the median number of loops per network was 4 (IQR:2-9).

Table 3.1. Size characteristics of the networks. Medians with interquartile ranges (in parentheses) are reported. Numbers in square brackets show the number of networks reporting the respective information.

Characteristics	Full networks	Star networks	Full & star networks
Number of studies per network	22 (13-38.5) [68]	18.5 (12.5-29) [20]	21.5 (13-37.5) [88]
Number of treatments per network	6 (4-9) [68]	5 (4-7) [20]	6 (4-9) [88]
Sample size per network	8,491 (4,587–27,659) [62]	2,995 (1,829–12,499) [20]	7,729 (3,043–24,987) [82]
Sample size per comparison	576 (185–1,785) [61]	600 (366–1,217) [19]	577 (208–1,707) [80]
Number of studies per comparison	2 (1–4) [68]	3 (2–6) [20]	2 (1–4) [88]

3.4.3 Clinical characteristics of networks

Out of the total 88 included networks, more than one in two networks (58 networks, 66%) had a harmful primary outcome, while in nearly half of the networks (36 networks, 41%) the primary outcome was subjective (**Table 3.2**). More frequently networks (60 networks, 68%) measured dichotomous primary outcomes and less often continuous (19 networks, 22%) or other types of outcomes.

Regarding the type of treatment comparison, 61 networks (69%) were classified in the category of pharmacological vs. control followed by 14 (16%) networks in the group of non-pharmacological vs. any intervention. In 48 (55%) networks the control intervention was placebo and in 7 (8%) an alternative non-active intervention.

Table 3.2. Clinical characteristics of the networks. Total numbers and the respective percentages are reported.

		Full networks	Star networks	Full & star networks
Type of outcome	*Objective*	20 (30%)	7 (35%)	27 (31%)
	Semi-objective	24 (35%)	1 (5%)	25 (28%)
	Subjective	24 (35%)	12 (60%)	36 (41%)
Type of data	*Dichotomous*	45 (66%)	15 (75%)	60 (68%)
	Continuous	16 (24%)	3 (15%)	19 (22%)
	Time-to-event	5 (7%)	2 (10%)	7 (8%)
	Rate	2 (3%)	0	2 (2%)
Effect measure	*Odds ratio*	29 (42%)	2 (10%)	31 (35%)
	Risk ratio	17 (25%)	13 (65%)	30 (34%)
	Risk difference	1 (1%)	0	1 (1%)
	Hazard ratio	5 (7%)	2 (10%)	7 (8%)
	Rate ratio	2 (3%)	0	2 (2%)
	Mean Difference	11 (16%)	2 (10%)	13 (15%)
	Standardized mean difference	5 (7%)	1 (5%)	6 (7%)

3.4.4 Methodological characteristics of networks

The most popular approach to NMA was the hierarchical model fitted within a Bayesian environment (43 networks, 49%) followed by the meta-regression

approach (assuming univariate distributions) (17 networks, 19%) and Bucher's method (16 networks, 18%) (see Section 2.2 for a description of the methods). The choice of the analysis method for indirect comparisons varied according to the network structure. More specifically, the hierarchical model was employed in 36 (53%) full networks and 7 (35%) star networks, the meta-regression approach in 15 (22%) full and 2 (30%) stars and the Bucher's method in 12 (18%) full and 4 (20%) star networks. There were also 13 (15%) networks (9% of the full and 35% of the stars) that did not report the method of analysis.

Figure 3.4 shows that the hierarchical model started being the most popular method for NMA since 2009, whereas the majority of networks published earlier employed the Bucher's method. However, the number of networks not reporting the method of analysis did not appear to be declining over years (e.g. 17% in 2007, 18% in the first three months of 2011). In addition, 41 full networks (60%) did not report whether and how the presence of inconsistency was assessed and 5 full networks (7%) used inappropriate methods for the evaluation of inconsistency. These findings raise concerns about the proper application of the methodologies in a respectable proportion of publications.

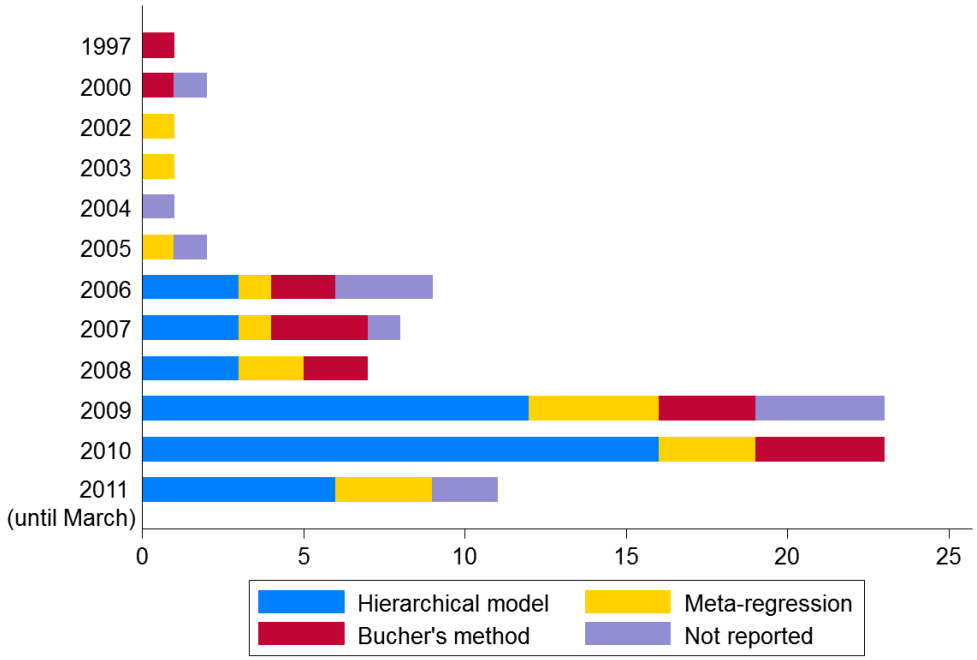

Figure 3.4. Number of eligible published networks by year and method of analysis.

3.5 Conclusions

This compilation of networks of interventions demonstrates that the advantages of NMA and the many available treatment options for almost any healthcare condition have increased its acceptance by the medical research community. The observed variation in the choice of the method of analysis can be partly explained by the stucture of the networks (e.g. differences between full and star-spahed networks).

The typical network of this database is a full network that includes 22 studies and compares 5 pharmacological treatments vs. a control intervention with respect to a dichotomous subjective outcome.

The findings of this network collection might be limited by the lack of a thorough search strategy; namely I used only one database (i.e. PubMed) and a not very sensitive search algorithm. However, the included networks cover a wide range of medical fields and the aforementioned systematic review process gave comparable results with two recently published systematic reviews that collected networks of interventions (19–21). An extended version of this database that included networks published until the end of 2012 yielded similar conclusions (22).

4 Methodology for network meta-regression & network meta-epidemiology

4.1 Introduction

Patient characteristics might differ across studies and/or pairwise comparisons challenging the assumptions of homogeneity (13) and transitivity (11,26) respectively. These discrepancies can be seen as differences in the distribution of potential effect modifiers within and across the pairwise comparisons in a network of interventions.

It is often the case that the estimated heterogeneity or inconsistency in a network of interventions can be explained using a *network meta-regression* model; that is a NMA model that incorporates one or more covariates. Standard meta-regression models for pairwise meta-analysis often lack of power to detect associations between treatment effects and study-level characteristics (17). An advantage of network meta-regression is the potential to borrow strength across the different comparisons, when such an assumption is reasonable; in this way the regression coefficients are estimated with increased power.

Meta-epidemiological methods rely on the comparability of the impact of effect modifiers across different meta-analyses (in terms of magnitude and direction) to overcome the issue of low power. In this way, a sizable amount of data is expected to be available allowing the adequate estimation of a relationship between treatment effects and several characteristics. A description of models that pool parameters across many pairwise meta-analyses can be found in Section 1.3. However, considering that these parameters are constant or even related over a range of meta-analyses from different clinical fields sometimes might not be a plausible assumption.

Data from networks of interventions usually involve a larger evidence base and may offer a promising alternative way to investigate the mechanisms of effect modification and bias in meta-epidemiological research; this methodology may be called *network meta-epidemiology* (89). Pooling parameters (e.g. regression coefficients) across many network meta-analyses can accommodate the assumption for these parameters being more similar within a network rather than across networks (89,90).

The following Section describes how the NMA models of Section 2.2.2 can be modified to incorporate one or more covariates as well as different assumptions for the regression coefficients that have been suggested in the literature. Then, Section 4.3 describes how the estimated coefficients can be pooled across several networks of interventions to derive an overall parameter of effect modification.

4.2 Models & assumptions for network meta-regression

4.2.1 Models for network meta-analysis incorporating covariates

The network meta-regression models are extensions of the standard NMA models that account for the impact of one or more covariates on the treatment effects estimates. More specifically, consider that p characteristics have been specified *a priori* as potential effect modifiers in a network of interventions. The extension of each NMA approach into a network meta-regression model is described below.

Network meta-regression model for the 'multivariate meta-regression' approach

When NMA is fitted as a multivariate meta-regression model the incorporation of covariates requires the extension of the design matrix $\boldsymbol{X}$. More specifically, according to Section 2.2.2 in a standard NMA model the entries of the design matrix express the comparison(s) being made in each sudy via the consistency

equations; hence the estimated coefficients from this model represent the network estimates of the relative effects. Network meta-regression aims further to estimate how these relative effects may change for different values of the characteristics of interest. This implies that an additional set of covariates should be included in the design matrix expressing these characteristics for every study in the network. In this way an additioanl set of coefficients can be estimated showing how each characteristic affects the relative effects. The extended design matrix that incorporates also information on p study-level characteristics that may act as effect modifiers is a matrix of dimension $\left[\sum_{i=1}^{S}(K_i-1)\right]\times[(T-1)\times(p+1)]$ denoted with $\boldsymbol{X}^*=\left(\boldsymbol{X},\boldsymbol{z_p}\right)$. The $\boldsymbol{z_p}=\left(\boldsymbol{z_{i1}},\dots,\boldsymbol{z_{ip}}\right)^T$ is the sub-matrix that contains for each study i $(i=1,\dots,S)$ the values of the covariates

$$\boldsymbol{z_{ip}}=\begin{pmatrix} z_{i11}^{1},\dots,z_{i11}^{T-1},\dots,z_{i1p}^{1},\dots,z_{i1p}^{T-1} \\ \vdots \\ z_{i(K_i-1)1}^{1},\dots,z_{i(K_i-1)1}^{T-1},\dots,z_{i(K_i-1)p}^{1},\dots,z_{i(K_i-1)p}^{T-1} \end{pmatrix} \tag{4.1}$$

(K_i the number of arms in study i) representing the p characteristics for each basic comparison. Note that for the basic comparisons not reported in the study i (with $x_{i(K_i-1)b}=0,\ b=1,\dots,T-1$ the basic comparisons), the respective covariates $z_{i(K_i-1)j}^{b}$ are zero as well for every $j=1,\dots,p$. Then, Equation (1.8) becomes

$$\boldsymbol{y}=\boldsymbol{X}^*\begin{pmatrix}\boldsymbol{\mu}^*\\ \boldsymbol{\beta}\end{pmatrix}+\boldsymbol{\varepsilon}^*+\boldsymbol{\delta}^*$$

where $\boldsymbol{\mu}^*=(\mu_1^*,\dots,\mu_{T-1}^*)^T$ is the vector of the summary effects when all p covariates are equal to zero and $\boldsymbol{\beta}=\left(\boldsymbol{\beta_1},\dots,\boldsymbol{\beta_p}\right)^T$. Each vector $\boldsymbol{\beta_j}=\left(\beta_j^1,\dots,\beta_j^{T-1}\right)^T$ contains the comparison-specific coefficients corresponding to the characteristic j. Also, $\boldsymbol{\delta}^*=(\boldsymbol{\delta_1^*},\dots,\boldsymbol{\delta_S^*})$ and $\boldsymbol{\varepsilon}^*=(\boldsymbol{\varepsilon_1^*},\dots,\boldsymbol{\varepsilon_S^*})$ with $\boldsymbol{\varepsilon_i^*}\sim N(\boldsymbol{0},\boldsymbol{s_i^2})$.

The network estimates $\widehat{\boldsymbol{\mu}}^*$ derived from this model account for the impact of the p characteristics. The random effects $\boldsymbol{\delta_i^*}$ are assumed again normally distributed $\boldsymbol{\delta_i^*}\sim N(\boldsymbol{0},\boldsymbol{\tau}^{*2})$ with $\boldsymbol{\tau}^{*2}=diag(\boldsymbol{\tau_i^{*2}})$ and

$$\boldsymbol{\tau_i^{*2}}=\begin{pmatrix}\tau_1^{*2} & \cdots & cov\left(\theta_{i1}^*,\theta_{i(K_i-1)}^*\right)\\ \vdots & \ddots & \vdots \\ cov\left(\theta_{i1}^*,\theta_{i(K_i-1)}^*\right) & \cdots & \tau_{K_i-1}^{*2}\end{pmatrix} \tag{4.2}$$

When differences in the p characteristics across studies explain the between-study variance, it is expected that the estimated heterogeneity of the network meta-regression model would be smaller than that of the standard NMA model (i.e. $\tau^{*2} < \tau^2$ or $\tau_{XY}^{*2} < \tau_{XY}^2$ for $X, Y = 1, \dots, T, X \neq Y$).

Network meta-regression as a hierarchical model

Fitting network meta-regression as a hierarchical model requires the addition of a regression term in Equation (2.10) assuming that

$$\boldsymbol{y_i} \sim N(\boldsymbol{\theta_i}, \boldsymbol{s_i^2})$$

and

$$\boldsymbol{\theta_i} = \boldsymbol{\theta_i^*} + \boldsymbol{z_{ip}\beta_i} \tag{4.3}$$

where $\boldsymbol{\theta_i} = \left(\theta_{i1}, \dots, \theta_{iK_i-1}\right)^T$, $\boldsymbol{\beta_i} = \left(\boldsymbol{\beta_{i1}}, \dots, \boldsymbol{\beta_{ip}}\right)^T$, $\boldsymbol{\beta_{ij}} = \left(\beta_j^1, \dots, \beta_j^{K_i-1}\right)^T$ and $\boldsymbol{z_{ip}}$ is defined in (4.1) but here includes only the covariates that correspond to the $K_i - 1$ comparisons modelled for study i. Similarly, the subscript i in the vector of coefficients in the above Equation denotes that $\boldsymbol{\beta_i}$ contains the comparison-specific coefficients for the $K_i - 1$ comparisons. Equation (4.3) implies that every true underlying effect θ_{ik} $(k = 1, \dots, K_i - 1)$ of study i is different by β_j^k units compared to a study in which the covariate representing the characteristic j for comparison k is equal to 0.

The study-specific underlying effects $\boldsymbol{\theta_i^*} = \left(\theta_{i1}^*, \dots, \theta_{i(K_i-1)}^*\right)^T$ are assumed normally distributed

$$\boldsymbol{\theta_i^*} \sim N(\boldsymbol{\mu_i^*}, \boldsymbol{\tau_i^{*2}}) \tag{4.4}$$

with mean $\boldsymbol{\mu_i^*} = \left(\mu_1^*, \dots, \mu_{K_i-1}^*\right)^T$ the vector of the summary effects. The variance-covariance matrix $\boldsymbol{\tau_i^{*2}}$ is described by Equation (4.2).

Network meta-regression model for the multivariate meta-analysis approach

Using the multivariate meta-analysis approach, the network meta-regression model can be constructed similarly to the case of multivariate meta-regression

approach. The two models differ only in the definition of the basic parameters $b = 1, \dots, T-1$ expressed in $\boldsymbol{\mu}^*$ and the elements of the design matrix $\boldsymbol{X} = (\boldsymbol{X_1}, \dots, \boldsymbol{X_S})^T$. The definition of these matrices is described in Section 2.2.2. Under this approach, when x_{ikb} is a missing value (i.e. the basic comparison b is not reported in study i), z_{ikj}^b $(j = 1, \dots, p)$ will be missing as well.

The three approaches presented above for fitting a network meta-regression model are equivalent and the estimated coefficients are expected to be similar between them. The 'two-stage approach' is not considered in this Section for fitting network meta-regression, since it would probably give regression coefficients not comparable to those derived from the three above methods. This is because the coefficients would be estimated only at the first stage of the analysis as well as due to the fact that they would be both comparison-specific and design-specific.

Similarly to the case of pairwise meta-analysis, caution is needed in the choice of the p characteristics that define the $\boldsymbol{z_p}$ matrix. Due to the limited power of any meta-regression model only a small subset of the total number of potential effect modifiers should be included in the model. A rule of thumb is to allow one covariate for every 10 studies in a pairwise comparison (13). However, the power to detect associations between study-level characteristics and treatment effects also depends on the size of the studies and the amount of heterogeneity (17).

4.2.2 Assumptions for the regression coefficients in network meta-regression

The network meta-regression models presented above would give estimates for a set of $C = \binom{T}{2}$ coefficients β_j^{XY} for each characteristic j. Different assumptions can be employed for the comparison-specific coefficients of each j characteristic (91).

The plausibility of these assumptions in a network of interventions depends on the clinical setting and outcome as well as on the nature of each characteristic j.

For example, suppose that baseline severity is a potential effect modifier in a network comparing active treatments either with placebo or with each other. It might be reasonable to assume that differences in baseline severity across studies impact similarly in all comparisons of active treatments vs. placebo. This assumption can be reflected in a network meta-regression model by using *identical* or *exchangeable coefficients* for all these comparisons. On the other hand, one may believe that baseline severity affects in a different way head-to-head comparisons (with respect to magnitude and direction); in that case the respective coefficients for this type of comparisons could be assumed *independent* from the coefficients corresponding to comparisons of active treatments vs. placebo.

All possible assumptions about the regression coefficients are described in detail below.

Independent comparison-specific coefficients

The weaker assumption for the elements for the comparison-specific coefficients is assuming them as being fixed effects and totally independent, hence $\beta_j^{XY} \neq \beta_j^{YZ}$ for every $X, Y, Z = 1, \ldots, T$ with $X \neq Y \neq Z$. The drawback of this approach is that in the presence of very few studies for one or more comparisons the model would fail to estimate adequately the respective β_j^{XY} and the estimated coefficients would have extreme uncertainty.

Exchangeable comparison-specific coefficients

Alternatively, if it is clinically meaningful, all comparison-specific coefficients for the j study characteristic can be treated as random effects sharing a common distribution, hence

$$\beta_j^{XY} \sim N(B_j, \xi_j^2) \tag{4.5}$$

with mean B_j and between-comparison variance ξ_j^2. This exchangeability assumption implies that all β_j^{XY} are related but does not impose any further constraints on their estimation.

Consistent comparison-specific coefficients

The comparison-specific coefficients can be further restricted by forcing them to satisfy the consistency equations (Equation (2.5)), namely

$$\beta_j^{YZ} = \beta_j^{XZ} - \beta_j^{XY} \tag{4.6}$$

Although this assumption might be stronger than the previous two (i.e. independent or exchangeable coefficients), it is often considered as being more reasonable; this is because it reflects the belief that the same assumption (i.e. consistency) should underlie both relative effects and regression coefficients. Then, three further assumptions are possible for the *basic coefficients*; these are the coefficients corresponding to the basic parameters $\beta_j^b \equiv \beta_j^{Xt}$ where $t = 1, \dots, T$ with $t \neq X$:

- *Independent basic coefficients*

$$\beta_j^{b_1} \neq \beta_j^{b_2} \text{ for every } b_1, b_2 = 1, \dots, T-1 \text{ with } b_1 \neq b_2$$

- *Exchangeable basic coefficients*

$$\beta_j^b \sim N(B_j, \xi_j^2)$$

- *Identical basic coefficients*

$$\beta_j^{b_1} = \beta_j^{b_2} \text{ for every } b_1, b_2 = 1, \dots, T-1$$

Note that in this case Equation (4.6) gives that all non-basic coefficients are 0.

The assumption of the consistent coefficients has been exemplified recently by Achana et al. (92) for the application of network meta-regression with control group risk as covariate. The idea was that the 'unobserved' control group risk in studies that do not include a control arm is 'missing at random' (which is implied by the transitivity assumption) and it can be estimated using information from studies that include the control intervention assuming consistency in coefficients.

Identical comparison-specific coefficients

The last possible assumption that can be employed in a network meta-regression model is to assume that all comparison-specific coefficients are identical, which means that

$$\beta_j^{XY} = B_j \tag{4.7}$$

This assumption might be quite strong; however it can be plausible in cases where the impact of characteristic j does not depend on the treatments being compared, such as the impact of study design limitations (e.g. lack of blinding). The drawback of this model is that it does not allow for between-comparison variation. Nevertheless, it is advantageous compared to all previous assumptions with respect to the precision of estimating the coefficient B_j.

4.3 Models & assumptions for network meta-epidemiology

Network meta-epidemiology is the analogy to conventional meta-epidemiology (see Section 1.3) when the synthesis of parameters (e.g. regression coefficients) is performed across networks of interventions instead of pairwise meta-analyses. More specifically, consider that network meta-regression has been performed in N available networks for the same $j = 1, \dots, p$ characteristics using the methodology of Section 4.2. Then, from each network a coefficient B_j^n with $n = 1, \dots, N$ has been estimated and all these network-specific coefficients are assumed comparable with each other; that means that the impact of characteristic j on the treatment effects is assumed independent of the treatment comparison being made, the clinical condition of interest and the studied outcome. Note that deriving the parameters B_j^n requires that the comparison-specific coefficients β_j^{XY} within each network should be assumed either exchangeable (Equation (4.5)) or identical (Equation (4.7)). Otherwise each network would give a set of coefficients pertaining only to specific treatment comparisons and pooling these sets of coefficients would be meaningless.

Similarly to conventional meta-epidemiology, network meta-epidemiology can be performed either in one or in two stages (see Section 1.1), but here the unit of analysis is the network. Using a one-stage network meta-epidemiological model has the advantage of borrowing strength across networks. In this way, the estimated overall coefficient is usually estimated with increased precision compared to the two-stage analysis. The one-stage model is constructed as an extension of the network meta-regression models as described below.

4.3.1 Models synthesizing parameters across network meta-analyses

Network meta-epidemiology using the 'multivariate meta-regression' or the 'multivariate meta-analysis' approach

The synthesis of regression coefficients using any of the two approaches is performed by extending Equation (1.9) into

$$\begin{pmatrix} \boldsymbol{\theta}'_{\boldsymbol{n,1}} \\ \vdots \\ \boldsymbol{\theta}'_{\boldsymbol{n,S_n}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X_{n,1}, z_{n,ip}} \\ \vdots \\ \boldsymbol{X_{n,S_n}, z_{n,S_n p}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}'_{\boldsymbol{n}} \\ \boldsymbol{\beta}'_{\boldsymbol{n,1}} + B_1^{overall} \\ \vdots \\ \boldsymbol{\beta}'_{\boldsymbol{n,p}} + B_p^{overall} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\delta}'_{\boldsymbol{n,1}} \\ \vdots \\ \boldsymbol{\delta}'_{\boldsymbol{n,S_n}} \end{pmatrix} \tag{4.8}$$

where S_n is the total number of studies in each network n with $n = 1, \dots, N$ (N the total number of networks) and $B_j^{overall}$ $(j = 1, \dots, p)$ is the *overall coefficient* showing the impact of the j characteristic on treatment effects and is assumed to pertain to the entire network collection. A normal distribution is assumed for the vector of coefficients $\boldsymbol{\beta}'_{\boldsymbol{n}} = \left(\boldsymbol{\beta}'_{\boldsymbol{n,1}}, \dots, \boldsymbol{\beta}'_{\boldsymbol{n,p}}\right)$, namely

$$\boldsymbol{\beta}'_{\boldsymbol{n}} \sim N(\mathbf{0}, \boldsymbol{\omega}^2)$$

Note that the definition of the basic parameters expressed in $\boldsymbol{\mu}'_{\boldsymbol{n}}$ and the design matrix $\boldsymbol{X_n} = \left(\boldsymbol{X_{n,1}}, \dots, \boldsymbol{X_{n,S_n}}\right)$ depends on the choice between the two approaches (see Section 2.2.2).

Network meta-epidemiology as a hierarchical model

Alternatively, the network meta-regression hierarchical model can be extended to allow for one additional hierarchy. In this case Equations (4.3) and (4.4) would be

$$\boldsymbol{\theta_{n,i}} = \boldsymbol{\theta'_{n,i}} + \boldsymbol{z_{n,ip}} \times \boldsymbol{\beta'_n}$$

$$\boldsymbol{\theta'_{n,i}} \sim N(\boldsymbol{\mu'_{n,i}}, \boldsymbol{\tau'^2_{n,i}})$$

with

$$\boldsymbol{\beta'_n} \sim N(\boldsymbol{B^{overall}}, \boldsymbol{\omega}^2)$$

The vector $\boldsymbol{B^{overall}} = \left(B_1^{overall}, \dots, B_p^{overall}\right)$ contains the overall coefficients for the p characteristics and $\boldsymbol{\omega}^2$ is the variance-covariance matrix of these coefficients.

Both approaches can assume that the network-specific coefficients are fixed and identical by setting $\boldsymbol{\omega}^2 = \mathbf{0}$.

As already mentioned in Section 1.3 for conventional meta-epidemiology, network meta-epidemiology can result in valuable benefit in the precision of coefficients only when a small number of characteristics (usually only 1 or 2) are included in the model.

4.4 Discussion

The presence of heterogeneity and/or inconsistency in a network of interventions raises concerns for the validity and applicability of the findings. The exploration of the possible sources for this observed extra variability within and across the pairwise comparisons should consider several patient and study characteristics that may act as effect modifiers.

The impact of these characteristics on the treatment effects can be explored via network meta-regression and different assumptions for the coefficients can be employed. The choice of the model assumption should be based primarily on clinical criteria; considering, for example, whether it is clinically meaningful to assume a common regression coefficient across all comparisons in the network. On

the other hand, computational reasons may imply the use of strong assumptions since for comparisons with very few studies the regression coefficients might not be estimable.

To date some empirical and simulation studies have been conducted to investigate the properties of the methodology for network meta-analysis (19,20,22,27,65,88,89,93,94). However, designing simulation studies based on networks of interventions is a computationally intensive procedure and only three-treatment networks have been simulated so far. Network meta-epidemiology can be considered as a valuable alternative to investigate several factors that might affect the results of NMA under certain conditions.

The exploration of heterogeneity and inconsistency can be an important source of imformation for decision-making. Protocols for NMA should consider network meta-regression as an additional analysis to investigate the impact of potential effect modifiers. However, researchers should be aware of the lack of power of the methodology and interpret the results always with a clinical insight and in light of the empirical evidence.

5 Assessing the impact of study characteristics in network meta-analysis results

5.1 Introduction

Several empirical studies have examined the possible association between the estimated treatment effects and the size of the trials suggesting that small studies tend to give larger effect estimates than do larger studies (95–98). Such a phenomenon, which is usually called *small-study effects*, is quite frequent in evidence synthesis and can be caused by several reasons.

A possible explanation is that small studies without significant treatment effects are less likely to be published, causing *publication bias*. An associated to publication bias reason for small-study effects, especially for secondary or safety outcomes is *selective outcome reporting* (i.e. selectively presenting only the significant results derived from a study). Genuine heterogeneity between small and large trials might also play a role (99,100); for example small studies are more likely to recruit high-risk patients that may benefit more from the treatment (101). In addition, study size (measured by sample size or precision) is often considered as a proxy for the quality of trials and any discrepancies between smaller and larger trials may be attributable to differences in study design. For instance, the lack of appropriate allocation concealment and blinding have been found to be associated with effect size in individual trials and meta-analyses (102–105).

For pairwise meta-analysis several regression-based methods have been suggested for investigating the association between treatment effect estimates and study size, which have been evaluated in empirical and simulation studies (96,106–108). These methods model the relationship between effect sizes and a measure of their precision, such as the variance, the standard error or a function of them. However,

these approaches are limited, as any meta-regression model is, by the lack of power to detect existing associations in the presence of few studies and substantial heterogeneity (17,109).

The presence of small-study effects in NMA can be assessed graphically by extending the standard funnel plot to allow the incorporation of multiple treatment comparisons (45,110). Network meta-regression models (see Section 4.2) can be used also to extend the statistical models accounting for small-study effects into the context of NMA. Models that account for the impact of other study characteristics on NMA results have been suggested in the literature as well (90,111–113). An important fact of the network meta-regression models that aim to control for potentially biased treatment effects is that they require making assumptions about the directionality of bias. This may be less problematic when analysing star-shaped networks; in these networks the common comparator intervention is usually an inactive or old intervention and it is not expected to be favored when bias is present.

The aim of this Chapter is to investigate whether certain characteristics of study design impact on the treatment effects estimated from NMA. The following Section describes a new graphical tool that I developed, known also as *comparison-adjusted funnel plot*, for assessing the presence of small-study effects in a network of interventions simultaneously for all direct comparisons. In Section 5.3, I provide a framework on how the network meta-regression models can incorporate different assumptions for the direction of potential bias due to differences in study design. In Sections 5.4 and 5.5, I use the methodology of network meta-epidemiology to evaluate the effect of four RoB items (random sequence generation, allocation concealment, blinding of participants, blinding of outcome assessors) and study precision in a collection of star-shaped networks (part of the database described in Chapter 3); the advantage of including only star network is that they do not require making strong assumptions about directionality of bias (89). Finally, Section 5.6 presents a case study, in which I explored the presence of small-study effects in two full networks (110).

5.2 A new graphical tool for assessing small-study effects in network meta-analysis

The presence of small-study effects in a pairwise meta-analysis is commonly assessed graphically using funnel plots, which are scatterplots of the observed relative effects in studies versus a measure of study precision. When the points in a funnel plot lie symmetrically around the summary effect, the assumption of no differences between small and large studies is likely to hold.

In a NMA that involves multiple treatment comparisons, putting all studies together in a single funnel plot is not helpful since there are multiple summary effects. Studies within each pairwise comparison form a comparison-specific reference line of symmetry that represents the respective summary effect. Hence, using conventional funnel plots, symmetry could only be judged separately for each available direct comparison. However, very often there are few studies per comparison (e.g. less than 10) and asymmetry cannot be assessed adequately. The inclusion of all studies of a NMA in a funnel plot needs somehow to account for the fact that different set of studies evaluate different treatment comparisons. To resolve this issue, I developed the comparison-adjusted funnel plot, in which the horizontal axis is modified to contain the value $(y_{iXY} - \hat{\mu}^{D}_{XY})$ for every study i in the network comparing any treatments X and Y (where y_{iXY} the observed relative effect in study i) (45,110). Similarly to the standard funnel plot, the direct estimate $\hat{\mu}^{D}_{XY}$ should more appropriately be derived from the fixed effect model; that is because the random effects model gives relatively more weight to small studies, which is undesirable when small-study effects are likely to operate.

It is important to note that the comparison-adjusted funnel plot can yield meaningful conclusions only if all comparisons have been defined in a consistent direction; for example active treatment vs. inactive, newer treatment vs. older, sponsored treatment vs. non-sponsored, etc. Then, a symmetrical plot around the zero line suggests the absence of small-study effects in the network.

Figure 5.1 shows an example comparison-adjusted funnel plot for the rheumatoid arthritis network described in Section 2.4 (64). This is a star-shaped network and all comparisons in the graph have been estimated as placebo vs. active treatment. The graph indicates a tendency of small studies to show the active treatments more effective than the respective summary effect for each pairwise comparison (since $y_{iXY} - \hat{\mu}_{XY} < 0$ for small studies).

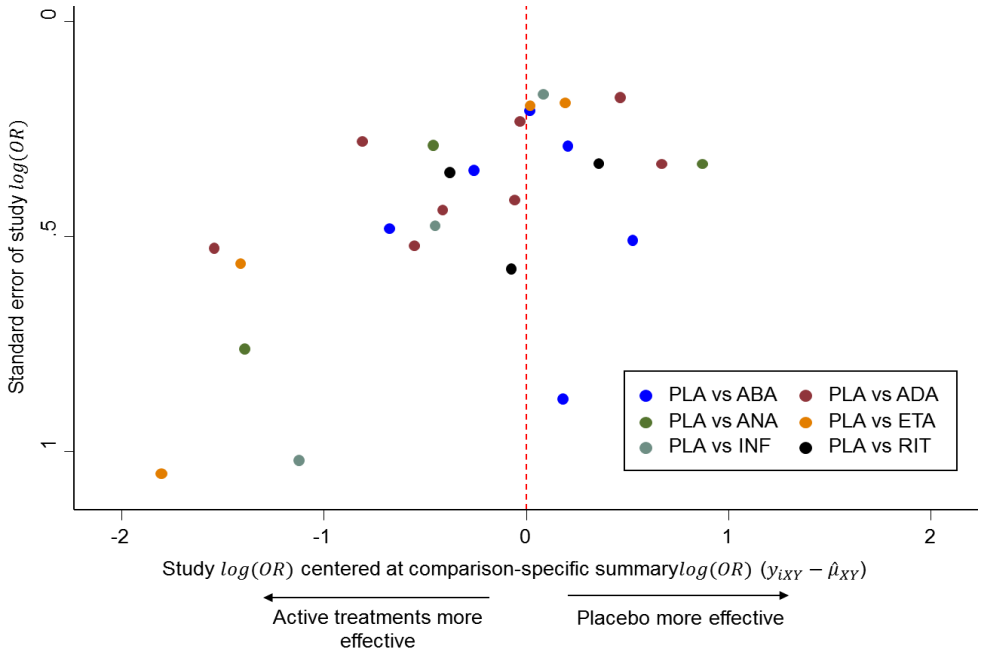

Figure 5.1. Comparison-adjusted funnel plot of a network network comparing the effectiveness of six biologic agents and placebo for rheumatoid arthritis. The dashed red line represents the null hypothesis that the observed effects in studies and the respective summary effects are in agreement. Points with different colors correspond to different pairwise comparisons.

I implemented the comparison-adjusted funnel plot can in STATA via the `netfunnel` command, which is included in the `network_graphs` package (see also Section 2.5.3). By default, the command estimates the observed effects for all comparisons as 'treatment earlier in alphabetical or numerical order vs. treatment later in the order'. This means that researchers should consider this direction when they assign names or codes to the competing treatments. Alternatively, adding the option `noalphabetical` specifies that the observed relative effects have been estimated *a priori* in a consistent direction; in this case the relative effects are not rearranged based on the treatment names. Also, the option `bycomparison` can be

used to present the different comparisons with different colors. Note that the command can be used also via a dialog box (**Figure 5.2**).

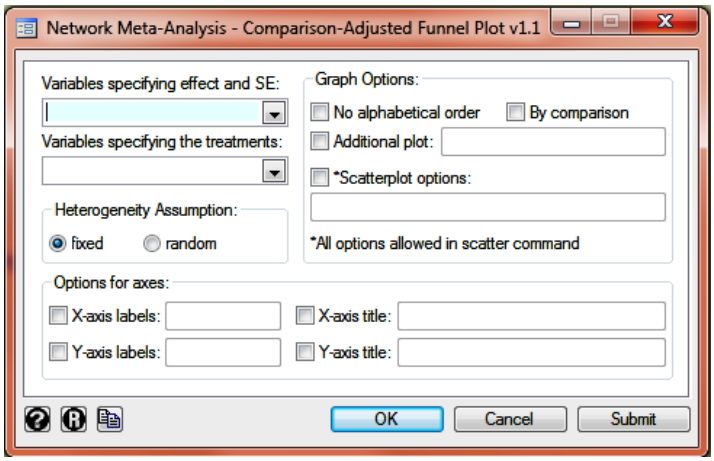

Figure 5.2. Dialog box of the `netfunnel` command in STATA.

5.3 Modeling the direction of potential bias in network meta-regression

5.3.1 Assumptions underlying network meta-regression when accounting for the impact of small-study effects

To develop all models provided in this Section, I rely on the assumption that the mechanism causing small-study effects is the same across treatment comparisons (110). This assumption implies that:

- If small-study effects are the consequence of differences in quality or, in general, due to clinical and methodological heterogeneity between smaller and larger studies, this possibly applies to any trial comparing a control

intervention (that is no treatment, placebo or standard care) versus an active (experimental) intervention.

- Similarly, when small-study effects are explained by publication bias or selective outcome reporting, this is likely to affect an entire research field exaggerating the effect of experimental interventions compared with controls. The magnitude of bias, though, might be different depending on the experimental comparator treatment.

Applying the above scenarios in the case of head-to-head trials requires stronger but reasonable assumptions; for example newer or sponsored treatments are often expected to be favored when compared with older or non-sponsored treatments respectively.

A major advantage of these assumptions is that the impact of study design characteristics (e.g. study size and RoB items) can be assessed with increased precision within a network of interventions rather than by looking at each pairwise treatment comparison separately.

5.3.2 Possible assumptions about the direction of potential bias

An overview of the network meta-regression models and the possible assumptions for the comparison-specific coefficients is available in Section 4.2.

The presence of multiple treatment comparisons in a network of trials implies the need to distinguish in each comparison the 'favored' arm from those that are not favored to ensure a meaningful interpretation of the estimated coefficients. For example, suppose a harmful dichotomous outcome. Drawing a funnel plot with a consistent direction for all effect sizes (e.g. favored vs. non-favored intervention) can reveal whether the summary effect indeed seems to exaggerate the effect of the treatment assumed favored; in that case missing studies are expected to lie on the right hand side of the graph. To make this distinction between the favored and non-favored interventions, I use a (dummy) *direction variable* as described below.

Note that in this Chapter all presented and applied network meta-regression models include only one covariate and follow the approach of the hierarchical model; this means that $p = 1$ and $\boldsymbol{z_{ip}} = \boldsymbol{z_{i1}} \equiv \boldsymbol{z_i}$. I write each covariate z_{iXY} in this matrix as

$$z_{iXY} = z^*_{iXY} \times d_{iXY} \tag{5.1}$$

where z^*_{iXY} is the part of the covariate that represents the study characteristic of interest (e.g. a measure of study precision) and d_{iXY} is the direction variable that takes values 1, 0 or -1 depending on the type of outcome and on whether treatment X or Y is expected to be favored by the effect of z^*_{iXY}. More specifically, for a harmful outcome

$$d_{iXY} = \begin{cases} 1, & Y \text{ is expected to be favored} \\ -1, & X \text{ is expected to be favored} \end{cases}$$

while for beneficial outcomes

$$d_{iXY} = \begin{cases} 1, & X \text{ is expected to be favored} \\ -1, & Y \text{ is expected to be favored} \end{cases}$$

The definition of d_{iXY} reflects the prior belief regarding which is the favored treatment in every comparison.

On the top of giving fixed values to the direction variable, I also employ a probabilistic approach to allows for some level of uncertainty in the prior belief about the favored treatment (112). In this case d_{iXY} would be 1 or 0 with some unknown probability P_{XY} and can be estimated as

$$d_{iXY} = \begin{cases} b_{XY}, & X \text{ is expected to be favored} \\ -b_{XY}, & Y \text{ is expected to be favored} \end{cases}$$

with

$$b_{XY} \sim Bernoulli(P_{XY})$$

where P_{XY} is the probability of the treatment assumed to be favored being indeed favored.

I combine both the *fixed* and the *probabilistic direction approaches* with several assumptions about the directionality of bias. These include the following scenarios:

- *The covariate z^*_{iXY} only impacts comparisons of control versus active (experimental) interventions ('active-favored').* This scenario might be more plausible when

small-study effects are caused either by publication bias or by differences in study quality; in both cases the effect of active interventions is expected to be exaggerated.

- *The covariate z^*_{iXY} impacts every comparison favoring the newer treatments ('new-favored').* In this scenario, the potential for small-study effects is allowed also in the head-to-head comparisons, where the newer treatment between X and Y is expected to be favored. The definition of newer and older treatments can be based on the date of licensing or on the publication of the first trial evaluating each treatment. The control intervention (no treatment, placebo or standard care) is considered here the older treatment in the network. In case that treatments X and Y are equally new, then $d_{iXY} = 0$ (i.e. bias is not likely to operate in this comparison).
- *The covariate z^*_{iXY} impacts every comparison favoring the sponsored treatments ('sponsored-favored').* This scenario is similar to the previous, but assumes that in each comparison XY the sponsored treatment is expected to be favored, while if no information about sponsoring is available, then $d_{iXY} = 0$.
- *The covariate z^*_{iXY} impacts every comparison favoring either the sponsored or the newer treatments ('sponsored/new-favored').* This is a combination of the two previous scenarios. More specifically, in every XY comparison the sponsored treatment is expected to be favored, if information about sponsorship exists, or the newer treatment otherwise.

Then, I infer about whether the employed scenario is likely to hold based on the sign of the coefficients β_{XY}. The way that I have defined the direction variable to take values 1 or -1 implies that a positive coefficient β_{XY} suggests that the direction of prior belief seems plausible, whereas a negative β_{XY} suggests that small-study effects are more likely to operate in the opposite direction. Note that all assumptions presented in Section 4.2.2 can be employed to model the comparison-specific coefficients β_{XY}.

It is important to note that the potential for bias in a network of interventions depends on the clinical setting. There might be several types of bias working with

unknown mechanisms; such biases could not be accounted for in a statistical model.

5.4 A network meta-epidemiological study assessing the impact of trial design limitations

5.4.1 Selection of networks of interventions

A detailed description of the search strategy and the selection process for networks of interventions published until March 2011 is available in Section 3.2. In this network meta-epidemiological study, which investigated the impact of RoB components, I considered only star-shaped networks. All types of outcome measures (dichotomous, continuous, etc.) and effect sizes (OR, MD, etc.) appeared in the individual studies were eligible.

5.4.2 Data extraction

The extracted data from the included networks are presented in detail in Section 3.3. In addition to the outcome data, RoB information was extracted from each network regarding the appropriate conduct of four items; random sequence generation, allocation concealment, blinding of participants and blinding of outcome assessors. If such information was not available, it was sought from the authors of each network publication. For networks for which RoB data could not be obtained either from the published paper or from the authors, this was assessed by two reviewers independently for every included RCT based on the Cochrane RoB Tool (59). All studies were classified into three categories; being at low, high or unclear RoB according to following criteria (59):

- *Random generation of allocation sequence.* Methods that were considered adequate to suggest the random sequence generation included coin tossing, random number table, dice throwing, computer random number generator, restricted randomization methods (minimization technique, random permuted blocks, etc.) or equivalent. Trials that used other inappropriate methods were considered as being at high RoB. Studies were classified into the unclear RoB group when no or insufficient information was provided from the publications to allow judgment of low or high risk.
- *Allocation concealment*. Appropriate methods of allocation concealment implying that both participants and investigators were not likely to foresee the treatment assignment included sequentially numbered opaque and sealed envelopes, central allocation, sequentially numbered drug containers of identical appearance or equivalent. High RoB was assumed for trials using any other non-adequate method and unclear RoB for trials that did not describe the employed method explicitly.
- *Blinding of participants & outcome assessors*. Studies were considered as being at low RoB when the authors described the study as double-blinded and reported the use of identical pills, identical containers, etc. For networks with hard (e.g. death) or objective (e.g. lab outcomes) outcomes all trials were classified into the low risk category regarding the blinding of outcome assessors. Studies described as double-blinded without providing further details on how blinding was achieved were classified into the unclear risk category. For networks with self-assessed outcomes the judgment for blinding of participants and blinding of outcome assessors was the same.

Any disagreements between the two reviewers were resolved by discussion.

5.4.3 Statistical analysis

I initially analyzed each network including at least ten studies (to ensure sufficient power of meta-regression (13)) using a hierarchical NMA model (see Section 2.2.2)

and network meta-regression (see Section 4.2) models assuming a common heterogeneity parameter (τ^2) across all comparisons. I used arm-level outcome data in the analyses when they were available and study-level data otherwise.

The covariate z^*_{iXY} in Equation (5.1) was an indicator variable, which showed whether study i was assessed being at high, unclear or low RoB. More specifically, I defined z^*_{iXY} as

$$z^*_{iXY} = \begin{cases} 0\,, & \text{study } i \text{ is at low RoB} \\ 1\,, & \text{study } i \text{ is at high/unclear RoB} \end{cases}$$

The fixed direction approach was used for the direction variable d_{iXY}, which was combined with the 'active-favored' assumption; that is the only sensible direction scenario in star-shaped networks.

The comparison-specific coefficients β_{XY} within each star network were assumed to be identical implying that the impact of each bias component was similar irrespective of the active treatment being compared with the control intervention (i.e. the common comparator intervention). I also employed the exchangeable coefficients assumption as a sensitivity analysis to allow for variability in the bias parameters across the different comparisons (see Section 4.2.2). For every network n a positive coefficient B_n suggested that studies at high or unclear RoB tended to estimate larger treatment effects than low-risk studies favoring the active treatments.

After analyzing each network separately, I used a network meta-epidemiological model to link all the network-specific coefficients B_n $(n = 1, \dots, N)$. This model allowed for between-network variability assuming that all B_n share a common normal distribution, hence $B_n \sim N(B, \omega^2)$. As a sensitivity analysis, the assumption of a common overall coefficient across networks ($B_n = B$) was employed as well. A detailed description of these models can be found in Section 4.3 and a graphical representation for the case of three networks in **Figure 5.3**. In this way networks including few studies borrowed strength from the larger networks and the overall coefficient B attained increased precision.

To ensure the comparability of bias parameters across the NMA and yield a meaningful overall B, I synthesized in the network meta-epidemiological model

only networks using the same effect size measure; this was OR for dichotomous outcomes, MD for continuous and HR for time-to-event outcomes. Networks using different effect sizes (RR, SMD, etc.) were analysed only separately.

The association between the four RoB items and the relative treatment effects was measured as the ratio of odds ratios (ROR) for dichotomous outcomes, which is estimated as $exp(B_n)$ within each network or $exp(B)$ across networks.

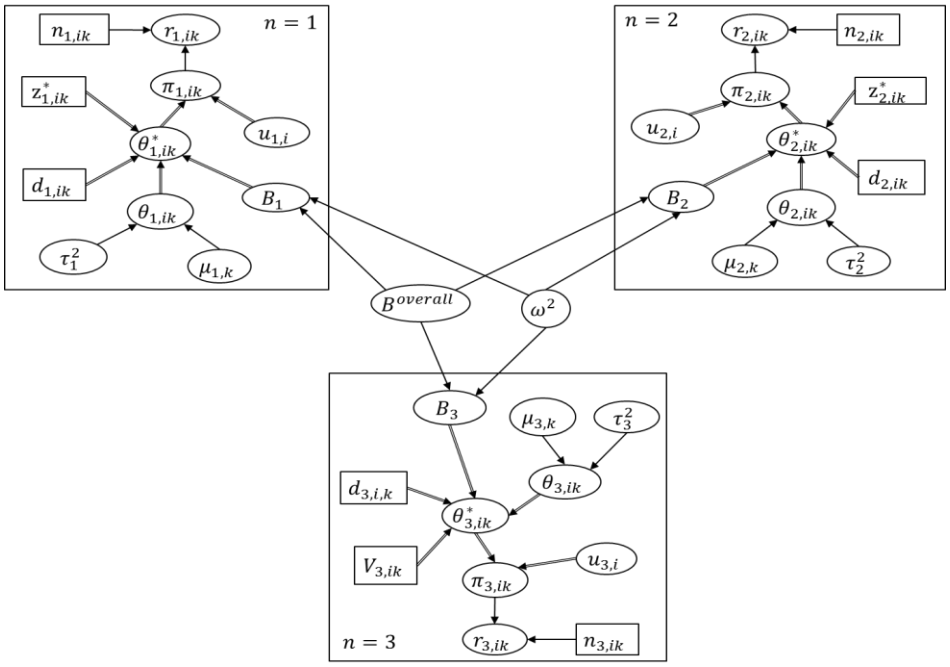

Figure 5.3. Graphical representation of the network meta-epidemiological model (three-network example). Stochastic nodes (associated with distributions) and deterministic nodes (logical functions of parameters) are presented in oval shapes and data are presented in rectangular shapes. Single-line arrows represent distributions and double-line arrows represent logical functions.

5.4.4 Subgroup & sensitivity analyses

I classified the included meta-analyses according to their primary outcome into mortality and non-mortality networks and I also synthesized the two subgroups separately using the network meta-epidemiological model (103).

An additional analysis was performed to check the sensitivity of results to different criteria for RoB classification that might have been used by the original authors of the publications. Therefore, this sensitivity analysis included only

networks for which RoB data were not available (in the published networks or after request from the authors) and thus were extracted from the two reviewers using the criteria described in Section 5.4.2. A second sensitivity analysis aimed to check whether the type of competing interventions affected the results by excluding the networks with non-pharmacological interventions.

5.4.5 Model selection & implementation

I used the deviance information criterion (DIC) as measure of model parsimony and I considered a three-unit decrease in DIC to suggest a better compromise between model fit and complexity (114).

All applied models were fitted in WinBUGS 1.4.3 (71) using Markov chain Monte Carlo simulations. Normal vague prior distributions $N(0,10^4)$ were given to the basic parameters μ_{XY}, the network-specific and overall mean coefficients, B_n and B, and the effect of the control intervention (e.g. $logit(\pi_{iX})$ or φ_{iX} in Equation (1.6)) for networks for which arm-level data were used. A half-normal prior distribution was assumed for the heterogeneity standard deviation, $\tau \sim N(0,1)$ with $\tau \geq 0$ and a uniform $U(0,3)$ for the between-network standard deviation ω in the network meta-epidemiological model. Two alternative prior distributions were employed for the between-network variability as sensitivity analysis; a uniform $U(0,10)$ on ω and a normal $N(0,10^3)$ on $ln(\omega)$.

For all analyses two Markov chains were run after a burn-in period of 10,000 simulations. Achievement of convergence was judged by a visual inspection of the two chains. Results are reported in terms of posterior medians with 95% credible intervals (CrI).

5.4.6 Results

Eligible networks

The selection process (**Figure 3.1**) identified 32 eligible star-shaped networks that involved 613 individual trials. The characteristics of the eligible networks are presented in **Table 5.1** and the judgment about the RoB of the included studies in each network in **Table 5.8**. The independent network-specific meta-regression analysis included 22 networks, which included at least ten studies (545 trials in total); thus 22 network-specific coefficients B_n were obtained.

The network meta-epidemiological model included 20 star networks (after excluding networks with overlapping studies) reporting dichotomous outcomes measured with OR, whereas the remaining networks were not synthesized due to the small number of networks suitable to be linked together; those were 5 star networks with continuous data, 4 with time-to-event data, 1 with rate data and 2 measuring dichotomous outcomes with RR.

Accounting for the impact of risk of bias components

The estimated network-specific coefficients $B_1, \dots, B_{22}$ suggested that none of the four RoB items substantially affected the treatment effects derived from network meta-analysis. From the 22 star networks included in the independent meta-regression analyses, 55%, 60%, 56% and 61% yielded positive coefficients for random sequence generation, allocation concealment, blinding of participants and blinding of outcome assessors respectively (**Table 5.3**). However, none of these coefficients reached statistical significance. Differences in study design with respect to the four bias items did not appear to explain the estimated heterogeneity in the networks; the relative change in heterogeneity standard deviation (τ) did not exceed the 3.5% of the estimated heterogeneity from the NMA model without covariates for any of the networks.

Similar results were derived from the network meta-epidemiological model that linked 20 networks with dichotomous data. The four overall coefficients (B) showing the estimated association between random sequence generation, allocation concealment, blinding of participants and outcome assessors (assuming fixed coefficients within networks and exchangeable across networks) and treatment effects are presented in **Figure 5.4**. The graph does not suggest a tendency of high or unclear risk studies to estimate larger treatment effects compared to low risk trials. The between-network standard deviation estimates (ω) were (on $log(OR)$ scale) 0.91 with CrI (0.75,1.09) for random sequence generation, 0.98 (0.83,1.18) for allocation concealment, 1.16 (0.95,1.43) for blinding of participants and 1.15 (0.83,1.60) for blinding of outcome assessors. Assuming a fixed coefficient across networks without accounting for the variability across networks did not alter substantially the results of the network meta-regression model.

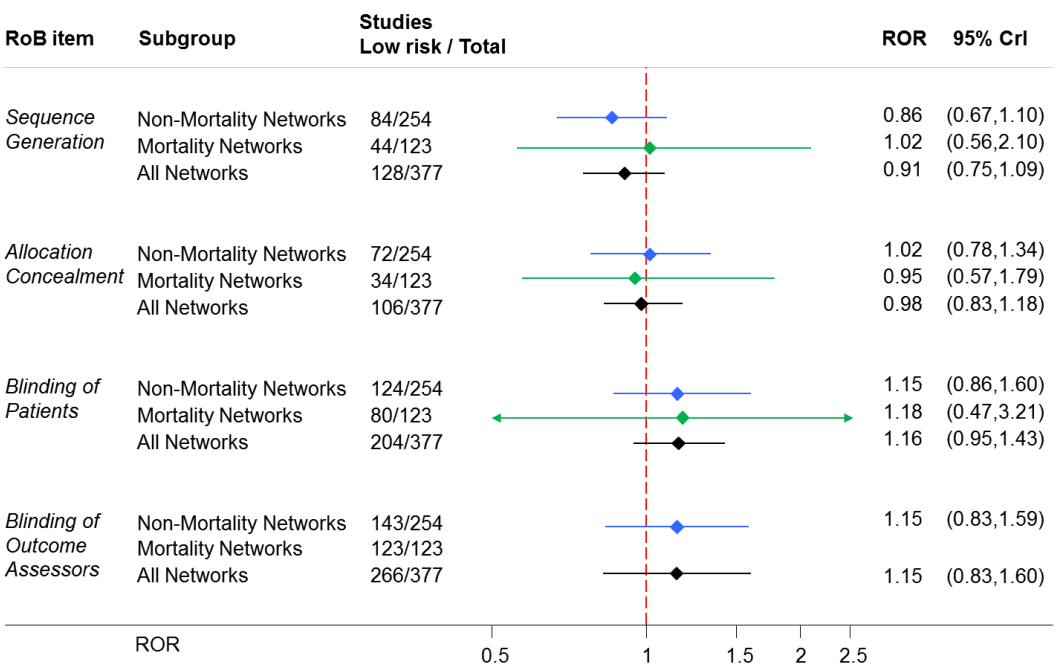

Figure 5.4. Overall ratios of odds ratios (ROR) for each risk of bias component derived from the joint analysis. Results are reported also in subgroups of mortality and non-mortality networks. Outcome assessors were defined as being blinded for all mortality networks. (CrI=Credible Interval)

Table 5.1. Characteristics of eligible star-shaped networks

Network ID	Reference	Studies	Treatments	Control	Topic & Outcome	Effect Measure
1	Abdullah 2008	18	8	Placebo	effectiveness in chronic asthma (oral corticosteroids elimination)	Odds ratio
2	Berner 2006	14	4	Placebo	effectiveness of phosphodiesterase-5 (PDE-5) inhibitors for erectile dysfunction (International Index of Erectile Function (IIEF) score)	Mean difference
3	Edwards 2009	30	4	Imipenem/Cilastatin	effectiveness of beta-lactams for the treatment of hospitalized patients with infection (clinical response)	Odds ratio
4	Fakhoury 2008	5	3	Evening NPH (neutral protamine Hagedorn)	reducing weight gain in patients with type 2 diabetes	Mean difference
5	Gafter-Gvili 2005	37	5	Placebo	antibiotic prophylaxis in neutroneutropenic patients (all-cause mortality)	Odds ratio
6	Jansen 2011	8	6	Placebo	effectiveness of bisphosphonates in the prevention of vertebral fractures (new vertebral fractures)	Odds ratio
7	Kumar 2010	6	3	Melphalan + Prednisone	first-line therapy for patients with multiple myeloma (overall survival)	Hazard ratio
8	Lim 2003	5	3	Placebo	effectiveness of medium or low dose aspirin in preventing occlusion of vein grafts (one or more occlusions)	Odds ratio
9	Lim 2009	33	3	No Treatment/ Placebo	Non-Small Cell Lung Cancer chemotherapy (overall survival)	Hazard ratio
10	Loke 2011	9	3	Enoxaparin	effectiveness for prevention of venous thromboembolism (total venous thromboembolisms (VTE))	Odds ratio
11	Maas 2009	14	3	Control	effectiveness of interventions for prevention of asthma in children at high risk (current diagnosis of asthma in	Odds ratio

					childhood)	
12	Makani 2011	25	4	CCB (calcium channel blocker)	effectiveness of renin-angiotensin system blockade on calcium channel blocker-associated peripheral edema (incidence)	Odds ratio
13	Mason 2004	29	7	Placebo	effectiveness of topical Non-aspirin, Non-steroidal Anti-inflammatory Drugs (NSAIDs) for acute pain (50% or more pain reduction)	Odds ratio
14	McLeod 2007	8	4	Placebo	effectiveness of drugs for the treatment of ankylosing spondylitis (Assessment in Ankylosing Spondylitis (ASAS 20) - 20% improvement at 12 weeks)	Odds ratio
15	Mills 2008	19	5	Placebo	effectiveness of statins in cardiovascular disease (mortality)	Odds ratio
16	Mills 2009	5	4	Interferon-A	relative effectiveness of new therapies for metastatic renal cell cancer (Progression-free survival)	Hazard ratio
17	Mills 2009	14	5	Amphotericin B Deoxycholate	effectiveness of antifungal treatments for invasive Candida infections (all-cause mortality)	Odds ratio
18	Mills 2011	58	7	Control	effectiveness of statins for cardiovascular disease (all-cause mortality)	Odds ratio
19	Moreno 2009	50	13	Placebo	effectiveness of antidepressants	Standardized mean difference
20	Nelson 2006	26	6	Placebo	effectiveness of non-hormonal therapies for menopausal hot flashes	Mean difference
21	Peterson 2008	17	4	Placebo	effectiveness of treatments for adults with attention-deficit hyperactivity disorder (clinical response)	Risk ratio
22	Piccini 2009	7	3	Placebo	effectiveness of treatments for the prevention of recurrent atrial fibrillation (recurrence)	Odds ratio

23	Playford 2006	7	4	Placebo	antifungal agents for preventing fungal infections in liver transplant recipients (total mortality)	Odds ratio
24	Quilici 2008	7	3	Placebo	effectiveness of the non-ergot-derived dopamine agonists in restless legs syndrome (Clinical Global Impressions Improvement (CGI-I) scale responders)	Odds ratio
25	Rheims 2011	58	20	Placebo	effectiveness of antiepileptic drugs in adult refractory partial epilepsy (response)	Risk ratio
26	Rice 1999	15	3	No intervention	effectiveness of nurse-delivered interventions on smoking behaviour in adults (smoking cessation)	Odds ratio
27	Singh 2009	27	7	Placebo	effectiveness of biologics for rheumatoid arthritis (American College of Rheumatology Criteria (ACR50 - 50% improvement	Odds ratio
28	Stettler 2006	10	3	Placebo	effectiveness of drug eluting stents in patients with and without diabetes (in-stent restenosis)	Rate ratio
29	Sultana 2008	11	5	Gemcitabine	effectiveness of treatments for advanced pancreatic cancer (overall survival)	Hazard ratio
30	Uthman 2010	14	6	Placebo	effectiveness for treating anxiety disorders in children and adolescents (improvement)	Odds ratio
31	Welton 2008	13	3	Placebo	antiviral treatments for influenza A & B (patients without symptoms)	Odds ratio
32	Yazdanpanah 2004	14	3	Dual therapy	effectiveness of antiretroviral combination therapy (progression to AIDS or death)	Odds ratio

Accounting for the impact of the RoB items did not appear to influence model parsimony in both independent and meta-epidemiological analyses; differences in DIC were smaller that the three-unit threshold which can suggest improvement in model parsimony (**Table 5.4**).

Table 5.2. Number of studies assessed as being at low, unclear and high risk of bias for each star network (L=low, U=unclear, H=high).

Network ID	Random Sequence Generation			Allocation Concealment			Blinding of participants			Blinding of outcome assessors		
	L	*U*	*H*	*L*	*U*	*H*	*L*	*U*	*H*	*L*	*U*	*H*
1	2	16	-	-	18	-	5	13	-	5	13	-
2	1	13	-	1	13	-	2	11	-	2	11	-
3	4	25	1	3	26	1	1	10	19	2	10	18
4	2	3	-	4	1	-	-	-	5	-	-	5
5	13	24	-	10	25	2	14	-	23	37	-	-
6	3	5	-	2	6	-	6	2	-	8	-	-
7	6	-	-	2	4	-	-	6	-	6	-	-
8	-	5	-	-	5	-	5	-	-	5	-	-
9	3	30	-	14	19	-	-	33	-	33	-	-
10	7	2	-	7	2	-	9	-	-	9	-	-
11	9	5	-	14	-	-	2	3	9	13	-	1
12	6	19	-	1	24	-	10	15	-	9	16	-
13	6	23	-	5	24	-	13	16	-	11	18	-
14	3	5	-	2	6	-	8	-	-	1	6	1
15	8	10	1	9	1	9	15	1	3	19	-	-
16	5	-	-	1	4	-	-	5	-	-	5	-
17	4	8	2	3	10	1	1	3	10	14	-	-
18	22	36	-	15	43	-	58	-	-	58	-	-
19	1	49	-	-	50	-	18	32	-	17	33	-
20	15	11	-	15	11	-	16	10	-	16	10	-
21	1	15	1	4	13	-	11	6	-	9	8	-
22	7	-	-	4	3	-	7	-	-	6	-	1
23	2	5	-	2	5	-	5	-	2	3	4	-
24	4	3	-	4	3	-	5	2	-	5	2	-
25	12	46	-	12	46	-	21	37	-	21	37	-
26	5	8	2	3	10	2	-	-	15	10	-	5
27	7	20	-	11	16	-	20	6	1	21	5	1
28	6	4	-	6	4	-	8	-	2	8	2	-
29	6	5	-	8	3	-	11	-	-	11	-	-
30	7	7	-	2	12	-	6	8	-	9	6	-
31	4	9	-	3	10	-	7	6	-	7	6	-
32	5	9	-	6	8	-	7	6	1	14	-	-
Total	**186**	**420**	**7**	**173**	**425**	**15**	**291**	**231**	**90**	**389**	**192**	**32**

I compared the relative ranking of interventions within each network between the standard NMA model without covariates and the meta-epidemiological model. Small differences in the ranking of treatments were present in three, two and one network when the results accounted for the impact of blinding of participants, blinding of outcome assessors or sequence generation and allocation concealment respectively.

Table 5.3. Network-specific coefficients from the separate analysis. Missing values are for networks in which all studies are at low or unclear/high RoB.

Network ID	Random sequence generation	Allocation Concealment	Blinding of participants	Blinding of outcome assessors	Measure of Effect
1	12.68 (0.52,237.46)	-	2.14 (0.16,24.29)	2.14 (0.16,24.29)	Ratio of odds ratios
3	0.88 (0.37,1.99)	1.17 (0.46,2.86)	1.39 (0.33,6.11)	1.36 (0.33,5.64)	
5	1.62 (0.57,5.37)	1.73 (0.55,5.53)	1.40 (0.47,4.06)	-	
11	1.03 (0.62,1.68)	-	1.77 (0.84,3.63)	1.75 (0.67,4.90)	
12	0.73 (0.49,1.07)	1.15 (0.66,1.99)	1.01 (0.62,1.55)	1.08 (0.68,1.65)	
13	0.41 (0.11,1.65)	1.36 (0.33,6.49)	1.00 (0.23,4.57)	2.18 (0.54,9.58)	
15	1.04 (0.53,1.62)	0.95 (0.53,1.77)	0.95 (0.58,1.95)	-	
17	1.11 (0.36,3.25)	0.92 (0.41,2.08)	-	-	
18	1.06 (0.81,1.32)	0.96 (0.77,1.19)	1.27 (0.95,1.82)	-	
26	0.61 (0.23,1.65)	1.03 (0.33,3.22)	-	1.31 (0.53,3.35)	
27	1.68 (0.89,3.25)	2.10 (0.95,4.66)	1.65 (0.82,3.35)	1.65 (0.75,3.71)	
30	0.84 (0.13,4.85)	1.28 (0.23,9.03)	0.92 (0.31,3.32)	0.51 (0.18,1.26)	
31	0.54 (0.31,0.94)	0.61 (0.31,1.13)	0.57 (0.24,1.27)	0.57 (0.24,1.28)	
32	0.82 (0.55,1.25)	0.82 (0.55,1.22)	0.97 (0.61,1.60)	-	
21	0.76 (0.19,3.13)	1.06 (0.57,2.08)	1.14 (0.64,1.92)	1.21 (0.68,2.01)	Ratio of risk ratios
25	1.34 (0.94,1.92)	0.86 (0.58,1.26)	0.97 (0.68,1.38)	0.97 (0.68,1.38)	
28	0.94 (0.10,9.30)	0.86 (0.05,15.03)	0.66 (0.02,17.12)	1.09 (0.04,27.11)	Ratio of rate ratios
9	1.09 (0.81,1.51)	1.09 (0.90,1.31)	-	-	Ratio of hazard ratios
29	0.92 (0.68,1.26)	1.01 (0.76,1.36)	-	-	
19	0.02 (-0.26,0.32)	-	-0.03 (-0.15,0.10)	-0.01 (-0.14,0.11)	Difference of standardized mean differences
2	0.73 (-2.01,3.47)	0.72 (-2.02,3.49)	0.19 (-1.78,2.16)	0.19 (-1.78,2.16)	Difference of mean differences
20	0.07 (-0.97,1.16)	0.41 (-0.65,1.57)	0.68 (-0.38,1.82)	0.68 (-0.38,1.82)	

Table 5.4. DIC values of the applied models for star networks. Missing values are for networks in which all studies are at low or unclear/high RoB.

Network ID	Model without covariates	Model with allocation concealment as covariate	Model with random sequence generation as covariate	Model with blinding of participant as covariate	Model with blinding of outcome assessors as covariate
1	72.23	-	71.10	70.52	70.52
2	18.73	20.07	20.06	20.29	20.29
3	101.76	102.29	102.69	102.17	102.10
5	142.99	148.38	147.21	148.93	-
9	44.16	44.88	44.71	-	-
11	41.63	-	43.62	40.97	42.43
12	81.45	83.03	80.63	83.61	83.28
13	112.63	112.70	114.20	112.76	111.71
15	63.36	67.90	67.48	67.28	-
17	40.48	42.57	42.60	-	-
18	218.3	228.77	229.36	225.04	-
19	48.44	-	50.19	50.25	50.22
20	43.25	43.59	43.83	42.87	42.87
21	27.78	28.52	29.28	28.55	28.36
25	82.72	83.42	81.89	84.33	84.31
26	55.7	56.40	56.59	-	56.27
27	100.72	100.27	100.21	101.05	100.94
28	5.62	7.49	7.48	7.45	7.50
29	13.91	15.72	15.40	-	-
30	54.27	54.46	54.46	54.82	54.30
31	41.61	40.47	37.87	41.11	41.21
32	38.97	40.21	40.36	41.54	-

Subgroup & sensitivity analyses

Figure 5.4 shows that when the networks were analysed in subgroups of mortality and non-mortality outcomes the results remained similar to the primary analysis.

The meta-epidemiological model that included only the 12 networks for which risk of data were not available from the publications but extracted according to the criteria of Section 5.4.2 gave RORs close to those from the model including all 20 networks; 0.83 (0.63,1.12) for random sequence generation, 0.97 (0.74, 1.30) for allocation concealment, 1.07 (0.80, 1.46) for blinding of participants and 1.22 (0.83, 1.88) for blinding of outcome assessors. This means that the potentially different criteria used by the authors for the RoB assessment did not affect the results.

Excluding 2 networks that compared non-pharmacological intervention did not alter the results of the primary analysis.

5.5 A network meta-epidemiological study assessing the impact of small-study effects

5.5.1 Selection of networks of interventions & data extraction

I also used the collection of star networks assembled as described in Section 5.4 to investigate the impact of small-study effects on NMA results. Subsequently, I used an enriched database with all full networks including at least 4 competing treatments in a sensitivity analysis. These full networks were compiled according to Section 3.2 with the additional inclusion criterion of including an 'obvious' control intervention; that is no treatment, placebo or standard care.

5.5.2 Statistical analysis

The process of analysis was similar to that described in Section 5.4.3, namely I analyzed the networks both independently and then synthesized them in a network meta-epidemiological model. I also used the same criteria to define the eligible networks for each type of meta-regression analysis.

The network meta-regression model that accounted for the impact of small-study effects on the results included as covariate the variance of the observed relative effects (e.g. $logOR$, MD, $logHR$) in individual studies. This means that

$$z^{*}_{iXY} = s^{2}_{iXY}$$

where for networks with arm-level dichotomous data

$$s^{2}_{iXY} = \frac{1}{r_{iX}} + \frac{1}{n_{iX} - r_{iX}} + \frac{1}{r_{iY}} + \frac{1}{n_{iY} - r_{iY}}$$

Formulae for estimating the variance of other effect sizes (RR, MD, etc.) can be found in Borenstein et al. (115,116). Using this covariate in every network n and the 'active-favored' definition for d_{iXY} in Equation (5.1), a positive network-specific coefficient B_n suggested that small studies tended to estimate larger treatment effects in favor of active treatments than larger studies. Note that study variance was selected for the primary analysis because a simulation study suggested that using this measure is less likely to give biased estimates compared to other measures of study precision (96).

A limitation of this approach for dichotomous outcomes is the inherent correlation between OR and its variance. This phenomenon, known as 'regression to the mean', will suggest the presence of small-study effects even when such an association does not truly exist. However, the use of the exact binomial likelihood, whenever arm-level data were available, mitigates this dependence.

5.5.3 Sensitivity analyses

I used alternative measures of study precision (standard error, inverse variance and square root of inverse variance) as covariates in sensitivity analyses. The impact of trial sample size on treatment effects was assessed using the enriched database with the full networks. I defined the studies as small if they included less than 200 participants (following Zhang et al. (117)) and less than 300 participants using the criteria described in the following Section. Then, I used a binary covariate taking values 1 and 0 to indicate that each study was small or large, respectively.

Considering that the exchangeability assumption might be violated by networks resulting in very small or large coefficients, such networks were excluded in a sensitivity analysis. To check whether different effect sizes were affected more by small-study effects, I alao used the network meta-epidemiological model on the RR scale instead of the OR scale.

5.5.4 Definition of small studies

The number of participants that might constitute a small RCT was defined based on previously suggested criteria by Nüesch et al. (95) and Zhang et al. (117). More specifically, I considered a trial being large if it could attain 80% power to detect a 'moderate' effect size. The definition of a moderate effect size for continuous outcomes is not straightforward as it depends on the measurement scale. In the case of dichotomous outcomes, I defined a moderate OR as the 1st quartile from the distribution of the summary pairwise $ln(OR)$ derived from the star networks with dichotomous beneficial outcomes. The rationale for this arbitrary choice was that the majority of star networks were expected to have estimated 'large' treatment effects due to the nature of the comparisons (i.e. active-experimental vs. control interventions). Then, a control group risk (p_X) was further assumed by taking the mean $log(odds)$ from all control groups included in these networks. The total number of participants that a trial needed to detect this reduction in risk was calculated using the formula (118)

$$N = 2 \times \frac{\left(z_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + z_{\beta}\sqrt{p_X(1-p_X) + p_Y(1-p_Y)}\right)^2}{\Delta^2} \quad (5.2)$$

where $\Delta = \left|\frac{r_X}{n_X} - \frac{r_Y}{n_Y}\right|$ is the required absolute risk reduction, $\bar{p} = \left(\frac{r_X}{n_X} + \frac{r_Y}{n_Y}\right)/2$, $z_{\alpha/2} \cong 1.96$ for a two-sided significance level at $\alpha = 5\%$ and $z_{\beta} \cong 0.84$ for 80% power.

5.5.5 Model selection & implementation

All applied models were fitted in WinBUGS 1.4.3 (71). For a description of the prior distributions, the assessment of convergences and the measures used for model selection see Section 5.4.5.

5.5.6 Results

Eligible networks

A description of the eligible star-shaped networks for this study is given in Section 5.4.6. The selection process further identified 48 full networks that met all inclusion criteria, which involved 934 RCTs; 34 full networks with dichotomous data that used OR and 13 full networks with continuous outcomes that used MD were synthesized with the network meta-epidemiological model as a sensitivity analysis. After excluding networks with overlapping studies, the aforementioned full networks were combined with 18 dichotomous and 2 continuous star networks respectively. The 7 networks with time-to-event data (3 full and 4 star networks) were not synthesized because of their small number. The characteristics of the included full networks are available in **Table 5.5**.

Accounting for the impact of small-study effects

Out of the 22 networks that included 10 or more studies 18 (81.8%) gave positive coefficients (B_n) suggesting that smaller studies tend to show the active treatments more effective or safer (compared to the control) than do larger. Eight of these positive coefficients as well as one from the four negatives were large in magnitude and also statistically significant (**Table 5.6**). Improvement in model parsimony after controlling for the impact of small-study effects was present in 4 networks with significant positive coefficients, where the DIC was considerably reduced.

Differences in precision across studies appeared to partly explain the estimated heterogeneity in 9 (40.9%) networks; the relative drop in heterogeneity standard deviation (τ) ranged from 7.1% to 39.5% compared to the NMA without covariates. 9 from the remaining networks resulted in a relative increase from 1.4% to 9.1% (**Figure 5.5**).

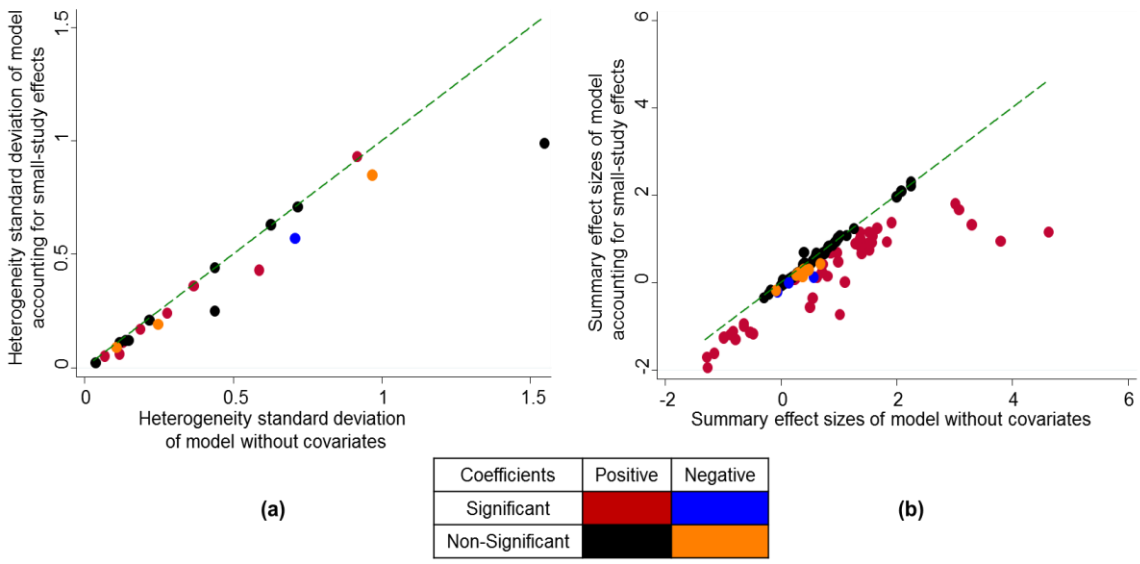

Figure 5.5. Comparison of (a) the heterogeneity standard deviations, and (b) the summary relative effect sizes of all treatments vs. control for all networks with at least 10 studies between the model without covariates and the model controlling for small-study effects. The green dashed lines represent equality of effects between the two models.

Figure 5.5 presents the change in all relative effects (active vs. control) of the 22 networks between the model without covariates and the model controlling for small-study effects. According to the graph most relative effects were reduced after including study variance as covariate suggesting that small studies might exaggerate the effect of active treatments when they are compared to the control. Note that I transformed all relative effects in this graph so as positive values show the active treatments more effective/safer than the control.

The network meta-epidemiological analysis yielded compatible findings with the independent analyses. The overall summary ROR estimated from the model that assumed exchangeability across networks was 1.84 (1.09,3.32) with between-network standard deviation (ω) equal to 0.83 (0.41,1.48) (**Figure 5.6**). When a fixed coefficient B was estimated without allowing for variability across networks, the summary ROR was smaller in magnitude but more precise (1.38 with CrI 1.11 to 1.70). The DIC of the network meta-epidemiological model was 1276, which was considerably smaller than the sum of DICs from the respective models without covariates (1291). This implies that the models accounting for small-study effects might be more parsimonious. In 3 of the networks included in

the meta-epidemiological analysis treatment ranking changed when small-study effects were taken into account. In these networks, the control intervention was placed in a higher rank when the network meta-regression model was applied.

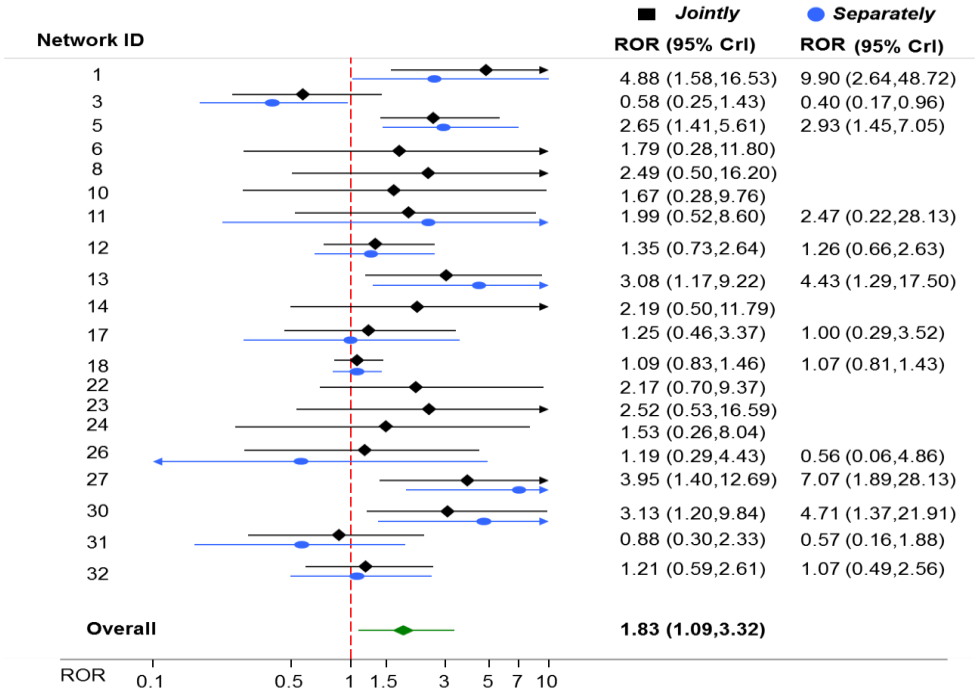

Figure 5.6. Network-specific small-study effect coefficients for all networks with dichotomous data based on separate and meta-epidemiological analyses. (CrI=Credible Interval)

Definition of small studies for the sensitivity analysis

Based on the approach described in Section 5.4.4 an OR=1.88 was considered as a moderate effect size and assuming a control group risk of 36% the required absolute reduction in risk (D) was 16%. Then, Equation (5.2) gives that a trial should include 300 participants to estimate this risk reduction with 80% power (two-sided $\alpha = 5\%$). Thus, the threshold of 300 was used to distinguish between small and large trials in the sensitivity analysis that considered the sample size as explanatory variable.

Table 5.5. Characteristics of eligible full networks for the sensitivity analysis.

Network ID	Reference	Studies	Treatments	Control	Topic & Outcome	Effect Measure
1	Ades 2010	15	9	Placebo	effectiveness of antipsychotic treatments for schizophrenia (relapse)	Odds ratio
2	Anothaisintawee 2012	11	4	Placebo	chronic prostatitis (pain score)	Mean difference
3	Ara 2009	12	5	Placebo	adverse event (treatment related) leading to drug discontinuation (no. of patients)	Odds ratio
4	Baker 2009	39	8	Placebo	exacerbation episodes in Chronic Obstructive Pulmonary Disease (COPD>=1)	Odds ratio
5	Ballesteros 2005	9	4	Placebo	effectiveness of antidepressants in dysthymia	Odds ratio
6	Bangalore 2011	49	8	Placebo	association between antihypertensive drugs and cancer risk (cancer risk)	Odds ratio
7	Bansback 2009	22	8	Placebo	effectiveness for the treatment of psoriasis (Psoriasis Area & Severity Index (PASI) 75 response score)	Odds ratio
8	Bottomley 2011	10	7	Placebo	effectiveness of topical therapies for scalp psoriasis in adults (Investigator Global Assessment (IGA) response after 4 weeks)	Odds ratio
9	Brown 2006	40	6	Placebo	effectiveness for the prevention of Non-aspirin, Non-steroidal Anti-inflammatory Drugs NSAID induced GI toxicity (serious Induced Gastrointestinal (GI) complications)	Odds ratio
10	Buscemi 2007	54	4	Placebo	effectiveness of treatments for chronic insomnia in adults (sleep onset latency - sleep diary)	Mean difference
11	Bucher 1997	18	4	Aerosolized pentamidine	prophylaxis against Pneumocystis carinii in Human Immunodeficiency Virus (HIV) infected patients (number of Pseudocystis Carinii pneumonia)	Odds ratio
12	Elliott 2007	22	6	Placebo	effectiveness of antihypertensive on incidence of diabetes mellitus (proportion of patients who developed diabetes)	Odds ratio

13	Goudswaard 2004	13	4	Insulin once daily	insulin therapies in patients with Type 2 diabetes (HbA1c change)	Mean difference
14	Hofmeyr 2009	24	4	Placebo/Other uterotonics	effectiveness for the prevention of postpartum hemorrhage (blood loss ≥ 1000 ml)	Odds ratio
15	Imamura 2010	38	13	No treatment	effectiveness of non-surgical treatments for women with stress urinary incontinence (cure)	Odds ratio
16	Jansen 2006	12	4	No self-monitoring	self-monitoring of glucose in type 2 diabetes mellitus (change in HbA1c level from baseline)	Mean difference
17	Lam 2007	12	7	Medical therapy	combined resynchronization and implantable defibrillator therapy in left ventricular dysfunction (all-cause mortality)	Odds ratio
18	Lapitan 2009	22	9	Conservative interventions	open retropubic colposuspension for urinary incontinence in women (number not cured within first year)	Odds ratio
19	Lu 2009	24	4	No contact	smoking cessation	Odds ratio
20	Lu 2009	40	6	Placebo	gastroesophageal reflux disease	Odds ratio
21	Macfayden 2005	13	4	No treatment	persistent discharge for chronically discharging ears with underlying eardrum perforations (treatment failure)	Odds ratio
22	McDaid 2010	56	4	Placebo	non-selective non-steroidal anti-inflammatory drugs (NSAIDs) for the reduction of morphine-related side effects after major surgery (24-hour morphine consumption)	Mean difference
23	Mills 2009	89	4	Control	smoking Abstinence at approximately 4 weeks (post-target quit date (TQD))	Odds ratio
24	Phung 2010	20	7	Placebo	effectiveness of non-insulin anti-diabetic drugs added to metformin therapy (change in HbA1c level from baseline)	Mean difference
25	Phung 2011	13	4	Control/ Placebo	effectiveness of thromboprophylaxis strategies (deep venous thrombosis)	Odds ratio
26	Phung 2011	20	6	Control/	effectiveness of oral anti-diabetic drugs for the	Odds ratio

				Placebo	prevention of type 2 diabetes (new cases)	
27	Picard 2009	43	8	No treatment	analgesic effectiveness of prophylactic interventions for the prevention of pain on injection with propofol (no pain)	Odds ratio
28	Playford 2004	10	5	Placebo	antifungal agents for preventing fungal infections in solid organ transplant recipients (mortality)	Odds ratio
29	Pollock 2009	9	5	No treatment/ Placebo	recovery of postural control and lower limb function in patients with stroke (global dependency scale)	Mean difference
30	Psaty 1997	28	7	Placebo/Untreated/ Usual care	antihypertensive Therapy-Coronary heart disease (CHD) including fatal and nonfatal events	Odds ratio
31	Puhan 2009	34	5	Placebo	exacerbation in patients with Chronic Obstructive Pulmonary Disease (COPD)	Odds ratio
32	Roskell 2011	12	10	Placebo	effectiveness for the treatment of fibromyalgia (30% improvement in pain response)	Odds ratio
33	Soares-Weiser 2007	14	8	Placebo	pharmacological and/or psychosocial interventions for the prevention of relapse in people with bipolar disorder (admission to hospital)	Odds ratio
34	Stowe 2011	29	4	Placebo	off-time reduction to levodopa therapy in Parkinson's Disease patients with motor complications	Mean difference
35	Thijs 2008	24	5	Placebo	effectiveness of antiplatelet regimens in the prevention of serious vascular events after transient ischaemic attack or stroke	Odds ratio
36	Trikalinos 2009	63	4	Medical therapy	effectiveness of percutaneous coronary interventions for non-acute coronary artery disease (death)	Odds ratio
37	Tropeano 2011	28	6	Placebo	decrease of carotid intima-media thickness (CIMT)	Mean difference
38	Tu 2010	28	6	EMD + Bone Grafts	treatment effectiveness of enamel matrix derivatives with other regenerative materials	Mean difference

					for infrabony lesions (probing pocket depth reduction)	
39	van der Valk 2009	24	9	Placebo	effectiveness of drugs in patients with primary open-angle glaucoma or ocular hypertension (lowering intraocular pressure at though)	Mean difference
40	Virgili 2011	10	5	Control	effectiveness of treatments for neovascular age-related macular degeneration (visual acuity loss)	Odds ratio
41	Walsh 2010	75	9	Placebo	Decayed, (missing), Filled tooth Surfaces (D(M)FS) caries increment (prevented fraction)	Mean difference
42	Wandel 2010	10	4	Placebo	absolute pain intensity in patients with osteoarthritis of hip or knee	Mean difference
43	Wang 2010	43	9	Standard catheter	effectiveness of venous catheters for catheter-related infections (colonization)	Odds ratio
44	Welton 2009	36	16	Control (usual care)	effectiveness of psychological interventions in coronary heart disease (total mortality)	Odds ratio
45	Wilhelmus 2000	60	24	Placebo	effectiveness for dendritic or geographic Herpes Simplex Virus (HSV) epithelial keratitis (proportion healed)	Odds ratio
46	Woo 2010	19	10	Placebo	relative efficacy of nucleotides in the treatment of chronic hepatitis B (Hepatitis B Virus (HBV) DNA levels)	Odds ratio
47	Yu 2006	14	6	Control	effectiveness of inhaled anesthetics in reducing post-operative myocardial infractions after cardiac surgery (Acute Myocardial Infarction (AMI))	Odds ratio

Table 5.6. Network-specific coefficients and DIC values from the separate network meta-regression analyses for small-study effects.

Network ID	Small-study effects RORs	Measure of effect	DIC of model without covariates	DIC of model with study variance as covariate
1	9.87 (2.66,49.90)	Ratio of odds ratios	72.23	63.27
3	0.40 (0.17,0.97)		18.73	20.05
5	2.10 (1.08,4.44)		101.76	101.55
11	2.48 (0.24,27.66)		142.99	136.65
12	1.26 (0.66,2.61)		44.16	40.64
13	4.71 (1.48,17.81)		41.63	42.97
15	1.35 (0.79,2.41)		81.45	82.67
17	0.99 (0.28,3.53)		112.63	114.6
18	1.07 (0.80,1.43)		63.36	62.99
26	0.57 (0.06,4.85)		40.48	42.62
27	6.96 (1.88,27.39)		218.3	219.62
30	4.66 (1.34,22.20)		48.44	45.12
31	0.57 (0.16,1.88)		43.25	44.89
32	1.07 (0.50,2.56)		27.78	25.82
21	6.96 (1.04,47.47)	Ratio of risk ratios	82.72	81.86
25	2.10 (0.97,4.44)		55.7	56.47
28	1.04 (0.84,1.30)	Ratio of rate ratios	100.72	100.25
9	4.06 (1.28,12.43)	Ratio of hazard ratios	5.62	7.51
29	2.44 (0.00,110.95)		13.91	15.50
19	4.56 (0.63,8.58)	Difference of standardized mean differences	54.27	53.1
2	0.48 (-0.86,1.84)	Difference of mean differences	41.61	42.66
20	0.05 (-0.42,0.49)		38.97	41.04

Sensitivity analyses

The model that used the standard error of the observed effects as covariate instead of the variance resulted in similar but less precise estimates of the summary ROR (2.01 with CrI 1.02 to 12.22). The alternative models using the $\frac{1}{s_{iXY}^2}$ and the $\sqrt{\frac{1}{s_{iXY}^2}}$ as measure of study precision gave summary ROR estimates close to 1; however the DIC of these two models suggested that they might be much worse than the primary model in terms of model parsimony. One (ID=1) out of the 20 star networks of the network meta-epidemiological analysis gave a quite

larger coefficient compared to the other networks. When I excluded this network, the estimated summary ROR slightly reduced into 1.62 (1.00,2.86). Using the RR as effect measure (instead of OR) did not materially change the results of primary analysis.

The inclusion of the full networks with dichotomous outcomes increased the precision of the analysis and resulted in a summary ROR equal to 1.88 (1.43,2.51). In the analysis that examined the effect of the total study sample size, I used the enriched database. The thresholds of 200 and 300 participants used to distinguish between small and large studies yielded overall RORs of 1.26 (1.10,1.46) and 1.13 (1.00,1.28) respectively. The meta-epidemiological model that synthesized the 15 networks (13 full and 2 star networks) with continuous outcomes resulted in an overall coefficient (difference of MD) equal to 0.02 (-0.34,2.38).

Changing prior distributions for the between-networks standard deviation parameter ω (Section 5.4.5) gave similar results to the primary analysis.

Association between study precision & risk of bias items

I further used the RoB information described in Section 5.4.2 and **Table 5.2** to check whether small studies were more prone to be designed improperly. **Figure 5.7** presents for each RoB item the distribution of study variance for low and high/unclear risk studies. The graph suggests that for blinding of participants and blinding of outcome assessors, studies with larger variances are more often appropriately blinded than studies with smaller variances. On the other hand, less precise studies appear to be more often at low RoB with respect to the conduct of random sequence generation and allocation concealment. These findings imply that in this network collection larger studies seem to have been conducted more appropriately than small studies prior to the treatment allocation but not after that point.

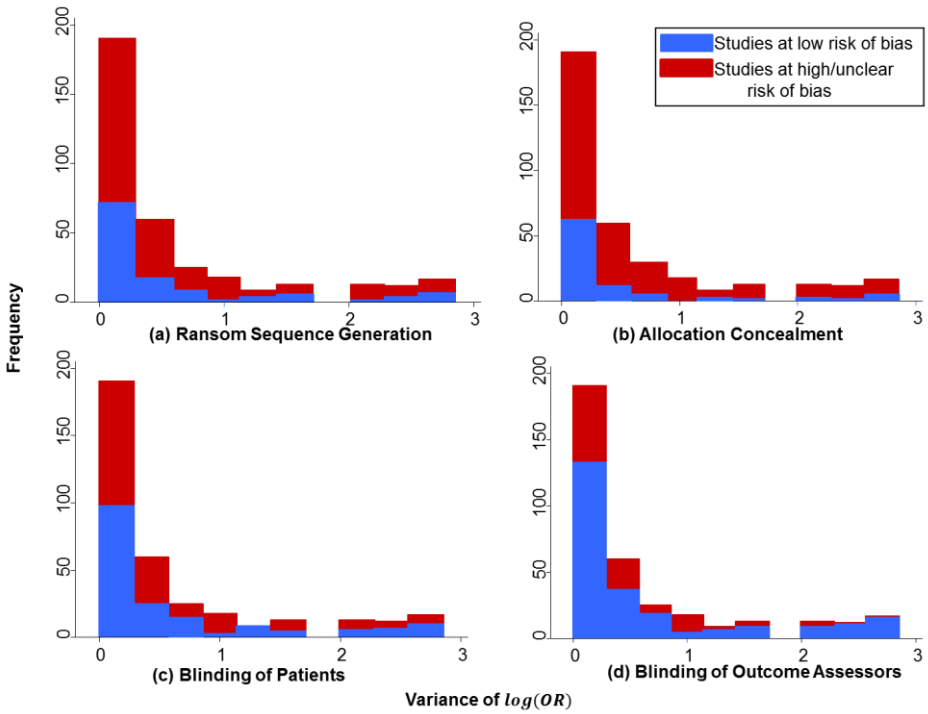

Figure 5.7. Histograms showing the distribution of the study variances (variance of $logOR$ in horizontal axis) by risk ob bias category for sequence generation (a), allocation concealment (b) and blinding of patients (c) and outcome assessors (d).

5.6 A case study exploring the impact of study precision in full networks

5.6.1 Motivating examples

To exemplify the methods of this Section, I used two full networks both evaluating typically secondary (safety) outcomes; these are failure of vascular graft or arterial patency with aspirin, dipyridamole or placebo, and incidence of diabetes with antihypertensive drugs. Such outcomes are more likely to be subject to selective reporting, which might cause the presence of small-study effects in a meta-analysis. In both networks there are trials not including placebo (i.e. the control intervention) and assumptions based on the novelty or sponsorship (Section 5.3) of treatments need to be employed for these studies. Note that substantial heterogeneity is present in both networks, while the

diabetes network has also a couple of inconsistent loops. All these factors imply that small-study effects are expectable in some of the pairwise comparisons. The two examples are described below.

Failure of vascular graft or arterial patency

The first network consists of 31 studies that assess the safety of aspirin, dipyridamole and placebo regarding the failure of vascular graft or arterial patency (119). The network plot is presented in **Figure 5.8** and the data are available in **Table 5.7**. **Figure 5.9** shows the contour-enhanced funnel plots for the three pairwise comparisons in the network, which appear quite asymmetric; particularly the plot corresponding to the comparison aspirin versus placebo. The shaded contours in these graphs serve as a way to figure out whether the asymmetry can be explained by publication bias (when missing studies lie in regions with $p>0.05$) or by other reasons such as heterogeneity (when missing studies lie in regions with $p<0.05$) (120). Note that no statistically significant inconsistency was found in the network by comparing the direct and indirect estimates.

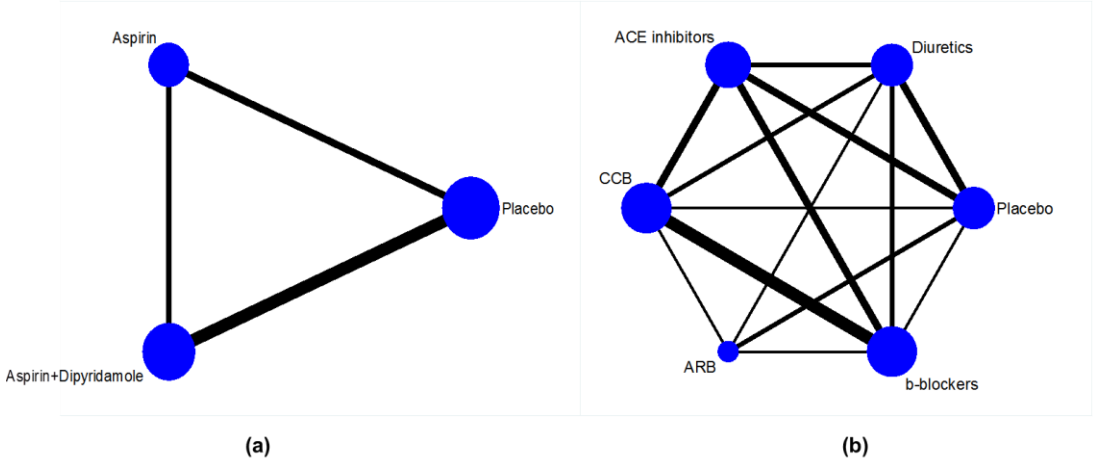

Figure 5.8. Network plots for (a) first and (b) second network examples. The size of the nodes is proportional to the number of studies that evaluate each intervention, and the thickness of the lines to the frequency of each comparison in the network. (CCB=calcium-channel blockers, ARB=angiotensin-receptor blockers, ACE=angiotensin-converting enzyme)

More specifically, the direct OR estimate of dipyridamole vs. aspirin was 0.92 (0.45, 1.90) and 1.29 (0.47, 2.81) from the 4 two-arm and the 6 three-arm studies respectively and the indirect estimate 1.32 (0.77, 2.29) (favoring placebo).

Table 5.7. Outcome data for the network comparing placebo, aspirin, aspirin+dipyridamole with respect to the failure of vascular graft or arterial patency.

Study	**Placebo**		**Aspirin**		**Aspirin+Dipyridamole**	
	events	*total*	*events*	*total*	*events*	*total*
1	18	51	10	47	15	49
2	47	153	37	155	35	162
3	114	671	85	676	83	668
4	39	100	16	100	23	100
5	12	17	2	16	6	16
6	12	100	6	100	0	100
7	22	64	-	-	20	60
8	27	317	-	-	26	313
9	6	40	-	-	10	41
10	15	55	-	-	8	55
11	37	160	-	-	33	160
12	81	205	-	-	37	202
13	9	30	-	-	4	18
14	20	63	-	-	17	62
15	24	64	-	-	8	61
16	27	46	-	-	13	47
17	14	35	-	-	21	34
18	15	68	-	-	11	72
19	13	189	-	-	6	187
20	86	263	-	-	86	286
21	15	32	-	-	4	33
22	12	50	-	-	15	50
23	19	31	-	-	7	22
24	13	67	-	-	15	132
25	16	71	15	71	-	-
26	15	31	6	29	-	-
27	17	69	7	68	-	-
28	47	213	24	215	-	-
29	28	150	19	148	-	-
30	18	25	6	19	-	-
31	11	45	2	47	-	-

Incidence of diabetes

The second network includes 22 studies and compares the safety of 5 antihypertensive drugs and placebo for incidence of diabetes mellitus (121). The presence of few studies in each direct comparison (**Table 5.8**) does not allow making judgments about asymmetry in conventional funnel plots.

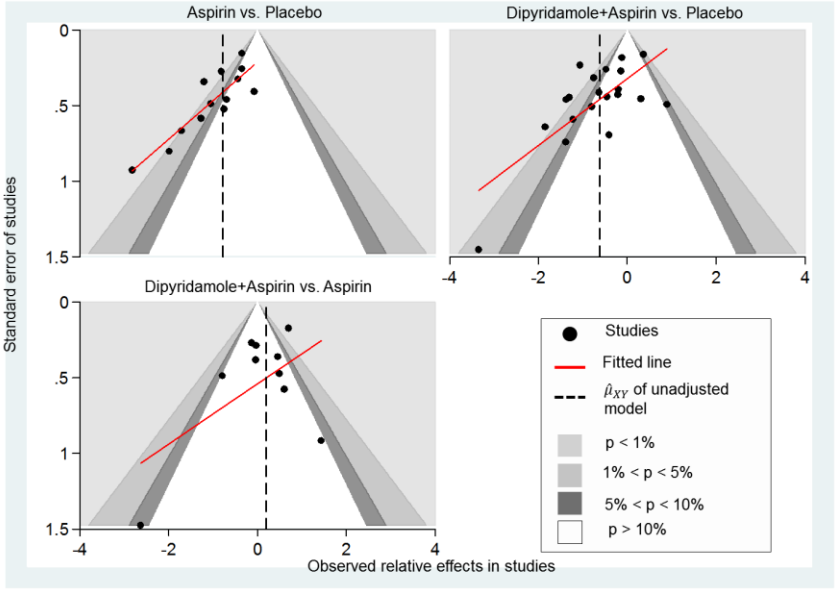

Figure 5.9. Contour-enhanced funnel plots for the first network example by comparison. Points represent the $log(OR)$ derived from all available direct estimates versus their standard errors. Shaded contours correspond to levels of statistical significance defined by the p-value of a z-test for $log(OR)$. Red solid lines are the estimated regression lines and black dashed lines represent the summary comparison-specific $log(OR)$ estimated from the network meta-analysis model without covariates.

The comparison-adjusted funnel plot in **Figure 5.10** shows all comparisons as older versus newer treatment assuming that angiotensin receptor blockers (ARB) is the newest treatment, followed by angiotensin-converting enzyme (ACE), calcium-channel blockers (CCB), b-blockers and diuretics (assumed equally new) and then placebo. In this graph missing points on the right of the 0 value suggest that small studies favoring the older interventions are missing. The graph seems symmetric implying the absence of small-study effects. However, there might be other mechanisms of bias directionality operating and masking the association between relative effects and study precision when a consistent direction (old vs. new) is assumed. Note also that incidence of diabetes is not very often a pre-specified outcome in trials comparing antihypertensive drugs (121); an earlier NMA that compared the same treatments identified 5 studies reporting the outcome of efficacy but not on incidence of diabetes (122).

Table 5.8. Outcome data for the network comparing placebo and 5 antihypertensives with respect to incidence of diabetes.

Study	Placebo		Diuretics		ACE inhibitors		CCB		ARB		β-blockers	
	events	*total*	*events*	*total*	*events*	*total*	*events*	*total*	*events*	*total*	*events*	*total*
1					45	410	32	202			70	405
2			302	6766	119	4096	154	3954				
3			8	196					1	196		
4			200	2826	138	2800						
5							567	7072			799	7040
6					337	5183					380	5230
7	202	2721							163	2715		
8	489	2646			449	2623						
9	20	424	29	416								
10	154	4870					177	4841				
11			75	3272							86	3297
12	155	2883			102	2837						
13			176	2511			136	2508				
14							569	8098			665	8078
15									242	4020	320	3979
16	34	2213	43	1081							37	1102
17							216	5095			251	5059
18	399	3472			335	3432						
19	115	2175							93	2167		
20	118	1578	140	1631								
21					93	1970	95	1965			97	1960
22							845	5074	690	5087		

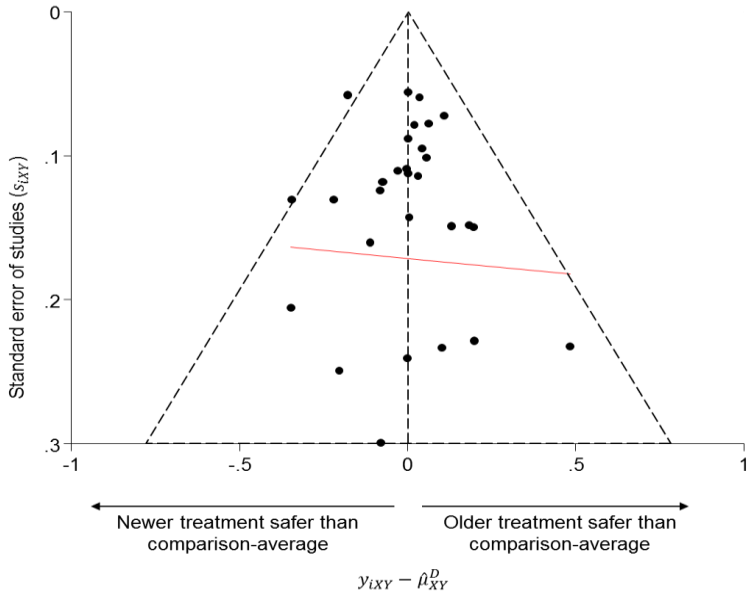

Figure 5.10. Comparison-adjusted funnel plot for the network comparing placebo and 5 antihypertensives with respect to incidence of diabetes. The red solid line is the estimated regression line.

5.6.2 Methods

The methods I used to investigate the possible presence of small-study effects in both examples are presented explicitly in Section 5.3. I used the variance of the observed effects in studies as explanatory variable in the network meta-regression models controlling for differences in precision across studies. I employed different assumptions about the direction of small-study effects (see Section 5.3) as well as about the respective coefficients (Section 4.2.2).

5.6.3 Model selection and implementation

All applied models were fitted in WinBUGS 1.4.3 (71). Vague normal prior distributions $N(0,10^4)$ were assumed for the network-specific coefficients B_n and the comparison-specific coefficients β_{XY}. A half-normal prior distribution was employed for both the heterogeneity standard deviation and the coefficients' standard deviation, hence $\tau, \varphi \sim N(0,1)$ with $\tau, \varphi > 0$. Alternative priors for τ (a log-normal distribution $log(\tau) \sim N(0,10^3)$ and a uniform $\tau \sim U(0,10)$) were used as sensitivity analysis. A non-informative prior distribution $Beta(1,1)$ was assumed for the probability P_{XY} in the probabilistic direction approach reflecting the absence of strong prior belief for the direction of small-study effects. 100,000 simulations were run with a burn-in period of 30,000 draws and convergence was assessed by visual inspection of three Markov Chains with different initial values.

The parsimony and fit of all models was evaluated using the DIC (see Section 5.4.5) and the values of the posterior mean of the residual deviance $\overline{D}$ (114). $\overline{D}$ is a measure of model fit and should approximate the number of data points.

5.6.4 Results

Failure of vascular graft or arterial patency

The positive and statistically significant regression coefficients B (**Table 5.10**) from all network meta-regression models for this network suggest that there is an association between study precision and relative effects; less precise studies gave on average larger treatment effects than more precise studies. **Table 5.9** shows also that model fit and parsimony were improved after the adjustment according to the DIC and $\overline{D}$ values. The slight reduction in the between-study standard deviation (τ) implies that heterogeneity can be explained partly from differences in precision across studies. All network meta-regression models gave similar results regarding the summary ORs, which show the active interventions less effective compared to the model without covariates.

The model that assumed consistency in the regression coefficients (Model 1.6) suggested that small studies tended to favor the active treatments when these are compared with placebo. According to the magnitude of the coefficients this association was larger for the comparison with aspirin rather than dipyridamole. The negative and non-significant coefficient $\hat{\beta}_{XY}$ suggests that dipyridamole is not favored over aspirin when compared with each other (**Table 5.10**). The probabilistic model yielded similar conclusions giving posterior probabilities of 85% and 75% for aspirin and aspirin+dipyridamole being favored over placebo from small studies respectively. The posterior probability for aspirin+dipyridamole being favored over aspirin monotherapy in small studies was 50%.

Meta-regression models using the standard error or the inverse of variance as covariate showed poorer model parsimony (i.e. larger DIC) and thus were not further investigated.

It is interesting that the DICs of the unrelated mean effects model (see Section 2.3.3) and the consistency model without covariates were the same; this implies that the consistency assumption might be plausible in this network. However, the

comparability of the direct and indirect estimates was improved after including study variance as covariate. This was explored by excluding the direct evidence about dipyridamole vs. aspirin; hence the four head-to-head trials comparing the two active treatments and the three-arm placebo-controlled studies after selecting randomly one of their two active treatment arms. This approach gave an indirect OR for dipyridamole vs. aspirin 1.32 (0.77,2.29). Analysing this dataset with the 'new favored' meta-regression model (with exchangeable coefficients) resulted in an OR estimate 1.17 (0.64,2.09), which is closer to the respective direct estimate (OR=0.92 with CrI from 0.45 to 1.90).

Table 5.9. Posterior medians and 95% credible intervals for the odds ratios (OR) and heterogeneity (τ) for the first network example. DIC and posterior mean residual deviance ($\bar{D}$) are also reported. (A=aspirin, A+D=aspirin+dipyridamole, P=placebo)

Model	Assumptions	DIC	$\bar{D}$[1]	OR_{AvsP}	$OR_{(A+D)vsP}$	$OR_{(A+D)vsA}$	τ
1.1	No covariates	130.9	75.5	0.46 (0.31,0.62)	0.54 (0.40,0.71)	1.21 (0.86,1.72)	0.50 (0.31,0.76)
1.2	*'active-favored'* & identical coefficients	114.8	66.2	0.75 (0.53,1.09)	0.90 (0.66,1.26)	1.20 (0.90,1.61)	0.36 (0.20,0.58)
1.3	*'active-favored'* & exchangeable coefficients	115.5	66.2	0.76 (0.53,1.12)	0.89 (0.65,1.24)	1.17 (0.89,1.88)	0.36 (0.2,0.58)
1.4	*'new-favored'* & identical coefficients	115.2	66.7	0.67 (0.49,0.93)	0.91 (0.67,1.25)	1.35 (1.02,1.82)	0.35 (0.19,0.57)
1.5	*'new-favored'* & exchangeable coefficients	115.8	66.8	0.54 (0.39,0.74)	0.82 (0.59,1.15)	1.52 (1.08,2.18)	0.35 (0.19,0.57)
1.6	*'active-favored'* (for basic coefficients) & consistent coefficients	116.4	66.3	0.78 (0.54,1.16)	0.87 (0.63,1.24)	1.11 (0.77,1.60)	0.37 (0.20,0.58)
1.7	*probabilistic 'new-favored'* & identical coefficients	114.9	64.5	0.67 (0.47,0.96)	0.84 (0.61,1.14)	1.25 (0.90,1.72)	0.32 (0.16,0.54)

[1]Should be compared with the 68 data points.

Table 5.10. Posterior medians and 95% credible intervals for small-study effects mean ($\hat{B}$) and comparison-specific coefficients ($\hat{\beta}$) and coefficients' variance ($\hat{\xi}$) (in $log(OR)$ scale) for the first network example. (A=aspirin, A+D=aspirin+dipyridamole, P=placebo)

Model	Assumptions	$\widehat{B}$	$\hat{\xi}$	$\widehat{\beta}_{AvsP}$	$\widehat{\beta}_{(A+D)vsP}$	$\widehat{\beta}_{(A+D)vsA}$
1.2	*'active-favored'* & identical coefficients	2.61 (1.40,4.02)	-	-	-	-
1.3	*'active-favored'* & exchangeable coefficients	2.62 (0.96,4.47)	0.53 (0.02,1.94)	2.76 (1.42,4.33)	2.47 (1.19,4.02)	0
1.4	*'new-favored'* & identical coefficients	2.39 (1.32,3.62)	-	-	-	-
1.5	*'new-favored'* & exchangeable coefficients	2.37 (0.93,3.88)	0.46 (0.02,1.76)	2.47 (1.27,3.86)	2.37 (1.18,3.83)	2.26 (0.34,3.89)
1.6	*'active-favored'* (for basic coefficients) & consistent coefficients	-	-	3.03 (1.46,4.76)	2.28 (0.98,4.02)	-0.71 (-2.57,1.19)
1.7	*probabilistic 'new-favored'* & identical coefficients	2.69 (1.42,4.10)	-	-	-	-

Incidence of diabetes

I excluded one study (123) from this network for being associated with very poor model fit compared to the other studies. More specifically, the standard error of the remaining 21 studies ranged between 0.06 and 0.30, whereas the standard error of that study was 1.07. When the study by Lindholm et al. was excluded the posterior mean deviance of the NMA model was $\overline{D}$=49.2 closer to the 46 data points than the posterior deviance of the full dataset ($\overline{D}$=53.2 compared to 48 data points). The remaining 21 studies were included in all analyses and involved 29 comparisons.

Table 5.11 shows the results of all applied models for this network. The models assuming identical comparison-specific coefficients (Models 2.2-2.5) suggest that small studies tend to exaggerate the safety of placebo or of the older treatments. However, there is large uncertainty and the estimated coefficients are not

statistically significant. The disagreement between the '*sponsored-favored*' and the '*sponsored/new-favored*' models (opposite sign of the estimated coefficients) implies the presence of a possible interaction between the direction of small-study effects and study sponsorship. When the comparison-specific coefficients were assumed exchangeable the results did not change materially.

Table 5.11. Posterior medians and 95% credible intervals for heterogeneity (τ) and small-study effects coefficients ($\hat{B}$) in $log(OR)$ scale for the second network example. DIC and posterior mean residual deviance ($\bar{D}$) are also reported.

Model	Assumptions	DIC	$\bar{D}$[1]	$\hat{\tau}$	$\hat{B}$
2.1	No covariates	86.1	49.2	0.13 (0.06,0.23)	-
2.2	'*active-favored*' & identical coefficients	84.7	47.6	0.12 (0.05,0.22)	-6.69 (-13.59,0.15)
2.3	'*new-favored*' & identical coefficients	85.1	48.0	0.12 (0.05,0.22)	-4.51 (-9.45,0.41)
2.4	'*sponsored-favored*' & identical coefficients	85.1	47.5	0.13 (0.06,0.24)	10.55 (-4.27,27.71)
2.5	'*sponsored/new-favored*' & identical coefficients	83.8	47.1	0.11 (0.05,0.21)	-6.40 (-12.03,-0.80)
2.6	'*active-favored*' (for basic coefficients) & consistent coefficients	86.3	47.8	0.11 (0.04,0.21)	-
2.7	'*sponsored-favored*' & subgroups by sponsoring	83.1	46.2	0.11 (0.04,0.20)	B_{spon}=4.16 (-9.80,18.92) B_{new}=-6.88 (-12.50,-1.25)
2.8	'*sponsored/new-favored*' & subgroups by year	82.8	46.4	0.10 (0.03,0.19)	$B_{\geq 2000}$=7.59 (-8.82,23.82) $B_{<2000}$=-6.90 (-12.40,-1.37)
2.9	*probabilistic* '*sponsored/new-favored*' & subgroups by year	84.1	45.7	0.10 (0.03,0.19)	$B_{\geq 2000}$=6.69 (-34.05,48.61) $B_{<2000}$=-9.45(-23.39,10.36)

[1]Should be compared with the 46 data points.

These findings imply that there might not be a 'fixed pattern' in the direction of small-study effects. In this case the assumption of identical or exchangeable

coefficients might not be reasonable. I employed the consistent coefficients model (Model 2.6) to explore this possibility. Despite the increased uncertainty of the five basic coefficients of the active treatments versus placebo, their different signs suggest that assuming a consistent direction of small-study effects might not be applicable for this network (**Table 5.12**).

Table 5.12. Posterior medians and 95% credible intervals for small-study effects coefficients (from Model 2.6) in $log(OR)$ scale for the second network example. (ARB=angiotensin-receptor blockers, ACE=angiotensin converting-enzyme, BB=β-blockers, CCB=calcium-channel blockers, D=diuretics, P=placebo)

$\widehat{\boldsymbol{\beta}}_{\boldsymbol{BBvsP}}$	-6.26 (-15.77,3.49)
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{DvsP}}$	-5.28 (-12.71,2.07)
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{CCBvsP}}$	-9.53(-43.84,26.61)
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{ACEvsP}}$	18.77 (-11.58,48.78)
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{ARBvsP}}$	3.30 (-21.47,29.59)

I incorporated the possible interaction between direction and sponsorship (i.e. that direction might be different between sponsored and non-sponsored studies) in an alternative model (Model 2.7) that estimated two different (fixed) coefficients B_{spon} and B_{new} . The first coefficient (B_{spon}) accounted for small-study effects in all sponsored studies assuming that the sponsored treatments are favored; the second (B_{new}) controlled for small-study effects in all non-sponsored studies assuming that newer treatments are favored. This model gave a negative and statistically significant coefficient B_{new}=-6.88 (-12.5,-1.25) suggesting that small trials tend to show the older treatments safer than do the larger trials. However, such an association was not confirmed for the sponsored studies as the coefficient B_{spon} was estimated with large uncertainty. Since these studies involve 21 treatment comparisons, possibly the observed uncertainty is not the consequence of low power. The findings of this model can be attributed to the presence of selective reporting bias; this might be an issue more often for non-sponsored trials which, unlike sponsored studies, do not necessarily follow a registered protocol. An alternative explanation which seems more plausible is that sponsored studies are, on average, newer trials; hence not as prone as non-

sponsored studies to substantial variability in study design between small and large trials.

This was investigated by fitting a model with two fixed coefficients $B_{<2000}$ and $B_{\geq 2000}$ for studies published before and after 2000 respectively (Model 2.8). Note that all trials published after 2000 were sponsored. The comparison-adjusted funnel of **Figure 5.11** shows separately the two subgroups with their regression lines. The estimated coefficients (**Table 5.11**) from this model suggest that small-study effects are more likely to operate in older trials, while the direction of bias might not be constant over time. These findings are further supported by the probabilistic model (Model 2.9). More specifically, the posterior probability that studies before 2000 tend to favor the older treatments and placebo is 66%, whereas the probability for studies after 2000 of favoring the newer and sponsored treatments is 39%.

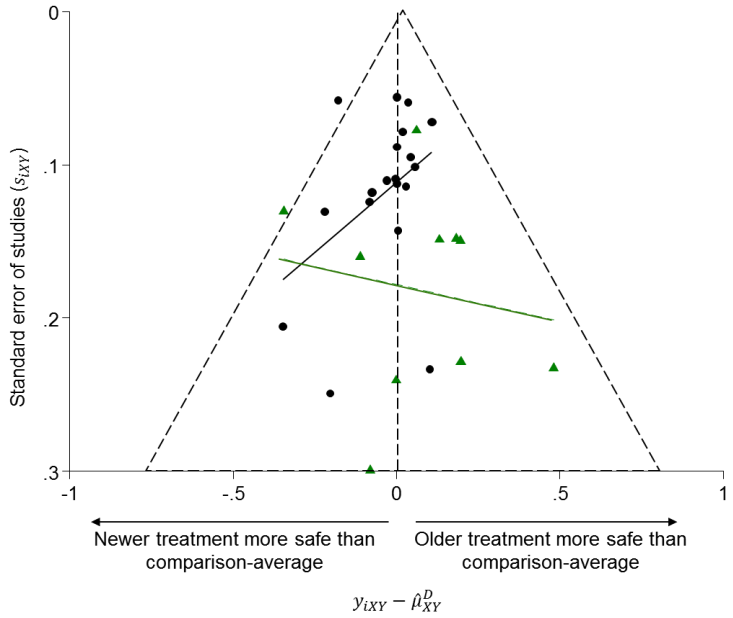

Figure 5.11. Comparison-adjusted funnel plot for the network comparing placebo and 5 antihypertensives with respect to incidence of diabetes in two subgroups; for studies published before and after 2000.

It is interesting that the meta-regression models that consider the year of publication of the included studies are the best models in terms of model fit and parsimony. Also, these two models have lower heterogeneity compared to any

other model; 23% relative drop in heterogeneity compared to the model without covariates. The differences in relative ranking between Model 2.8 and the model without covariates are presented in **Figure 5.12**. According to the estimated cumulative probability curves, after the accounting for the impact of small-study effects β-blockers and ARB appear slightly less safe, whereas placebo, diuretics and ACE inhibitors appear a bit safer. However, there were no important differences in the relative ranking and the relative effects as estimated from the two models.

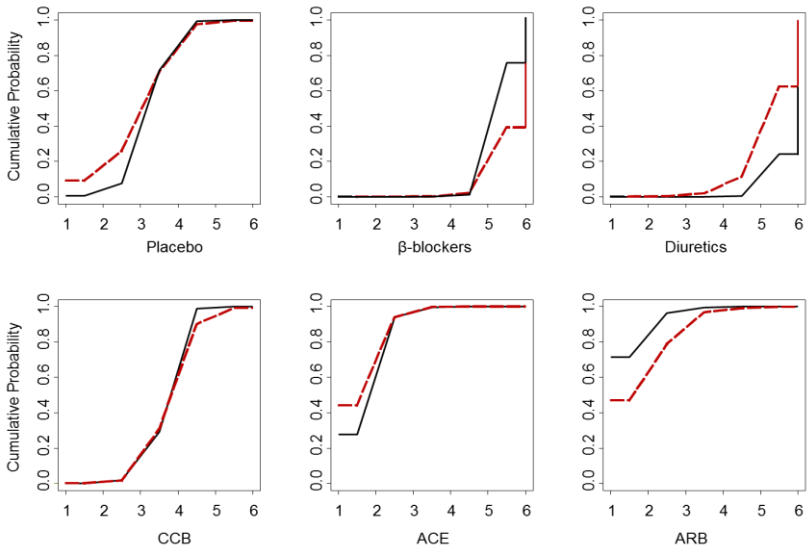

Figure 5.12. Plots of the cumulative ranking curves for antihypertensive drugs. The black solid lines correspond to the estimates from the model without covariates (Model 2.1) and the dashed lines to Model 2.8.

I evaluated the presence of inconsistency in the network using the loop-specific approach (see Section 2.3.2), which resulted in two loops (placebo versus β-blockers versus ACE inhibitors and placebo versus β-blockers versus ARB) with statistically significant inconsistency. The two inconsistency factors were 0.82 (0.28,1.36) and 0.71 (0.18,1.24) respectively. To explore whether differences in precision across studies can partly explain the inconsistency in these loops, I applied an unrelated mean effects model accounting for small-study effects similar to Model 2.8 (using the sponsored/new assumption and considering the subgroups before and after 2000). Then, using the direct and indirect estimates

from this model, inconsistency became statistically non-significant in the two loops with inconsistency factors 0.12 (-0.19,0.43) and 0.14 (-0.35,0.64) respectively. This means that disagreement between direct and indirect estimates may be attributed to some degree to differences between small and large trials as well as between old and new studies.

It is possible that the observed discrepancies in estimation between older and newer studies are caused by other trial characteristics related to the year of publication. For example, studies published before and after 2000 probably differ with respect to the criteria used for the diagnosis of diabetes. More specifically, in 1999 the diagnostic criterion changed from $\geq$ 7.8 mmol/L glucose level to $\geq$ 7.0 mmol/L. Identifying possible explanations requires examination of the study protocols and evaluation of the quality of studies. However, the inclusion of only 21 trials does not allow the adequate investigation of multiple interactions between several study characteristics.

Sensitivity analyses

To check the sensitivity of results to the choice of prior distribution for the heterogeneity parameter, I performed additional analyses for all models with identical comparison-specific coefficients. When I assumed a uniform $U(0,10)$ prior on τ, the results were similar to the primary analysis regarding the estimated parameters and fit of the different models. The use of a log-normal distribution (i.e. $\log(\tau) \sim N(0,10^3)$) resulted in a slight reduction in the heterogeneity standard deviation (from 0.01 to 0.02 units).

5.7 Discussion

This Chapter illustrates the use of network meta-regression and network meta-epidemiology to investigate the impact of five trial characteristics (four RoB components and study precision) on treatment effect estimates. All applied

models require assumptions for the direction of bias. These assumptions are simplified in the meta-epidemiological studies, where I considered primarily only star-shaped networks. I exemplify the several possible assumptions for a full network in the case study of Section 5.6.

I did not find the inappropriate conduct of random sequence generation, allocation concealment, blinding of participants and blinding of outcome assessors to influence substantially the estimated relative effects in trials. A recent study that synthesized data from seven (standard) meta-epidemiological studies resulted in stronger associations (105). The different findings between the two studies might be due to differences in power (31% more trials in Savovic et al.). An alternative explanation could be the fact that NMA is a time-consuming and resource-demanding procedure; hence researchers undertake NMA mainly in areas where they expect to find many high-quality trials. This is in agreement with the fact that very few studies were assessed at high RoB for the four RoB components (see **Table 5.2**)**.** This apparent lack of association could be also attributed to random misclassification because of the different criteria the original authors of the networks may have used to evaluate the RoB in individual trials. However, the sensitivity analysis that considered only networks, for which RoB data were extracted subsequently using the criteria of Section 5.4.2, did not yield different results from the primary analysis. Another reason that may explain this finding is that the RoB assessement relies also on the quality of reporting. For example, older studies might be more likely to be classified at high/unlear RoB due to poor reporting of the methods.

In this empirical study, trials being at high or unclear RoB were assumed to have similar effect modification due to their design limitations. The use of weaker assumptions is possible via models that assign a probability to the unclear studies of being at high RoB (112). However, the absence of many high risk studies in this database did not allow the use of such probabilistic models, since they would have very limited power in the estimation of coefficients. Study precision sometimes is considered as a proxy for study quality, if information of

RoB characteristics is not available. Nevertheless, this network collection did not confirm a clear association between study precision and the four RoB items, which was suggested by Nuesch et al. (95) and Kjaergard et al. (124). In addition, other types of biases (e.g. attrition bias or selective outcome reporting), which were not taken into account in the analyses, may operate as well.

I found a strong association between study precision and treatment effects for dichotomous outcomes using the same collection of star-shaped networks. This implies that important differences in the distribution of study precision across comparisons within a network can cause intransitivity and/or inconsistency. Hence, researchers should routinely consider the differences in precision across studies in the evaluation of transitivity and exploration of inconsistency. The use of star networks in this empirical study makes the evaluation of statistical inconsistency impossible.

Although continuous outcomes might be expected more often to be subject to selective outcome reporting (i.e. continuous measures are more frequently used for subjective and secondary outcomes), small-study effects did not appear to affect the treatment effects for such outcomes. This finding might be explained by limited power (only 16 networks with continuous outcomes were synthesized) or by the presence of larger heterogeneity across the different continuous outcome measures.

Accounting for the potential of biased treatment effects within a full network requires stronger assumptions for the direction of bias. Models assuming that bias favors one of the treatments in each comparison with an unknown probability can be employed, if there is no prior belief about the direction of bias. I found presence of small-study effects in both examples of full networks. In the network of incidence of diabetes the direction of bias did not consistently favor the newer or older interventions in all studies. However, since information on other trial characteristics was not available, study precision and year of publication partly explained the estimated heterogeneity and inconsistency in this network.

Despite the above findings, exploration of heterogeneity and inconsistency should not be based solely on the investigation of differences in precision across studies (99). Other trial characteristics associated with both study size or study precision and relative effects may operate as well. For example, small and large trials can be different with respect to their design limitations, population baseline severity of disease or other characteristics. When such information is not readily available, study precision may be used as a good proxy for these characteristics. However, this was not the case in my empirical studies, where study precision was not clearly associated with the four studied RoB items. Also, information from trial registries can be used to identify unpublished studies to explore whether small-study effects are caused by publication bias (96,125).

The findings of this Chapter illustrate that NMA may suffer the same biases as pairwise meta-analysis. However, the presence of multiple treatment comparisons in NMA is likely to differentiate the mechanisms of bias within the same network and this should be taken into account in the analysis. In the presence of small-study effects, results should be presented using the relative effects from network meta-regression models (extrapolated to the largest study in the network) (113). Other approaches, such as selection models (126), can be employed, if small-study effects are caused by publication bias.

Both the empirical and case studies that I conducted highlight that there is large potential for drawing misleading conclusions when the impact of several factors is ignored in a NMA. Therefore, an overall final judgement of the quality of evidence and the risk for several types of bias is crucial for the interpretation of the results (60).

6 Summary

6.1 Summary

Systematic reviews and meta-analyses have been established as an integral part of comparative effectiveness research and are widely used by international health-care institutions to inform policy-making. However, the increasing number of competing interventions in many medical fields has led to the development of NMA. This new evidence synthesis tool integrates direct with indirect evidence to infer about the relative effect of any pair of treatments included in the network. In this way, network estimates usually have increased precision in comparison to the respective direct estimates, while inference can be drawn also for comparisons not evaluated in individual studies. The inclusion of all competing treatments in the same meta-analytic model allows the estimation of their relative ranking. These advantages have rendered NMA an increasingly popular statistical tool.

NMA can be seen from several perspectives that are, in principal, equivalent, but differ with respect to the ease of implementation in standard software packages. All approaches rely on the fundamental assumption of transitivity, which suggests that the common comparator treatment X is similar (e.g. administered the same way and in similar populations) in the XY and XZ studies. At the level of model parameters transitivity is reflected by the consistency equations, which imply that direct and indirect evidence in the network are in agreement regarding the true (summary) underlying relative effects. If the consistency assumption does not hold in parts or in the entire network, the results of NMA might be questionable.

The rapid development of NMA methodology underlines the need for flexible and user-friendly software options that would facilitate the appropriate conduct

and comprehensive reporting of the analysis. To aid the ease of presentation and interpretation of NMA, I developed new and modified existing graphical presentation tools, which I implemented in the STATA software. These graphs can be used to present the complex evidence base, assumptions and results from NMA and aim to make these outputs understandable to researchers that are less familiar with advanced statistical methods.

The fact that NMA is a relatively new statistical tool implies the need for several empirical and simulation studies to investigate the properties of each approach. An overview of the characteristics of published networks of interventions is a useful resource of information for methodologists that aim to update the current knowledge on appraising NMA methods. To this end, I compiled networks of interventions published until March 2011. This collection demonstrates that the advantages of NMA have increased its acceptance by the medical research community. Also, it suggests that the observed variation in the choice of the method of analysis can be partly explained by the size characteristics of the networks. The typical network of this database includes 22 studies and compares 5 pharmacological treatments vs. a control intervention with respect to a dichotomous subjective outcome.

Heterogeneity and inconsistency can be seen as differences in the potential effect modifiers within and across the pairwise comparisons in a network of interventions. It is often the case that the estimated heterogeneity or inconsistency in a network of interventions can be explained using a network meta-regression model; that is a NMA model that incorporates one or more covariates. A usual issue of any meta-regression model is the lack of power to detect associations between treatment effects and study-level characteristics. To date, researchers have usually relied on the comparability of effect modification parameters, in terms of both magnitude and direction, across different meta-analyses to overcome the issue of low power. In this way, they expect that a sizable amount of data would be available to allow the adequate estimation of a relationship between treatment effects and other factors. Data from networks of

interventions usually involve a larger evidence base and may offer a promising alternative way to investigate the mechanisms of effect modification and bias in network meta-epidemiological research. Pooling parameters across many network meta-analyses can accommodate the assumption for these parameters being more similar within a network rather than across networks.

Several empirical studies have examined the possible association between the estimated treatment effects and the size of the trials suggesting that small studies tend to give larger effect estimates than do larger studies. Such a phenomenon, which is usually called small-study effects, is quite frequent in evidence synthesis research and can be caused by several reasons, such as publication bias, selective outcome reporting, differences in study quality or genuine heterogeneity. To assess graphically the presence of small-study effects in NMA, I extended the standard funnel plot to allow the incorporation of multiple treatment comparisons. Network meta-regression models can be used also to extend the statistical models accounting for small-study effects into the context of NMA. An important issue of the network meta-regression models that aim to control for potentially biased treatment effects is that they require making assumptions about the directionality of bias. This may be less problematic when analysing star-shaped networks; in these networks the common comparator intervention is usually an inactive or old intervention and it is not expected to be favored when bias is present. Inference on the plausibility of the assumed direction of bias is based on the estimated coefficients.

I investigated the impact of five trial characteristics (four RoB components and study precision) on treatment effect estimates using the methodology of network meta-regression and meta-epidemiology on a collection of 32 star-shaped networks. In this analysis, inadequate random sequence generation, allocation concealment, blinding of participants and blinding of outcome assessors were not found to influence substantially the estimated relative effects in trials. Study precision sometimes is considered as a proxy for study quality, if information of

RoB characteristics is not available. Nevertheless, my network collection did not confirm a clear association between study precision and the four RoB items.

I found a strong association between study precision and treatment effects for dichotomous outcomes using the same collection of star-shaped networks. This implies that important differences in the distribution of study precision across comparisons within a network may cause intransitivity and/or inconsistency. Additional empirical studies are necessary to confirm such a conclusion. However, this finding implies that researchers should routinely consider the differences in precision across studies in the evaluation of transitivity and exploration of inconsistency. Although continuous outcomes might be expected more often to be subject to selective outcome reporting (i.e. continuous measures are more frequently used for subjective and secondary outcomes), small-study effects did not appear to affect the treatment effects for such outcomes.

Accounting for the potential of biased treatment effects within a full network requires stronger assumptions for the direction of bias. I exemplified such assumptions using two full networks both evaluating typically secondary (safety) outcomes; these are failure of vascular graft or arterial patency with aspirin, dipyridamole or placebo, and incidence of diabetes with antihypertensive drugs. I also employed models assuming that bias favors one of the treatments in each comparison with an unknown probability, when there is not prior belief about the direction of bias. I found presence of small-study effects in both examples. In the network of incidence of diabetes the direction of bias did not consistently favor the newer or older interventions in all studies. However, since information on other trial characteristics was not available, study precision and year of publication partly explained the estimated heterogeneity and inconsistency in this network.

Despite the above findings, exploration of heterogeneity and inconsistency should not be based solely on the investigation of differences in precision across studies. Other trial characteristics associated with both study size or study precision and relative effects may operate as well. For example, small and large

trials can be different with respect to their design limitations, population baseline severity of disease or other characteristics. Also, information from trial registries can be used to identify unpublished studies to explore whether small-study effects are caused by publication bias.

The findings of this Thesis illustrate that NMA may suffer the same biases as pairwise meta-analysis. However, the presence of multiple treatment comparisons in NMA is likely to differentiate the mechanisms of bias within the same network and this should be taken into account in the analysis. The empirical and case studies that I present in Chapter 5 highlight that there is large potential for drawing misleading conclusions when the impact of several factors is ignored in a NMA. Therefore, an overall final judgment of the quality of evidence and the risk for several types of bias is crucial for the interpretation of the results.

6.2 Περίληψη

Οι συστηματικές ανασκοπήσεις και οι μετα-αναλύσεις έχουν καθιερωθεί ως αναπόσπαστο κομμάτι στην έρευνα που σχετίζεται με την συγκριτική αποτελεσματικότητα των θεραπειών (comparative effectiveness research). Συγκεκριμένα, χρησιμοποιούνται πλέον ευρέως από διεθνείς οργανισμούς υγείας με σκοπό να συμβάλλουν στη διαμόρφωση πολιτικής για τη δημόσια υγεία. Παρόλα, αυτά ο συνεχώς αυξανόμενος αριθμός των ανταγωνιστικών θεραπειών σε πολλούς τομείς της ιατρικής οδήγησε στην ανάπτυξη της *μετα-ανάλυσης δικτύων* (ΜΑΔ) (network meta-analysis), γνωστή και ως μετα-ανάλυση πολλαπλών παρεμβάσεων. Η μετα-ανάλυση δικτύων συνθέτει το σύνολο της πληροφορίας που μπορεί να προέρχεται είτε άμεσα από μεμονωμένες μελέτες είτε έμμεσα χρησιμοποιώντας μελέτες που συνδέονται σε ένα δίκτυο θεραπειών. Με αυτόν τον τρόπο οι εκτιμήσεις της ΜΑΔ συνήθως είναι πιο ακριβείς από τις αντίστοιχες

άμεσες εκτιμήσεις, ενώ μπορούν να εξαχθούν συμπεράσματα για τη σχετική αποτελεσματικότητα θεραπειών που δεν έχουν συγκριθεί σε κάποια μελέτη. Επίσης, το γεγονός ότι όλες οι ανταγωνιστικές θεραπείες συγκρίνοται ταυτόχρονα στο ίδιο στατιστικό μοντέλο δίνει τη δυνατότητα να εκτιμηθεί η σχετική κατάταξή τους. Τα πλεονεκτήματα αυτά έχουν καταστήσει τη ΜΑΔ ένα πολύ δημοφιλές στατιστικό εργαλείο.

Η ΜΑΔ μπορεί να προσεγγιστεί από διαφορετικές οπτικές γωνίες, οι οποίες είναι θεωρητικά ισοδύναμες, αλλά διαφέρουν στην δυνατότητα εφαρμογής τους στα συνήθη λογισμικά. Όλες οι προσεγγίσεις βασίζονται στη θεμελιώδη υπόθεση της *μεταβατικότητας* (transitivity), που υπαγορεύει ότι η κοινή θεραπεία X είναι όμοια (π.χ. χορηγείται με τον ίδιο τρόπο σε όμοιους πληθυσμούς) στις μελετες XY και στις μελέτες XZ. Σε επίπεδο παραμέτρων του μοντέλου, η μεταβατικότητα εκφράζεται μέσω των *εξισώσεων συνέπειας* (consistency equations), που συνεπάγονται ότι οι άμεση πληροφορία συμφωνεί με την έμμεση, όσον αφορά στις 'πραγματικές' συνοπτικές σχετικές επιδράσεις. Αν η *υπόθεση της συνέπειας* δεν ισχύει στο δίκτυο, τα αποτελέσματα της ΜΑΔ πιθανόν να μην είναι αξιόπιστα.

Η ραγδαία ανάπτυξη της μεθοδολογίας για τη ΜΑΔ υπογραμμίζει την ανάγκη για ευέλικτες και φιλικές στο χρήστη επιλογές λογισμικού που θα διευκολύνουν την κατάλληλη διεξαγωγή και τη κατανοητή αναφορά της ανάλυσης. Προκειμένου να συνεισφέρω στην ευκολία της παρουσίασης και ερμηνείας της ΜΑΔ, ανέπτυξα καινούρια ή τροποποίησα ήδη υπάρχοντα γραφήματα, τα οποία έκανα διαθέσιμα στο ευρύ κοινό προγραμματίζοντας τα στο λογισμικό STATA. Αυτά τα γραφήματα μπορούν να χρησιμοποιηθούν για την παρουσίασης της πολύπλοκης βάσης δεδομένων, των υποθέσεων και των αποτελεσμάτων από τη ΜΑΔ. Σκοπός τους είναι να κάνουν τη διαδικασία της ΜΑΔ πιο κατανοητή σε ερευνητές χωρίς ισχυρές στατιστικές γνώσεις και ικανότητες.

Το γεγονός ότι η ΜΑΔ είναι ένα σχετικά καινούριο στατιστικό εργαλείο συνεπάγεται την ανάγκη για εμπειρικές μελέτες και μελέτες προσομοίωσης που θα διερευνούν τις ιδιότητες των διαφορετικών μεθόδων. Η επισκόπηση των χαρακτηριστικών δημοσιευμένων δικτύων θεραπειών είναι χρήσιμη πηγή

πληροφόρησης για ερευνητές που σκοπεύουν να αναβαθμίσουν την υπάρχουσα γνώση όσον αφορά στην αποτίμηση των μεθόδων της ΜΑΔ. Για το σκοπό αυτόν, δημιούργησα μια συλλογή δικτύων θεραπειών που δημοσιεύτηκαν μέχρι το Μάρτιο του 2011. Η σθλλογή αυτή δείχνει ότι τα πλεονεκτήματα της ΜΑΔ έχουν αυξήσει την αποδοχή της από την ιατρική ερευνητική κοινότητα. Επίσης, συστήνει ότι οι διαφορές που παρατηρούνται στην επιλογή της μεθόδου ανάλυσης μπορεί να εξηγηθεί μερικώς από διαφορές στο μέγεθος και τη δομή των δικτύων. Ένα τυπικό δίκτυο της βάσης αυτής περιλαμβάνει 22 μελέτες και συγκρίνει 5 φαρμακευτικές θεραπείες έναντι μιας θεραπείας ελέγχου σχετικά με μια διχότομη υποκειμενική έκβαση.

Η *ετερογένεια* (heterogeneity) και η *ασυνέπεια* (inconsistency) εμφανίζονται ως διαφορές στους πιθανούς τροποποιητές επίδρασης (effect modifiers) μέσα και ανάμεσα στις συκρίσεις σε ένα δίκτυο θεραπειών. Συχνά υπάρχουν περιπτώσεις που η εκτιμώμενη ετερογένεια ή ασυνέπεια σε ένα δίκτυο μπορεί να εξηγηθεί χρησιμοποιώντας *μετα-παλινδρόμηση δικτύου* (network meta-regression), δηλαδή ένα μοντέλο ΜΑΔ που περιλαμβάνει επεξηγηματικές μεταβλητές (covariates). Ένα σύνηθες ζήτημα της μετα-παλινδόμησης είναι η έλλειψη επαρκούς ισχύος για τον εντοπισμό σχέσεων ανάμεσα στην επίδραση των θεραπειών και χαρακτηριστικών των μελετών. Μέχρι σήμερα οι ερευνητές για να ξεπεράσουν το πρόβλημα αυτό, συνήθως βασίζονται στη συγκρισιμότητα παραμέτρων, λαμβάνοντας υπόψη το μέγεθος και την κατεύθυνση της επίδρασης, ανάμεσα σε διαφορετικές μετα-αναλύσεις. Με αυτόν τον τρόπο, αναμένουν ότι θα έχουν διαθέσιμα αρκετά δεδομένα που θα επιτρέπουν επαρκή εκτίμηση της σχέσης των θεραπευτικών επιδράσεων με άλλους παράγοντες. Τα δεδομένα δικτύων θεραπειών συνήθως συνεπάγονται μεγαλύτερη βάση δεδομένων έτσι ίσως προσφέρουν ένα υποσχόμενο εναλλακτικό μέσο για τη διερεύνηση των μηχανισμών πιθανών συστηματικών σφαλμάτων *στη μετα-επιδημιολογική έρευνα δικτύων* (network meta-epidemiology). Η σύνθεση παραμέτρων από πολλές ΜΑΔ μπορεί να υποστηρίξει την υπόθεση ότι οι παράμετροι αυτοί είναι πιο όμοιοι μέσα στο ίδιο δίκτυο παρά ανάμεσα στα δίκτυα.

Διάφορες εμπειρικές μελέτες έχουν εξετάσει την πιθανή σχέση μεταξύ των εκτιμώμενων θεραπευτικών επιδράσεων και και του μεγέθους των μελετών, συστήνοντας ότι οι μικρότερες μελέτες έχουν την τάση να δίνουν μεγαλύτερες επιδράσεις από τις μεγαλύτερες μελέτες. Αυτό το φαινόμενο, που συνήθως ονομάζεται *επίδραση των μικρών μελετών* (small-study effects) είναι αρκετά συχνό στις μετα-αναλύσεις και μπορεί να προκληθεί από πολλές αιτίες, όπως το συστηματικό *σφάλμα δημοσίευσης* (publication bias), η *επιλεκτική αναφορά αποτελεσμάτων* (selective outcome reporting), διαφορές στην ποιότητα των μελετών ή ετερογένεια. Προκειμένου να καταστήσω δυνατή την γραφική αξιολόγηση της παρουσίας επίδρασης μικρών μελετών στη ΜΑΔ, επέκτεινα τα συνήθη σχετικά γραφήματα (funnel plot) ώστε να ενσωματώσουν τις διαφορετικές συγκρίσεις θεραπειών. Μοντέλα μετα-παλινδρόμησης δικτύων μπορούν, επίσης, να χρησιμοποιηθούν για να επεκτείνουν τα υπάρχοντα στατιστικά μοντέλα που λαμβάνουν υπόψη την επίδραση των μικρών μελετών στο πλαίσιο της ΜΑΔ. Ένα σημαντικό ζήτημα των μοντέλων μετα-παλινδρόμησης δικτύων που σκοπεύουν να διορθώσουν πιθανώς εσφαλμένες εκτιμήσεις επιδράσεων είναι ότι απαιτούν τη χρήση υποθέσεων για την κατεύθυνση του σφάλματος. Αυτές οι υποθέσεις ίσως είναι λιγότερο ισχυρές στην περίπτωση των δικτύων σε σχήμα αστεριού, καθώς σε αυτά τα δίκτυα η κοινή παρέμβαση είναι συνήθως μια ανενεργή ή παλιά θεραπεία που δεν αναμένεται να ευνοείται λόγω συστηματικών σφαλμάτων. Τα συμπεράσματα όσον αφορά στην σωστή επιλογή υπόθεσης για την κατεύθυνση του σφάλματος βασίζονται στους εκτιμώμενους συντελεστές από τη μετα-παλινδρόμηση.

Διερεύνησα την επίδραση πέντε χαρακτηριστικών των μελετών (τεσσάρων στοιχείων σχεδιασμού των μελετών (RoB items) και την ακρίβειά τους (study precision)) στις θεραπευτικές επιδράσεις χρησιμοποιώντας τη μεθοδολογία μετα-παλινδρόμησης και μετα-επιδημιολογίας δικτύων σε μια συλλογή 32 δικτύων-αστεριών. Στην ανάλυση αυτή, η μη-κατάλληλη διεξαγωγή της δημιουργίας τυχαίας αλληλουχίας, της απόκρυψης της κατανομής, της τυφλοποίησης των συμμετεχόντων και της τυφλοποίησης των αξιολογητών της έκβασης (random

sequence generation, allocation concealment, blinding of participants and blinding of outcome assessors) δε φάνηκε να επηρεάζει σημαντικά τα αποτελέσματα των αποτελεσμάτων της ΜΑΔ. Πολλές φορές το μέγεθος των μελετών θεωρείται ως προβλεπτικός παράγοντας για την ποιότητά τους, όταν δεν υπάρχει πληροφορία σχετικά με την καταλληλότητα του σχεδιασμού τους. Παρόλα αυτά, η συγκεκριμένη βάση δικτύων θεραπειών που δημιούργησα δεν επιβεβαίωσε ξεκάθαρα τη συσχέτιση μεταξύ σωστού σχεδιασμού και μεγέθους των μελετών.

Χρησιμοποιώντας την ίδια συλλογή δικτύων-αστεριών, βρήκα ισχυρή συσχέτιση ανάμεσα στην ακρίβεια των μελετών και στις θεραπευτικές επιδράσεις. Το γεγονός αυτό ίσως δείχνει ότι σημαντικές διαφορές στην κατανομή της διασποράς των εκτιμήσεων από τις μεμονωμένες μελέτες ανάμεσα στις συγκρίσεις ενός δικτύου μπορεί να απειλήσει την υπόθεση της μεταβατικότητας και/ή της συνέπειας. Περισσότερες εμπειρικές μελέτες θα πρέπει να διεξαχθούν για να επιβεβαιωθεί το παραπάνω συμπέρασμα. Οι ερευνητές όμως θα πρέπει πάντα να λαμβάνουν υπόψη την επίδραση των μικρών μελετών κατά την αξιολόγηση της μεταβατικότητας και τη διερεύνηση της ασυνέπειας. Παρόλο που οι συνεχείς εκβάσεις αναμένονται να είναι πιο συχνά αντικείμενο της επιλεκτικής αναφοράς αποτελεσμάτων (διότι συνεχή μέτρα επίδρασης χρησιμοποιούνται πιο συχνά για υποκειμενικές και δευτερεύοντες εκβάσεις), δε βρέθηκε σημαντική επίδραση των μικρών μελετών σε τέτοιου είδους δεδομένα.

Η διόρθωση των αποτελεσμάτων για πιθανόν εσφαλμένες εκτιμήσεις σε ένα ‘πλήρες’ δίκτυο θεραπειών (full network) απαιτεί τη χρήση πιο ισχυρών υποθέσεων για την κατεύθυνση του συστηματικού σφάλματος. Διερεύνησα αυτές τις υποθέσεις χρησιμοποιώντας δύο πλήρη δίκτυα που αξιολογούν δευτερεύουσες εκβάσεις (ασφάλεια). Οι εκβάσεις αυτές είναι: αποτυχία του αγγειακού μοσχεύματος ή αρτηριακής βατότητας με χρήση ασπιρίνης, διπυριδαμόλης ή εικονικού φάρμακου, και η συχνότητα εμφάνισης του διαβήτη με αντιϋπερτασικά φάρμακα. Επιπρόσθετα, χρησιμοποίησα μοντέλα που υποθέτουν ότι μια θεραπεία ευνοείται με μια άγνωστη πιθανότητα στην περίπτωση που δεν

υπάρχει ισχυρή πεποίθηση για την κατεύθυνση του σφάλματος. Και στα δύο παραδείγματα βρήκα παρουσία επίδρασης των μικρών μελετών. Στο δίκτυο της εμφάνισης διαβήτη η επίδραση αυτή δεν ευνοούσε πάντα τις νεότερες ή τις παλιότερες θεραπείες σε όλες τις μελέτες. Παρόλα αυτά, εφόσον δεν υπήρχαν πληροφορίες για άλλα χαρακτηριστικά των μελετών, το μέγεθος και η χρονολογία δημοσίευσης φάνηκαν να εξηγούν μερικώς την ετερογένεια και την ασυνέπεια στο δίκτυο.

Παρά τα παραπάνω αποτελέσματα, η διερεύνηση της ετερογένειας και της ασυνέπειας δε θα πρέπει να βασίζεται μόνο στον έλεγχο του διαφορετικού μεγέθους των μελετών. Άλλα χαρακτηριστικά που σχετίζονται και με το μέγεθος ή την ακρίβεια των μελετών και τις θερπευτικές επιδράσεις ίσως να λειτουργούν επίσης ως τροποποιητές επίδρασης. Για παράδειγμα, οι μικρές οι μικρότερες μελέτες μπορεί να διαφέρουν από τις μεγαλύτερες στο σχεδιασμό, το βασικό κίνδυνο (baseline risk), ή άλλα χαρακτηριστικά. Επίσης, οι καταγραφές των τυχαιοποιημένων μελετών θα πρέπει να ελέγχονται για τον εντοπισμό μη δημοσιευμένων μελετών που μπορεί να προκαλούν την επίδραση των μικρών μελετών μέσω σφάλματος δημοσίευσης.

Τα ευρύματα της παρούσας διατριβής τονίζουν ότι ΜΑΔ συχνά πάσχει από τα ίδια συστηματικά σφάλματα με την απλή μετα-ανάλυση (pairwise meta-analysis). Παρόλα αυτά, η παρουσία πολλών διαφορετικών συγκρίσεων μεταξύ θεραπειών είναι πιθανό να διαφοροποιεί το μηχανισμό δράσης των σφαλμάτων αυτών μέσα στο ίδιο δίκτυο. Το γεγονός αυτό θα πρέπει να λαμβάνεται υπόψη στην ανάλυση. Οι εμπειρικές και οι μεμονωμένες μελέτες (case studies) που παρουσιάζω στο κεφάλαιο 5 δείχνουν ότι υπάρχει μεγάλη πιθανότητα διεξαγωγής παραπλανητικών συμπερασμάτων όταν αγνοείται η επίδραση διάφορων παραγόντων σε μια ΜΑΔ. Για το λόγο αυτόν, είναι πολύ σημαντικό για την ερμηνεία των αποτελεσμάτων να πραγματοποιείται μια γενική κρίση όσον αφορά στην ποιότητα των δεδομένων και το κίνδυνο για διάφορα είδη σφαλμάτων.

Appendix

A. Proof of the mathematical relationship between SUCRA and mean rank

$$SUCRA_t = \frac{\sum_{r=1}^{T-1} cp_{rt}}{T-1} \Rightarrow SUCRA_t = \sum_{r=1}^{T-1}\left[\frac{(T-r)\times p_{rt}}{T-1}\right] mr_t = \sum_{r=1}^{T}(p_{rt}\times r)$$

Subtracting the two above equations gives that

$$mr_t - SUCRA_t = \sum_{r=2}^{T}\left[\left(r - \frac{T-r}{T-1}\right)\times p_{rt}\right] = \frac{T}{T-1}\sum_{r=2}^{T}[(r-1)\times p_{rt}]$$

$$\Rightarrow SUCRA_t = mr_t + \frac{T}{T-1}\sum_{r=2}^{T}[(r-1)\times p_{rt}] = mr_t + \frac{T}{T-1}\times\left(mr_t - \sum_{r=1}^{T} p_{rt}\right)$$

$$\Rightarrow SUCRA_t = mr_t + \frac{T}{T-1}\times(mr_t - 1)$$

$$\Rightarrow SUCRA_t = \frac{(2T-1)\times mr_t - T}{T-1}$$

B. WinBUGS codes

Network meta-epidemilogical model accounting for small-study effects linking networks with dichotomous arm-level data

```
# nn=number of networks
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment

model {
```

```
for (z in 1:nn) {
for(i in 1:ns[z]) {
w[z,i,1]<- 0
theta[z,i,t[z,i,1]]<- 0
mu[z,i] ~ dnorm(0,.0001)
for (k in 1:na[z,i]) {
r[z,i,t[z,i,k]]~dbin(p[z,i,t[z,i,k]],n[z,i,t[z,i,k]])}
logit(p[z,i,t[z,i,1]])<- mu[z,i]
for (k in 2:na[z,i]) {
lamda.1[z,i,k]<- equals(r[z,i,t[z,i,k]],0)
lamda.2[z,i,k]<- equals(n[z,i,t[z,i,k]],r[z,i,t[z,i,k]])
lamda.3[z,i,k]<- equals(r[z,i,t[z,i,1]],0)
lamda.4[z,i,k]<- equals(n[z,i,t[z,i,1]],r[z,i,t[z,i,1]])
lamda.a[z,i,k]<- max(lamda.1[z,i,k],lamda.2[z,i,k])
lamda.b[z,i,k]<- max(lamda.3[z,i,k],lamda.4[z,i,k])
lamda[z,i,k]<-max(lamda.a[z,i,k],lamda.b[z,i,k])
var[z,i,k]<-
1/(r[z,i,t[z,i,k]]+(0.5*lamda[z,i,k]))+1/(r[z,i,t[z,i,1]]+(0.5*lamda[z,i
,k]))+1/(n[z,i,t[z,i,k]]-
r[z,i,t[z,i,k]]+(0.5*lamda[z,i,k]))+1/(n[z,i,t[z,i,1]]-
r[z,i,t[z,i,1]]+(0.5*lamda[z,i,k]))
logit(p[z,i,t[z,i,k]])<- mu[z,i] + theta1[z,i,t[z,i,k]]
I[z,i,k]<-0.5*(ind[z,i,k]-ind[z,i,1])
theta1[z,i,t[z,i,k]]<- theta[z,i,t[z,i,k]] + B[z]*var[z,i,k]*I[z,i,k]
theta[z,i,t[z,i,k]] ~ dnorm(md[z,i,t[z,i,k]],precd[z,i,t[z,i,k]])
precd[z,i,t[z,i,k]]<- prec[z] *2*(k-1)/k
md[z,i,t[z,i,k]]<-d[z,t[z,i,k]]-d[z,t[z,i,1]]+sw[z,i,k]
w[z,i,k]<- (theta[z,i,t[z,i,k]]  - d[z,t[z,i,k]] + d[z,t[z,i,1]])
sw[z,i,k]<- sum(w[z,i,1:k-1])/(k-1)}}
tau[z] ~ dnorm(0,1)I(0,)
prec[z]<- 1/pow(tau[z],2)
d[z,ref[z]]<- 0
for(k in 1:(ref[z]-1)) {d[z,k] ~ dnorm(0,.0001)}
for(k in (ref[z]+1):nt[z]) {d[z,k] ~ dnorm(0,.0001)}
B[z] ~ dnorm(Boverall,W)}
Boverall ~ dnorm(0,.0001)
W<- 1/(prec.Boverall*prec.Boverall)
prec.Boverall ~ dunif(0,3)
for (z in 1:nn) {for(i in 1:ns[z]) {for (k in 1:na[z,i]) {
Darm[z,i,k]<--2*(r[z,i,t[z,i,k]]
*log(n[z,i,t[z,i,k]]*p[z,i,t[z,i,k]]/r[z,i,t[z,i,k]])+(n[z,i,t[z,i,k]] -
r[z,i,t[z,i,k]])*log((n[z,i,t[z,i,k]]-n[z,i,t[z,i,k]]*
p[z,i,t[z,i,k]])/(n[z,i,t[z,i,k]]- r[z,i,t[z,i,k]])))}
Dstudy[z,i]<- sum(Darm[z,i,1:na[z,i]])}
Dnetwork[z]<- sum(Dstudy[z,1:ns[z]])}
D.bar<- sum(Dnetwork[])
}
```

Network meta-regression accounting for small-study effects in full networks (Section 5.6)

Example 1. Failure of vascular graft or arterial patency

Model 1.1: No covariates

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
logit(p[i,t[i,k]])<- u[i] + theta[i,t[i,k]]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {
for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {
for (k in 1:na[i]) {
```

```
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 1.2: 'Active-favored' & identical coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k  }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
B ~ dnorm(0,0.0001)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {
for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
```

```
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 1.3: 'Active-favored' & exchangeable coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + beta[t[i,1],t[i,k]]*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
```

```
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
beta[2,3] <- 0
for (j in 2:nt) {beta[1,j] ~ dnorm(B, prec.B) }
B ~ dnorm(0,0.0001)
prec.B <- 1/(sigma*sigma)
sigma ~ dnorm(0,1)I(0,)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 1.4: 'New-favored' & identical coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
```

```
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))
+1/(n[i,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]] - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
B ~ dnorm(0,0.0001)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {
for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {
for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 1.5: 'New favored' & exchangeable coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
```

```
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + beta[t[i,1],t[i,k]]*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]] - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
for(i in 1:(nt-1)) {
for (j in (i+1):nt){
beta[i,j] ~ dnorm(B, prec.B) }}
B ~ dnorm(0,0.0001)
prec.B <- 1/(sigma*sigma)
sigma ~ dnorm(0,1)I(0,)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {
for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
```

```
}
```

Model 1.6: 'Active-favored' (for the basic coefficietns) & consistent coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + beta1[t[i,k]]*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
for(i in 1:(nt-1)) {
for (j in (i+1):nt){
beta[i,j] <- beta1[j] - beta1[i] }}  # consistency assumption in betas
beta1[1] <- 0
for (j in 2:nt) {beta1[j] ~ dnorm(0,0.0001)}
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {
for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
```

```
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 1.7: Probabilistic 'new-favored' & identical coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
b[i,1] <- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B*var[i,k]*x[i,t[i,k]]
x[i,t[i,k]]<- b[i,t[i,k]]*I[i,k] # probabilistic direction
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
```

```
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
B ~ dnorm(0,0.0001)
for (i in 1:6) {b[i,2] ~ dbern(P1)}          # 3-arm trials
for (i in 1:6) {b[i,3] ~ dbern(P2)}
for (i in 7:10) {b[i,3] ~ dbern(P3)}        # head-to-head
for (i in 11:24) {b[i,3] ~ dbern(P2)}     # placebo-controlled
for (i in 25:31) {b[i,2] ~ dbern(P1)}
P1 ~ dbeta(1,1)   # favoring A compared to P
P2 ~ dbeta(1,1)   # favoring D compared to P
P3 ~ dbeta(1,1)   # favoring D compared to A
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {
for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {
for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Example 2. Incidence of Diabetes

Model 2.1: No covariates

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
```

```
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
logit(p[i,t[i,k]])<- u[i] + theta[i,t[i,k]]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]] - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {
for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {
for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 2.2: Active favored only - Fixed coefficient

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
```

```
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
B ~ dnorm(0,0.0001)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative    #
change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 2.3: 'New-favored' & identical coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
```

```
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))
+1/(n[i,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
B ~ dnorm(0,0.0001)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 2.4: 'Sponsored-favored' & identical coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))
+1/(n[i,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
B ~ dnorm(0,0.0001)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
```

```
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 2.5: 'Sponsored/New-favored' & identical coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
B ~ dnorm(0,0.0001)
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
```

```
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {
for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 2.6: 'Active-favored' (for the basic coefficients) & consistent coefficients

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + beta1[t[i,k]]*var[i,k]*I[i,k]
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
```

```
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
for(i in 1:(nt-1)) {
for (j in (i+1):nt){
beta[i,j] <- beta1[j] - beta1[i] }}  # consistency assumption in betas
beta1[1] <- 0
for (j in 2:nt) {beta1[j] ~ dnorm(0,0.0001)}
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 2.7: 'Sponsored/new-favored' & subgroups by sponsoring

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'
# s=shows sponsored and non-sponsored studies

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
```

```
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B[s[i]]*var[i,k]*I[i,k] # two
different betas for #sponsored and non-sponsored studies
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
for (i in 1:2) { B[i] ~ dnorm(0,0.0001) }
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 2.8: 'Sponsored/new-favored' & subgroups by year

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'
```

```
# y=shows studies published before and after 2000

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))
+1/(n[i,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<-0.5*(Z[i,1]-Z[i,k])
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + B[y[i]]*var[i,k]*I[i,k] # two
different betas for #studies published before and after 2000
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
for (i in 1:2) { B[i] ~ dnorm(0,0.0001) }
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] -r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

Model 2.9: Probabilistic 'Sponsored/new-favored' & subgroup by year

```
# ns=number of studies
# r=number of events
# n=sample size
# t=treatment
# nt=number of treatments
# na=number of arms
# ref=reference treatment
# I='direction variable'
# y=published before or after 2000

model {
for(i in 1:ns) {
w[i,1]<- 0
theta[i,t[i,1]]<- 0
b[i,1]<-0
##binomial likelihood
for (k in 1:na[i]) {r[i,t[i,k]] ~ dbin(p[i,t[i,k]],n[i,t[i,k]])}
##parameterization
logit(p[i,t[i,1]])<- u[i]
for (k in 2:na[i]) {
lamda.1[i,k]<- equals(r[i,t[i,k]],0)
lamda.2[i,k]<- equals(n[i,t[i,k]],r[i,t[i,k]])
lamda.3[i,k]<- equals(r[i,t[i,1]],0)
lamda.4[i,k]<- equals(n[i,t[i,1]],r[i,t[i,1]])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])
var[i,k]<-
1/(r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(r[i,t[i,1]]+(0.5*lamda[i,k]))+1/(n[i
,t[i,k]]-r[i,t[i,k]]+(0.5*lamda[i,k]))+1/(n[i,t[i,1]]-
r[i,t[i,1]]+(0.5*lamda[i,k]))
I[i,k]<- (y[i]-1.5)*(Z[i,1]-Z[i,k]) # assumption of opposite direction
for          # studies before and after 2000
logit(p[i,t[i,k]])<- u[i] + theta1[i,t[i,k]]
theta1[i,t[i,k]]<- theta[i,t[i,k]] + beta[y[i]]*var[i,k]*x[i,t[i,k]] #
two different betas # for studies before and after 2000
x[i,t[i,k]]<- b[i,t[i,k]]*I[i,k]    # probabilistic direction
theta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],precd[i,t[i,k]])
md[i,t[i,k]]<- d[t[i,k]] - d[t[i,1]] + sw[i,k]
w[i,k]<- theta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]
sw[i,k]<- sum(w[i,1:k-1])/(k-1)
precd[i,t[i,k]]<- prec *2*(k-1)/k }}
##priors
for (i in 1:ns) {u[i] ~ dnorm(0,.0001)}
tau ~ dnorm(0,1)I(0,)
prec<- 1/pow(tau,2)
for (i in 1:2) {beta[i] ~ dnorm(0,0.001) }
B1 <- -beta[1]          # (notation) opposite direction should give
opposite betas
B2 <- beta[2]
for (i in 1:ns) {
for (k in 2:na[i]) {
b[i,t[i,k]] ~ dbern(P[y[i]]) }}
for (i in 1:2) {P[i] ~ dbeta(1,1) }
d[ref] <- 0
for(k in 1:(ref-1)) {d[k] ~ dnorm(0,.0001)}
for(k in (ref+1):nt) {d[k] ~ dnorm(0,.0001)}
```

```
##estimates
for(i in 1:(nt-1)) {for (j in (i+1):nt) {
OR[j,i]<- exp(d[j] - d[i])
LOR[j,i]<- d[j] - d[i] }}
##ranking
for(k in 1:nt) {
order[k]<- rank(d[],k)  # this is when the outcome is negative
# change to 'order[k]<- nt+1-rank(d[],k)' if the outcome is positive
most.effective[k]<-equals(order[k],1)
for(j in 1:nt) {
effectiveness[k,j]<- equals(order[k],j)
cumeffectiveness[k,j]<- sum(effectiveness[k,1:j]) }}
for(k in 1:nt) { SUCRA[k]<- sum(cumeffectiveness[k,1:(nt-1)]) /(nt-1)}
##model fit
for(i in 1:ns) {for (k in 1:na[i]) {
Darm[i,k]<- -2*( r[i,t[i,k]] *log(n[i,t[i,k]]*p[i,t[i,k]]/
r[i,t[i,k]])+(n[i,t[i,k]] - r[i,t[i,k]])*log((n[i,t[i,k]]-n[i,t[i,k]]*
p[i,t[i,k]])/(n[i,t[i,k]]- r[i,t[i,k]]))) }
D[i]<- sum(Darm[i,1:na[i]]) }
D.bar<- sum(D[])
}
```

C. References of published networks used in empirical analyses

Star-shaped networks

1. Abdullah AK, Khan S. Relative oral corticosteroid-sparing effect of 7 inhaled corticosteroids in chronic asthma: a meta-analysis. Ann Allergy Asthma Immunol 2008; 101(1):74-81.
2. Berner MM, Kriston L, Harms A. Efficacy of PDE-5-inhibitors for erectile dysfunction. A comparative meta-analysis of fixed-dose regimen randomized controlled trials administeringthe International Index of Erectile Function in broad-spectrum populations. Int J Impot Res 2006; 18:229-235.
3. Edwards SJ, Clarke MJ, Wirdsworth S, et al. Carbapenems versus other beta-lactamsin the treatment of hospitalized patients with infection: a mixed treatment comparison. CMRO 2009; 25(1):251-261.
4. Fakhoury W, Lockhart I, Kotchie RW, et al. Indirect comparison of once daily insulin determir and glargine in reducing weight gain hypoglycaemic episodes when administered in addition to convetional oral anti-diabetic therapy in patients with type-2 diabetes. Pharmacology 2008; 82:156-163.
5. Gafter-Gvili A, Fraser A, Paul M, et al. Meta-Analysis: Antibiotic Prophylaxis Reduces Mortality in Neutropenic Patients. Ann Intern Med 2005; 142:979-995.
6. Jansen JJ, Bergman GJD, Huels J, et al. The Efficacy of Bisphosphonates in the Prevention of Vertebral, Hip, and Nonvertebral-Nonhip Fractures in Osteoporosis: A Network Meta-Analysis. Semin Arthrit Rheumat 2011; 40(4):275-284.
7. Kumar A, Hozo I, Wheatley K, et al. Thalidomide versus bortezomib based regimens as first-line therapy for patients with multiple myeloma: a systematic review. Amer J Hematol 2010; 86(1):18-24.
8. Lim E, Ali Z, Ali A, et al. Indirect comparison meta-analysis of aspirin therapy after coronary surgery. BMJ 2003; 327.

9. Lim E, Harris G, Patel A, et al. Preoperative versus Postoperative Chemotherapy in Patients with Resectable Non-small Cell Lung Cancer Systematic Review and Indirect Comparison Meta-Analysis of Randomized Trials. J Thorac Oncol 2009; 4.
10. Loke YK, Kwok CS. Dabigatran and rivaroxaban for prevention of venous thromboembolism – systematic review and adjusted indirect comparison. J Clin Pharm Therap 2011; 36:111-124.
11. Maas T, Kaper J, Sheikh A, et al. Mono and multifaceted inhalant and/or food allergen reduction interventions for preventing asthma in children at high risk of developing asthma. Cochrane Database Syst Rev 2009; (3).
12. Makani H, Bangalore S, Romero J, et al. Effect of Renin-Angiotensin System Blockade on Calcium Channel Blocker-Associated Peripheral Edema. Amer J Med 2011; 124:128-135.
13. Mason L, Moore AR, Edwards JE, et al. Topical NSAIDs for acute pain: a meta-analysis. BMC Family Practice 2004; 5(10).
14. McLeod C, Bagust A, Boland A, et al. Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation. Health Technol Assess 2007; 11(28):1-iv.
15. Mills EJ, Rachlis B, Wu P, et al. Primary prevention of cardiovascular mortality and events with statin treatments: a network meta-analysis involving more than 65,000 patients. J Am Coll Cardiol 2008; 52(22):1769-1781.
16. Mills EJ, Rachlis B, O' Regan C, et al. Metastatic renal cell cancer treatments: An indirect comparison meta-analysis. BMC Cancer 2009; 9(34).
17. Mills EJ, Perri D, Cooper C, et al. Antifungal treatment for invasive Candida infections: a mixed treatment comparison meta-analysis. Ann Clin Microbiol Antimicrob 2009; 8:23.
18. Mills EJ, Wu P, Chong G, et al. Efficacy and safety of statin treatment for cardiovascular disease: a network meta-analysis of 170,255 patients from 76 randomized trials. QJM 2011; 104(2):109-124.
19. Moreno SG, Sutton AJ, Ades AE, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Med Res Methodol 2009; 9:2.
20. Nelson HD, Vesco KK, Haney E, et al. Nonhormonal therapies for menopausal hot flashes: systematic review and meta-analysis. JAMA 2006; 295(17):2057-2071.
21. Peterson K, McDonagh MS, Fu R. Comparative benefits and harms of competing medications for adults with attention-deficit hyperactivity disorder: a systematic review and indirect comparison meta-analysis. Psychopharmacology 2008; 197:1-11.
22. Piccini JP, Hasselblad V, Peterson ED, et al. Comparative Efficacy of Dronedarone and Amiodarone for the Maintenance of Sinus Rhythm in Patients With Atrial Fibrillation. JAmer Coll Cardiol 2009; 54(12).
23. Playford EG, Webster AC, Sorell TC, et al. Systematic review and meta-analysis of antifungal agents for preventing fungal infections in liver transplant recipients. Europ J Clin Microbiol & Infect Dis 2006; 25:549-561.
24. Quilici S, Abrams KR, Nicolas A, et al. Meta-analysis of the efficacy and tolerability of pramipexole versus ropinirole in the treatment of restless legs syndrome. Sleep Med 2008; 9(7):715-726.
25. Rheims S, Perucca E, Cucherat M, et al. Factors determining response to antiepileptic drugs in randomized controlled trials. A systematic review andmeta-analysis. Epilepsia 2011; 52(2):219-233.
26. Rice VH. Nursing intervention and smoking cessation: A meta-analysis. Heart & Lung 1999; 28(6).
27. Singh JA, Christensen R, Wells GA, et al. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. CMAJ 2009; 181(11):787-796.
28. Stettler C, Allemann S, Egger M, et al. Efficacy of drug eluting stents in patients with and without diabetes mellitus: indirect comparison of controlled trials. Heart 2006; 92:650-657.

29. Sultana A, Ghaneh P, Cunningham D, et al. Gemcitabine based combination chemotherapy in advanced pancreatic cancer-indirect comparison. BMC Cancer 2008; 8(192).
30. Uthman O, Abdulmalik J. Comparative efficacy and acceptability of pharmacotherapeutic agents for anxiety disorders in children and adolescents: a mixed treatment comparison meta-analysis. CMRO 2010; 26(1):53-59.
31. Welton NJ, Cooper NJ, Ades AE, et al. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: Evaluation of antivirals for treatment of influenza A and B. Stat Med 2008; 27:5620-5639.
32. Yazdanpanah Y, Sissoko D, Egger M, et al. Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials. BMJ 2004; 328(7434):249.

Full networks

1. Ades AE, Mavranezouli I, Dias S, et al. Network meta-analysis with competing risk outcomes. Value Health 2010; 13(8):976-983.
2. Anothaisintawee T, Attia J, Nickel JC, et al. Management of Chronic Prostatitis/Chronic Pelvic Pain Syndrome A Systematic Review and Network Meta-analysis. JAMA 2012; 305(1).
3. Ara R, Pandor A, Stevens J, et al. Early high-dose lipid-lowering therapy to avoid cardiac events: a systematic review and economic evaluation. Health Technol Assess 2009; 13(34):1-118.
4. Baker WL, Baker EL, Coleman CI. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. Pharmacotherapy 2009; 29(8):891-905.
5. Ballesteros J. 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 2005; 25(2):127-131.
6. Bangalore S, Kumar S, Kjeldsen SE, et al. Antihypertensive drugs and risk of cancer: network meta-analyses and trial sequential analyses of 324,168 participants from randomised trials. Lancet Oncol 2011; 12(1):65-82.
7. Bansback N, Sizto S, Sun H, et al. Efficacy of systemic treatments for moderate to severe plaque psoriasis: systematic review and meta-analysis. Dermatology 2009; 219(3):209-218.
8. Bottomley JM, Taylor RS, Ryttov J. The effectiveness of two-compound formulation calcipotriol and betamethasone dipropionate gel in the treatment of moderately severe scalp psoriasis: a systematic review of direct and indirect evidence. CMRO 2011; 27(1):251-268.
9. Brown TJ, Hooper L, Elliott RA, et al. A comparison of the cost-effectiveness of five strategies for the prevention of non-steroidal anti-inflammatory drug-induced gastrointestinal toxicity: a systematic review with economic modeling. Health Technol Assess 2006; 10(38):iii-xiii.
10. Buscemi N, Vandermeer B, Friesen C, et al. The efficacy and safety of drug treatments for chronic insomnia in adults: a meta-analysis of RCTs. J Gen Intern Med 2007; 22(9):1335-1350.
11. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50(6):683-691.
12. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet 2007; 369(9557):201-207.

13. Goudswaard AN, Furlong NJ, Rutten GE, et al. Insulin monotherapy versus combinations of insulin with oral hypoglycaemic agents in patients with type 2 diabetes mellitus. Cochrane Database Syst Rev 2004; (4):CD003418.
14. Hofmeyr G J, Gülmezoglu A M, Novikova N, et al. Misoprostol to prevent and treat postpartum haemorrhage: a systematic review and meta-analysis of maternal deaths and dose-related effects. Bull World Health Organ 2009; (87):645-732.
15. Imamura M, Abrams P, Bain C, et al. Systematic review and economic modeling of the effectiveness and cost-effectiveness of non-surgical treatments for women with stress urinary incontinence. Health Technol Assess 2010; 14(40):1-iv.
16. Jansen JP. Self-monitoring of glucose in type 2 diabetes mellitus: a Bayesian meta-analysis of direct and indirect comparisons. Curr Med Res Opin 2006; 22(4):671-681.
17. Lam SK, Owen A. Combined resynchronisation and implantable defibrillator therapy in left ventricular dysfunction: Bayesian network meta-analysis of randomised controlled trials. BMJ 2007; 335(7626):925.
18. Lapitan MC, Cody JD, Grant A. Open retropubic colposuspension for urinary incontinence in women. Cochrane Database Syst Rev 2009; (2):CD002912.
19. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics 2009; 10(4):792-805.
20. Macfadyen CA, Acuin JM, Gamble C. Topical antibiotics without steroids for chronically discharging ears with underlying eardrum perforations. Cochrane Database Syst Rev 2005; (4):CD004618.
21. McDaid C, Maund E, Rice S, et al. Paracetamol and selective and non-selective non-steroidal anti-inflammatory drugs (NSAIDs) for the reduction of morphine-related side effects after major surgery: a systematic review. Health Technol Assess 2010; 14(17):1-iv.
22. Mills EJ, Wu P, Spurden D, et al. Efficacy of pharmacotherapies for short-term smoking abstinance: a systematic review and meta-analysis. Harm Reduct J 2009; 6:25.
23. Phung OJ, Scholle JM, Talwar M, et al. Effect of noninsulin antidiabetic drugs added to metformin therapy on glycemic control, weight gain, and hypoglycemia in type 2 diabetes. JAMA 2010; 303(14):1410-1418.
24. Phung OJ, Kahn SR, Cook DJ, et al. Dosing frequency of unfractionated heparin thromboprophylaxis: A meta-analysis. Chest 2011; 140(2):374-81.
25. Phung OJ, Sood NA, Sill BE, et al. Oral anti-diabetic drugs for the prevention of Type 2 diabetes. Diabet Med 2011; 28(8):948-64.
26. Picard P, Tramer MR. Prevention of pain on injection with propofol: a quantitative systematic review. Anesth Analg 2000; 90(4):963-969.
27. Playford EG, Webster AC, Sorell TC, et al. Antifungal agents for preventing fungal infections in solid organ transplant recipients. Cochrane Database Syst Rev 2004; (3):CD004291.
28. Pollock A, Baer G, Pomeroy V, Langhorne P. Physiotherapy treatment approaches for the recovery of postural control and lower limb function following stroke. *Cochrane Database Syst Rev* 2007;(1):CD001920.
29. Psaty BM, Smith NL, Siscovick DS, et al. Health outcomes associated with antihypertensive therapies used as first-line agents. A systematic review and meta-analysis. JAMA 1997; 277(9):739-745.
30. Puhan MA, Bachmann LM, Kleijnen J, et al. Inhaled drugs to reduce exacerbations in patients with chronic obstructive pulmonary disease: a network meta-analysis. BMC Med 2009; 7:2.
31. Roskell NS, Beard SM, Zhao Y, et al. A meta-analysis of pain response in the treatment of fibromyalgia. Pain Pract 2011; 11(6):516-527.
32. Soares-Weiser K, Bravo VY, Beynon S, et al. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder. Health Technol Assess 2007; 11(39):iii-206.

33. Stowe R, Ives N, Clarke CE, et al. Meta-analysis of the comparative efficacy and safety of adjuvant treatment to levodopa in later Parkinson's disease. Mov Disord 2011; 26(4):587-598.
34. Thijs V, Lemmens R, Fieuws S. Network meta-analysis: simultaneous meta-analysis of common antiplatelet regimens after transient ischaemic attack or stroke. Eur Heart J 2008; 29(9):1086-1092.
35. Trikalinos TA, Alsheikh-Ali AA, Tatsioni A, et al. Percutaneous coronary interventions for non-acute coronary artery disease: a quantitative 20-year synopsis and a network meta-analysis. Lancet 2009; 373(9667):911-918.
36. Tropeano AI, Saleh N, Hawajri N, et al. Do all antihypertensive drugs improve carotid intima-media thickness? A network meta-analysis of randomized controlled trials. Fundam Clin Pharmacol 2011; 25(3):395-404.
37. Tu YK, Woolston A, Faggion CM, et al. Do bone grafts or barrier membranes provide additional treatment effects for infrabony lesions treated with enamel matrix derivatives? A network meta-analysis of randomized-controlled trials. J Clin Periodontol 2010; 37(1):59-79.
38. van der Valk, Webers CA, Lumley T, et al. A network meta-analysis combined direct and indirect comparisons between glaucoma drugs to rank effectiveness in lowering intraocular pressure. J Clin Epidemiol 2009; 62(12):1279-1283.
39. Virgili G, Novielli N, Menchini F, et al. Pharmacological treatments for neovascular age-related macular degeneration: can mixed treatment comparison meta-analysis be useful? Curr Drug Targets 2011; 12(2):212-220.
40. Walsh T, Worthington HV, Glenny AM, et al. Fluoride toothpastes of different concentrations for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2010; (1):CD007868.
41. Wandel S, Juni P, Tendal B, et al. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. BMJ 2010; 341:c4675.
42. Wang H, Huang T, Jing J, et al. Effectiveness of different central venous catheters for catheter-related infections: a network meta-analysis. J Hosp Infect 2010; 76(1):1-11.
43. Welton NJ, Caldwell DM, Adamopoulos E, et al. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. Am J Epidemiol 2009;169(9):1158-1165.
44. Wilhelmus KR. The treatment of herpes simplex virus epithelial keratitis. Trans Am Ophthalmol Soc 2000; 98:505-532.
45. Woo G, Tomlinson G, Nishikawa Y, et al. Tenofovir and entecavir are the most effective antiviral agents for chronic hepatitis B: a systematic review and Bayesian meta-analyses. Gastroenterology 2010; 139(4):1218-1229.
46. Yu CH, Beattie WS. The effects of volatile anesthetics on cardiac ischemic complications and mortality in CABG: a meta-analysis. Can J Anaesth 2006; 53(9):906-918.

Bibliography

1. Borenstein M, Hedges LV, Higgins JPT, et al. Why Perform a Meta-Analysis. Introduction to Meta-Analysis. John Wiley & Sons, Ltd; 2009.

2. Brand R, Kragt H. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. Stat Med 1992; 11(16):2077–82.

3. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994; 309(6965):1351–5.

4. Borenstein M, Hedges LV, Higgins JPT, et al. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 2010; 1(2):97–111.

5. Borenstein M, Hedges LV, Higgins JPT, et al. Meta-Regression. Introduction to Meta-Analysis. John Wiley & Sons, Ltd; 2009.

6. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999; 18(20):2693–708.

7. Sterne JAC, Jüni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in "meta-epidemiological" research. Stat Med 2002; 21(11):1513–24.

8. Siersma V, Als-Nielsen B, Chen W, et al. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med 2007; 26(14):2745–58.

9. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50:683–91.

10. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005; 331:897–900.

11. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Meth 2012; 3(2):80–97.

12. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011; 64:163–71.

13. Deeks JJ, Higgins JP, Altman DG. Analysing Data and Undertaking Meta-Analyses. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, Ltd; 2008.

14. Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. Evid Based Ment Health. 2014 [in press].

15. Dias S, Sutton AJ, Ades AE, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making 2013; 33:607–17.

16. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002; 21(11):1559–73.

17. Hempel S, Miles JN, Booth MJ, et al. Risk of bias: a simulation study of power to detect study-level moderator effects in meta-analysis. Syst Rev. 2013; 2(1):107.

18. Welton NJ, Ades AE, Carlin JB, et al. Models for potentially biased evidence in meta-analysis using empirically based priors. J R Stat Soc Ser A 2009; 172(1):119–36.

19. Bafeta A, Trinquart L, Seror R, et al. Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review. BMJ 2013; 347:f3675.

20. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J Clin Epidemiol 2013 [in press].

21. Sobieraj DM, Cappelleri JC, Baker WL, et al. Methods used to conduct and report Bayesian mixed treatment comparisons published in the medical literature: a systematic review. BMJ Open 2013; 3.

22. Nikolakopoulou A, Chaimani A, Veroniki AA, et al. Characteristics of Networks of Interventions: A Description of a Database of 186 Published Networks. PLoS One. 2014 Jan 22; 9(1):e86754.

23. Stata Corp. Stata Statistical Software. College Station, TX: StataCorp LP; 2013.

24. Donegan S, Williamson P, Gamble C, et al. Indirect comparisons: a review of reporting and methodological quality. PLoS One 2010; 5:e11054.

25. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. J Amer Statist Assoc 2006; 101(474):447–59.

26. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC Med 2013; 11:159.

27. Song F, Altman DG, Glenny AM, et al. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003; 326:472.

28. Higgins JPT, Jackson D, Barrett JK, et al. Consistency and insconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Meth 2012 ;3(2):98–110.

29. Krahn U, Binder H, Konig J. A graphical tool for locating inconsistency in network meta-analyses. BMC Med Res Methodol 2013; 13:35.

30. Van Valkenhoef G, Lu G, de Brock B, et al. Automating network meta-analysis. Res Synth Methods 2012; 3(4):285–99.

31. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011; 14:417–28.

32. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002; 21:2313–24.

33. White IR, Barrett JK, Jackson D, et al. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Meth 2012 ;3(2):111–25.

34. Rucker G. Network meta-analysis, electrical networks and graph theory. Res Synth Meth 2012; 3(4):312–24.

35. Rucker G, Schwarzer G. netmeta: An R package for network meta-analysis. R Proj Website 2013; Available from: http://www.r-project.org

36. Lu G, Welton NJ, Higgins JPT, et al. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. Res Synth Meth 2011; 2(1):43–60.

37. König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Stat Med 2013; 32(30):5414–29.

38. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23(20):3105–24.

39. Salanti G, Higgins JP, Ades AE, et al. Evaluation of networks of randomized trials. Stat Methods Med Res 2008; 17:279–301.

40. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. Stat Med 2011; 30(20):2481-2498.

41. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. Stat Methods Med Res 2013; 22(2):133–58.

42. Cipriani A, Higgins JP, Geddes JR, et al. Conceptual and technical challenges in network meta-analysis. Ann Intern Med 2013; 159:130–7.

43. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996; 15:2733–49.

44. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics 2009; 10:792–805.

45. Chaimani A, Higgins JPT, Mavridis D, et al. Graphical tools for network meta-analysis in STATA. PLoS One 2013; 8(10):e76654.

46. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011; 342:d549.

47. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A 2009; 172(1):137–59.

48. Cooper H, Hedges LV, Valentine JC. The Handbook of Research Synthesis and Meta-Analysis. Russell Sage Foundation; 2009. 644 p.

49. Turner RM, Davey J, Clarke MJ, et al. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012; 41(3):818–27.

50. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21(11):1539–58.

51. Jackson D, White IR, Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Stat Med 2012; 31(29):3805–20.

52. Jackson D, Barrett JK, Stephen R, et al. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. Stat Med 2014 [in press].

53. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol 2010; 63:875–82.

54. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010; 29:932–44.

55. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making 2013; 33:641–56.

56. Anzures-Cabrera J, Higgins JPT. Graphical displays for meta-analysis: An overview with suggestions for practice. Res Synth Meth 2010; 1(1):66–80.

57. Bax L, Ikeda N, Fukui N, et al. More than numbers: the power of graphs in meta-analysis. Am J Epidemiol 2009; 169(2):249–55.

58. Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. Lancet 2011; 378(9799):1306–15.

59. Higgins JP, Altman DG. Assessing Risk of Bias in Included Studies. In: Fellow JPHSSV, Director SGF, editors. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. John Wiley & Sons, Ltd; 2008

60. Salanti G, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from network meta-analysis. PLoS One 2014 [in press].

61. Hoaglin D, Welsch R. The hat matrix in regression and ANOVA. Am Stat. 1978;32(1):17–22.

62. Trikalinos TA, Alsheikh-Ali AA, Tatsioni A, et al. Percutaneous Coronary Interventions for Nonacute Coronary Artery Disease: A Quantitative Twenty-Year Synopsis. Lancet 2009; 373(9667):911–8.

63. Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. J Clin Epidemiol 2009; 62:857–64.

64. Singh JA, Christensen R, Wells GA, et al. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. CMAJ 2009; 181(11):787–96.

65. Veroniki AA, Vasiliadis HS, Higgins JP, et al. Evaluation of inconsistency in networks of interventions. Int J Epidemiol 2013; 42:332–45.

66. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. CRC Press; 1994. 456 p.

67. Saporta G. The Analysis of Proximity Data. B. S. Everitt and S. Rabe-Hesketh, Arnold, London, 1997.

68. Efthimiou O, Mavridis D, Cipriani A, et al. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. Stat Med 2014 [in press].

69. Madan J, Chen Y-F, Aveyard P, Wang D, et al. 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 2014; 177(1):295–314.

70. Relethford JH. Cluster analysis for researchers. By H.C. Romesburg. Belmont, CA: Lifetime Learning Publications. 1984.

71. Lunn DJ, Thomas A, Best N, et al. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stat Comput 2000; 10(4):325–37.

72. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2013. Available from: http://www.R-project.org

73. SAS Institute. The SAS system for Windows. SAS Institute, Cary, NC; 2013.

74. Miladinovic B, Hozo I, Chaimani A, et al. Indirect treatment comparison. The Stata Journal 2014; 14(1):1–11.

75. Wells GA, Sultan SA, Chen L, et al. Indirect treatment comparison. Version 1.0. Ott Can Agency Drugs Technol Health 2009;

76. Harbord RM, Higgins JPT. Meta-regression in Stata. The STATA Journal 2008; 8(4):493–519.

77. Viechtbauer W. metafor: Meta-Analysis Package for R. R Proj Website. Available from: http://www.r-project.org

78. Schwarzer G. meta: Meta-Analysis with R. R Proj Website [Internet]. Available from: http://www.r-project.org

79. Dias S, Sutton AJ, Welton NJ, et al. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. Med Decis Making 2013; 33:618–40.

80. White IR. Multivariate random-effects meta-regression: Updates to mvmeta. The STATA Journal 2011; 11(2):255–70.

81. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat Med 2012 Dec 20; 31(29):3821–39.

82. Gasparrini A. mvmeta: Multivariate and univariate meta-analysis and meta-regression. R Proj Website. Available from: http://www.r-project.org

83. Jansen JP, Trikalinos T, Cappelleri JC, et al. Indirect Treatment Comparison/Network Meta-Analysis Study Questionnaire to Assess Relevance and Credibility to Inform Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report. Value Health 2014; 17(2):157–73.

84. Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 2013; 382(9896):951–62.

85. Filippini G, Del Giovane C, Vacchi L, et al. Immunomodulators and immunosuppressants for multiple sclerosis: a network meta-analysis. Cochrane Database Syst Rev. 2013; 6:CD008933.

86. Li T, Puhan MA, Vedula SS, et al. Network meta-analysis-highly attractive but more methodological research is needed. BMC Med 2011; 9(1):79.

87. Song F, Loke YK, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009; 338:b1147.

88. Song F, Xiong T, Parekh-Bhurke S, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. BMJ 2011; 343:d4909.

89. Chaimani A, Vasiliadis HS, Pandis N, et al. Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. Int J Epidemiol 2013; 42:1120–31.

90. Salanti G, Dias S, Welton NJ, et al. Evaluating novel agent effects in multiple-treatments meta-regression. Stat Med 2010; 29:2369–83.

91. Cooper NJ, Sutton AJ, Morris D, et al. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med 2009; 28:1861–81.

92. Achana FA, Cooper NJ, Dias S, et al. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. Stat Med 2013; 32:752–71.

93. Song F, Clark A, Bachmann MO, Maas J. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. BMC Med Res Methodol 2012; 12(1):138.

94. Mills EJ, Kanters S, Thorlund K, et al. The effects of excluding treatments from network meta-analyses: survey. BMJ 2013; 347:f5195.

95. Nüesch E, Trelle S, Reichenbach S, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. BMJ. 2010; 341:c3515.

96. Moreno SG, Sutton AJ, Ades AE, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Med Res Methodol 2009; 9(1):2.

97. Egger M, Juni P, Bartlett C, et al. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7(1):1–76.

98. Mavridis D, Salanti G. Exploring and accounting for publication bias in mental health: a brief overview of methods. Evid Based Ment Health. 2014 Feb; 17(1):11–5.

99. Peters JL, Sutton AJ, Jones DR, et al. Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. J R Stat Soc Ser A 2010; 173(3):575–91.

100. Rücker G, Carpenter JR, Schwarzer G. Detecting and adjusting for small-study effects in meta-analysis. Biom J 2011; 53(2):351–68.

101. Borenstein M, Hedges LV, Higgins JPT, et al. Simpson's Paradox. Introduction to Meta-Analysis. John Wiley & Sons, Ltd; 2009.

102. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998; 352(9128):609–13.

103. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008; 336(7644):601–5.

104. Herbison P, Hay-Smith J, Gillespie WJ. Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. J Clin Epidemiol 2011; 64(10):1070–5.

105. Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012 Sep 18;157(6):429–38.

106. Peters JL, Sutton AJ, Jones DR, et al. Comparison of two methods to detect publication bias in meta-analysis. JAMA 2006; 295(6):676–80.

107. Schwarzer G, Carpenter J, Rücker G. Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. J Clin Epidemiol 2010; 63(3):282–8.

108. Carpenter JR, Schwarzer G, Rücker G, et al. Empirical evaluation showed that the Copas selection model provided a useful summary in 80% of meta-analyses. J Clin Epidemiol 2009; 62(6):624–631.e4.

109. Terrin N, Schmid CH, Lau J, et al. Adjusting for publication bias in the presence of heterogeneity. Stat Med 2003; 22(13):2113–26.

110. Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Res Synth Meth 2012; 3(2):161–76.

111. Dias S, Welton NJ, Ades AE. Study designs to detect sponsorship and other biases in systematic reviews. J Clin Epidemiol 2010; 63:587–8.

112. Dias S, Welton NJ, Marinho VCC, et al. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. J Roy Statist Soc Ser A 2010; 173(3):613–29.

113. Moreno SG, Sutton AJ, Ades AE, et al. Adjusting for publication biases across similar interventions performed well when compared with gold standard data. J Clin Epidemiol 2011; 64:1230–41.

114. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. J R Stat Soc Ser B 2002; 64(4):583–639.

115. Borenstein M, Hedges LV, Higgins JPT, et al. Effect Sizes Based on Binary Data (2×2 Tables). Introduction to Meta-Analysis [Internet]. John Wiley & Sons, Ltd; 2009.

116. Borenstein M, Hedges LV, Higgins JPT, et al. Effect Sizes Based on Means. Introduction to Meta-Analysis. John Wiley & Sons, Ltd; 2009.

117. Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. Crit Care Lond Engl 2013; 17(1):R2.

118. Eng J. Sample size estimation: how many individuals should be studied? Radiology 2003; 227(2):309–13.

119. Chootrakool H, Shi JQ, Yue R. Meta-analysis and sensitivity analysis for multi-arm trials with selection bias. Stat Med 2011; 30(11):1183–98.

120. Peters JL, Sutton AJ, Jones DR, et al. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol 2008; 61(10):991–6.

121. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet 2007; 369(9557):201–7.

122. Psaty BM, Lumley T, Furberg CD, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. JAMA 2003; 289(19):2534–44.

123. Lindholm LH, Persson M, Alaupovic P, et al. Metabolic outcome during 1 year in newly detected hypertensives: results of the Antihypertensive Treatment and Lipid Profile in a North of Sweden Efficacy Evaluation (ALPINE study). J Hypertens 2003;21(8):1563–74.

124. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135(11):982–9.

125. Turner EH, Matthews AM, Linardatos E, et al. Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. N Engl J Med 2008; 358(3):252–60.

126. Mavridis D, Sutton A, Cipriani A, et al. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. Stat Med 2013; 32:51–66.