UNIVERSITY OF IOANNINA
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE
SECTION OF SOCIAL MEDICINE & MENTAL HEALTH
DEPARTMENT OF HYGIENE & EPIDEMIOLOGY

Multivariate extension of meta-analysis

Orestis Efthimiou

DOCTORAL THESIS

Ioannina 2017, Greece

UNIVERSITY OF IOANNINA
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE
SECTION OF SOCIAL MEDICINE & MENTAL HEALTH
DEPARTMENT OF HYGIENE & EPIDEMIOLOGY

Multivariate extension of meta-analysis

Orestis Efthimiou

DOCTORAL THESIS

Ioannina 2017, 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 του κ. Ευθυμίου Ορέστη: 15-10-2012

Ημερομηνία ορισμού Τριμελούς Συμβουλευτικής Επιτροπής: 747[α]/31-1-2013

Μέλη Τριμελούς Συμβουλευτικής Επιτροπής:

Επιβλέπουσα

Σαλαντή Γεωργία, Επίκουρη Καθηγήτρια Επιδημιολογίας του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων.

Μέλη

Μαυρίδης Δημήτριος, Λέκτορας του Παιδαγωγικού Τμήματος Δημοτικής Εκπαίδευσης του Πανεπιστημίου Ιωαννίνων.

Μπάγκος Παντελής, Επίκουρος Καθηγητής του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Στερεάς Ελλάδας.

Ημερομηνία ορισμού θέματος: 19-2-2013

«Πολυπαραγοντική Επέκταση της Μετα-ανάλυσης»

ΟΡΙΣΜΟΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ 798[α]/6-6-2017

Δημολιάτης Ιωάννης	Αναπληρωτής Καθηγητής Υγιεινής με έμφαση στην Ιατρική Εκπαίδευση του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων.
Μπάγκος Παντελής	Αναπληρωτής Καθηγητής Βιοπληροφορικής και Βιοστατιστικής του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας.
Ντζάνη Ευαγγελία	Αναπληρώτρια Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων.
Σαλαντή Γεωργία	Αναπληρώτρια Καθηγήτρια Βιοστατιστικής και Επιδημιολογίας του Τμήματος Ιατρικής του Πανεπιστημίου Ιωαννίνων.
Ευαγγέλου Ευάγγελος	Επίκουρος Καθηγητής Υγιεινής με έμφαση στην Κλινική και Μοριακή Επιδημιολογία του Τμήματος Ιατρικής Πανεπιστημίου Ιωαννίνων.
Μαυρίδης Δημήτριος	Επίκουρος Καθηγητής Στατιστικής στην Εκπαίδευση του Παιδαγωγικού Τμήματος Δημοτικής Εκπαίδευσης του Πανεπιστημίου Ιωαννίνων.
Cipriani Andrea	Καθηγητής του Πανεπιστημίου της Οξφόρδης.

Έγκριση Διδακτορικής Διατριβής με βαθμό **«ΑΡΙΣΤΑ»** στις **8-6-2017**

ΠΡΟΕΔΡΟΣ ΤΟΥ ΤΜΗΜΑΤΟΣ ΙΑΤΡΙΚΗΣ

Μηνάς Πασχόπουλος

Καθηγητής Μαιευτικής-Γυναικολογίας

Η Γραμματέας του Τμήματος

ΜΑΡΙΑ ΚΑΠΙΤΟΠΟΥΛΟΥ

Acknowledgements

This PhD thesis is lovingly dedicated to my son, Dimitris, and to my wife, Maria, for her support, encouragement, and constant love, which have been the greatest source of inspiration and energy for me.

I particularly want to thank my supervisor and friend, Professor Georgia Salanti, for the invaluable help she has provided. Without her guidance and constant feedback this PhD would not have been achievable. I would also like to thank my colleagues Dimitris Mavridis, Anna Chaimani and Adriani Nikolakopoulou for their assistance and support throughout this dissertation. My sincere thanks also goes to Professors Stefan Leucht, Andrea Cipriani and Pantelis Bagos for their scientific advice and their valuable input on some of the projects presented in this dissertation. I am also grateful to my parents for their never-ending support and love.

I will never forget the good times I have had during this PhD: the fun at the office with the rest of the team during our breaks; our dinners at Mistagogia; the countless drinks we had and the fun activities we did during conferences and meetings, including (but not limited to) participating in Bollywood and Waltz dance competitions and winning fabulous prizes.

The work presented in this PhD Thesis was supported partly by the European Research Council (IMMA 260559) and partly by the Innovative Medicines Initiative Joint Undertaking under grant agreement no.115546, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution.

Table of Contents

1 Introduction

1.1 General aims of this thesis

Meta-analysis of randomized control trials (RCTs) is a key ingredient in today's comparative effectiveness research in evidence-based medicine. International health organizations such as the World Health Organization or the Cochrane Collaboration recognize their value and use meta-analyses routinely while Agencies such as the Canadian Agency for Drugs and Technologies in Health (CADTH), the Agency for Healthcare Research and Quality (AHRQ) and the National Institute for Health and Clinical Excellence (NICE) use them to produce guidelines for clinical practice. The institute of Medicine in the United States set a goal that, by the year 2020, 90 percent of clinical decisions will be evidence-based (1).

Traditional meta-analytical techniques, however, can only compare two treatments (i.e. can only perform a 'pairwise' comparison) and thus their usefulness is limited when three or more competing treatments for the same condition are present. In addition, even though the interests of policy-makers lie in the comparison of active agents, new treatments are commonly compared only to placebo. In such cases pairwise meta-analysis cannot give a definite answer as to which treatment works best for a specific condition, setting hurdles to the decision-making process (2).

This situation drove the interest of researchers and funding bodies towards a new framework for synthesizing information from studies comparing different subsets of competing treatments. Network Meta-Analysis (NMA, sometimes also called 'multiple treatment meta-analysis' or 'mixed-treatment comparison') was developed to address this issue (2–6). NMA is a statistical tool which can combine information across a network of randomized trials, and which produces inferences concerning the relative effectiveness of multiple interventions.

In the last few years NMA has become increasingly popular (7–12) and its usefulness has been recognized by various organizations. For example, the Decision Support Unit of NICE provides extensive guidance on performing an NMA (13) and the Cochrane Collaboration has established the 'Comparing Multiple Interventions Group' for promoting the methodology for comparing multiple interventions, http://cmimg.cochrane.org/. Moreover, there have been many papers discussing the advantages and limitations of the

method (14–30), which have also been explored in empirical assessments (31–35) and simulation studies (9,34,36). The advantages include a potential increase in the precision from the estimates of an NMA compared to an estimate based on direct evidence alone and that it allows comparing treatments that have never been compared in head-to-head experiments. This is particularly valuable when active agents are only compared to placebo or standard care for regulatory purposes but not to each other (37). In addition, NMA can be used to answer policy-relevant questions by providing a ranking of all competing treatments (38) and to reduce the uncertainty in cost-effectiveness analyses (39).

Despite the aforementioned advantages, the implementation of NMA in practice may be hindered because of several reasons. First, the methodology of NMA rests on the assumption of transitivity, i.e. that different sources of evidence (direct and indirect evidence for the same treatment comparison) are in agreement. This assumption is often viewed as an important limitation of the method because it may be difficult to assess its plausibility in practice, and because if it does not hold NMA results may be invalid. Moreover, the field of NMA is swiftly evolving; during the last few years there has been an abundance of published methodological articles presenting alternative approaches to deal with issues related to NMA.

The first objective of this thesis is to give a comprehensive account of the currently available methods for NMA and discuss in depth conceptual and statistical ways for evaluating the underlying assumptions of the model while providing guidance for researchers that set out to perform an NMA. To this end we performed a systematic review of the methodology, to ensure that interested researchers use state-of-the-art methods for practical applications and when conducting further methodological research.

The second objective of this PhD thesis is the extension of NMA methods to the case of multiple correlated outcomes. Studies typically report on more than one outcome, and multiple outcomes can be correlated. For example, a study on antihypertensives may report systolic and diastolic blood pressure. These two outcomes are correlated because they are measured in the same patients. Moreover, on the meta-analysis level, there may be between-study correlations of the true outcome effects across studies. These correlations will reflect the way that the true outcome effects depend on each other when measured in different settings.

Currently available models for performing a multiple-outcomes meta-analysis of randomized trials are limited to the case of studies that compare only two treatments. In this

thesis we present new methods for performing an NMA for the case of multiple, correlated outcomes. We discuss a range of different modeling approaches to perform such an analysis, depending on the nature of the outcomes (e.g. binary/continuous) and the availability of information regarding the correlations.

In the next section of this introductory chapter we give a brief account of the basic concepts and statistical models used in simple (pairwise) meta-analysis. In the following chapters we will see how these methods are generalized for the case of NMA, and also multiple outcomes NMA.

1.2 A brief outline of methods for pairwise meta-analysis

1.2.1 General concepts in meta-analysis

Let us start by considering a collection of N_S studies, which compare two interventions for the same disease in terms of a specific outcome. Let us also assume that the populations of patients are similar across the studies. Each study provides an estimate of the magnitude of relative treatment effect (y_i), along with the corresponding measure of the uncertainty of this estimate (e.g. this could be the observed variance s_i^2 of y_i). Relative effects can be expressed for example in terms of odds ratio, risk ratio, risk difference (for binary outcomes), mean difference, standardized mean difference (for continuous outcomes), hazard ratios (for time-to-event outcomes), etc.

The basic assumption behind all meta-analysis methods is that these distinct – but conceptually similar – studies aim to estimate a common underlying truth regarding relative treatment effects. Thus, the scope of meta-analysis is to synthesize the N_S different answers into a single, pooled estimate of this treatment effect. There are several different statistical approaches to meta-analysis, but most are variations of a weighted average, where the result obtained in each study is assigned a study-specific weight (40). These weights usually relate to the precision of the studies, where more precise studies receive larger weights. The advantage of this pooling is that it leads to a higher statistical power, an increase in precision as compared to the individual studies' results, and the chance to settle controversies arising from conflicting results in the individual studies (40,41).

The two most popular approaches to meta-analysis are the fixed (or common) effect and the random effects models, and we describe them in brief in the following paragraph.

1.2.2 Fixed vs. random effects meta-analysis

Fixed effects models assume that there is a single true treatment effect that underlies all studies in the analysis. Observed differences between the estimates of the studies are only due to random (sampling) error. This implies that the studies are similar in all aspects that might potentially modify the relative treatment effect. These include population characteristics (e.g. age of the participants), study design characteristics (e.g. duration of follow-up), intervention characteristics (e.g. dose) etc. (42). If we denote the true treatment effects in study i by θ_i (where $i = 1,2, \dots N_S$), under the fixed effect assumption all θ_i are equal, i.e. $\theta_i = \mu$. The observed effects in each study are $y_i = \mu + \varepsilon_i$, where ε_i is the random error.

By contrast, random effect models assume that the true effect size is different in each study. For example effect sizes might be larger in studies with older or more severely ill patients, or when more intensive variants of the treatment were used (43). In most cases studies are expected to have at least some variability in terms of patient or care-taker characteristics, implementations of the treatments etc., so that there may be different true effect sizes in each of the different studies. If N_S was infinitely large we could reconstruct the distribution of the study-specific effect sizes. In a random effects model the observed effect sizes in the studies are assumed to be a random sample of this underlying distribution (43). The variability of this distribution of effects is typically termed heterogeneity. Thus, in a random effects meta-analysis model the observed differences of the estimates of the studies can be attributed to two factors: random (sampling) error, and random effects (due heterogeneity). The most common assumption used to model the distribution of study-specific true effects is to assume a normal distribution. The observed effects in each study are $y_i = \mu + \delta_i + \varepsilon_i$, where ε_i are is the random effect. We will denote the standard deviation of random effects by τ in this dissertation. Setting $\tau = 0$ corresponds to assuming no variation in the study-specific effects, and in that case the random effects meta-analysis models simplifies to a fixed effects model.

Regarding the choice between the two models, fixed effect vs. random effects: if researchers expect the identified studies to share a common effect size and also they are only interested in identifying the treatment effect for a specific population, then a fixed effect meta-analysis is more appropriate to use. In all other circumstances the random effects assumption is much more suitable (41) and should be considered.

1.2.3 Inverse variance meta-analysis

Perhaps the most common approach to defining weights in a meta-analysis is the inverse variance method. According to this method each study i is assigned a weight w_i, where, for a fixed effects meta-analysis we assume:

$$w_i = \frac{1}{s_i^2} \tag{1}$$

The pooled treatment effect is estimated to be:

$$\hat{\mu} = \frac{\sum_i w_i\, y_i}{\sum_i w_i}, \tag{2}$$

and the corresponding variance:

$$var(\hat{\mu}) = \frac{1}{\sum_i w_i}, \tag{3}$$

For a random effects meta-analysis the weights are defined in a way similar to Equation (1), but they now also include the heterogeneity variance:

$$w_i^* = \frac{1}{s_i^2 + \tau^2} \tag{4}$$

The pooled estimate and corresponding variance is still given by Equations (2) and (3), with the only change being the replacement of w_i with w_i^*.

In order to use Equation (4) one first needs to obtain an estimate of heterogeneity, τ^2. The most widely used approach to estimating τ^2 is the DerSimonian and Laird method (44). In recent years, however, a plethora of alternative method have been proposed. Among these, an estimator proposed by Paule and Mandel (for both continuous and dichotomous outcomes) and the restricted maximum likelihood estimator (for continuous outcomes) have been shown to perform better (45).

1.2.4 Estimating pairwise meta-analysis in a Bayesian framework

Meta-analysis can also be formulated in the form of a hierarchical model, and then be fitted using Bayesian machinery. For a random effects meta-analysis, we assume:

$$\begin{aligned} y_i &\sim N(\theta_i, s_i^2) \\ \theta_i &\sim N(\mu, \tau^2) \end{aligned} \tag{5}$$

Prior distributions then need to be assigned to μ and τ. The pooled treatment effects follow from the posterior distribution of μ. Note that equation (5) can be used for meta-analysing

continuous outcomes, but also for dichotomous or time-to-event outcomes, if one uses measures that can be assumed to follow a normal distribution, e.g. log-odds ratio, log-risk ratio, log-hazard ratio etc.

A better approach would be to take into account the exact likelihood of the data. Note that this requires arm-level data to be available from the original studies. Let us focus on the case of a dichotomous outcome. Let us assume that we have a number of studies comparing treatments A to B, and that study i reports the number of events and number of randomized patients per treatment arm, i.e. $r_{i,A}$, $r_{i,B}$ and $n_{i,A}$, $n_{i,B}$. These are assumed to follow a binomial distribution:

$$r_{i,A} \sim Bin(p_{i,A}, n_{i,A})$$
$$r_{i,B} \sim Bin(p_{i,B}, n_{i,B})$$

The arm-specific probabilities $p_{i,A}$ and $p_{i,B}$ can be used to estimate the treatment effects in that study. For instance, in order to use log-odds ratios we set $logOR_i = log\left(\frac{p_{i,A}(1-p_{iB})}{(1-p_{i,A})p_{i,B}}\right)$. These study-specific effects can then be assumed exchangeable across studies, e.g. by setting $logOR_i \sim N(\mu, \tau^2)$. A detailed account of the various hierarchical models one can use depending on the likelihood of the data can be found in a paper by Dias et al. (46).

In the next, final section of this introductory chapter, we provide a brief outline of this dissertation.

1.3 Outline of the dissertation

In Chapter 2 we present the results of the systematic review on the methodology of NMA. We present our search strategy in Section 2.1. In Section 2.2 we provide an in-depth discussion of some conceptual issues and assumptions that underlie NMA. We discuss statistical methods for fitting NMA in Section 2.3. We present approaches for evaluating the underlying assumptions of NMA in Section 2.4. We summarize the currently available methods for fitting NMA in Section 2.6 and we discuss the use of alternative effect measures in Sections 2.7 and 2.8. We then present extensions of the model for adding covariates in the analysis (Section 2.9), and for investigating potential sources of bias 2.10. In Section 2.11 we discuss the reporting of NMA results. In Section 2.12 we review methods for synthesizing repeated measurements and multiple outcomes in NMA. In Section 2.13 we

discuss the issue of deciding which treatments to include in a NMA. In Sections 2.14 and 2.15 we summarize recent advances in incorporating individual patient data (IPD) and non-randomised studies in NMA. In Section 2.16 we discuss the issue of planning future studies.

In Chapter 3 we propose a new model for performing a joint network meta-analysis, for the case of multiple, correlated, dichotomous outcomes. In Section 3.2 we describe a motivating clinical example, borrowed from a systematic review aiming to compare 14 different drugs and placebo for acute mania, in terms of efficacy and acceptability. In Section 3.3 we present our model in detail. One of the important features of the model is that it requires external input in the form of information elicited from clinical experts. There we discuss methods that can be employed for obtaining such information. In Section 3.4 we apply our methods to the network of treatments for acute mania, and obtain relative treatment effects for all comparisons in the network, for both outcomes. In Section 3.5 we summarize our findings.

In Chapter 4 we present two additional models that can be used for the network meta-analysis of multiple correlated outcomes, for the general case of analyzing either dichotomous, continuous, or time-to-event correlated outcomes. In Section 4.2 we present the mathematical details models. In Section 4.3 we apply the two models to the acute mania dataset and we present our results. In Section 4.4 we compare the two models and we discuss how to choose between the two in real-life clinical applications

Finally, in Chapter 5 we present the most important findings of this dissertation. We start by summarizing our recommendations for performing NMA as they emerged from our systematic review. We discuss the best practices and highlight the most appropriate methods for NMA, aiming to provide guidance to future researchers. We also give an overview of the new models we propose for performing NMA for multiple outcomes. We summarize the advantages and limitations of each model and discuss how to choose between them in practical applications. We also highlight some areas of future research.

2 Systematic review of the methodology of network meta-analysis

2.1 Search strategy

For the purposes of our review of the methodology we have searched for published articles that presented new methods for NMA or articles evaluating existing methodology. We based our search on a previous review of the literature in NMA performed by the '*Comparing Multiple Interventions Methods Group*' of the Cochrane Collaboration. We also used the results from a recent literature review performed by Donegan et al. (47) where 116 papers on methods for assessing the homogeneity and consistency assumptions of NMA were identified (referred to as "key paper" in Figure 1) In addition, we searched the PUBMED database for relevant hits using the following terms:

(network OR mixed treatment OR multiple treatment* OR mixed comparison* OR indirect comparison* OR umbrella OR simultaneous comparison*) AND (meta-analysis).*

This query produced 1789 hits (14 March 2014), 88 of which were deemed relevant. Articles that have appeared in two methodological journals, namely Journal of Research Synthesis Methods (RSM) and Journal of the Royal Statistical Society (JRSS), series A, B and C, are not indexed by the PUBMED database; for this reason we performed a hand-search for relevant publications in these two journals.

Inclusion criteria

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

Exclusion Criteria

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

- full text of the publication was not available
- published in a language other than English
- conference posters

- applications of NMA without a methodological focus

In Figure 1 we present the flow chart of the papers identified in our search. The identified articles were organized according to their context and are discussed in the relevant sections of this review. An online database of all included articles tagged by topic can be found at https://www.zotero.org/groups/wp4_-_network_meta-analysis/items. This database has been shared with experts in the field to identify missing relevant articles.

Aiming to make our results more accessible to the interested reviewers, each identified article was assigned one or more tags according to the type of research presented, one or more tags according to the methodological topics addressed and one or more tags according to the software it used to implement the methods it presented.

We used 5 contribution tags:

- *Methodology development*: this will be assigned to papers presenting a novel methodology.
- *Didactical/good practice/recommendations*: for papers giving guidance or advice.
- *Methodology overview*: to be assigned to papers presenting a summary of the existing methodology for NMA
- *Simulation*: for papers using a simulated dataset to make assessments
- *Empirical assessment*: for papers presenting an assessment based on published NMAs.

There were 14 methodology tags:

- *Basic Methodology*: for papers presenting novel methodology for addressing fundamental issues of NMA
- *Definition of nodes:* which will be assigned to articles presenting methodology regarding the definition of treatment nodes included in an NMA
- *Effect Sizes*: for papers addressing issues on the different effect sizes that can be used in an NMA
- *Conceptual issues/Assumptions underlying NMA*: to be assigned to papers that elucidate the conceptual issues of an NMA and discuss the assumptions that need to hold in order for an NMA to give valid results.
- *Statistical inconsistency:* for papers discussing methods for quantifying statistical inconsistency in NMA, for papers presenting ways for addressing inconsistency or for papers examining the prevalence of inconsistency in published networks.

- *Risk of Bias*: for papers presenting methods for addressing the risk of bias in an NMA
- *Non-randomized and observational studies*: assigned to papers suggesting ways to include data from non-randomized and observational studies in an NMA
- *Publication Bias*: for papers presenting methods for addressing the risk of publication bias in an NMA
- *Multiple outcomes/repeated measures/survival analysis*: for papers presenting methods for the joint analysis of multiple correlated outcomes, repeated measures and analysis of survival data
- *NMA meta-regression:* for papers discussing the use of covariates in an NMA
- *IPD in NMA studies*: for papers presenting ways to include evidence from studies reporting individual patient data in an NMA
- *Sensitivity analyses*: for papers presenting some form of sensitivity analysis
- *Planning future studies*: for papers discussing methods for planning future trials
- *Reporting NMA*: for papers discussing methods for reporting the results of an NMA

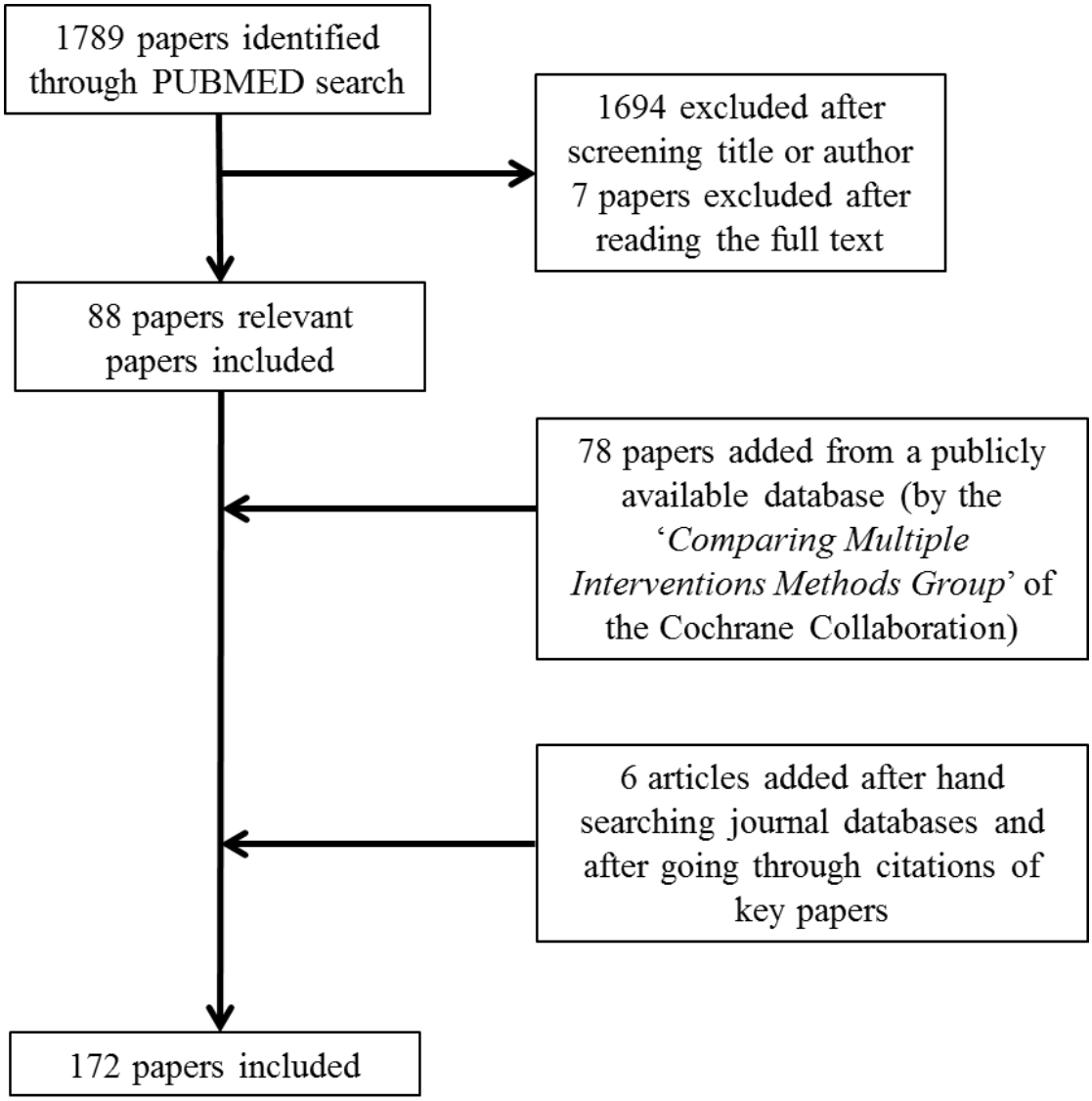
Finally we used 4 software implementation tags:

- *BUGS*: for papers using either WinBUGS or OpenBUGS
- *R*: to be assigned to papers using the R programming language
- *STATA:* for papers using the STATA software package
- *SAS:* for papers using the SAS software package

Since our focus is on methodology, completeness of the search is less of an issue: a more extensive search might provide some additional articles, but it is unlikely that it will provide any new insights or further methodological perspectives. This effect is termed theoretical saturation (48).

In the remainder of this Chapter we provide an overview of best practices and methods for NMA, as they were identified by our systematic review.

Figure 1: Flow chart of included and excluded methodological papers for the systematic review



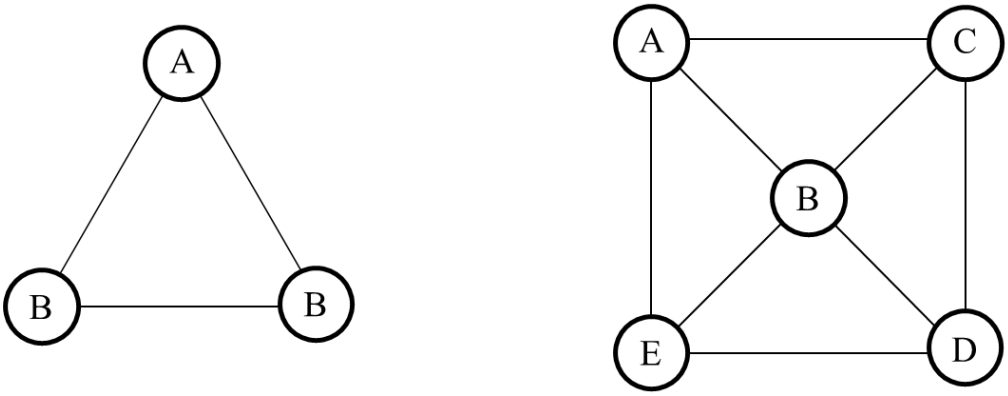
2.2 Conceptual issues and assumptions underlying network meta-analysis

The key feature of NMA is that it allows the synthesis of direct and indirect estimates for the relative effects of many competing treatments for the same health condition. Two treatments A and B may have been *directly* compared in head-to-head (A vs. B) studies. An *indirect* estimate may also be obtained from studies comparing these two treatments with a common comparator treatment C, i.e. AC and BC studies (49), as shown in the left panel of Figure 2. If both direct and indirect estimates are available, they can be combined into a *mixed* treatment effect.

In practice, for most health conditions there is a plethora of interventions being compared in randomized control trials, forming a network of evidence. For a given treatment comparison within such a network there may be direct and many different indirect estimates, obtained via many different comparators, as shown in the example of the right panel of Figure 2. Using NMA one can synthesize all these different pieces of information to produce an internally consistent overall estimate of all treatments' relative effects.

Despite the benefits of NMA discussed in the Introduction Chapter, there is still controversy among the scientific community about the validity of using indirect treatment comparisons (*indirect evidence*) for decision making. The use of such evidence may be particularly challenged when direct treatment comparisons (*direct evidence*) are also available (50–52). One focal point of criticism is the nature of evidence NMA provides. Even though patients within an RCT are randomized to receive one of the treatments being compared, the treatments are not randomized across the included trials. Therefore, indirect comparisons are non-randomized comparisons. In fact, indirect comparisons provide observational, rather than randomized, evidence. As a consequence indirect treatment comparisons may be more subject to biased treatment effect estimates, due, for example, to confounding, when randomized AB and AC studies are systematically different than BC (3); also, due to selection bias, when the choice of comparator in a study is dependent on the relative treatment effect (53). Such considerations are also closely related to the underlying assumptions of NMA; in what follows we discuss these assumptions in detail.

Figure 2: Each circle represents an intervention. A line connecting two interventions represents the availability of studies performing the corresponding comparison. Left panel: Three interventions A, B, and C form a simple triangular network. The indirect AB comparison is estimated via C, i.e. using the direct AC and BC comparisons. The mixed relative treatment effect for AB is estimated by combining the direct comparison and the indirect comparison. Right panel: A network of five interventions and eight direct comparisons. Overall, one direct comparison and four indirect comparisons contribute evidence to A versus B (indirect comparisons are via C, via E, via C and D, and via E and D).



2.2.1 Transitivity

The aim of a NMA is to enhance the decision-making process regarding the choice among alternative treatments for a certain disease and a target population. NMA adopts the same set of assumptions as a pairwise meta-analysis (54) but it also employs one additional assumption which can be hard to assess (55) called transitivity (56) (also termed similarity (11,32) or exchangeability (57)). Transitivity implies that information for the comparison between treatment A and B can be obtained via another treatment C, using the comparisons A vs. C and B vs. C. This assumption cannot be tested statistically, but its validity can be evaluated in a conceptual and epidemiological manner (24).

The transitivity assumption implies that we can combine the direct evidence from AC and BC studies to learn (indirectly) about the comparison AB. This, however, will be questionable if there are important differences in the distribution of the effect modifiers

(variables or characteristics which modify the observed relative effects, e.g. mean age of the participants, treatment dosage etc.) across the AC and BC trials which inform the indirect comparison (56,58). An effect modifier might differ across studies of the same comparison (e.g. mean participant age might be different across the AC trials) but if it has a similar distribution across comparisons (AC and BC) the transitivity assumption may still be valid (24). Consequently, the plausibility of the transitivity assumption can be evaluated by putting the collection of studies under scrutiny for important differences in the distribution of effect modifiers. If the studies are deemed to be similar then the transitivity assumption might be realistic, provided that there are no unknown modifiers of the relative treatment effect (59). Obviously, such an evaluation of transitivity may be impossible when the effect modifiers are not reported or when there are few studies per treatment comparison (60). If important differences are identified and there are enough data available, a network meta-regression can be used to improve the transitivity of the network (see also Section 2.9).

This implies, for example, that the common comparator treatment C must be similar in the AC and in the BC studies in terms of dose, modes of administration, duration etc. In an NMA of studies comparing fluoride treatments for the prevention of dental carries, the definition of placebo was different between studies of fluoride toothpaste and studies of fluoride rinse (61), casting doubt about the plausibility of the transitivity assumption and thereby challenging the reliability of NMA results. In another example, Julious and Wang (62) discussed how the use of placebo as an intermediate comparator might bias the results of indirect comparisons due to changes in the placebo response of the population over the years; for example, when studies comparing treatment A to placebo are older than studies comparing B to placebo the indirect estimate for A vs. B via placebo may be biased.

Other ways of formulating the transitivity assumption is to assume that regardless of the treatments being compared in each study the true relative effect of A vs. B is the same in a fixed effects model or exchangeable across studies in a random effects model (57,63), that the 'missing' treatments in each trial are missing at random (64) or, equivalently, that the choice of treatment comparisons in the trials is not associated either directly or indirectly with the relative effectiveness of the interventions (24). Finally, an alternative way of postulating this assumption is to state that the included patients could in principle be randomized to any of the treatments included in the network (24).

2.2.2 Consistency

The statistical manifestation of transitivity is called consistency (60). Checking the network for consistency constitutes an additional method of inferring indirectly about the plausibility of the transitivity assumption. Consistency refers to the statistical agreement between the observed direct and the (possibly many) indirect sources of evidence. A simple network may only include three treatments A, B and C. The transitivity assumption then implies that $\mu_{AB} = \mu_{AC} - \mu_{BC}$ (also termed consistency equation), where μ_{AB} denotes the true relative effect of treatment B over C; likewise for μ_{AC}, μ_{BC}. When this equation does not hold for the (direct) estimates, the network is said to be inconsistent (64) or incoherent (65). If this is the case, results from an NMA will be more difficult to interpret and become less reliable. In a following section we review various statistical methods and models that have been suggested for identifying inconsistency and thus assessing the transitivity assumption in NMA.

Statistical inconsistency can be thought of as another form of heterogeneity: heterogeneity results from the variation of effect modifiers within a treatment comparison, while inconsistency results from the variation of effect modifiers across treatment comparisons (58). Researchers should keep in mind, though, that the consistency of a network can only be assessed statistically when there is both direct and indirect evidence for one or more treatment comparisons. This situation only occurs when there are *closed loops* in the network (i.e. when three or more interventions are connected by a polygon, the edges of which represent head-to-head comparisons between the corresponding treatments). When there are no closed loops present in the network, a statistical assessment of inconsistency will not be possible. In these situations there cannot be inconsistency by definition. This, however, does not imply that the transitivity assumption will necessarily hold. It should also be noted that the absence of statistical inconsistency does not provide proof for the validity of the transitivity assumption, which, as discussed in the previous section is essentially an untestable assumption. Thus, next to statistical tests for inconsistency, a conceptual/theoretical assessment of the transitivity assumption should always take place before an NMA is conducted (60) and the studies included in an NMA should always be scanned for important differences in terms of patients, interventions, outcomes, study design, methodological characteristics and reporting biases (3,7,9,15,19,26,28,35,49,59,66–70).

2.3 Statistical models for network meta-analysis

A simple network may include three treatments of interest, A, B and C. An estimate of the indirect treatment effect of A vs. B can then be obtained by utilizing the direct observations A vs. C and B vs. C as $\hat{\mu}_{AB}^{Ind} = \hat{\mu}_{AC}^{Dir} - \hat{\mu}_{BC}^{Dir}$ (49). This result is sometimes also referred to as "adjusted indirect comparison". The variance of the indirect estimate is the sum of the variances of the two direct ones. When direct evidence is also available for the A vs. B comparison it can be combined with the indirect estimate using the usual inverse variance method to produce a mixed estimate. Note that this method for obtaining indirect estimates is only valid for 'triangular networks', where three treatments have been compared in a number of two-arm trials and for 'star-shaped' networks, where all treatments are compared to a common comparator (e.g. placebo) but not to each other. For complex networks there will be multiple sources of indirect information, and thus more advanced models need to be used.

Popular implementations of NMA models adopt meta-regression (Section 2.3.1), hierarchical modelling (Section 2.3.2) or a multivariate meta-analysis approach (Section 2.3.3). A common feature of all of these models is that the use of the consistency equations minimizes the number of parameters that need to be estimated. The minimum set of parameters needed to model the relative treatment effects is usually termed as the set of "basic parameters" or "basic contrasts"; these parameters are in number equal to the number of treatments minus one and can be used to generate estimates for all possible treatment comparisons, via the consistency equations. The basic parameters can be chosen arbitrarily as long as they form a "spanning tree" of the evidence (64); if this condition is satisfied the actual choice of basic parameters does not affect the NMA results. These parameters are commonly taken to be the relative effects of each treatment versus a reference (e.g. the placebo, if present in the network). For example, for a network of four treatments A, B, C and D three basic parameters are needed. These can be chosen to correspond to the relative treatment effects of all other treatments versus A, i.e. AB, AC and AD. All other treatment effects can be generated from these 3 parameters, e.g. the relative treatment effect for BD can be estimated using the AB and AD parameters. Choosing instead BA, BC and BD as the basic parameters would have no impact on the NMA results.

In what follows we describe the most popular approaches for performing an NMA.

2.3.1 Network meta-analysis as a meta-regression

In the meta-regression approach, first proposed by Lumley (65) the various treatment comparisons are treated as covariates in a meta-regression model (6). The usual NMA meta-regression model can be summarized in the following equation: $\boldsymbol{y} = \boldsymbol{X\mu} + \boldsymbol{\varepsilon} + \boldsymbol{\delta}$, with $\boldsymbol{y}$ being the vector of observed relative treatment effects, $\boldsymbol{\mu}$ the vector of basic parameters, $\boldsymbol{\varepsilon}$ the vector of random errors, and $\boldsymbol{\delta}$ the vector of random effects. Note that for a study i comparing T_i different treatments, only $T_i - 1$ observations on treatment comparisons need to enter the model. For a parallel randomized three-arm ABC trial, for example, we only need to include two of the three comparisons, e.g. AB and AC; the BC comparison is just a linear function of the other two. This means that $\boldsymbol{y}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ have a length equal to $\sum(T_i - 1)$. Random errors are assumed to follow a multivariate normal distribution, $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ being the (block-diagonal), within-study variance covariance matrix. A study i with T_i treatments arms will contribute a $(T_i - 1) \times (T_i - 1)$ matrix to $\boldsymbol{\Sigma}$; a two-arm AB study, for example, will only contribute to $\boldsymbol{\Sigma}$ the variance of the relative treatment effect of A vs. B. A three-arm trial ABC will contribute to $\boldsymbol{\Sigma}$ a 2×2 matrix with the variances and the covariance of the 2 relative treatment effects chosen to be included in $\boldsymbol{y}$, eg. AB and AC. Similarly, $\boldsymbol{\delta} \sim N(0, \boldsymbol{\Delta})$ for the random effects, with $\boldsymbol{\Delta}$ being the heterogeneity variance-covariance matrix. Matrix $\boldsymbol{X}$**,** the *design matrix*, has as elements 1, –1 and 0 and describes the structure of the network, providing information on which comparison is being performed in each study (6). If for example the network is built by an AB study (study 1), an AC study (study 2) and a BC study (study 3), the model would be written as:

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

The basic parameters can be estimated as $\hat{\boldsymbol{\mu}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{y}$, with variance $var(\hat{\boldsymbol{\mu}}) = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-1}$, where $\boldsymbol{W}$ is the weight matrix, $\boldsymbol{W} = (\boldsymbol{\Sigma} + \boldsymbol{\Delta})^{-1}$. The within-study variance-covariance matrix $\boldsymbol{\Sigma}$ can be estimated from the observed data (4,71), while for the between-study variance-covariance matrix $\boldsymbol{\Delta}$ one can use various ways of estimation including likelihood methods or the methods of moments (72–74). Estimating $\boldsymbol{\Delta}$ may be difficult especially when the data are sparse or in the presence of multi-arm studies. For this reason, it is common to introduce additional assumptions to reduce the number of parameters in $\boldsymbol{\Delta}$ and simplify the estimation. The most common approach is to assume equal

heterogeneity variances across comparisons, i.e. the between-study heterogeneity of the relative treatment effects is the same for all treatment comparisons (4,65). This assumption is, however, quite strong and may often be unrealistic. Lu & Ades (63) discussed how the consistency equations impose restrictions in the heterogeneity of each comparison, based on the (different) heterogeneity variances of each of the basic parameters. Thorlund et al. (75) presented models for exchangeable heterogeneity variances and also discussed the use of informative prior distributions in the context of a Bayesian analysis.

In a different approach, Lu et al. proposed a two-stage method for performing an NMA as a meta-regression (76). At the first stage a meta-analysis is performed in each group of trials comparing the same treatments, e.g. all two-arm trials that compare A vs. B, all three-arm trials that compare A vs. B vs. C, etc. This provides the direct estimates on treatment comparisons. At the second stage of the meta-analysis, a weighted linear regression is performed with the direct estimates as dependent variables. This provides inference for the basic parameters. This two-stage method can be used to investigate how the first-stage (direct) evidence influences the network estimates and may therefore help to assess the consistency of the network (see next section).

2.3.2 Network meta-analysis as a hierarchical model

Hierarchical NMA models (5,6) seem to be implemented most often (7,8). An important advantage of this approach is that if arm-level data are available, their exact likelihood can be used (46).

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

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

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

$$logit(p_{iA}) = u_i,$$
$$logit(p_{iB}) = u_i + \theta_{iAB}.$$

Here u represents the log-odds of the outcome for treatment A and θ_{iAB} the log-odds ratio of A versus B; the event probabilities for each arm parameterize the binomial

likelihood, $r_{iT} \sim Bin(p_{iT}, n_{iT})$, with r_{iT} denoting the events and n_{iT} the total number of randomized patients in each treatment arm ($T = A, B$). We then allow $\theta_{iAB} \sim N(\mu_{AB}, \tau_{AB}^2)$ for a random-effects meta-analysis. If two non-reference treatments are compared in a study, e.g. treatments B and C, we utilize the consistency equations by setting $\theta_{iBC} \sim N(\mu_{AC} - \mu_{AB}, \tau_{BC}^2)$. In the presence of multi-arm studies multivariate normal distributions should be used instead, where the within- and between-study variances are replaced by the corresponding variance-covariance matrices $\boldsymbol{S_i}$ and $\boldsymbol{\Delta_i}$. Details on how to model other types of data can be found in (46). Note that the issues discussed in the previous section regarding the estimation of the between-trial heterogeneity hold for the hierarchical models as well.

NMA can be fitted as a hierarchical model also if only contrast-level data are available from the studies (i.e. when the reported data is on the relative treatment effects of the treatments being compared, but not on the specific arms). For a two-arm study i comparing A (reference treatment) and B the model is written as $y_{iAB} \sim N(\theta_{iAB}, s_i^2)$. Note here that the normality assumption can be justified even if the underlying patient-level distributions are skewed, due to the central limit theorem (46).

Hierarchical models can also be fitted when a combination of arm-level and contrast-level data is available, using the exact likelihood for the arm-level data and the normal approximation for the contrast-level data in a so-called shared parameter model (46).

2.3.3 Network meta-analysis as a multivariate meta-analysis mode

White et al. (77) suggested a method of performing NMA as a multivariate meta-analysis by treating the basic comparisons as different outcomes and by employing standard multiple-outcome meta-analytical techniques (78). For this model to work all studies need to report on the reference treatment; if this is not the case for some studies, a data-augmentation technique is required to impute a minimally informative reference treatment arm. The model is written as $\boldsymbol{y} = \boldsymbol{X}^*\boldsymbol{\mu} + \boldsymbol{\varepsilon} + \boldsymbol{\delta}$, with $\boldsymbol{X}^*$ being a matrix with all elements either 0 or 1, depending on which 'outcomes' are reported in each study.

Assume for example that treatments A, B and C are compared in a number of studies, and also assume that treatment A is taken to be the reference treatment. In this approach $\boldsymbol{\mu}$ will be a 2×1 vector of the basic parameters, AB and AC. A study comparing A vs. B will contribute an element 1 in the first column of $\boldsymbol{X}^*$ and 0 in the second, since in this study only the first 'outcome' is reported. An A vs. C study will report the second outcome only, thus

the relevant elements in $\boldsymbol{X}^*$ will be 0 and 1 respectively. For a B vs. C study, however, an A arm must be imputed; this study becomes three-arm, and reports on both 'outcomes'. The model for these three studies is as follows:

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

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

The models described in this section should be considered equivalent; the choice between them should be primarily dictated by the availability of software packages for implementing them and by the technical expertise of the researchers. We discuss the currently available software options for fitting all models presented in this review in a following section. Alternative models for performing an NMA have also been recently proposed in the literature. Rucker (79) described the analogy between network meta-analysis and electrical networks and applied graph-theoretical methods to perform a fixed-effects NMA. Also, Yang et al. (80) introduced a confidence distribution approach for performing an NMA. In this approach, instead of combining point estimates from each study, the authors combine confidence distributions.

For more in-depth reviews of the methodology that include the statistical details of the models we presented we refer the reader to (15,81,82).

2.4 Detecting inconsistency in networks of interventions

As we have previously discussed (Section 0), transitivity is a central assumption of NMA. A statistical assessment of this assumption can be made by checking whether the various sources of evidence fit together in a coherent way. This assessment is vital for ensuring that the NMA results are valid and interpretable for clinical decision making (83), but may be difficult to do in practice, especially in the case of complex networks or when multi-arm studies are included in the network.

Statistical consistency can be assessed only in closed loops of evidence; there are two general approaches to do this: either locally (by focusing on the inconsistency of a specific treatment comparison) or globally (by checking for inconsistency in the entire network of

evidence). In what follows we discuss methods and models that have been proposed for both of these approaches.

2.4.1 Local methods to detect inconsistency

A straightforward approach for evaluating the presence of inconsistency in a network is to apply a loop-specific approach; in this approach we examine each loop of the network in isolation with the rest of the network. For an ABC loop in the network, for example, we choose one of the comparisons (e.g. B vs. C) and compute the direct ($\hat{\mu}_{BC}^{Dir}$) and indirect estimates ($\hat{\mu}_{BC}^{Ind}$). Their absolute difference measures inconsistency and is usually termed inconsistency factor (64): $\widehat{w}_{ABC} = \left|\hat{\mu}_{BC}^{Ind} - \hat{\mu}_{BC}^{Dir}\right|$, with variance $var(\widehat{w}_{ABC}) = var(\hat{\mu}_{BC}^{Ind}) + var(\hat{\mu}_{BC}^{Dir})$. A 95% confidence interval can be obtained as $\widehat{w}_{ABC} \pm 1.96\sqrt{var(\widehat{w}_{ABC})}$ and a Z-statistic for the null hypothesis of consistency, i.e. $\widehat{w}_{ABC} = 0$, can be constructed as $z_{ABC} = \widehat{w}_{ABC}/\sqrt{var(\widehat{w}_{ABC})}$; this can be compared with the standard normal distribution to obtain a p-value (49). Inconsistency is a property of the loop, in the sense that choosing a different treatment comparison of the loop and repeating the computations would give the exact same results (57,64); thus we denote the inconsistency factor with an ABC subscript.

The loop-specific approach can be applied for each loop in a network to point out hotspots for inconsistency. The major advantage of this approach is that it is easy to implement, it suffers however from important limitations: when a treatment comparison is a part of more than one loop this method does not compare direct evidence for this comparison to all available indirect information, but to evidence from only one loop at a time; also in this case the tests for different loops sharing this comparison will not be independent. In addition, for networks with many loops there are multiple-testing issues.

It is possible to extend the loop-specific approach by accounting for more than one indirect estimates for a treatment comparison (composite test for inconsistency (57,84)). Suppose that there are *L* loops that provide independent indirect information for the A vs. B comparison; these can be combined with the direct information using the usual inverse variance method to obtain a pooled, overall estimate of the relative treatment effect of A vs. B. Under the null hypothesis that the $L + 1$ different estimates are in agreement, a test statistic following a chi-squared distribution with *L* degrees of freedom can be constructed to check for inconsistency. One should keep in mind, however, that the presence of multi-

arm studies induces correlations among the estimates for the treatment effects, which this method, as well as the loop-specific approach, fail to account for.

Dias et al. also proposed two additional methods for locally checking inconsistency (54). The first method ("back-calculation") can be applied when the only available data are the pooled summaries of the pairwise meta-analyses. In the first step, the data is used to obtain a network estimate for each pairwise comparison in the network. It is then assumed that this estimate is a weighted average of the direct and the indirect evidence, coming from the rest of the network. This allows a back-calculation of the indirect estimate and its variance, which in turn can be used to construct a Z-test for the difference of direct and indirect evidence. Note that this method is problematic for a random-effects meta-analysis, as the posterior distribution of the heterogeneity variance will in general be different between the NMA model and the model for the pairwise meta-analysis.

The second method proposed by Dias et al. (54), the node-splitting approach, can be used when trial-level data are available. In this method the direct evidence for a specific treatment comparison is excluded from the rest of the network and is used to obtain a direct estimate. The remaining information in the network is used to obtain an indirect estimate for this comparison, after fitting an NMA model. The two estimates, direct and indirect are then used to evaluate inconsistency with a Z-test. The main drawbacks of this approach (as well as the back-calculation approach) are that they might be computationally intensive, especially for large networks with many treatment comparisons, and that they cannot properly handle multi-arm studies.

2.4.2 Global methods to detect inconsistency

Lu and Ades introduced a model (64), in which the consistency equations are 'bent' by including extra terms, the inconsistency factors. For an ABC loop, for example, the consistency equation is written as $\mu_{BC} = \mu_{AC} - \mu_{AB} + w_{ABC}$, where the w parameter measures the discrepancy of direct and indirect evidence. For networks comprising many loops a different inconsistency factor needs to be included in each loop. When the network only includes two-arm studies the number of independent inconsistency factors (the 'inconsistency degrees of freedom', $ICDF$) is $ICDF = C - T - 1$, with C being the number of available pairwise comparisons in the data and T the number of different treatments. The inconsistency factors can be assumed to follow a common distribution in order to increase the power in their estimation. A χ^2 test can be used to assess the inconsistency of the whole

network, under the null assumption that all inconsistency factors are zero. In the presence of multi-arm studies, however, this model is problematic (85). Higgins et al. (86) showed that different parameterizations of the model may lead to different results. Thus, when multi-arm studies are present, the use of the Lu and Ades model should be avoided.

Higgins et al. (86) and White et al. (77) introduced an alternative inconsistency model, the 'design-by-treatment' interaction model, which encompasses both loop and design inconsistencies. The latter corresponds to the possible discrepancies in the treatment effects across designs, where 'design' refers to the treatments being compared in a study. For example, a study comparing treatments A and B is considered to be an AB design. The A vs. B estimate coming from such a study may be different than the A vs. B estimate coming from a three-arm study comparing treatments A, B and C (ABC design); this difference is referred to as design inconsistency. In the absence of multi-arm studies the Lu and Ades model is equivalent to the design-by-treatment model. Similarly to the Lu and Ades model, when multi-arm studies are present, the estimates of the inconsistency factors depend on the parameterization. Unlike the Lu and Ades model, however, the global statistic for inconsistency in the design-by-treatment interaction model is invariant under re-parameterization. The main drawback of this approach is that the definition of inconsistency seems artificial, as it is mainly dictated by methodological rather than clinical considerations. A model similar to the design-by-treatment was also proposed by Piepho et al. (87).

An alternative method was introduced by Dias et al. (57). In this model the consistency equations are completely removed, and the network meta-analysis model is equivalent to a series of separate, independent meta-analyses for each pairwise contrast, sharing, however, a common heterogeneity variance (88). The fit of the model is then compared to the standard consistency NMA model using the posterior deviance and the deviance information criterion (DIC) (89). In the presence of multi-arm studies, however, a re-parameterization will affect the results of a random-effects meta-analysis. In addition, estimating the contribution to posterior mean deviance for each data point can help identify possibly 'problematic' studies, i.e. studies not fitting well with the rest of the evidence. Each data point is expected to have a contribution of about 1 to the posterior mean deviance. A larger value will suggest a poor fit to the model, pointing out possibly inconsistent pieces of evidence; also, the use of leverage plots was suggested as a diagnostic tool for identifying inconsistency (54).

A different approach for globally assessing inconsistency is by using the Q statistic for inconsistency (90), which is analogous to the Q statistic for heterogeneity in simple meta-

analysis. This approach is based on the two-stage method for fitting NMA (76). On the first stage we perform a pairwise meta-analysis for the studies of each design available in the dataset, and obtain the direct relative treatment effect. From the AB studies, for example, we estimate $\hat{\mu}_{AB}$; this can be used to compute the Q_{AB}^{het} statistic for heterogeneity as $Q_{AB}^{het} = \sum_i \frac{1}{s_i^2}(y_i - \hat{\mu}_{AB})^2$, where i runs through all AB studies. Similarly, on the second stage of the analysis we obtain network estimates for all pairwise comparisons; using these estimates and the direct estimates of the first stage, a Q statistic for the inconsistency of the whole network can be obtained and the null hypothesis of consistency in the network can be tested using a χ^2 distribution with $C - T - 1$ degrees of freedom. This approach can be generalized to account for the presence of multi-arm studies. Rucker has also suggested a Q statistic for inconsistency (79), but it is only applicable for fixed-effects NMA and cannot handle multi-arm studies.

In a another, graphical approach proposed by Chung and Lumley (91) the multidimensional scaling method is used to infer about inconsistency in a network. For each pairwise comparison a usual inverse variance meta-analysis is performed; the magnitude of the relative treatment effects is considered to be a measure of the observed pairwise 'dissimilarity' of the treatments. The pairwise estimates are summarized in a dissimilarity matrix, to which a weighted multidimensional scaling is applied in order to obtain the 'fitted dissimilarities'. Important differences between observed and fitted dissimilarities are an indicator of possible inconsistencies. Note that this method cannot properly handle multi-arm studies.

2.4.3 Empirical studies and simulations on inconsistency

Empirical studies show that the prevalence of inconsistency in published networks is non-trivial. Song et al. (92) performed a meta-epidemiological study that included 112 published triangular networks, 16 of which were found to be statistically inconsistent. Veroniki et al. (93) evaluated inconsistency in 40 published networks including a total of 303 loops. They found that 2-9% of the loops were inconsistent, depending on the effect measure used and the assumptions for heterogeneity; also, approximately one eighth of the networks were found to be inconsistent using the design-by-treatment method.

The various methods for assessing inconsistency have been rarely and poorly applied in published NMAs (11). In a meta-epidemiological study by Nikolakopoulou et al. (8) it

was found that in 24% of the published NMAs the authors did not use appropriate methods to evaluate inconsistency, while in 44% the authors did not report using any method at all.

Song et al. (36) performed simulations in order to evaluate the statistical properties of various methods for inferring about network inconsistency. They explored the use of the loop-specific approach, the node-splitting technique and the Lu & Ades model. They found that even though all methods are unbiased, they have little power in detecting inconsistency. It is also important to note that inferences on inconsistency heavily depend on the extent of heterogeneity and the method used to evaluate it (93). Thus, analysts should keep in mind that a statistically non-significant estimate for inconsistency should not be interpreted as proof of consistency. In addition, even when statistically significant inconsistency is found, its magnitude should be interpreted in terms of clinical relevance; thus, a statistically-significant inconsistency in a certain loop might be clinically unimportant.

2.5 Choosing between the methods for evaluating inconsistency

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

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

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

If statistically significant inconsistency is detected, researchers are advised to explore potential sources of it and try to explain it. Local methods for assessing inconsistency can indicate outlying studies, which should be checked for data extraction errors, important differences in the distribution of effect modifiers or other possible biases. In Section 2.9 we present various models for adding covariates and adjusting for suspected biases in the analysis. If sufficient studies are available, such models can be applied to explain and possibly eliminate inconsistencies, while, if inconsistency persists, researchers can consider splitting up the network (see discussion in Section 2.13). Finally, in the case of unexplained inconsistency, researchers may choose not to synthesize the evidence in an NMA at all, or to present the results from the appropriate inconsistency model (Lu & Ades model when all studies are two-armed; design-by-treatment model when multi-arm trials are present) along with the direct evidence and a warning to the readers of the limitations of the analysis).

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

Approach to inconsistency	**Method**	**Limitations**
Local methods	Loop-specific	• Does not compare direct to all indirect evidence for each comparison • Different loops sharing a comparison are not independent • Multiple-testing issues
	Composite test	Fails to account for correlations induce by multi-arm studies
	Back-calculation	• Problematic for a random-effects meta-analysis • Cannot properly handle multi-arm studies
	Node splitting	• Computationally intensive • Cannot properly handle multi-arm studies
Global methods	Lu and Ades model	Depends on parameterization when multi-arm studies are included
	Design-by-treatment model	Non-intuitive definition of inconsistency.
	Unrelated mean effects model	Depends on parameterization when multi-arm studies are included
	Q statistic for inconsistency	Based on the notion of design-by-treatment inconsistency model (non-intuitive definition of inconsistency)
	Multidimensional scaling	Cannot properly handle multi-arm studies

2.6 Software options for fitting network meta-analysis and statistically evaluating inconsistency

Our literature search showed that BUGS software is a popular choice for implementing new methods in NMA, the majority of the articles included in our database reported using WinBUGS, OpenBUGS (94,95) or JAGS (96): (3–5,14,19,26–28,31,34–36,38,46,54,57,59,61,63,64,67,68,71,75,77,82,84,88,97–147). An alternative option for implementing Bayesian statistical inference is Stan, a recently developed programming language (148). However, we did not identify any articles using Stan.

Also, there were many articles that reported the use of R (149): (36,54,61,63,65,69,76,79,83,90,91,93,111,122,129,135–137,139,141,147,150–154); some papers used STATA (155): (34,35,71,77,86,93,140,142,156–158) and a few papers reported using SAS software (159): (100,115,34,118,87,160–162). Finally, Van Valkenhoef et al. (141) presented GeMTC, a freely-available, open-source program for performing NMA.

Neupane et al. performed a review of the available automated packages for performing an NMA in R aiming to summarize the key features and functionality of each package (163). In Table 2 we describe possible software solutions for some of the models presented in this review.

Table 2: Available software solutions for NMA

Method	Software
NMA as meta-regression	• BUGS • Stata (metareg) • R (rma command from metaphor; netmeta package)
Hierarchical NMA model	• BUGS codes available at: http://www.mtm.uoi.gr and http://www.bris.ac.uk/social-community-medicine/projects/mpes/ • GeMTC software
NMA as a multivariate meta-analysis	• Stata (mvmeta, network) • BUGS • R (mvmeta)
Loop-specific approach for inconsistency	• BUGS • R (routines available at http://mtm.uoi.gr/index.php/how-to-do-an-mtm) • Stata (network_graphs, available at http://mtm.uoi.gr/index.php/stata-routines-for-network-meta-analysis)
Node splitting approach	• BUGS (codes available at http://www.bristol.ac.uk/cobm/research/mpes) • GeMTC software • Stata (mvmeta; network)
Lu & Ades model	• BUGS • Stata (mvmeta) • GeMTC software
Design-by-treatment model	• BUGS • Stata (mvmeta; network)
Q-statistics in NMA	• R (routine available at http://www.unimedizin-mainz.de/fileadmin/kliniken/imbei/Dokumente/Biometrie/Software/netheat.R) • R (netmeta) • Stata (mvmeta; network)
Graphical presentation	• Stata (network_graphs, available at http://mtm.uoi.gr/index.php/stata-routines-for-network-meta-analysis)

2.7 The use of different measures of effect size

There is a wide choice of summary effect measures that can be used for the meta-analysis of evidence on a binary outcome. The most common choices are the odds ratio (OR), risk ratio for harmful or beneficial outcomes (RR_H and RR_B), risk difference (RD) and hazard ratio (HR) for time-to event data. Donegan et al. (11) reported that the majority of the published NMAs of dichotomous data used OR and RR (50% and 40% respectively), with RD being used in only 10% of the analyses. Veroniki et al. (93) analyzed 40 published networks and showed that the choice of effect measure may have an impact on the inferences about the statistical inconsistency. This was also discussed by Coory and Jordan (66); using information from published networks they concluded that the use of OR and RR is preferable over RD. In addition, it has been demonstrated that the choice of the scale may have an impact on the results of an NMA (164). In particular, Eckermann et al. (165) showed that the use of RR may lead to inferential fallacies and advocate the use of OR. Norton et al. (158) discussed that different choices of scale may lead to differences in the ranking of the treatments in an NMA. They propose that researchers should explore how sensitive the NMA results are in the choice of effect measure. Van Valkenhoef and Ades (166) on the other hand discuss that a rank reversal is unlikely to take place unless the assumptions underlying NMA do not hold, or the data is very sparse.

Caldwell et al. (88) proposed the use of the posterior mean deviance and the deviance information criterion (DIC) to evaluate the model fit of the different effect measures in an NMA. The choice of the effect measure can also be guided by considering the estimates of between-study heterogeneity, with lower values being preferable; however this might be problematic when there are not enough data available, in which case the choice of scale may be driven by the ease of interpretation and the epidemiological understanding of the disease process (46,88). HR should be always considered as a suitable choice of scale for the case there is an underlying time-to-event process and the proportional hazards assumption is deemed plausible (88,133) (see following Section).

Note that the discussion of this paragraph pertains to the analysis of a binary outcome. When continuous data is available the analysts should avoid dichotomization since it translates into a loss of power and also because the choice of cut-off point may impact on the inferences of NMA (132).

2.8 Modelling time-to-event data in network meta-analysis

In many RCTs the outcome measured is the time to the occurrence of an event (e.g. death, disease progression etc.). Welton et al. (143) described a method for simultaneously synthesizing survival and disease progression outcomes in a single NMA analysis; also, Woods et al. (146) provided guidance on how to perform an NMA on the log-hazard scale when studies report different survival statistics.

Analysts, however, should keep in mind that the synthesis of time-to-event data in terms of hazard ratios relies on the proportional hazards assumption; treatment effects, however, may vary over time and this might threaten the validity of the meta-analysis results. For NMA this might have an extra impact, on the consistency of the results. Ouwens et al. (124) and Jansen (116) modeled the hazard functions using parametric survival curves and fractional polynomials respectively; in these models the hazard ratio is allowed to vary over time. Jansen and Cope (117) discussed methods for extending these models by including covariates that can reduce possible inconsistencies and bias. In another paper by the same authors (104), various alternative summaries were presented for summarizing the estimates of the relative treatment effects obtained from an NMA of survival data.

2.9 Extension of network meta-analysis to account for effect modifiers

In a pairwise meta-analysis a meta-regression on important effect modifiers (covariates) can explain the presence of between-study heterogeneity, which may hinder the interpretation of the results and may have important implications in decision making (106). In NMA interpreting results will be even more problematic in the presence of evidence inconsistency; meta-regression techniques in NMA adjust the treatment effects for possible effect-modifiers and can reduce heterogeneity and inconsistency in the results that may be present when these covariates are distributed unevenly among studies (28,34,61,81,97,103,106,123). The effect modifiers can be either categorical or continuous variables, and may represent either patient-level or trial-level characteristics.

2.9.1 General model for including covariates in network meta-analysis

Nixon et al. (123) first combined NMA and meta-regression techniques to develop models that allow the simultaneous comparison of multiple competing treatments while adjusting for study-level covariates, in an attempt to investigate and explain heterogeneity.

Salanti et al. (61) and Cooper et al. (103) proposed the use of meta-regression as a tool for eliminating inconsistency as well as heterogeneity in NMA. As an example of adding covariates in NMA, Salanti et al. (61) considered the year of randomization in each trial as a covariate in an NMA for topical fluoride treatments for the prevention of dental carries. The covariate adjusted for possible time trends in the placebo-controlled comparisons, and relative treatment effects were estimated for a pre-defined year of randomization (the year of the most recent study).

In general, there are three main approaches in the meta-regression of study-level covariates for NMA (103,106): using different and unrelated interaction terms (coefficients), using exchangeable interaction terms and using a common interaction term.

Unrelated interaction terms

In this approach there are a number of interaction terms for each covariate equal to the number of the basic parameters of the model. Let us assume for simplicity that we are only interested in one study-level covariate x_i. We can augment the hierarchical random-effects model previously presented as follows: for a study i, comparing treatments B versus C, we allow $\theta_{iBC} \sim N(\mu_{BC} + x_i\beta_{BC}, \tau^2)$, assuming a common heterogeneity variance τ^2 for the treatment effects. If treatment A is chosen to be the reference treatment, we can utilize the consistency equations to write $\theta_{iBC} \sim N(\mu_{AC} - \mu_{AB} + x_i(\beta_{AB} - \beta_{AC}), \tau^2)$; in a Bayesian analysis the β_{AT} 'basic' coefficients (where T≠A) can be assigned unrelated vague prior distributions.

Exchangeable interaction terms

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

Common interaction term

The common interaction term model is the same as the exchangeable interaction model described in the previous paragraph, but now all basic interaction terms are assumed equal, $\beta_{AT} = \beta$, for all treatments T≠A. A vague prior is then assigned to β. This model implies that the relative treatment effects between the non-reference treatments are independent of

the covariate, since the interaction terms cancel out. In this case the choice of the reference treatment becomes important, as it might change the meta-regression results (106).

2.9.2 Network meta-analysis meta-regression for baseline risk

The underlying risk of the disease, usually termed as 'baseline risk', is a proxy for important patient characteristics that may be possible modifiers of the treatment effects and it can be included as a covariate in an NMA; however, care should be taken to account for its correlation with the treatment effects (97,106). Achana et al. (97) proposed a random-effects meta-regression model in which the effect of the reference treatment was used as a measure of the baseline risk. In order to include studies not reporting the reference treatment the authors proposed three alternative distributional assumptions for the 'true' unobserved baseline risk. Following Cooper et al. (103), the interaction terms were taken to be independent, exchangeable or common. The authors recommended that the goodness of fit of the various alternative configurations can be based on residual deviance.

2.9.3 Limitations of network meta-analysis meta-regression models

Dias et al. (106) advocate that even though the use of the models with exchangeable coefficients seems attractive, they are likely to lead to statistically insignificant interaction terms; when this is the case decision-making may be based in non-robust results. Therefore, even though the exchangeable coefficient model – or even more complex models – can be fitted, the authors suggest that their use should be limited to exploratory analyses. Also, analysts should keep in mind that NMA meta-regression inherits all the interpretation difficulties from regular meta-regression, most importantly the inability to infer causal relationships (106), and the risk of ecological bias if study-level covariates are used to infer about individuals. An additional drawback of meta-regression models for decision-making in general is that in order to assess the relative treatment effects the analyst must choose a value of the covariate at which to make the comparison (28).

2.10 Investigating potential sources of bias in network meta-analysis

When combining results from different studies researchers always run the risk of obtaining biased pooled estimates. This may be the case when some of the studies provide biased evidence ('internal bias'); for example when treatment effects are overestimated in studies of low methodological quality. The pooled result may be biased even if the estimates

of the included studies are not, ('external bias'); e.g. when studies without statistically significant results have not been published (167).

Dias et al. (106) discussed that when confronted with studies of mixed quality, researchers have three options: they can choose to analyze only the high-quality studies, thus ignoring a possibly important amount of information, they can choose to analyze all evidence, thus risking a bias in the pooled estimates, or they can include all studies after taking into account and adjusting for possible biases in the studies. In what follows we focus on the latter, presenting various available approaches for adjusting for suspected internal biases in the included studies, and also for adjusting for various forms of external bias.

2.10.1 Accounting for study limitations in network meta-analysis

A conceptually straightforward way to adjust for possible biases in the included studies is by eliciting bias distributions (167). In this approach a number of independent experts evaluate each study separately in terms of some predefined criteria and provide information that is used in order to construct an overall bias distribution. The parameters of this distribution are combined with the estimates of the studies in order to produce a bias-adjusted estimate of the treatment effect in each study. These estimates are then synthesized using standard NMA techniques. A disadvantage of this approach is that it is rather difficult and time-consuming to implement (106).

A class of models assumes that biased studies estimate $\mu + \beta$, where μ is the unbiased treatment effect and β is a bias parameter. If the study-specific bias parameters are assumed to be exchangeable across studies the unbiased treatment effects and the mean bias can be estimated from the network (106,109). Dias et al. (109) presented a model where exchangeable bias parameter with non-zero mean were included in studies that compared an active versus an inactive treatment and were considered to be of a high risk of bias (according to some predefined measure such as allocation concealment, blinding or other trial characteristics). They also explored the use of two different bias parameters, one for active versus inactive, and one for active versus active comparisons; note that in this approach some assumption on the direction of bias in the active versus active trials is necessary to be made. Salanti et al. (127) considered a similar model in which the newest treatments were favored, thus adjusting the treatment effects for possible 'optimism' or 'novelty' bias (108). Study size can be a proxy for the study's risk of bias and Chaimani and Salanti (102) presented a method for adjusting for the 'small study effects', the exaggeration of treatment effects in

smaller trials. This exaggeration might be due to methodological differences between smaller and larger trials that affect treatment effectiveness, due to publication bias or due to reporting bias. They proposed a network meta-regression model, where the bias parameter is multiplied with the observed variance of the treatment effects in each study; the standard error or the precision (inverse variance) can be used alternatively. Their model can also adjust for suspected 'sponsorship' bias, for the case when interventions are sponsored in some of the studies. A similar model was also presented by Trinquart et al. (139).

2.10.2 Selection model to account for publication bias

Mavridis et al. (121) proposed a Bayesian implementation of the Copas selection model (168) for addressing for possible publication bias in NMA. The idea behind selection models is that the observed set of published studies is considered to be a 'biased' sample, due to the nature of the publication process. This is addressed by introducing a latent variable for each study, the 'propensity of publication', which is assumed to be correlated with the study's effect size. Mavridis et al. (121) modeled propensity via a regression model, where it was assumed to be inversely proportional to the standard error of the effect size. They considered alternative scenarios of how the selection model parameters depend on the treatments being compared in each study. Trinquart et al. also presented a selection model (139) which modeled the propensity score of a trial as a linear function of the standard error. The effect sizes of the studies were weighted according to their propensity. Their model was shown to yield similar results to the model by Mavridis et al.

2.10.3 Accounting for ecological bias

The meta-analysis of aggregated data can lead to *ecological bias*. This refers to a bias that may arise when using aggregated data in order to make inferences about patient-level interactions. Govan et al. (113) proposed an NMA model to control for ecological bias by specifically modeling the effects of the covariates. Their model allows the inclusion of studies that provide information on all covariates, studies that report marginal data on some of the covariates and also studies not providing any covariate information at all. The model allows the joint estimation of both the treatment and the covariate effects.

2.10.4 Graphical approaches to assessing bias in network meta-analysis

In a different approach Salanti et al. (53) discussed how the geometry of the network can offer insight on the presence of a 'comparator preference' bias, i.e. when head-to-head comparisons between effective treatments are deliberately avoided, which in turn would imply that the treatment effects versus the reference treatments might be exaggerated. The authors utilized two indices from ecological literature: diversity, which is a measure of the number of treatments present in the network and how often they were tested, and co-occurrence, which measures whether specific treatment comparisons were preferred in the network while others were avoided. Limited diversity and statistically significant co-occurrence in a network is an indicator of possible preference bias in the network (18).

Jansen et al. (169) discussed the use of directed acyclic graphs (DAGs) as a graphical tool for conceptually evaluating the consistency assumption and also identifying possible sources of bias in indirect and mixed treatment estimates. By means of DAGs they showed that NMA estimates can be biased when relative treatment effect modifiers vary across comparisons and are not adjusted for in the analysis. They also showed that adjusting for covariates that are not effect modifiers is not only unnecessary, but that it can introduce bias.

2.10.5 Empirical assessments of the impact of bias in network meta-analysis

Chaimani et al. (101) performed a network meta-epidemiological study to explore the effect of trial characteristics and study precision in NMA. They analyzed 32 networks and found evidence that imprecise studies (studies reporting broader confidence intervals for their estimates) tend to report larger effects compared to more precise studies, thus altering the results of the NMA. However, they found no evidence of association between effect size and previously identified indicators of bias, including blinding, allocation concealment and random sequence generation. Trinquart et al. (138) used data from 74 FDA-registered placebo-controlled studies on 12 antidepressants along with 51 corresponding publications in order to assess the impact of publication bias. They found that the effect sizes derived from published studies differed from the ones derived from the FDA data by at least 100% for almost half of the pairwise comparisons. They concluded that reporting bias alters NMA estimates and changes the treatments' ranking. They also noted that the impact of reporting bias may be more important in NMA compared to classical meta-analyses, in the sense that

reporting bias in one treatment comparison may have an effect in the ranking of all treatments in NMA.

2.11 Reporting results from network meta-analysis

Although the implementation of NMA is increasingly gaining popularity, the quality of reporting has been rather low. Various meta-epidemiological studies of published NMAs showed that the methods used and the assumptions made were not routinely reported (8,11,170,171). Ohlssen et al. (82) presented a checklist of items that should be reported in a drug-safety Bayesian NMA while Ades et al. (172) and Mills et al. (173) give guidelines for those reviewing an NMA for the purposes of decision making.

One possible hurdle in the reporting of an NMA is that presenting all results can be a challenging task, especially for networks with many treatments and multiple outcomes. The literature offers a plethora of graphical and tabular methods for visualizing the evidence base (91,157,174), the assumptions made (90,157) and the results obtained from an NMA (38,91,154,157,174–176). In a meta-epidemiological study on the presentational approaches used, Tan et al. (177) examined NMAs published in the UK and found that there is no standardized presentational approach for reporting the results of NMA. The authors concluded that a standardization of reporting is required.

2.12 Modelling repeated measures and multiple outcomes

In some cases, studies may report on a single outcome for multiple time points, which leads to a series of correlated observations. Lu et al. (119) proposed a hierarchical NMA model for synthesizing repeated measures of a discrete outcome. Dakin et al. (105) suggested a model for a continuous outcome, but did not include in the analysis the correlations between the observations. Ding and Fu (110) also presented a model for a continuous outcome that automatically modeled the correlations between the observations at different time points. Madan et al. (120) presented methods for analyzing two dichotomous outcomes reported on multiple time points, for studies comparing complex interventions.

As we discussed in the Introduction chapter, RCTs commonly report on more than one outcome. These outcomes may be correlated within a study (due to the fact that observations come from the same set of patients) and in addition the true treatment effects on the outcomes can be correlated across studies (reflecting the way outcomes are related when measured in different settings). The usual meta-analytical approach on multiple outcomes is to analyze

each one separately, ignoring all possible correlations. On the other hand, a joint meta-analysis of all outcomes which incorporates possible correlations can increase precision by 'borrowing of strength' across outcomes and may reduce the impact of outcome reporting bias (178,179).

Welton et al. (145) described a method for performing an NMA of two correlated outcomes, but it can only be used for the case when all studies are two-armed. Schmid et al. (129) proposed a model for unordered categorical data that also allows the inclusion of studies with partially observed data Hong et al. (114) presented a model for multiple outcomes that does not take into account within-study correlations. Competing risks is a special case of multiple-outcome structure where the outcomes are mutually exclusive; Ades et al. (99) presented methods for performing a competing-risks NMA. Price et al. (125) discussed methods for an NMA in multi-state Markov models; a model averaging technique was also proposed (126) for combining estimates from alternative multi-state models. More details regarding methods for jointly synthesizing multiple outcomes in NMA can be found in Chapters 3 and 4 of this dissertation. There we also present a range of new models that we developed for the purposes of this PhD.

Using NMA results to decide which of all available treatments is optimal for a specific condition might be a non-trivial issue, when the treatments are compared for more than one outcome. In order to facilitate decision making in the presence of multiple outcomes, Van Valkenhoef et al. (69) proposed a method for multiple criteria benefit-risk assessment of all competing treatments in an NMA; also, Hong et al. (115) described a similar method for producing an overall ranking of the treatments in the network using a scoring system for combining efficacy and safety outcomes.

2.13 Definition of nodes in the treatment network

One important decision that analysts must make in the onset of an NMA regards the number of nodes (treatments) to be included in the network. A simple choice would be to include all relevant treatments; alternatively researchers might want to focus on just a subset of the treatments, the ones that are deemed to be clinically relevant (e.g. newer/more effective treatments). This, however, poses a dilemma, since including in the evidence-base studies that compare treatments that are not clinically interesting might provide additional indirect evidence for the clinically interesting ones, which in turn may increase the precision of the results (24,180).

Hawkins et al. (181) performed an empirical study that supported the use of all potentially relevant data; in another empirical study Mills et al. (122) concluded that the exclusion of treatments in an NMA might have an important effect on the results and might limit its usefulness, if important comparisons are excluded. On the other hand, obtaining all relevant evidence, including clinically uninteresting treatments, may be very time consuming and inefficient. To address this issue Hawkins et al. (182) presented two alternative iterative search strategies for identifying an efficient set of evidence, where the comparators included in each search is determined by the results of the previous iteration. In addition, Cooper et al. (14) showed that extending the network to include more treatments might lead to increased heterogeneity, which in turn will increase the uncertainty in the results despite the inclusion of additional information.

An additional issue that analysts might face regards the definition of treatments across studies. It is not uncommon for a treatment to be administered in different ways in the included studies, for example in different doses. This differential definition of the nodes will make the transitivity assumption less easy to defend and might cause inconsistency and/or heterogeneity in the results (24,67). Del Giovane et al. (112) and also Warren et al. (142) presented various alternative models to account for variability in treatment definition due to differences in the dose. In another frequently encountered scenario in which the definition of nodes can be of importance is when interventions are administered as a combination of more than one treatment; the simplest approach would be to analyze each combination as a different node in the NMA. Welton et al. (144) and Mills et al. (183) proposed possible scenarios for modeling how interventions interact with each other when combined into a complex intervention, with one of the approaches being the assumption of additive treatment effects. Thorlund et al. (137) performed a simulation study which showed that when the treatment effects are truly additive, the conventional NMA model performs poorly in comparison to the additive effects model.

In summary, even though there is no exact recipe available for setting up the network and defining the nodes, the choice should be guided by considerations of the transitivity assumption, the presence of statistical inconsistency, the possibility of bias and also practical constraints on the resources available for setting up the database. Ideally, whenever possible such decisions should be described a priori in the protocol in order to avoid selective use of data (24).

2.14 Incorporating individual patient data in network meta-analysis

The NMA models we have discussed so far can be used only for the analysis of aggregated data (AD) while the meta-regression approaches presented in Section 2.9.1 allow the exploration of the effects of only study-level covariates to the relative treatment effects. On the other hand the use of individual patient data (IPD) in an NMA (either exclusively or in combination with AD) is expected to increase precision and also allows the distinction between within-study from across-study associations to be made, so as to avoid possible ecological bias (150). Debray et al. extensively discuss the statistical methodology and the potential advantages of an IPD-MA when pooling head-to-head trials (184). These advantages also apply to NMA, and access to IPD is particularly relevant when the number of included studies is small and the validity of using meta-regression of study-level covariates becomes increasingly questionable. The use of patient-level covariates will allow a better evaluation of the heterogeneity and inconsistency in the network (19,28,81,185).

A few models for including IPD in an NMA have been recently proposed. Saramago et al. (128) developed a series of NMA models set in a Bayesian background that can be used for the simultaneous synthesis of IPD and AD while incorporating both study and individual-level covariates. Their models also allow the inclusion of studies with different designs (cluster and individual allocation). The authors found that the incorporation of IPD in the network resulted in an increase in the precision compared to an AD-only analysis, even when IPD are available only for a fraction of the studies. Donegan et al. (83,111) presented a model for combining IPD and AD in a single analysis with three alternative specifications (unrelated, exchangeable and common interactions; see also Section 2.9.1). The inclusion of both IPD and AD in the analysis was shown to lead to an increased precision of the estimates of the regression coefficients and a better assessment of the consistency assumption. A similar model was proposed by Jansen (150). In the same paper a second, alternative model was also suggested for the case of a binary covariate. The author performed a simulation study indicating that the second model is less affected by bias at the cost of larger uncertainty in the results. Finally Ali et al. (186) discussed the use of IPD in order to identify possible interactions between treatment effects and potential effect modifiers; when such modifiers are found to be unevenly distributed among studies, the authors suggest that NMA models need to account for these differences.

2.15 Utilizing data from non-randomized and observational studies

Ades and Sutton (187) discuss that results obtained from RCTs may not be necessarily generalizable to a wide population and that randomized studies' results could be combined with information from observational studies or patient registries, by adjusting for potential biases. Randomized and non-randomized evidence can be regarded as being complimentary, in the sense that observational studies can be considered to be reliable sources of information regarding the population baseline, while RCTs regarding intervention effects data. Dias et al. (107) describe how non-randomized studies can be used to inform a 'baseline natural history model'. Evidence from such studies can be used to estimate the absolute effect for a reference treatment. This can in turn be combined with NMA results for the relative effects of active treatments, in order to obtain an estimate of the absolute treatment effects.

Schmitz et al. (131) proposed three alternative models for jointly synthesizing information from RCTs as well as non-randomized studies: the simplest approach presented was be to perform a naïve pooling, disregarding differences in study design; the second approach was to utilize non-randomized studies as prior information, while adjusting for bias due to study design; the third was a three-level hierarchical model which accounts for bias and for heterogeneity between trial designs. The first of the models (naïve pooling) should only be used as the first step of the analysis, since it disregards potential biases in non-randomized trials. The second model (using observational evidence as prior information) allows adjusting for biases, but between-trial design heterogeneity is not taken into account, and it is not possible to include more than two different trial designs. The third model (three-level hierarchical model) addresses these issues and should be preffered.

Finally, Soares et al. (133) discussed the use of observational data for the case that there are sparse and few data in an NMA. In their approach such data were used to inform the baseline effects, but did not directly contribute to the relative treatment effects.

2.16 Planning future studies

The issue of planning future studies based on the results of an existing NMA has received little attention in the literature. Thorlund and Mills (188) and also Snapinn and Jiang (189) provided sample size considerations for determining the statistical power of indirect evidence and Mills et al. (151) performed a simulation study to estimate the power of indirect comparisons; however, there is to our knowledge currently little guidance on the design (i.e.

treatments to be compared) and the sample size needed for updating an existing NMA in an optimal manner.

Naci and O' Conor (190) alternatively suggest the design and conduct of prospective NMAs; this would go against the current practice of retrospective NMA, where each individual study is planned in isolation from others. They also suggest that the regulatory agencies should have an active role in the design of future trials, especially in the selection of comparators and in ensuring that the patient populations are comparable in terms of treatment effect modifiers.

2.17 Concluding remarks

The popularity of NMA has been increasing over the last few years; however, NMA is still a subject of controversy. Many concerns focus on the assumptions underlying the use of indirect evidence. These assumptions can be difficult to understand, hard to test, and may challenge the validity of the NMA results. Moreover, the mathematical and statistical complexity of the model and the lack of user-friendly software may deter researchers from using it. Even worse, it has been shown that a non-trivial amount of published reviews employed inappropriate methods, although the percentage has been decreasing over the years (8,191).

In our review we summarized the state-of-the-art in the field aiming to provide guidance to researchers interested in applying network meta-analytical techniques. We tried to shed light to the assumptions behind NMA and to present the statistical aspects of the model. We also discussed extensions of the basic NMA model and we summarized the currently available software options for fitting NMA.

Our review has several limitations. Pragmatic decisions needed to be made given the lack of a widely accepted terminology referring to network meta-analysis, the abundance of recently published articles and the complexity of new methods in order to ensure a timely publication of this review. Thus, there may have been articles that presented methodological advances for NMA which we failed to identify by not including in our search more online databases and by not hand-searching additional journals. We believe, however, that even if the identified set of articles might not be complete, it is representative of the currently available methods for NMA and that the most important methodological aspects, challenges and solutions of NMA are covered. Moreover, although we present some of the mathematical

features of the various models and methods, we do it in a descriptive manner and we do not provide all relevant details. Hence, this review serves as a roadmap for researchers: the keen reader should refer to the original articles for details, keeping also in mind that NMA is still an active, rapidly developing research field.

The results of the research presented in this Chapter were published in the Research Synthesis Methods journal (192) .

3 Modelling correlated binary outcomes in network meta-analysis using odds ratios

3.1 Introduction

As we discussed in Chapter 1 of this dissertation, RCTs typically report on multiple outcomes, and these outcomes may be correlated. There are two types of correlations to consider:

i. *within-study correlations* of the multiple outcome effect estimates, reflecting the fact that the same patients report on each of the analyzed outcomes
ii. *between-study correlations* of the true outcome effects across studies, reflecting the way the true outcome effects depend on each other when measured in different settings.

For the case of simple (pairwise) meta-analysis, researchers typically disregard these correlations and perform a series of independent, univariate meta-analyses for each outcome. Ignoring the correlations between outcomes, however, has been shown to lead to a loss in precision for the estimated effect sizes and an increase in bias in the presence of selective outcome reporting (178,193–195).

A multiple outcomes meta-analysis (MOMA) model can account for the correlations between treatment effects on different outcomes. In the recent years, MOMA has gained in popularity and several methodological developments have taken place (73,196–202). Two recent papers offer a comprehensive review of multivariate meta-analysis methods (78,179). A practical constraint frequently encountered in a MOMA framework is that the correlations between the effect sizes observed in the same study (within-study correlations) are rarely reported (78,179,194). Wei and Higgins (202) estimated the within-study correlations from the correlation coefficient between the outcomes, while Bujkiewicz et al.(203) used external evidence to inform correlations between dichotomous and continuous outcomes. While expert opinion could inform the unknown within-study correlation coefficients, it is not an easy task to elicit quantitative estimates for correlations from clinical experts (204,205). Focusing on dichotomous outcomes, in this chapter we suggest an alternative approach for eliciting expert opinion in a straightforward and easily understood manner.

In addition, most available MOMA models are applicable only for the case of pairwise treatment comparisons. However, as we discussed in Chapters 1 and 2 of this thesis, NMA

constantly gains in popularity and is often used to compare more than two interventions for the same outcome. It would be desirable to combine both methods (MOMA and NMA), so as to jointly synthesize data about multiple competing interventions on multiple outcomes (multiple outcomes network meta-analysis, MONMA).

There have been few attempts for a MONMA model (see also Section 2.12 of the previous Chapter). Welton et al. (145) described a MONMA model but is limited to the case of two armed studies. However, the majority of applications of NMA include at least one multi-armed study (8). Schmid et al. (129) proposed a MONMA model for analysing unordered categorical outcomes. This model also allows the inclusion of studies with partially observed data. However, it cannot be applied for the more frequent cases of meta-analyses of binary or continuous outcomes. Hong et al. (114) presented a model for multiple outcomes which, however, does not take into account within-study correlations. Madan et al. presented an approach for modeling multiple outcomes reported over multiple follow-up times; their models are applicable only for repeated measurements (120).

The primary aim of the research we present in this Chapter is to develop a model for synthesizing multiple dichotomous outcomes over a network of studies. In Section 3.2 we describe a clinical example from acute mania (206) which we use to illustrate our methods. In Section 3.3 we present a method for estimating the within-study correlation coefficients by utilizing a set of conditional probabilities. We show how these probabilities can be elicited from clinical experts through easily understood questions. We then present a new MONMA model. In Section 3.4 we discuss the application of our methods to the network of treatments for acute mania. In Section 3.5 we summarize our findings.

3.2 Example: the acute mania dataset

The dataset includes a network of 65 randomized controlled trials comparing 14 active antimanic drugs and placebo for acute mania, Cipriani et al. (206). Most of the studies have two arms (47 studies) and there are 18 three-arm studies.

The primary outcomes of interest were efficacy and treatment discontinuation (acceptability, or “dropout”) after 3 weeks. Acceptability was estimated as the number of patients leaving the study early for any reason, before or after having a response to the treatment, out of the total number of randomized patients. All-cause discontinuation from allocated treatment may be due to a number of reasons, such as: adverse effects, inefficacy, other reasons not related to treatment (e.g. moving away, protocol violation), or a

combination of the above. Efficacy was reported either as dichotomous outcome (number of patients who responded to treatment, defining response as a reduction of at least 50% in manic symptoms from baseline to week 3) or as continuous outcome (mean change scores on a standardized rating scale for mania after 3 weeks). Although we recognize that outcome dichotomization may lose some information, we used data on efficacy as a dichotomous outcome as it may be easier to interpret clinically and allows us to illustrate our methodology for two related binary outcomes, a frequent scenario encountered by researchers. Only a few patients did not provide data for response to treatment and their outcome was coded as treatment failure; an imputation assumption that has been shown to be sensible when the missing rates are low (134). Among the included studies, only 65 contributed with data for at least one of the outcomes of interest: 18 studies (28%) did not report usable data on response, while only one study did not report information on the number of dropouts (1.5%). Efficacy and acceptability outcomes are generally expected to be negatively correlated; although early full response to the treatment may be a cause for leaving the study prematurely, more often it is reasonable to assume that more efficacious treatments are associated with a lower dropout rate. Within-study correlations were not reported in any of the studies and individual patient data (*IPD*) which could be used to estimate within-study correlations were not available. The dataset included a total of 69 head-to-head comparisons for response and 100 for dropout. In Section I of the Supplementary Material we provide a table with all head to head comparisons for each outcome, along with the odds ratios and their 95% confidence interval. The initial analysis consisted of two independent network meta-analyses, one for each outcome (206). As both outcomes are crucially important for clinical decision making, the ranking of the treatments was presented for both efficacy and acceptability in a two-dimensional scatter plot (Figure 6 in Cipriani et al.) so that efficacious treatments with high tolerability could be identified. This is a suboptimal approach and the rankings of the treatments for each outcome can be better estimated jointly in a MONMA model to account for the correlation in the outcomes. This is especially important here as 19 studies provide data on only one of the two outcomes, and MONMA can 'borrow strength' from these studies even for the missing outcomes.

In Figure 3 and Figure 4 we present the network of evidence for response and dropout.

Figure 3: Network of pharmacological treatments for acute mania, for the response outcome. Nodes and edges are weighted according to the number of studies involved in each treatment or comparison respectively.

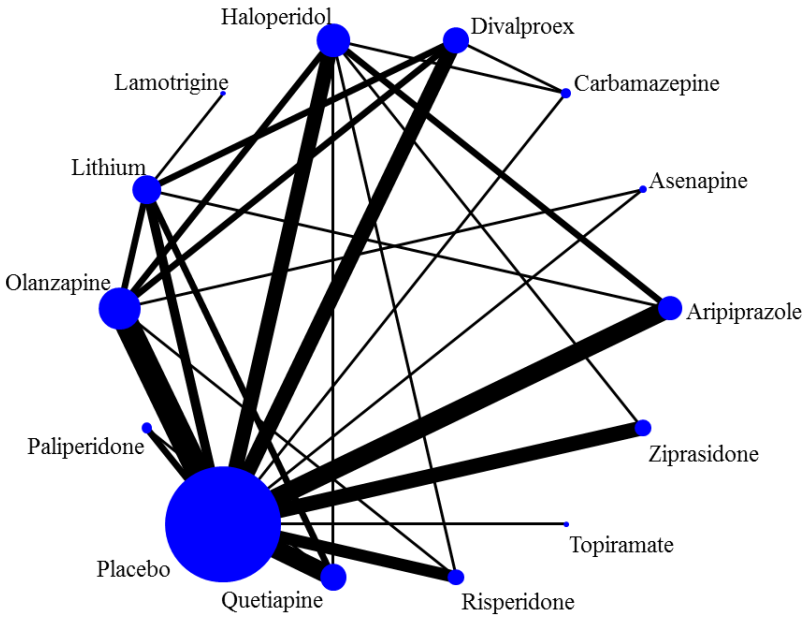
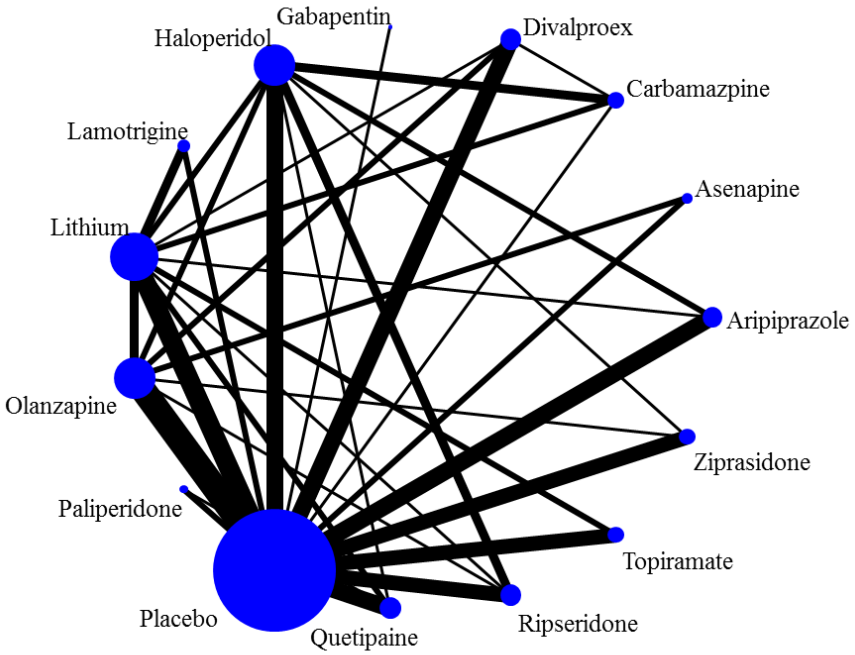


Figure 4: Network of pharmacological treatments for acute mania, for dropout.



3.3 Statistical methods

Here we start by revising and extending a MOMA model when only two treatments are compared. Emphasis is placed on estimating the within-study correlation coefficients. Subsequently we generalize the approach to a network meta-analysis with multi-arm studies.

3.3.1 Pairwise meta-analysis models for multiple outcomes

Suppose we have a total of N_S studies comparing two treatments with respect to two different, correlated outcomes, denoted by R and D. These two outcomes are identified as the response to the treatment (R) and dropout rate (D) in the acute mania example. Note that some studies may not report on both outcomes. We denote the observed treatment effects in study i for outcomes R and D with $y_{i,R}$ and $y_{i,D}$ respectively. A bivariate random effects meta-analysis model can be written as follows:

$$\begin{pmatrix} y_{1,R} \\ y_{1,D} \\ y_{2,R} \\ \vdots \\ y_{N_S,D} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} \mu_R \\ \mu_D \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,R} \\ \varepsilon_{1,D} \\ \varepsilon_{2,R} \\ \vdots \\ \varepsilon_{N_S,D} \end{pmatrix} + \begin{pmatrix} \delta_{1,R} \\ \delta_{1,D} \\ \delta_{2,R} \\ \vdots \\ \delta_{N_S,D} \end{pmatrix} \tag{6}$$

Equation (6) can be compactly written as $\boldsymbol{Y} = \boldsymbol{X\mu} + \boldsymbol{\varepsilon} + \boldsymbol{\delta}$, where $\boldsymbol{Y}$ is the ($2N_S$-dimensional) vector of the observed effects, $\boldsymbol{\mu}$ is the vector of the true relative effects for each outcome, $\boldsymbol{\mu} = (\mu_R, \mu_D)'$, $\boldsymbol{X}$ is the $(N_S \times 2)$ 'design matrix', $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are the vectors of random errors and random effects respectively. We assume multivariate normal distributions for $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$, so that $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma})$ and $\boldsymbol{\delta} \sim N(0, \boldsymbol{\Delta})$, with $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ denoting the within and between-study variance-covariance matrices. Note that letters in bold denote vectors and matrices.

The random errors within a study and the random effects across studies are in principle correlated, and this correlation is incorporated in $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ respectively. More specifically, the variance-covariance matrix for the random effects takes a block-diagonal form:

$$\boldsymbol{\Delta} = \begin{pmatrix} \tau_R^2 & \rho_\tau \tau_R \tau_D & 0 & 0 & \cdots \\ \rho_\tau \tau_R \tau_D & \tau_D^2 & 0 & 0 & \cdots \\ 0 & 0 & \tau_R^2 & \rho_\tau \tau_R \tau_D & \cdots \\ 0 & 0 & \rho_\tau \tau_R \tau_D & \tau_D^2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Delta}_{(2\times 2)} & 0 & \cdots \\ 0 & \boldsymbol{\Delta}_{(\mathbf{2\times 2})} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{7}$$

The above $2N_S \times 2N_S$ matrix involves the heterogeneity standard deviations for each

outcome, τ_R and τ_D, and the between-study correlation coefficient, ρ_τ. Note that this between-study variance-covariance matrix is block-diagonal with identical $\boldsymbol{\Delta}_{(2\times2)}$ matrices in its diagonal. The parameters ρ_τ, τ_R and τ_D need to be estimated from the data: this can be done either within a frequentist setting. using approaches like maximum likelihood, restricted maximum likelihood method and the generalized method of moments (72–74,198,207), or in a Bayesian setting using Markov Chain Monte Carlo. Similarly, the within-study variance-covariance matrix is also block diagonal:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{1,R}^2 & \rho_1\sigma_{1,R}\sigma_{1.D} & 0 & 0 & \cdots \\ \rho_1\sigma_{1,R}\sigma_{1.D} & \sigma_{1,D}^2 & 0 & 0 & \cdots \\ 0 & 0 & \sigma_{2,R}^2 & \rho_2\sigma_{2,R}\sigma_{2.D} & \cdots \\ 0 & 0 & \rho_2\sigma_{2,R}\sigma_{2.D} & \sigma_{2,D}^2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma_1} & 0 & \cdots \\ 0 & \boldsymbol{\Sigma_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (8)$$

In this matrix ρ_i is the within-study correlation coefficient and ${\sigma_{i,R}}^2$, ${\sigma_{i,D}}^2$ are the variances of the effect sizes in every study i. All entries in $\boldsymbol{\Sigma}$ can be estimated from the data. Sample estimates for the ${\sigma_{i,R}}^2$, ${\sigma_{i,D}}^2$ are often available, but few studies, if any, provide enough information to estimate the within-study correlation coefficient ρ_i. In the absence of sample estimates for ρ_i, a range of plausible values can be used in a sensitivity analysis, or one could try to elicit prior distributions for the correlation coefficient from clinical experts. However, obtaining a prior for the correlation coefficient is not straightforward. In the following sections we discuss how partial information reported in studies can be combined with external information to obtain estimates of ρ_i and incorporate them in the MOMA model.

3.3.2 Estimation of within-study correlation coefficient for two dichotomous outcomes

Studies that report on two or more dichotomous outcomes typically provide the number of successes and failures for every outcome in each arm. For two outcomes the data can be summarized in two independent 2×2 tables which we refer to as 'collapsed' tables. We refer to a 'full cross' table as the table that gives information about the cross-classification of the patients in both outcomes. Let us consider for example a study reporting on response and dropout: a full cross table provides information on the number of successes and failures among those who drop out as well as the number of successes and failures among those who

do not drop out for each arm. In a recent paper Bagos (208) showed how to compute the covariance of two correlated log odds ratios (logOR) when the full cross tables are available. Consequently, if all studies in a meta-analysis provide the full cross tables, the within variance-covariance matrix of Equation (8) can be estimated and a multivariate analysis can be readily performed. However, outcomes are routinely analyzed separately and only the row and column margins of the full cross tables are usually provided in the studies.

In this section we show how to reconstruct the full cross tables for every study given the usual 2×2 collapsed tables and external evidence. Having reconstructed the full cross tables we can then use the methods described in (208) and compute the correlation coefficient needed for the multivariate analysis.

3.3.2.1 Reconstruction of the full cross table and estimation of the correlation coefficient

Consider a study i comparing two treatments A and B for response (R) and dropout (D). The data are $e_{i,T,R}$, $f_{i,T,R}$, $e_{i,T,D}$, $f_{i,T,D}$ for treatments $T = A, B$, where $e_{i,T,R}$ denotes the number of patients that responded positively (R^+) to treatment T and $f_{i,T,R}$ the ones that did not (R^-); likewise $e_{i,T,D}$ denotes the patients randomized in group T that dropped out of the study early (D^+). Similarly, $f_{i,T,D}$ denotes those who did not drop out (D^-).

Let us denote by $\varphi_{i,T} = P(D^+|R^+)_{i,T}$ the probability that a patient who responded to the treatment would drop out; also let $\zeta_{i,T} = P(D^+|R^-)_{i,T}$ denote the probability of a non-responder to drop out. Table 3 shows how to compute the elements of the full cross table from the elements of the collapsed table ($e_{i,T,R}$, $f_{i,T,R}$, $e_{i,T,D}$, $f_{i,T,D}$) for every treatment. For example $\hat{\varphi}_{i,A} e_{i,A,R}$ patients received treatment A, had a positive response to the treatment but dropped out of the study and $(1 - \hat{\varphi}_{i,B}) e_{i,B,R}$ who received treatment B had a positive response and stayed in the study.

Table 3: Reconstructing the full cross table from the collapsed table for a study i comparing treatments A and B for response (R) and dropout (D). $\hat{\varphi}_{i,T}$ denotes the proportion of dropouts among responders and $\hat{\zeta}_{i,T}$ the proportion of dropouts among non-responders. R^+ (R^-) denotes a positive (negative) response to the treatment while D^+(D^-) denotes dropping out of (staying in) the study.

Treatment A	$\boldsymbol{R^+}$	$\boldsymbol{R^-}$	**TOTAL**
$\boldsymbol{D^+}$	$\hat{\varphi}_{i,A}e_{i,A,R}$	$\hat{\zeta}_{i,A}f_{i,A,R}$	$e_{i,A,D}$
$\boldsymbol{D^-}$	$(1-\hat{\varphi}_{i,A})e_{i,A,R}$	$(1-\hat{\zeta}_{i,A})f_{i,A,R}$	$f_{i,A,D}$
TOTAL	$e_{i,A,R}$	$f_{i,A,R}$	$N_{i,A}$
Treatment B	$\boldsymbol{R^+}$	$\boldsymbol{R^-}$	**TOTAL**
$\boldsymbol{D^+}$	$\hat{\varphi}_{i,B}e_{i,B,R}$	$\hat{\zeta}_{i,B}f_{i,B,R}$	$e_{i,B,D}$
$\boldsymbol{D^-}$	$(1-\hat{\varphi}_{i,B})e_{i,B,R}$	$(1-\hat{\zeta}_{i,B})f_{i,B,R}$	$f_{i,B,D}$
TOTAL	$e_{i,B,R}$	$f_{i,B,R}$	$N_{i,B}$

It can be shown that $\hat{\zeta}_{i,T}$ is dependent on $\hat{\varphi}_{i,T}$ given the marginal counts:

$$\hat{\zeta}_{i,T} = \frac{1}{f_{i,T,R}}(e_{i,T,D} - e_{i,T,R}\hat{\varphi}_{i,T}) \tag{9}$$

Thus, information about only one of $\hat{\varphi}$, $\hat{\zeta}$ is needed for every arm in order to reconstruct the table. Note that Equation (9) holds when the total sample size is the same for the two outcomes, that is $e_{i,T,R} + f_{i,T,R} = e_{i,T,D} + f_{i,T,D}$.

Having reconstructed the full cross tables, the correlation coefficient between the two log-odd ratios $y_{i,R}$ and $y_{i,D}$ can be estimated using the formula produced by Bagos (208), which after some algebra can be shown to be equal to:

$$\hat{\rho}_i = \frac{1}{\hat{\sigma}_{i,R}\hat{\sigma}_{i,D}} \sum_{T=A,B} \frac{\hat{\varphi}_{i,T}\left(e_{i,T,R} + f_{i,T,R}\right)^2 - e_{i,T,D}(e_{i,T,R} + f_{i,T,R})}{e_{i,T,D}f_{i,T,R}f_{i,T,D}} \tag{10}$$

Equation (10) allows us to estimate the correlation coefficient between log odds ratios of the different outcomes in study i, given the data typically reported ($e_{i,T,R}$, $f_{i,T,R}$, $e_{i,T,D}$, $f_{i,T,D}$) and $\hat{\varphi}_{i,T}$ for every treatment (assumed known), under the restriction that both outcomes were reported for every patient within the same study.

Table 4: Data from a two-arm study comparing Aripiprazole to placebo for response (R) and drop-out (D) (209). $e_{T,R}$, $e_{T,D}$ denote the number of patients who were positive in outcomes R and D, while $f_{T,R}$, $f_{T,D}$ denote the number of patients negative in outcomes R and D respectively.

Treatment (T)	$e_{T,R}$	$f_{T,R}$	$e_{T,D}$	$f_{T,D}$
Aripiprazole	155	98	54	199
Placebo	63	68	20	111

Equation (10) suggests that if the proportion of dropouts in the responders equals the proportion of dropouts in the total number of patients, i.e. $\hat{\varphi}_{i,T} = e_{i,T,D}/(e_{i,T,R} + f_{i,T,R})$, then the two outcomes are independent and the correlation coefficient in (10) becomes zero. If both $\hat{\varphi}_{i,T}$ are equal to zero, which suggests that all responders stayed in the study, we get $\rho_i = -1$. In the contrary, if both $\hat{\varphi}_{i,T}$ are equal to one (all responders dropped out) Equation (10) gives $\rho_i = 1$.

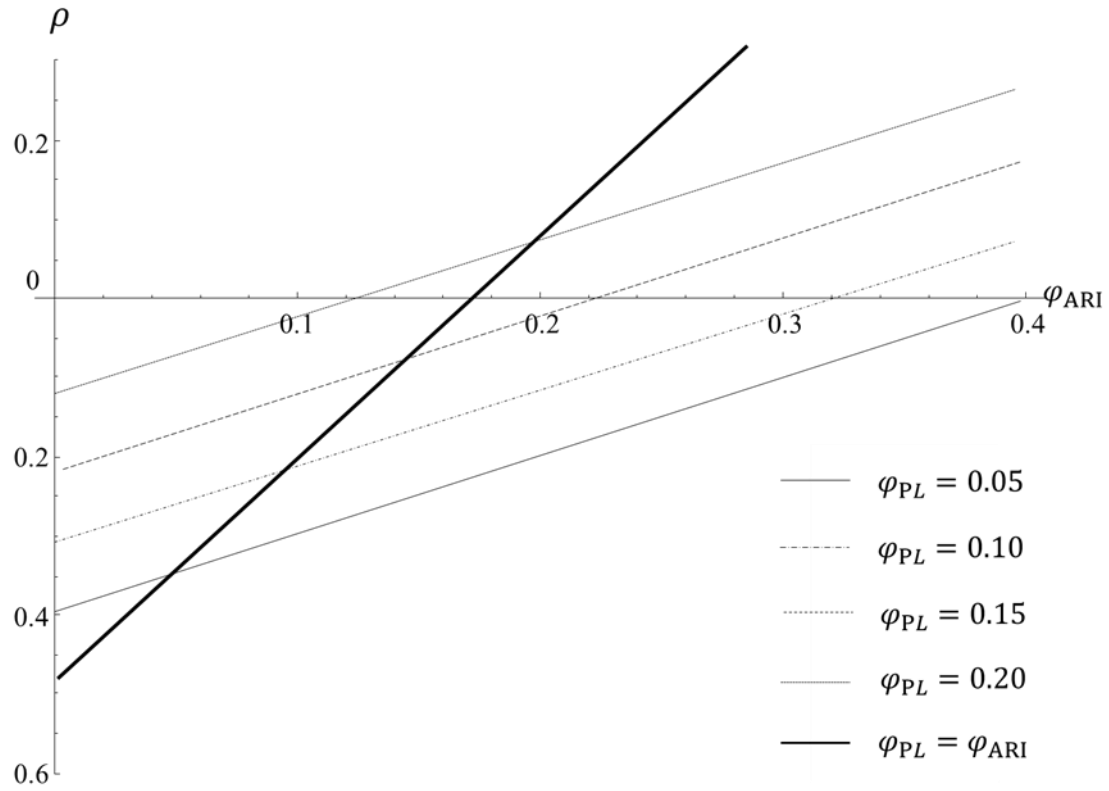
Note here that $\hat{\varphi}_{i,T}$ in each study can only take values that ensure $0 \leq \hat{\zeta}_{i,T} \leq 1$, that is:

$$\frac{e_{i,T,D} - f_{i,T,R}}{e_{i,T,R}} \leq \hat{\varphi}_{i,T} \leq \frac{e_{i,T,D}}{e_{i,T,R}} \tag{11}$$

Of course, $0 \leq \hat{\varphi}_{i,T} \leq 1$ must also hold. In Table 4 we present data from a two-arm study comparing Aripiprazole with Placebo for both response and drop-out rate (Vieta *et al.* (209)). We have dropped study index i since we refer to a single study.

Equation (11) implies that $0 \leq \hat{\varphi}_{ARI} \leq 0.34$ and $0 \leq \hat{\varphi}_{PL} \leq 0.31$, with $\hat{\varphi}_{ARI}$ and $\hat{\varphi}_{PL}$ denoting the proportion of dropouts among responders for the Aripiprazole and the placebo arm respectively. This means that less than 34% of those who responded positively to Aripiprazole could have dropped out. Using Equation (10) we plot ρ for various values of φ_{ARI} and φ_{PL} in Figure 5, in order to explore how the correlation coefficient depends on these proportions. We assume four different values for φ_{PL}, and plot the correlation coefficient versus φ_{ARI}. We also plot the corresponding ρ if the two proportions are assumed to be equal. Figure 5 suggests that the correlation coefficient for the logOR for response and

Figure 5: Within-study correlation coefficient ρ between the log odds ratios for response and dropout versus the φ_{ARI} for a study (209) comparing Aripiprazole vs Placebo. Four different values for φ_{PL} are presented and the line $\varphi_{ARI} = \varphi_{PL}$



drop-out rate remains negative for small values of both φ parameters, something which is readily understood: the smaller the proportion of responders dropping out, the more negatively correlated are the log odds ratios for response and drop-out rate.

In a recent paper Wei and Higgins (210) have also produced a formula for estimating the covariance between the logORs of two correlated outcomes. Their formula requires the correlation coefficient between the two dichotomous outcomes to be known. The motivation for their approach was that the value for the correlation between the outcomes is more likely to be available (or easier to guess) than the correlation between the treatment effects. It can be shown that our formula in Equation (10) is equivalent to the formula (8) derived by Wei and Higgins in (210), under the assumption that all patients report on all outcomes (the mathematical details can be found in Section II of the Appendix). However, we think it is more useful to express the correlation coefficient (or the covariance) of the log odds ratios

in terms of the parameters $\hat{\varphi}_{i,T}$. This is because, as we will show in the following section, these parameters are flexible in modeling and prior distributions can be obtained from clinical experts through easily understood questions.

3.3.2.2 Modeling the φ parameters, eliciting priors and synthesizing prior distributions

Assuming that the proportions $\hat{\varphi}_{i,T}$ are not reported in the studies we can use expert opinion to inform the 'true' conditional probabilities $\varphi_{i,T}$ they estimate. Then we can compute the correlations of the logORs using Equation (10). The parameters $\varphi_{i,T}$ can be assumed to be:

- study and treatment-specific $\varphi_{i,T}$
- fixed $\varphi_{i,T} = \varphi$
- treatment-specific $\varphi_{i,T} = \varphi_T$
- study-specific $\varphi_{i,T} = \varphi_i$
- or we can assume group-specific probabilities $\varphi_{i,T} = \varphi_{Group(T)}$, by identifying groups of treatments that share some common characteristic. For example we may assume that there are two parameters; one common for all active treatments and one for placebo.

Investigators could choose between these options after considering the nature of the clinical condition under investigation, the types of interventions and the outcomes of interest. The decision about the number of different φ parameters and their plausible values should be specified after consulting with clinicians experienced in randomized controlled trials in the field.

Having assigned a value to every $\varphi_{i,T}$, the full cross table for each study can be reconstructed, the within-study correlation coefficient can be estimated and the full multivariate analysis can be performed. Alternatively, we can treat φ as a random variable and elicit information about its distribution. Then the reconstruction of the full cross tables is carried out stochastically. For the acute mania example, we use the following question to elicit information about φ_T, which is assumed to be treatment-specific:

"If a number of people randomized to treatment T responded to the treatment, what proportion of them do you expect to leave the study early? Please provide a 95% confidence interval for this proportion."

Once external evidence is collected from experts, we need to combine their input into a single distribution of φ_T. To this end we will use an approach described in (204) which attributes a different weight to each expert's input.

Suppose we have a number N_E of clinical experts. The k-th expert ($k = 1,..,N_E$) provides an estimate for the 95% confidence interval of φ_T. Assuming the expert's opinion to be a beta distribution we can construct a prior $\pi_k(\varphi_T) = Beta(\varphi_T; a_k, b_k)$ from the provided confidence interval.

An overall prior can then be obtained as a combination of the individual expert opinions:

$$f(\varphi_T) \propto \left(\pi_1(\varphi_T)\right)^{w_1}\left(\pi_2(\varphi_T)\right)^{w_2}\cdots\left(\pi_{N_E}(\varphi_T)\right)^{w_{N_E}} = \prod_k \left(\pi_k(\varphi_T)\right)^{w_k} \quad (12)$$

The w_k parameters are weights ($\sum_{k=1}^{N_E} w_k = 1$) assigned to experts, and reflect the credibility attached to their opinions. In the acute mania example we define the weights based on the years of relevant clinical experience of each expert and the number of clinical trials he/she has been involved with.

Equation (12) suggests that the prior distribution for φ_T is a beta distribution:

$$\varphi_T \sim Beta\left(\sum w_k a_k\,, \sum w_k \beta_k\right) \quad (13)$$

If the parameter φ is believed to be trial-specific, experts should also be given information about relevant study characteristics (such as trial duration) and then synthesis of their opinions could be done as described above. Note that the prior distribution for φ might need to be truncated within each study to account for the plausible range of values as explained in Equation (11). In the infrequent case that the prior distribution for a φ parameter provided by the experts lies outside the permissible range of values for a study as given in Equation (11) a uniform uninformative distribution in the allowed values can be employed. Obtaining priors outside the permitted area could be prevented if the experts were provided with the range of permitted values for φ; however we do not recommend this as prior elicitation should not include any consideration of the data.

Instead of eliciting expert opinion one could assume both study- and treatment-specific φ parameters and employ vague priors for each, such as with a uniform distribution in the allowed range given by Equation (11). This would substantially increase the uncertainty about the parameters (as neither data nor informative priors would inform the correlation coefficients) but could be considered as a sensitivity analysis to complement the analysis with informative priors.

3.3.3 Network meta-analysis for two correlated outcomes

In the previous sections we presented a method for performing a pairwise meta-analysis for two outcomes. We now extend the method to network meta-analysis. We restrict our analysis to networks that contain only two-arm and three-arm studies and a maximum of two different dichotomous outcomes. We allow for random effects and we assume consistency in the network, i.e. there is no discrepancy between direct and indirect evidence (24).

Consider a network of studies reporting on outcomes R and D for a number N_T of different treatments. Assuming consistency, we need to estimate $N_T - 1$ independent (*basic*) parameters for every outcome. The model is a generalization of the simple meta-analysis model of Section 3.3.1, $\boldsymbol{Y} = \boldsymbol{X\mu} + \boldsymbol{\varepsilon} + \boldsymbol{\delta}$, with $\boldsymbol{Y}$ the vector of the observed log odd ratios, $\boldsymbol{X}$ the design matrix, $\boldsymbol{\mu}$ the vector of the basic parameters, $\boldsymbol{\varepsilon}$ the vector of random errors, and $\boldsymbol{\delta}$ the vector of random effects (6,46). The design matrix $\boldsymbol{X}$ describes the structure of the network, and the consistency assumption is embedded within it.

For a two-arm study i that compares treatments A and B the random errors are assumed to follow a multivariate normal distribution, i.e. $(\delta_{i,AB,R}\ ,\ \delta_{i,AB,D})' \sim N(0, \boldsymbol{\Delta}_{(2\times2)})$. In network meta-analysis it is often assumed that the amount of heterogeneity is independent of the treatment comparison; that is, for any two random treatments X and Y it is $\tau^2_{XY,R} = \tau^2_R$ and $\tau^2_{XY,D} = \tau^2_D$ (4,6). Under this assumption, the variance-covariance matrix of a two-arm study is exactly as in the case of a pairwise meta-analysis.

For a three-arm study i that compares treatments A, B and C, the random effects are assumed to follow a multivariate normal distribution: $(\delta_{i,AB,R}\ ,\delta_{i,AB,D}\ ,\ \delta_{i,AC,R}\ ,\delta_{i,AC,D})' \sim N(0, \boldsymbol{\Delta}_{(4\times4)})$. The assumption of consistency on the random effects (e.g. $\delta_{i,AB,R} = \delta_{i,AC,R} + \delta_{i,CB,R}$) and the equal heterogeneity parameters across comparisons suggest that the covariance between logOR of different comparisons for response is $\tau^2_R/2$ and for dropout is $\tau^2_D/2$. Consequently the (4×4) variance-covariance matrix is:

$$\boldsymbol{\Delta}_{(4\times4)} = \begin{pmatrix} \tau^2_R & \rho_\tau\tau_R\tau_D & \tau^2_R/2 & \chi_1\tau_R\tau_D \\ \rho_\tau\tau_R\tau_D & \tau^2_D & \chi_2\tau_R\tau_D & \tau^2_D/2 \\ \tau^2_R/2 & \chi_2\tau_R\tau_D & \tau^2_R & \rho_\tau\tau_R\tau_D \\ \chi_1\tau_R\tau_D & \tau^2_D/2 & \rho_\tau\tau_R\tau_D & \tau^2_D \end{pmatrix} \quad (14)$$

We further assume that the correlations $corr(\delta_{i,AB,R}, \delta_{i,AC,D}) = \chi_1$ and $corr(\delta_{i,AB,D}, \delta_{i,AC,R}) = \chi_2$ between logORs of different comparisons-different outcomes are all equal to χ. This is a plausible assumption to make if treatments B and C are comparable in terms of both efficacy and acceptability. For example, in a three-arm study with two active treatments and placebo, the assumption will be a plausible one as long as we identify placebo as treatment A. Although this assumption might be difficult to defend in practice, it will often be necessary to reduce the number of parameters to be estimated. In Section III of the Appendix we show that this assumption simplifies the $\Delta_{(4\times4)}$ matrix to:

$$\boldsymbol{\Delta}_{(4\times4)} = \tau_R^2 \begin{pmatrix} 1 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \tau_D^2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1 \end{pmatrix} + \rho_\tau \tau_R \tau_D \begin{pmatrix} 0 & 1 & 0 & 1/2 \\ 1 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1 \\ 1/2 & 0 & 1 & 0 \end{pmatrix} \quad (15)$$

Under these three assumptions (consistency, heterogeneities equal across comparisons and equal correlations between effects of different comparisons and different outcomes) there are only three between-study parameters to estimate: the heterogeneity for response (τ_R^2) and dropout (τ_R^2) and the between-study correlation coefficient (ρ_τ) just like in the case of pairwise comparison, Equation (7).

When a considerable amount of data is available and the network is very dense (that is many studies connecting pairs of interventions) then the assumptions we used to reduce the number of model parameters in $\boldsymbol{\Delta}$ might not be necessary, e.g. if there are at least three studies per comparison, then different heterogeneity variances can be used. However, real-life networks of interventions tend to be poorly connected and the median number of studies per comparison has been found to be low, equal to two studies (8). In Section III of the Appendix we present how $\boldsymbol{\Delta}$ is modelled when correlations between different treatments and different outcomes are not equal. Note that the variance-covariance matrix as defined above is always positive-definite.

For a three-arm study i that compares A, B and C the random errors are assumed to be distributed as $(\varepsilon_{i,AB,R}, \varepsilon_{i,AB,D}, \varepsilon_{i,AC,R}, \varepsilon_{i,AC,D})' \sim N(0, \boldsymbol{\Sigma}_i)$, with the variance-covariance matrix $\boldsymbol{\Sigma}_i$:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{i,AB,R} & . & . & . \\ \rho_{i,AB_RAB_D}\sigma_{i.AB,R}\sigma_{i.AB,D} & \sigma^2_{i,AB,D} & . & . \\ \kappa_{i,AB_RAC_R} & \rho_{i,AB_DAC_R}\sigma_{i.AB,D}\sigma_{i.AC,D} & \sigma^2_{i,AC,R} & . \\ \rho_{i,AB_RAC_D}\sigma_{i.AB,R}\sigma_{i.AC,D} & \kappa_{i,AB_RAC_R} & \rho_{i,AC_RAC_D}\sigma_{i.AC,R}\sigma_{i.AC,D} & \sigma^2_{i,AC,D} \end{pmatrix} \quad (16)$$

There are four different correlation coefficients entering this study-specific variance-covariance matrix. Two of them, ρ_{i,AB_RAB_D} and ρ_{i,AC_RAC_D} correlate logORs of the same treatment comparisons for different outcomes, while the other two, ρ_{i,AB_RAC_D} and ρ_{i,AC_RAB_D}, correlate different comparisons for different outcomes. The quantities σ and κ in $\boldsymbol{\Sigma}_i$ can be readily estimated from the data, e.g. the variance for the AB comparison for response (R) can be estimated as $\hat{\sigma}^2{}_{i,AB,R} = \frac{1}{e_{i,A,R}} + \frac{1}{f_{i,A,R}} + \frac{1}{e_{i,B,R}} + \frac{1}{f_{i,B,R}}$ and also $\hat{\kappa}_{i,AB_RAC_R} = \frac{1}{e_{i,A,R}} + \frac{1}{f_{i,A,R}}$. The data needed to compute these two quantities are typically available from the published articles while the four correlation coefficients can be estimated from the collapsed tables and using external evidence about the φ parameters as in Section 3.3.2. More specifically Equation (10) can be employed to estimate coefficient ρ_{i,AB_RAB_D} of Equation (16) as:

$$\hat{\rho}_{i,AB_RAB_D} = \frac{1}{\hat{\sigma}_{i,AB,R}\hat{\sigma}_{i,AB,D}} \sum_{T=A,B} \frac{\hat{\varphi}_{i,T}\left(e_{i,T,R} + f_{i,T,R}\right)^2 - e_{i,T,D}(e_{i,T,R} + f_{i,T,R})}{e_{i,T,D}f_{i,T,R}f_{i,T,D}} \quad (17)$$

and an analogous formula can be used to estimate ρ_{i,AC_RAC_D}. We show in Appendix, section IV that:

$$\hat{\rho}_{i,AB_RAC_D} = \frac{1}{\hat{\sigma}_{i,AB,R}\hat{\sigma}_{i,AC,D}} \frac{\hat{\varphi}_{i,A}\left(e_{i,A,R} + f_{i,A,R}\right)^2 - e_{i,A,D}(e_{i,A,R} + f_{i,A,R})}{e_{i,A,D}f_{i,A,R}f_{i,A,D}} \quad (18)$$

A similar formula holds for ρ_{i,AC_RAB_D}. Using Equations (17) and (18) we can use prior information on $\varphi_{i,T}$ to estimate all correlation coefficients in a three-arm study and perform a full MOMA. As we have already seen in Section 3.3.2.1the values of these parameters are bounded for every study according to Equation (11), which means that the values for the correlation coefficients of Equations (17) and (18) are bounded as well.

3.4 Application to acute mania dataset: network meta-analysis for response and dropout

3.4.1 Prior distributions and model fit

In Figure 6 we present the informative prior distributions we used for the conditional probabilities φ_T of dropping out given a positive response to the treatment T. These distributions were elicited from experts in the field following the method presented in Section 3.3.2.2. Experts were not provided access to the actual data with an aim to get unprejudiced results. In Section V of Appendix we provide details about the individual prior distribution elicited from each expert. Then, by using Equation (10) for the two-arm studies, and Equations (17) and (18) for the three-arm studies, we computed all within-study correlations. In Section VI of the Appendix we present a table with the estimated correlation coefficients for all two arm-studies.

After inspecting the data we could divide the treatments in categories according to their efficacy and dropout, and assume a common φ in each category. We chose, however, to assume a different φ for every treatment in order to present the most general case.

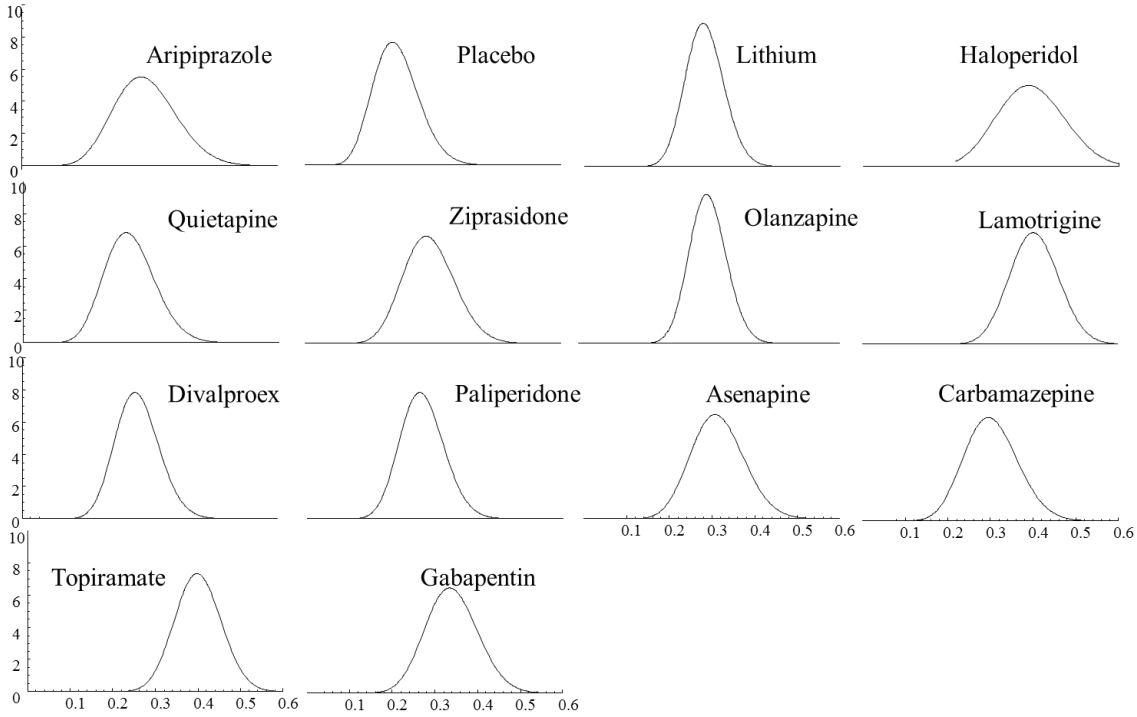
We used OpenBUGS software (94,95) to fit our model. When studies did not report on one of the outcomes we imputed data with very large variances and zero within-study covariances (179).

The heterogeneity standard deviations τ_R and τ_D were assigned a minimally informative prior distribution (211,212), $\tau_R, \tau_D \sim U(0,1)$. For the between-study correlation coefficient ρ_τ, an uninformative prior distribution $Unif(-1,1)$ can be used if there is no information about the correlation between the outcomes. In our example expert opinion suggested that the outcomes are expected to be negatively correlated, therefore a $Unif(-1,0)$ was chosen.

In order to assess the relative ranking of the treatments, we computed the surface under the cumulative ranking curve (SUCRA) for each treatment and outcome (38).

The R routine needed to estimate beta priors based on expert opinion can be found in Section V of the Appendix. The OpenBUGS code used to fit the model can be found in Section XI of the Appendix. All results pertain to 1,000,000 cycles and a thinning of 100 after a 5,000 burn-in period.

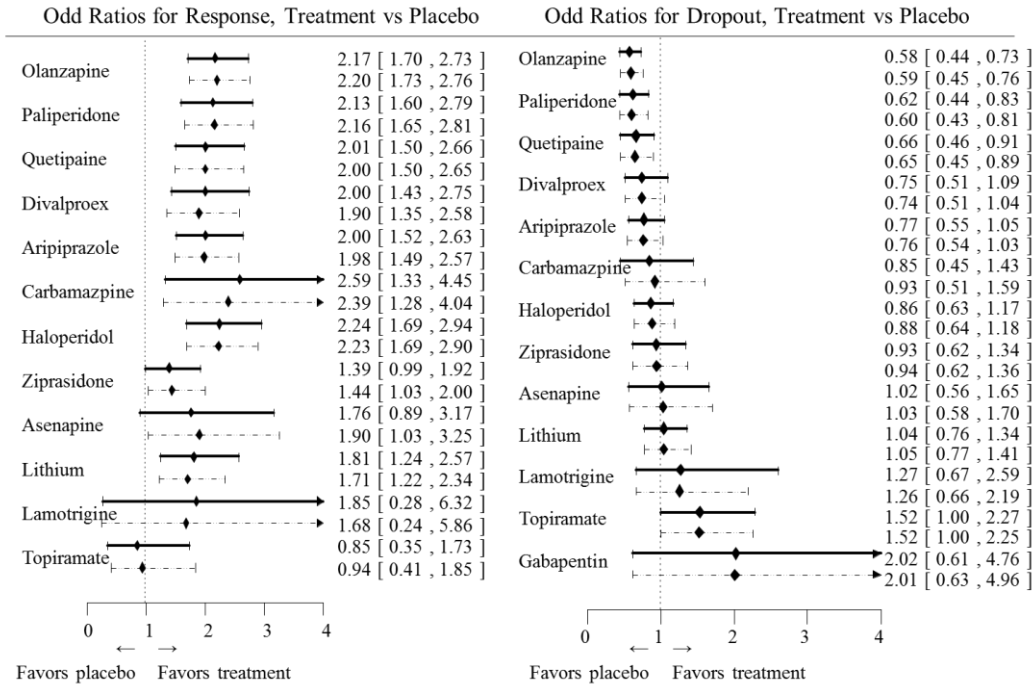
Figure 6: Prior distributions for the treatment-specific probability φ, as elicited from the experts.



3.4.2 Results

Figure 7 shows the odds ratios (ORs) for response and dropout for all active versus placebo comparisons as estimated in two independent NMAs and the MONMA model. The results from NMA and MONMA were comparable for the dropout outcome. This happened because all but one studies reported dropout and, consequently, the joint analysis of both outcomes did not have much impact on the dropout estimates. On the other hand, the ORs for response were estimated with higher precision due to the fact that 28% of the studies did not report on response. The relative decrease in the width of the OR confidence intervals with the MONMA model was 4% on average. The maximum relative decrease in the width of the confidence intervals was 15% and was observed for the case of Lithium. This can be attributed to the fact that more than half of the studies comparing Lithium (8 out of 15) did not report on efficacy. All results should be interpreted in the light of the high between-studies correlation coefficient, which was estimated to be -0.84 (credible interval -0.98 to -0.52).

Figure 7: Odds ratios for response and dropout for Treatment vs. Placebo. The thick lines present results from two independent NMA models (one for each outcome) and the thin lines from the MONMA model.



In Table 5 we present the point estimates and the 95% credible intervals for the heterogeneities and the between-study correlation parameter. Figure 8 shows the relative ranking of the treatments for both outcomes, based on their surface under the cumulative ranking curve ("SUCRA", (38)). For treatment A, outcome R, SUCRA is defined as $\sum_{k=1}^{N_T-1} cum_k^{A,R} /(N_T - 1)$, with $cum_k^{A,R}$ denoting the probability of A ranking among the best k treatments for outcome R. SUCRA values lie between 0 (when the treatment is certain to be the worst for the outcome) and 1 (when the treatment is certain to be the best for the outcome). It is a transformation of the mean rank which takes uncertainty of estimation into account. Treatments lying at the upper right corner of Figure 8 are the best in both acceptability and efficacy; those in the bottom left corner are the 'worst' treatments. The small changes in the point estimates and the precision of the ORs for response had also an effect on the relative ranking of the treatments. For instance, Carbamazepine ranked as the most efficacious treatment with the usual NMA model, while it fell to the third place for the MONMA model. The change in the OR, however, was not clinically important.

Table 5: Model parameters for the outcome-specific NMA model, and the MONMA model

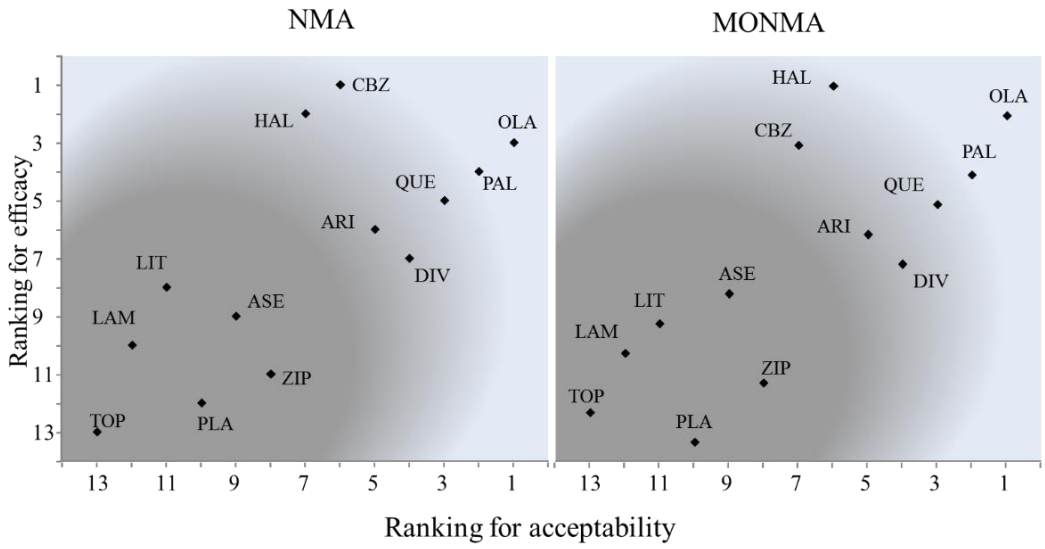
	τ_R^2	τ_D^2	ρ_τ
NMA	0.08 (0.02,0.17)	0.13 (0.06, 0.24)	-
MONMA	0.08 (0.03,0.17)	0.14 (0.07, 0.24)	-0.84 (-0.98, -0.52)

3.5 Concluding remarks

In this Chapter, we have presented a model for performing data synthesis in a network of competing interventions, with multi-arm studies reporting on multiple dichotomous outcomes. Both within and between-study correlations between the outcomes were taken into account. We proposed a method for eliciting expert opinion to inform within-study correlations. Motivated by the fact that questions about probabilities are better understood compared to questions about correlations, we proposed the use of a set of conditional probabilities to elicit information for the correlations. We showed how to construct prior distributions for these probabilities based on expert opinion and how to use these priors in order to estimate the within-study correlation coefficients needed. For between-study correlations we proposed a method of simplifying the variance-covariance matrix by making a set of assumptions. Our method was applied to the case of two correlated dichotomous outcomes in the presence of two-arm and three-arm studies. The methods presented can be extended for more than two outcomes and for networks that include studies with more than three arms. A generalization of our model is presented in Section VII of the Appendix.

We fitted our model within a Bayesian framework which allows for a direct incorporation of prior information and an easy way of including studies that report on some, but not all of the outcomes. Another advantage of the Bayesian approach is that it is free of the convergence problem often encountered in likelihood based methods when the number of studies is small or the within-study variation relatively large (213,214).

Figure 8: Ranking for antimanic drugs for response and dropout. Treatments located in the darker (brighter) areas of the plots have the lowest (highest) rankings. ARI = aripiprazole, ARI = aripiprazole. ASE = asenapine. CBZ = carbamazepine. VAL = divalproex. HAL = haloperidol. LAM = lamotrigine. LIT = lithium. OLZ = olanzapine, PBO = placebo. QTP = quetiapine. PAL = paliperidone. TOP = topiramate. ZIP = ziprasidone. Gabapentin does not feature in the graphs as its efficacy has not been studied in any of the trials.



We implemented our model for the case of a network of treatments for acute mania and two (negatively) correlated outcomes: response to the treatment and all-cause discontinuation (dropout rate). Our model gave similar results with the simple univariate model for the mean estimated treatment effects. However, it produced narrower confidence intervals, especially for response, since almost one third of the studies did not report on this outcome, thus allowing for a 'borrowing of strength' between the two outcomes. The precision gain for the dropout was marginal, since all studies except one reported the number of patients dropping out. In this particular example the change in the precision of the estimates for response had a small impact on the relative ranking of the treatments.

Our model is suitable for dichotomous outcomes but requires arm-level data and it is also subject to the assumptions we have made for the structure of the between-study variance-covariance matrix. Our method considered the case of correlated odd ratios;

however it can be extended to analyze risk ratios. Following the methodology presented in this Chapter one can derive formulas for correlated log risk ratios.

The research presented in this Chapter was published in the Statistics in Medicine journal (215).

4 Joint synthesis of multiple correlated outcomes in network meta-analysis

4.1 Introduction

In Chapter 3 of this dissertation we presented a new model that can be used to perform a network meta-analysis for the case of multiple dichotomous outcomes. In this chapter we describe two additional MONMA models, that can be used to synthesize multiple binary, continuous or time-to-event outcomes. The first model is based on making a set of simplifying assumptions for the within and between-studies variance-covariance matrices. The second model is a generalization of a bivariate pairwise meta-analysis model initially proposed by Riley et al. (213). This model includes a single correlation coefficient, which is used to model the overall correlation, i.e. an amalgam of the within-study and between-study correlations. In order to exemplify our methods we use the acute mania dataset, which was introduced in Section 3.2. We fit the two new MONMA models in a Bayesian framework, which offers flexibility in incorporating prior beliefs and allows for a straightforward inclusion of studies that do not report on all outcomes, as well as accounting for uncertainty in parameter estimates.

This chapter is organized as follows: in Section 4.2.1 we start by presenting a brief outline of the general framework for jointly meta-analyzing multiple outcomes for the case of two competing treatments. This framework was presented in detail in Section 3.3.1, but we also summarize it here in brief, for the reader's convenience. Then, in Section 4.2.2 we present an alternative MOMA model introduced by Riley et al. (213), which can be used for the pairwise meta-analysis, for the case of two outcomes. In Section 4.2.3 we generalize both these approaches for a network of interventions in the presence of multi-arm studies. In that section we also discuss the technicalities of fitting these models. In Section 4.3 we apply the new models to our data, in order to produce estimates for outcome-specific relative treatment effects, and evaluate the relative ranking of the treatments for each outcome. In Section 4.4 we summarize our findings. All mathematical and statistical details, as well as the software codes that were used for the analyses are presented in the Appendix.

4.2 Statistical methods

4.2.1 General framework for pairwise meta-analysis of multiple outcomes

Here we start by providing a brief account of a MOMA model. More details can be found in Section 3.3.1 of this thesis. Assume N_S studies comparing two treatments (for example a new treatment versus a placebo) with respect to two different but correlated outcomes, denoted with R and D. We use $\boldsymbol{Y}$ to denote the $2N_S$-dimensional vector of the observed effects; in our example, these are the log odds ratio for R and D, but in other situations they could be mean difference or log hazard ratio estimates, for example. The bivariate random effects meta-analysis model can be written, using matrix notation, as $\boldsymbol{Y} = \boldsymbol{X\mu} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{X}$ is the design matrix, $\boldsymbol{\mu}$ the vector of true mean relative treatment effects and $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are the vectors of random effects (reflecting between-study variability) and random errors (reflecting within-study sampling variability) respectively.

For a joint meta-analysis of both outcomes we must incorporate the correlations between the outcomes, both within as well as between-studies. We assume multivariate normal distributions for $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$, so that $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma})$ and $\boldsymbol{\delta} \sim N(0, \boldsymbol{\Delta})$, with $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ being the within and between study variance-covariance matrices. The variance-covariance matrix for the random effects takes a block-diagonal form, with identical $\boldsymbol{\Delta}_{(2\times2)}$ matrices in its diagonal (Equation (7), Section 3.3.1), and incorporates three parameters, ρ^{τ}, τ_R and τ_D. More specifically:

$$\boldsymbol{\Delta}_{(2\times2)} = \begin{pmatrix} \tau_R^2 & \rho^{\tau}\tau_R\tau_D \\ \rho^{\tau}\tau_R\tau_D & \tau_D^2 \end{pmatrix}$$

Note that this matrix is always positive-definite for $-1 < \rho^{\tau} < 1$. The corresponding parameters need to be estimated from the model. In a frequentist framework options include restricted maximum likelihood and methods of moments; here we focus on a Bayesian framework estimated using Markov Chain Monte Carlo (described in Section 4.3.1 later). The random errors variance-covariance matrix $\boldsymbol{\Sigma}$ is also block diagonal, see Equation (8), Section 3.3.1. In this matrix, ρ_i is the within-study correlation coefficient and $\sigma_{i,R}^2$, $\sigma_{i,D}^2$ are the variances of the effect sizes in each study i. All entries in $\boldsymbol{\Sigma}$ are estimated from the data. Sample estimates for $\sigma_{i,R}^2$ and $\sigma_{i,D}^2$ are often available, but few studies, if any, would provide enough information to estimate the within-study correlation coefficient ρ_i and the majority of meta-analyses do not have access to IPD that would enable its estimation.

Within a Bayesian framework we can give prior distributions to all the correlation

coefficients in order to perform a full multivariate meta-analysis. One can model these coefficients in a variety of ways, e.g. assume all ρ_i to be equal ($\rho_i = \rho \; \forall i$), assume a different coefficient depending on study characteristics, place a vague or informative prior on each ρ_i etc.

4.2.2 Riley's alternative multiple outcomes meta-analysis model

Following a different approach, Riley et al. proposed an alternative model for a bivariate, random-effects pairwise meta-analysis. The model allows for a single coefficient to model the overall correlation. This plays the role of an amalgam of the correlations within and between studies (213). Instead of modeling $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ separately, in this model the authors assume an overall variance-covariance matrix $\boldsymbol{\Omega}$, so that $\boldsymbol{Y} = \boldsymbol{X\mu} + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \sim N(0, \boldsymbol{\Omega})$. This matrix $\boldsymbol{\Omega}$ is again block diagonal, with each block corresponding to a study, so that $\boldsymbol{\Omega} = Diag(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}, \dots, \boldsymbol{\Omega_{N_S}})$.

For a study i this matrix takes the following form:

$$\boldsymbol{\Omega} = \begin{pmatrix} \psi_R^2 + \sigma_{i,R}^2 & \rho_i^h \sqrt{\psi_R^2 + \sigma_{i,R}^2}\sqrt{\psi_D^2 + \sigma_{i,D}^2} \\ \rho_i^h \sqrt{\psi_R^2 + \sigma_{i,R}^2}\sqrt{\psi_D^2 + \sigma_{i,D}^2} & \psi_D^2 + \sigma_{i,D}^2 \end{pmatrix} \quad (19)$$

The ρ_i^h coefficient in Equation (19) is the overall correlation in study i, a hybrid of the within-study and between-study correlation coefficients.

We can again model the different ρ_i^h in a variety of ways, depending on the nature of the data, e.g. $\rho_i^h = \rho \; \forall \; i$. The ψ parameters model for the variation additional to the sampling error that enters due to heterogeneity, and they are similar to the τ parameters that enter the $\boldsymbol{\Delta}_{(2\times 2)}$. But they are not directly equivalent, unless the within-study variances are small relative to the between-study variances. The clear advantage of model (19) is that the within-study correlations are no longer needed.

4.2.3 Network meta-analysis for two correlated outcomes

The models described in the two previous Sections cannot handle the case when studies comparing more than two treatments. Moreover, the model described in Chapter 3 of this dissertation focused on the case of binary outcomes.

In this section we present two models for performing a network meta-analysis of studies

with multiple arms reporting on two correlated outcomes. The outcomes can be binary (and relative treatment effect can be measured as log odds ratios or log risk ratios), continuous (effects measured as mean differences or standardized mean differences) or time-to-event (effects measured as log hazard ratios). Note that in order to use standardized mean difference for a continuous outcome a large sample approximation is required. For more details see Section III of the Appendix.

In the acute mania example the outcomes are identified as the binary response to the treatment (R) and dropout rate (D). We exemplify the methodology for the case of networks containing studies with a maximum of three arms. We assume a random effects model and that the consistency equations ($\beta_{XY,R} = \beta_{XZ,R} - \beta_{YZ,R}$) hold for all treatments X, Y, Z; similarly for outcome D.

4.2.3.1 Model 1: Simplifying the variance-covariance matrices

The first MONMA model we present is based on making assumptions that simplify the within and between-study variance-covariance matrices. These assumptions are needed in order to minimize the number of parameters that need to be estimated, thus easing the computational burden and potential estimation difficulties. Some of the considerations presented in this section were also discussed in Chapter 3, but they are also briefly summarized here for completeness.

Let us start by considering a network of studies reporting on the correlated outcomes R and D for a network of N_T different treatments. The model is $\boldsymbol{Y} = \boldsymbol{X\mu} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}$ with $\boldsymbol{Y}$ the vector of the observed effects, $\boldsymbol{X}$ the design matrix, $\boldsymbol{\mu}$ the vector of the basic parameters i.e. the $N_T - 1$ parameters for the comparison of each treatment versus the reference (6,64), $\boldsymbol{\delta}$ the vector of random effects and $\boldsymbol{\varepsilon}$ the vector of random errors (6,46). The design matrix $\boldsymbol{X}$ describes the structure of the network and embeds the consistency equations (6); it maps the observed comparisons into the basic parameters. For example, if A is chosen to be the reference treatment, a study comparing B to C for outcome R provides information for a linear combination of two basic parameters as $\beta_{BC,R} = \beta_{AC,R} - \beta_{AB,R}$.

For a two-arm study i that compares treatments A and B the random errors are assumed to follow a multivariate normal distribution, $(\delta_{i,AB,R}, \delta_{i,AB,D})' \sim N(0, \boldsymbol{\Delta}_{(2\times2)})$. In network meta-analysis it is often assumed that the heterogeneity is independent of the comparison being made, i.e. $\tau^2_{AB,R} = \tau^2_R$ and $\tau^2_{AB,D} = \tau^2_D$ for every pair of treatments A, B, and we also

assume this here. For a three-arm study i that compares treatments A, B and C, the random effects are again assumed to follow a multivariate normal distribution $(\delta_{i,AB,R}\ , \delta_{i,AB,D}\ ,\ \delta_{i,AC,R}\ ,\ \delta_{i,AC,D})' \sim N(0, \boldsymbol{\Delta}_{(4\times4)})$. Assuming equal heterogeneities between treatment comparisons and equal correlations between random effects of different comparisons and different outcomes, i.e. $(\delta_{i,AB,R}\ , \delta_{i,AC,D}) = corr(\delta_{i,AB,D}\ \ \delta_{i,AC,R}\ )$, as we show in Section III of the Appendix, the $\boldsymbol{\Delta}_{(4\times4)}$ matrix takes the form presented in Equation (15):

$$\boldsymbol{\Delta}_{(4\times4)} = \tau_R^2 \begin{pmatrix} 1 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \tau_D^2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1 \end{pmatrix} + \rho_\tau \tau_R \tau_D \begin{pmatrix} 0 & 1 & 0 & 1/2 \\ 1 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1 \\ 1/2 & 0 & 1 & 0 \end{pmatrix}$$

The random errors are also assumed to follow a multivariate normal distribution. For a three-arm study i that compares treatments A, B and C for response (R) and dropout (D) we assume $(\varepsilon_{i,AB,R}\ ,\ \varepsilon_{i,AB,D}\ ,\ \varepsilon_{i,AC,R}\ ,\ \varepsilon_{i,AC,D})' \sim N(0, \boldsymbol{\Sigma}_i)$. The variance-covariance matrix $\boldsymbol{\Sigma}_i$ is given by Equation (16). As we discuss in Section 3.3.3, the σ and κ coefficients in $\boldsymbol{\Sigma}_i$ can be readily estimated if arm level data are available.

In what follows we present a method for dealing with the remaining correlation terms within $\boldsymbol{\Sigma}_i$. We start by assuming that there are two different types of within-study correlation coefficient for every study i. The first we denote by ρ_i^*, and corresponds to the correlation of relative treatment effects of different outcomes for the same treatment comparison. This enters the variance-covariance matrices for both two-arm and three-arm studies. The second we denote by ρ_i^{**} and correlates the relative treatment effects for different comparisons and different outcomes within the same study. This enters only the (4×4) matrices of the three-arm studies. This means that:

$$\rho_{i,AB_RAB_D} = \rho_{i,AC_RAC_D} \equiv \rho_i^*,\ \rho_{i,AC_RAB_D} = \rho_{i,AB_RAC_D} \equiv \rho_i^{**}, \quad \textit{(Assumption 1)}$$

The within-study variance-covariance matrix for a two-arm study i comparing treatments A and B for two outcomes is:

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_{i,AB,R}^2 & \rho_i^* \sigma_{i,AB,R} \sigma_{i,AB,D} \\ \rho_i^* \sigma_{i,AB,R} \sigma_{i,AB,D} & \sigma_{i,AB,D}^2 \end{pmatrix} \tag{20}$$

For a three-arm study comparing treatments A, B and C for two outcomes the $\boldsymbol{\Sigma}_i$ matrix

of Equation (16) now becomes:

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma^2_{i,AB,R} & . & . & . \\ \rho_i^{*}\sigma_{i,AB,R}\sigma_{i,AB,D} & \sigma^2_{i,AB,D} & . & . \\ \kappa_{i,AB_RAC_R} & \rho_i^{**}\sigma_{i,AB,D}\sigma_{i,AC,R} & \sigma^2_{i,AC,R} & . \\ \rho_i^{**}\sigma_{i,AB,R}\sigma_{i,AC,D} & \kappa_{i,AB_DAC_D} & \rho_i^{*}\sigma_{i,AC,R}\sigma_{i,AC,D} & \sigma^2_{i,AC,D} \end{pmatrix}$$

It is very often the case that study arms are balanced in numbers of patients randomized. Then, for treatments that are not very different in efficacy and dropout (e.g. drugs from the same class) we can assume that:

$$\sigma_{i,BC,R} = \sigma_{i,AB,R} = \sigma_{i,AC,R} \text{ and } \sigma_{i,BC,D} = \sigma_{i,BA,D} = \sigma_{i,AC,D} \quad \textit{(Assumption 2)}$$

This assumption will not be reasonable if trials are imbalanced or compare very different treatments. Insight on the validity of this assumption can be obtained from the data after scanning for important differences among the estimated variances across studies. If we choose to employ this assumption the model is considerably simplified as it implies that $\rho_i^{**} = {}^1\!/_2\, \rho_i^{*}$ (see Section VIII of the Appendix). Consequently, an estimate of the variance-covariance matrix for the three-arm study i after Assumptions 1 and 2 is as follows:

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \hat{\sigma}^2_{i,AB,R} & . & . & . \\ 0 & \hat{\sigma}^2_{i,AB,D} & . & . \\ \hat{\kappa}_{i,AB_RAC_R} & 0 & \hat{\sigma}^2_{i,AC,R} & . \\ 0 & \hat{\kappa}_{i,AB_DAC_D} & 0 & \hat{\sigma}^2_{i,AC,D} \end{pmatrix} + \rho_i \begin{pmatrix} 0 & . & . & . \\ \hat{\sigma}_{i,AB,R}\sigma_{i,AB,D} & 0 & . & . \\ 0 & \hat{\sigma}_{i,AB,D}\sigma_{i,AC,R}/2 & 0 & . \\ \hat{\sigma}_{i,AB,R}\sigma_{i,AC,D}/2 & \kappa_{i,AB_DAC_D} & \hat{\sigma}_{i,AC,R}\sigma_{i,AC,D} & 0 \end{pmatrix} \equiv \widehat{\boldsymbol{\Sigma}}_{i,1} + \rho_i \widehat{\boldsymbol{\Sigma}}_{i,2} \tag{21}$$

In the last line we have renamed ρ_i^{*} to ρ_i, in order to simplify notation and to highlight that the correlation coefficient is equivalent to the one presented in Equation (16). It is important to note that Assumption 2 does not mean that we force all study variances to be equal: the diagonal elements of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}$ are distinct and are estimated from the studies. We employ this assumption only for the off-diagonal elements of the variance-covariance matrix so that all correlations are functions of a single parameter ρ_i. Consequently all elements of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,1}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,2}}$ in equation (23) can be estimated when arm-level data are available. The assumption of equal variances within a multi-arm study can be omitted, if it is deemed inappropriate. In Section VIII of the Appendix we present the most general form of the variance-covariance matrix for different variances, and compute general relations between the correlation

coefficients it contains. This, however, results in a rather complicated structure for the $\boldsymbol{\Sigma_i}$ matrix and we will not consider any further.

To summarize, we have expressed all within-study variance-covariance matrices utilizing a set of correlation coefficients ρ_i, one for every study i, that measure the correlation between the relative treatment effects of the two outcomes R and D for the same treatment comparison. These coefficients might be available in study reports. Alternatively, they can be deducted from empirical evidence (194) or expert opinion; in Chapter 3 we discussed how this can be achieved for the case of MONMA for binary outcomes. If IPD are available then the correlation coefficient can be estimated (203). A joint network meta-analysis of the two outcomes can be performed within a Bayesian framework after assigning prior distributions to the ρ_i. These priors can be either uninformative or can be defined after consulting with clinicians (210). We have a number of options on how to model these coefficients. The simplest one is to assume $\rho_i = \rho$, common correlation for all studies. We could alternatively assume correlation coefficients across studies to share a common distribution. Another choice would be to have different ρ_i's for different group of studies. For example we could assume a coefficient ρ_{Act-Pl} for placebo-controlled studies, and another $\rho_{Act-Act}$ for head-to-head studies that compare only active treatments; this would be based on the assumption that the two relative effect measures are differently correlated when one of the treatments compared is the placebo.

One technical implication that comes up is that the positive-definiteness of the within-study variance-covariance matrix is not guaranteed for three-arm studies. The estimated matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}$ for the random errors in Equation (21) is not always positive-definite, as it depends on the data and on an arbitrary parameter ρ_i. One way to overcome this problem is to compute the four eigenvalues $\lambda_{i,j}$ of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}$ for every study i, with $j = 1, 2, 3, 4$, and truncate them to zero, replacing $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i}} = \sum_j \max(0, \lambda_{i,j}) \boldsymbol{v_{i,j}} \boldsymbol{v'_{i,j}}$, with $\boldsymbol{v_{i,j}}$ the corresponding eigenvectors as in Jackson et al. (73). This, however, might be difficult to implement, particularly if a Bayesian software is used. Here we propose a different way of dealing with this problem: we can truncate the correlation coefficient for every study so that the positive-definiteness of the variance-covariance matrix is ensured. If for example we assume a uniform $(-1,1)$ prior distribution for each ρ_i, we must truncate: $\rho_i \sim Unif(-1,1) I(l_i, u_i)$. The limits l_i and u_i are the lowest and highest values of ρ_i that lead to a positive definite matrix. That means that we need to compute those values for all three-arm studies: it can be easily achieved by checking

the corresponding eigenvalues of the variance-covariance matrix, as a positive-definite matrix has only positive eigenvalues. In Section 4 of the Supplementary Material we provide a program in R software that computes the limits l_i and u_i for every three-arm study. Wei and Higgins discuss other approaches to ensure positive-definite matrices including Cholesky paramaterisation and spherical decomposition (202).

4.2.3.2 Model 2: Extending the alternative MOMA model

In this Section we discuss a second method for performing a multiple-outcomes network meta-analysis, by extending Riley's et al alternative model (213). The model described in Section 4.2.2 is $\boldsymbol{Y} = \boldsymbol{X\mu} + \boldsymbol{\eta}$, with $\boldsymbol{\eta} \sim N(0, \boldsymbol{\Omega})$, where, as in the case of pairwise meta-analysis the matrix $\boldsymbol{\Omega}$ is block diagonal. For a two-arm study the variance-covariance matrix is as given in Equation (19). As we show in Section X of the Appendix, if we are willing to employ *Assumption 2* for a three-arm study i comparing treatments A, B and C for two outcomes, then its variance - covariance matrix $\boldsymbol{\Omega_i}$ is given by:

$$\boldsymbol{\Omega_i} = \begin{pmatrix} \zeta_{i,AB,R} & . & . & . \\ \rho_i^h\sqrt{\zeta_{i,AB,R}\zeta_{i,AB,D}} & \zeta_{i,AB,D} & . & . \\ \frac{1}{2}\sqrt{\zeta_{i,AB,R}\zeta_{i,AC,R}} & \frac{\rho_i^h}{2}\sqrt{\zeta_{i,AB,D}\zeta_{i,AC,R}} & \zeta_{i,AC,R} & . \\ \frac{\rho_i^h}{2}\sqrt{\zeta_{i,AB,R}\zeta_{i,AC,D}} & \frac{1}{2}\sqrt{\zeta_{i,AC,R}\zeta_{i,AC,D}} & \rho_i^h\sqrt{\zeta_{i,AC,R}\zeta_{i,AC,D}} & \zeta_{i,AC,D} \end{pmatrix} \quad (22)$$

Here we have defined $\zeta_{i,AB,R} = \sigma^2_{i,AB,R} + \psi^2_R$, $\zeta_{i,AB,D} = \sigma^2_{i,AB,D} + \psi^2_D$, etc. Equation (22) extends the model presented by Riley et al. for three-arm studies with two outcomes. The σ parameters can again be estimated from the data as the standard errors of the effect sizes, and assuming a common correlation coefficient across studies there are three parameters left to estimate: ψ_R, ψ_D and ρ^h. One of the advantages of this approach is that the variance-covariance matrix is always positive-definite, so a multivariate meta-analysis can be readily performed without further complications. As described in the previous Section, the equal variance assumption (Assumption 2) can be omitted if the studies are imbalanced or the treatments have significant differences in the measured effects, leading, however, to a much more complicated $\boldsymbol{\Omega_i}$ variance-covariance matrix.

4.3 Application to the acute mania example

4.3.1 Description of the analyses and model fit

We fit the two models we presented in Section 4.2.3 in a Bayesian framework, using the OpenBUGS software (94,95). Prior distributions need to be assigned to all model parameters. The parameters τ_R, τ_D of the first model and ψ_R, ψ_D of the second can be assigned minimally informative prior distributions. If there is no prior information on the correlation of the outcomes, an uninformative $U(-1,1)$ prior can be used on all correlation coefficients. If external information is available on these coefficients, e.g. elicited from experts in the field, it can be used to inform ρ or ρ^h. In the acute mania example, the correlation between response and dropout rate is expected to be negative so we assigned appropriate negative priors to parameters ρ_i (the within-study correlations between outcomes, assumed equal across studies), ρ^τ (the between-study correlation in outcomes) and ρ^h (the overall correlation). However, the robustness of conclusions to this assumption could be checked if desired. In order to rank the treatments with respect to the response and the dropout rate, we computed the surface under the SUCRA value (38), for each treatment and for each outcome. All results pertain to 1,000,000 iterations and thinning of 100 after a 5,000 burn-in period; the thinning was deemed necessary since a preliminary analysis showed a high auto-correlation in the chains. The code we used is provided in Sections XII and XIII of the Appendix. We explored the following analysis scenarios:

I. Univariate (independent) NMA of response and dropout rate separately, assuming $\tau_R, \tau_D \sim U(0,1)$. This corresponds to setting all correlations equal to zero.

II. Multiple outcome network meta-analysis (MONMA) following the approach of Section 4.2.3.1. We used minimally informative priors for the heterogeneity parameters: $\rho^\tau \sim U(-1,0)$, $\tau_R, \tau_D \sim U(0,1)$, and: (a) we assumed a negative common $\rho_i = \rho$ with $\rho \sim U(-1,0)$; (b) we assumed a strongly informative, negative, common $\rho \sim U(-0.7,-0.5)$; (c) we assumed a common fixed $\rho_i = \rho$ with $\rho = -0.7$; (d) we assumed two different within-studies correlation coefficients ρ_i: one for the studies comparing two active treatments, which we denote as $\rho_{Act-Act}$, and another for the studies comparing active treatments to placebo, ρ_{Act-Pl}. This distinction could be based on the assumption that the two

relative treatment effects are differently correlated when one of the treatments compared is the placebo. For both parameters we used a uniform negative, $U(-1,0)$, prior distribution

III. MONMA following the approach in Section 4.2.3.2, assuming a common correlation coefficient and the following prior distributions for the parameters of the model: $\rho^h \sim U(-1,0)$, $\psi_R \sim U(0,1)$, $\psi_D \sim U(0,1)$.

In order to evaluate our assumption of a negative correlation coefficient within and across studies we fitted MONMA model following the approach of Section 4.2.3.1. with $\rho_i = \rho$ with $\rho \sim U(-1,1)$ and $\rho^\tau \sim U(-1,1)$.

4.3.2 Results

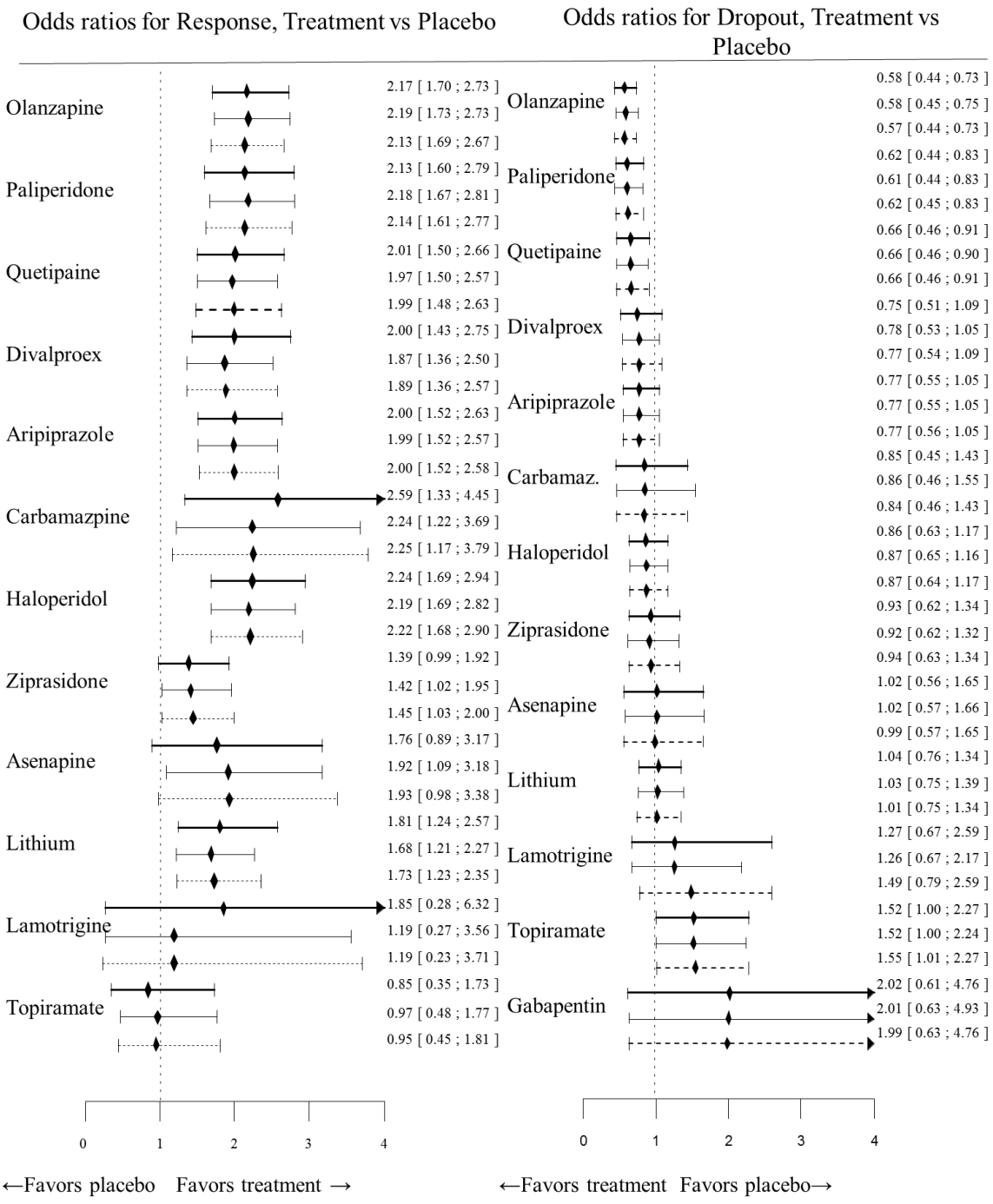
The median posterior values for ρ and ρ^τ when non-informative $U(-1,1)$ priors are used were -0.33 and -0.84 with 95% credible intervals [-0.66; 0.14] and [-0.99; -0.38] respectively. These values corroborate our prior belief of a negative association between dropout and efficacy. In Table 6 we present the median posterior estimates and 95% credible intervals for the parameters in each model. An interesting observation is that the heterogeneity variances τ_R^2 and τ_D^2 are invariant across the different models. This may be due to the large number of studies available in this meta-analysis. The mean estimates for the correlation coefficients are well below zero (e.g. the between-study correlation ranges from –0.56 for scenario II.c up to -0.82 for scenario II.a). The posterior median value for the overall correlation ρ^h in model III is -0.51 (95% Cr.I. [-0.68; -0.29]), a value lying between the estimates of the two correlation coefficients for the multivariate model II.a (-0.34 for ρ and -0.82 for ρ^τ). This is reasonable since ρ^h is an overall correlation coefficient that amalgamates the within and between-studies correlations measured by ρ and ρ^τ.

Table 6: Median posterior estimates and 95% credible intervals for the heterogeneity variance and correlation parameters in MONMA models.

Scenario	τ_R^2	τ_D^2	ρ	ρ^{τ}
I	0.08 [0.02;0.17]	0.13 [0.06;0.24]	-	-
II.a	0.07 [0.02;0.16]	0.13 [0.06;0.24]	-0.34 [-0.66;-0.04]	-0.82 [-0.99;-0.38]
II.b	0.07 [0.02;0.15]	0.13 [0.06;0.23]	-0.56 [-0.68;-0.50]	-0.68 [-0.93;-0.23]
II.c	0.07 [0.02;0.16]	0.13 [0.06;0.24]	-	-0.56 [-0.83;-0.12]
II.d	0.08 [0.02;0.16]	0.13 [0.06;0.23]	$\rho_{Act-Act}$: -0.31 [-0.71;-0.02] ρ_{Act-Pl}: -0.39 [-0.77;-0.04]	-0.80 [-0.99;-0.33]
	ψ_R^2	ψ_D^2	ρ^h	
III	0.07 [0.02;0.16]	0.12 [0.04;0.22]	-0.51 [-0.68;-0.29]	

In Figure 9 we present the summary odds ratios for both outcomes for each treatment vs. placebo and for models I, II.b and III. In XI of the Appendix we present the results from fitting each model in detail. The multivariate approach has a minimal effect on the summary results for the dropout outcome compared to the univariate. That is expected (200) since this outcome was reported in all studies except one, and thus inferences do not gain much through the joint analysis in terms of the posterior estimates and precision for this outcome. In contrast, the posterior summary ORs for the response to treatment outcome have considerable gain in precision when we use a multivariate rather than univariate model. This gain arises because 28% of the studies did not report on response, and thus the multivariate models additionally borrow strength from the correlated dropout outcome in these studies (200). The gain in precision is larger as within-study correlation coefficient moves away from zero; the decrease in the width of the confidence intervals of the ORs compared to the results from the univariate approach is on average 8.4% for analysis II.a, 12% for II.b, 12.1% for II.c, 8.2% for II.d, and 10.8% for model III. Note that apart from differences in precision gain there are small changes in the point estimates for most odds ratios among the MONMA models (see Section XI of the Appendix).

Figure 9: Summary odds ratios for response and dropout, for active treatment vs. placebo. The thick lines correspond to scenario I (univariate model), the slim lines to scenario II.b (MONMA model assuming strong correlation coefficient ρ~U(-0.7,-0.5)) and the dashed lines to scenario III (alternative MONMA model assuming ρ^h~U(-1,0))



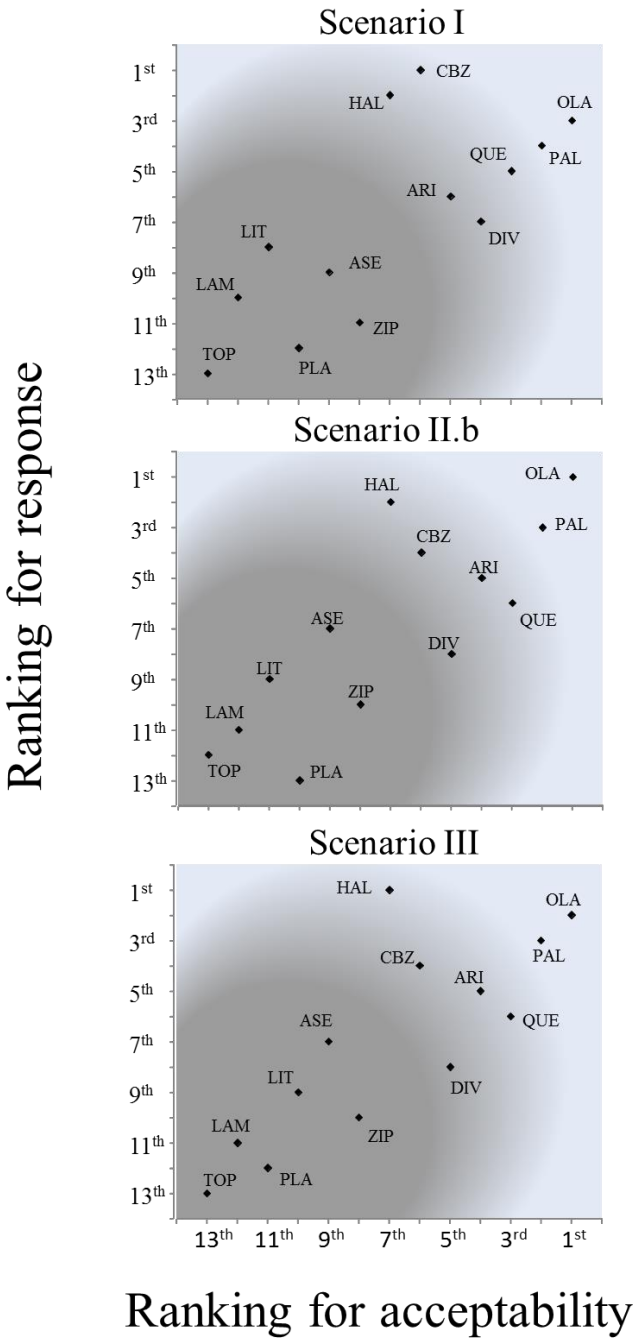
In Figure 10 we present the relative ranking of treatments for response and dropout, for models I, II.b and III, based on the SUCRA value for each outcome. Treatments near the

upper right corner are the best when both acceptability and efficacy outcomes are considered jointly important; those near the bottom left corner (dark areas of the plots) are the worst. Note that Gabapentin is not present in the graph, since it was only reported for dropout. Regardless of the choice of model, OLA has the highest ranking across both outcomes jointly. However, the ranking of some other treatments is affected by the choice of multivariate rather than univariate, especially in regard the response outcome which (through correlation) is able to borrow strength from the more complete acceptability outcome. This use of additional information leads to (small) differences in the multivariate and univariate mean posterior estimates and precision of the summary ORs for response, and this has an impact on the relative ranking of the treatments for this outcome. For example, Carbamazepine ranks as the best treatment in terms of response with the univariate model but it falls to the fourth place when we consider a within-study correlation coefficient $\rho = -0.7$.

4.4 Concluding remarks

In this Chapter we have presented two models for meta-analyzing evidence from multi-arm studies reporting multiple correlated outcomes in a network of interventions. Our models require minimum aggregated-level information and are applicable to any NMA with multiple continuous, dichotomous or time-to-event outcomes; that is the majority of the NMA applications (8). The set of models we present provides a unified way of handling multiple outcomes in the presence of multi-arm studies using only a handful of parameters. Choice between the two models may be informed by various factors. The first MONMA model accounts for within-study variances (sampling error), between-study variance (heterogeneity) as well as within and between-studies correlation.

Figure 10: Ranking of antimanic drugs for response and acceptability. Treatments located in the darker (brighter) areas of the plots have the lowest (highest) rankings. ARI = aripiprazole, ARI = aripiprazole. ASE = asenapine. CBZ = carbamazepine. VAL = divalproex. HAL = haloperidol. LAM = lamotrigine. LIT = lithium. OLZ = olanzapine, PBO = placebo. QTP = quetiapine. PAL = paliperidone, TOP = topiramate. ZIP = ziprasidone



The second (alternative) model includes both within-study and between-study variances, but uses a single correlation parameter ρ^h. Thus, the second model can be viewed as an approximation of the first MONMA model, with the latter having a more detailed likelihood structure. The second model can be used in the common situation when within-study covariances (the κ parameters of $\boldsymbol{\Sigma}_i$, Equation (16) in Section 3.3.3) are not available from all studies or cannot be reliably obtained from external data or expert opinion.

Ease of application is another consideration when choosing between the two models. The first model is more difficult to implement as it has a richer structure and investigators need to ensure the positive-definiteness of the variance-covariance matrix. Our models perform better than the univariate one in terms of precision; this gain, however, does not come without a cost. The complexity of the multivariate analysis is an important limitation, and the difficulty in implementing the models rises as the number of outcomes of interest or the number of arms of the studies in the network grows. When only a small number of studies do not report on all outcomes the gain in precision can be trivial, rendering the use of multivariate methods redundant. The models are also limited by the assumptions we used to simplify the structure of the variance-covariance matrices; in the Appendix we offer guidance for the case the analyst is unwilling to employ these assumptions.

Despite their limitations, the two presented models are to our knowledge the first attempts for meta-analyzing data from networks of interventions comprising multi-arm studies that report on multiple, correlated outcomes. In Chapter 3 we have presented a framework that utilizes expert clinical opinion about quantities easily understood by clinicians (such as proportions) to impute unreported correlation parameters. However, that method is only applicable for binary outcomes measured with odds ratios. In the present approach we provide two general models for all types of outcomes assuming that the within-study correlations are known or directly informed by external evidence (model 1) or completely unknown (model 2).

The research presented in this Chapter was published in the Biostatistics journal (216).

5 Summary

Standard methods for meta-analysis are limited to the case of comparing two interventions. In real life clinical practice, however, there are usually many alternative competing interventions that can be used to treat the same disease, while studies may contrast different sets of these interventions, thus forming a network of evidence. In such complicated cases of data availability pairwise meta-analyses cannot give a definite answer as to which intervention works best for the target condition. NMA is an extension of the standard, pairwise meta-analysis, and can be used to jointly analyze evidence regarding multiple interventions in order to produce clinically relevant estimates. It does so by utilizing the totality of the available information. For each comparison of two interventions there may be direct evidence (obtained from studies that compare the two interventions head-to-head) and/or indirect evidence (coming from the rest of the network). NMA combines these two types of information in a single analysis, which results in increased precision as compared to the usual pairwise meta-analysis. In addition, NMA allows the comparison of interventions that have never been compared in a clinical trial directly.

For these reasons, NMA methods are becoming increasingly popular, and there is an almost exponential growth in the number of published applications during the last few years. However, the underlying assumptions of the model may sometimes be difficult to assess while the mathematical complexity of the model, combined with the lack of easy-to-use computer software often result in researchers using suboptimal or even inappropriate methods (8,191).

In Chapter 2 of this dissertation we described an updated review of methods for NMA, which we performed in order to summarize the state-of-the-art in the field. Our scope was to provide a comprehensive account of the currently available methods, which can be used by researchers interested in assessing the quality of published NMAs, in applying NMA to answer new clinical questions, or in conducting further methodological research. In this final section of the dissertation we also provide a brief summary of recommendations regarding the implementation of NMA, as they emerged from our review.

When setting off to perform an NMA, researchers should start by considering whether the treatments they plan to compare can be thought to be ‘jointly randomizable’. This means that in principle any patient could be randomized to receive any of the treatments in the network. This is a key assumption and should always be considered when setting up the

network.

Next, researchers need to perform a systematic review to identify studies that address the clinical question at hand. This should be followed by a critical appraisal of all the available evidence. After identifying relevant studies to be included in the NMA, researchers should check whether there are differences in the definition of the treatments across comparisons in terms of dosage, duration, means of administration (e.g. pill vs. injection) etc. The existence of systematic differences in the definition of treatments in the studies may shed doubts regarding the validity of the transitivity assumption, which is a fundamental assumption of NMA.

Then researchers need to check the distribution of potential effect modifiers across comparisons, to make sure there are no important differences. Effect modifiers are study characteristics that may influence the relative effectiveness of interventions. Checking for differences in the effect modifiers, however, might be hindered by limited accessibility to information on relevant covariates or by the small number of studies contributing to the analysis. All these considerations should be described in detail in the review, to allow readers to conceptually evaluate the validity of the assumptions of NMA on their own.

Researchers then need to decide on the model they will use to perform the NMA. In Chapter 2 of this dissertation we discussed a range of alternative (but equivalent) models that can be employed. The choice between the different models should be primarily driven by the availability of technical expertise in the research team regarding the various software packages. If a Bayesian framework is adopted, it is important to discuss the choice of prior distributions. Particular caution is warranted when modelling variance parameters (such as heterogeneity), as they typically cannot be assigned non-informative prior distributions. This implies that the estimated heterogeneity may vary depending on the chosen prior distribution, and thereby influence network consistency and precision of relative treatment effects. Sensitivity analyses are crucial to understand the potential impact of key assumptions in the modelling process, and a minimal set of sensitivity analyses should be always pre-specified, to avoid data dredging.

If the network structure allows it, i.e. if there are closed loops in the network, a statistical assessment of inconsistency should follow the fitting of the model. Inconsistency refers to the statistical difference between direct and indirect information for a given treatment comparison. Evaluating inconsistency may be a challenging task, especially in the presence of multi-arm studies. In Chapter 2 we have presented a variety of methods and models

currently available for statistically checking the network for consistency and we have discussed the advantages and limitations of each method. We recommend the application of as many of the presented methods and models as possible in order to gain a better understanding on the validity of the consistency assumption and possible sources of inconsistencies. Researchers should bear in mind, however, that the absence of a statistically significant finding for inconsistency does not necessarily mean that the transitivity assumption holds: all tests for inconsistency are expected to have low power, while large values of heterogeneity may mask important inconsistency.

If statistically significant inconsistency is detected, researchers are advised to explore potential sources of it and try to explain it. Local methods for assessing inconsistency can point out possibly problematic studies, which should then be checked for data extraction errors, important differences in the distribution of effect modifiers or other possible biases. We have also presented various models for adding covariates and adjusting for suspected biases in the analysis. If sufficient studies are available, such models can be applied to explain and possibly eliminate inconsistencies. If however inconsistency persists, researchers can consider splitting up nodes in the network (e.g. high dose – low dose) or they can present the results from the appropriate inconsistency model (Lu & Ades model when all studies are two-armed; design-by-treatment model when multi-arm trials are present) along with the direct evidence.

Even though a statistical significant finding for inconsistency will imply that the NMA results may not be valid, observing such an inconsistency may provide additional insight and generate additional research questions about modifiers of the relative treatment effects. It can motivate further analyses, such as combining individual participant data and aggregated data or including information from observational studies. Thus, for the purposes of the decision-making process, NMA may be used not only as a method for comparing treatments, but may also serve as a tool for gaining insight on the drivers of real-life effectiveness.

The second aim of this dissertation was to advance the statistical methodology for jointly analyzing multiple correlated outcomes in NMA. In Chapter 3 we introduced a MONMA model which focused on the case of analyzing multiple dichotomous outcomes while accounting for the correlations between them. The model synthesizes information from RCTs augmented by external evidence, which can be obtained from expert clinicians. We highlighted the mathematical details of this model and we discussed in depth the elicitation process for obtaining expert information. In Chapter 4 we presented two additional MONMA

models. Both models can be used to synthesize multiple dichotomous, continuous, or time-to-event outcomes. The first of them has a richer structure. The second model is an approximation of the first, and can be used in cases of limited data availability. We provided the software codes needed to run all models and discussed possible extensions in the Appendix.

In order to illustrate our methods, we applied all our MONMA models to a network of antimanic drugs, where 15 drugs and placebo were compared in terms of efficacy and acceptability. We found that our models provided more precise estimates for most treatment comparisons, for both outcomes. This increase in precision was more pronounced when larger correlation was assumed between the outcomes. In addition, multivariate meta-analysis might provide more powerful and less biased results in the presence of selective outcome reporting in the original studies (178). This refers to the case when in some studies researchers choose to not present results that were statistically non-significant, or that were deemed to be clinically not so interesting.

Although our MONMA models were shown to perform well and might be preferable to a series of independent univariate analyses, they also have their drawbacks and limitations. The complexity of the analyses increases as the number of outcomes or the number of arms in the included studies increases. The gain in precision may be small if the correlation coefficients are close to zero. In such instances, the added benefit of joint modeling of correlated outcomes might be too small to justify the increased modeling complexity (195). The benefit of a performing a joint analysis of multiple outcomes will also depend on the fraction of studies not reporting one of the outcomes. There is a critical balance of having enough studies reporting both outcomes to capture the correlation, and having enough studies not reporting both outcomes, in order to benefit from the borrowing of strength that could result from the model. A future simulation study may help explore the gains in precision for different values of the correlation coefficients, for different numbers of studies not reporting some of the outcomes, in order to pinpoint the cases where complicated modeling will result in considerably more precise estimates.

A limitation of our models is that they make the assumption that there are no missing outcome data in the studies. In the acute mania example response data were not available for a small proportion of patients (less than 10%) and the missing entries were imputed as failures. This imputation analysis has been shown to not materially impact the result from network meta-analysis when the total missing rate is small (134). When imputations are

needed for multiple, correlated outcomes, the impact of the imputation process on the correlation between the outcomes should be considered. For example, if the same strategy (e.g. to impute missing data as failures) is followed for two efficacy outcomes, then this is expected to increase their correlations.

To summarize, based on our findings we recommend researchers to consider both univariate and multivariate approaches when possible, to ascertain if clinical conclusions about the ranking of treatments for each outcome remain consistent under different model assumptions. Finally, since multiple outcomes network meta-analysis is a new, largely unexplored area, there are still many open areas for research. A possible extension would be to include IPD, either exclusively or in a combination with aggregated data. Furthermore, our models could be implemented in popular statistical software making MONMA more easily accessible to review authors.

As a final, concluding remark, we believe that the research presented in this dissertation is an important advancement in the field of NMA. We also think that our models constitute the best available method for the network meta-analysis of multiple correlated outcomes, and that their implementation is in practice straightforward.

6 Περίληψη

Οι κλασικές μέθοδοι μετά-ανάλυσης περιορίζονται στην περίπτωση της σύγκρισης δύο θεραπευτικών παρεμβάσεων. Στην κλινική πράξη, ωστόσο, υπάρχουν συνήθως πολλές εναλλακτικές παρεμβάσεις που μπορούν να χρησιμοποιηθούν για τη θεραπεία της ίδιας νόσου. Παράλληλα, όταν προσπαθήσει κάποιος να απαντήσει στο ερώτημα ποια από αυτές τις παρεμβάσεις δουλεύει καλύτερα για την αντιμετώπιση της συγκεκριμένης νόσου, μπορεί να εντοπίσει μια πληθώρα τυχαιοποιημένων κλινικών δοκιμών που έχουν συγκρίνει διαφορετικά υποσύνολα αυτών των παρεμβάσεων. Σε τέτοιες τις περιπτώσεις, στις οποίες τα διαθέσιμα τεκμήρια σχηματίζουν ένα δίκτυο θεραπευτικών παρεμβάσεων, η συνηθισμένη μετά-ανάλυση δεν μπορεί να δώσει μια σαφή απάντηση ως προς το ποια παρέμβαση είναι προτιμότερη. Η μετά-ανάλυση δικτύου (ΜΑΔ) είναι μια επέκταση της κλασικής μετά-ανάλυσης, και μπορεί να χρησιμοποιηθεί για να αναλύσει από κοινού δεδομένα για πολλαπλές παρεμβάσεις, προερχόμενα από πολλές διαφορετικές κλινικές μελέτες, με σκοπό την παραγωγή κλινικά χρήσιμων εκτιμήσεων. Η ΜΑΔ συνθέτει το σύνολο της πληροφορίας σε μία ενιαία ανάλυση. Η υπάρχουσα πληροφορία για την σύγκριση δύο συγκεκριμένων παρεμβάσεων μπορεί να είναι είτε άμεση (από κλινικές μελέτες που συγκρίνουν τις εν λόγω παρεμβάσεις) είτε έμμεση (μέσω του δικτύου). Συνδυάζοντας άμεση και έμμεση πληροφορία η ΜΑΔ επιτυγχάνει μια αύξηση της ακρίβειας σε σύγκριση με την κλασσική μετά-ανάλυση, ενώ επιτρέπει και την σύγκριση θεραπειών που δεν έχουν συγκριθεί ποτέ σε κάποια κλινική μελέτη.

Για αυτούς τους λόγους η ΜΑΔ γίνεται ολοένα και πιο δημοφιλής. Μάλιστα τα τελευταία χρόνια, έχει παρατηρηθεί μια σχεδόν εκθετική αύξηση του αριθμού των δημοσιευμένων εφαρμογών της ΜΑΔ. Παρόλα αυτά, για ένα αριθμό ερευνητών εξακολουθεί να μην είναι ξεκάθαρο το πως μπορεί κάποιος να αξιολογήσει τις βασικές παραδοχές του μοντέλου. Αυτό το γεγονός, σε συνδυασμό με την έλλειψη εύχρηστου λογισμικού για ΜΑΔ συχνά έχει ως αποτέλεσμα οι ερευνητές να χρησιμοποιούν ανεπαρκείς ή ακόμη και ακατάλληλες μεθόδους στις αναλύσεις τους (8,191).

Στο Κεφάλαιο 2 της παρούσας διατριβής περιγράψαμε μια συστηματική ανασκόπηση των μεθόδων για ΜΑΔ, η οποία πραγματοποιήθηκε με σκοπό να συνοψίσει τις σύγχρονες μεθόδους. Σκοπός μας ήταν να προσφέρουμε μια ολοκληρωμένη εικόνα για την σύγχρονη μεθοδολογία της ΜΑΔ, συνοψίζοντας την υπάρχουσα τεχνογνωσία. Το προϊόν αυτής της ανασκόπησης μπορεί να φανεί χρήσιμο σε ερευνητές που ενδιαφέρονται για την αξιολόγηση

της ποιότητας των δημοσιευμένων ΜΑΔ, για την εφαρμογή της ΜΑΔ στην απάντηση νέων κλινικών ερωτημάτων, ή και στη διεξαγωγή περαιτέρω μεθοδολογικής έρευνας σε μεθόδους για ΜΑΔ. Σε αυτό το σημείο θα παρουσιάσουμε ένα περίγραμμα των βέλτιστων πρακτικών για την εφαρμογή της ΜΑΔ, όπως αυτές προέκυψαν από την συστηματική ανασκόπηση της βιβλιογραφίας.

Το πρώτο βήμα σε μια ΜΑΔ έγκειται στο οι ερευνητές να εξετάσουν το αν και κατά πόσο οι θεραπείες που σκοπεύουν να εξετάσουν μπορούν να θεωρηθούν «από κοινού τυχαιοποιήσιμες». Αυτό σημαίνει ότι θα πρέπει ο κάθε ασθενής να μπορούσε – κατ' αρχήν – να έχει τυχαιοποιηθεί να λάβει οποιαδήποτε από τις παρεμβάσεις στο δίκτυο. Αυτή είναι μια βασική παραδοχή και θα πρέπει να λαμβάνεται πάντα υπόψη κατά την δημιουργία του δικτύου.

Στη συνέχεια, οι ερευνητές καλούνται να κάνουν μια συστηματική ανασκόπηση της βιβλιογραφίας ώστε να εντοπίσουν τις υπάρχουσες κλινικές μελέτες που απαντούν στο υπό εξέταση κλινικό ερώτημα. Στην συνέχεια πρέπει να λάβει χώρα μια κριτική αξιολόγηση της ποιότητας των μελετών αυτών. Αφού καταλήξουν στις μελέτες που θα περιληφθούν στην μετά-ανάλυση, οι ερευνητές θα πρέπει να ελέγξουν αν υπάρχουν διαφορές στο πως ορίζονται οι θεραπείες ανά ζεύγος συγκρίσεων όσον αφορά την δοσολογία, τη διάρκεια, τον τρόπο χορήγησης (π.χ. χάπι/ένεση) κλπ. Η ύπαρξη συστηματικών διαφορών στον ορισμό των θεραπειών μπορεί να δημιουργήσει αμφιβολίες σχετικά με την εγκυρότητα της υπόθεσης μεταβατικότητας (*transitivity*). Αυτή η υπόθεση είναι θεμελιώδους σημασίας για την ΜΑΔ, και όταν αυτή καταστρατηγείται τα αποτελέσματα μιας ΜΑΔ θα περιέχουν εν γένει μεροληψίες.

Στη συνέχεια, οι ερευνητές πρέπει να ελέγξουν την κατανομή των εν δυνάμει *τροποποιητών επίδρασης* στις μελέτες, για να βεβαιωθούν ότι δεν υπάρχουν σημαντικές διαφορές. Τροποποιητής επίδρασης ονομάζεται μια μεταβλητή (σε επίπεδο μελέτης) η οποία επηρεάζει την σχετική επίδραση των θεραπειών που συγκρίνονται. Πολλές φορές ο εν λόγω έλεγχος μπορεί να αποδειχτεί στην πράξη δύσκολος, ειδικά όταν δεν υπάρχει η σχετική πληροφορία σε όλες τις δημοσιευμένες μελέτες ή όταν ο αριθμός των μελετών που συμβάλλουν στην ανάλυση είναι μικρός. Όλες οι σχετικές λεπτομέρειες θα πρέπει να περιγράφονται λεπτομερώς σε τυχόν δημοσιεύσεις ή αναφορές των αποτελεσμάτων της ΜΑΔ, ώστε να μπορούν οι αναγνώστες να αξιολογήσουν την εγκυρότητα των υποθέσεων της ανάλυσης.

Το επόμενο στάδιο είναι να αποφασιστεί το ακριβές μοντέλο που θα χρησιμοποιηθεί

για την εφαρμογή της ΜΑΔ. Στο Κεφαλαιο 2 αυτής της διατριβής παρουσιάσαμε ένα σύνολο διαφορετικών (αλλά παραπλήσιων) μοντέλων που μπορούν να χρησιμοποιηθούν για αυτόν τον σκοπό. Η επιλογή μεταξύ των διαφόρων μοντέλων θα πρέπει να βασιστεί κυρίως στην υπάρχουσα τεχνογνωσία της ερευνητικής ομάδας σχετικά με τα διάφορα πακέτα λογισμικού που μπορούν να χρησιμοποιηθούν για την ανάλυση. Αν υιοθετηθεί ένα Μπαεζιανό πλαίσιο, είναι σημαντικό να συζητηθεί η επιλογή των εκ των προτέρων κατανομών (*prior distributions*) που χρειάζεται να χρησιμοποιηθούν για τις παραμέτρους. Ιδιαίτερη προσοχή απαιτείται κατά την μοντελοποίηση των παραμέτρων διασποράς (όπως η ετερογένεια), δεδομένου ότι γενικά για αυτού του είδους τις παραμέτρους οι εκ των προτέρων κατανομές πάντα συνεισφέρουν πληροφορία στο μοντέλο. Αυτό συνεπάγεται ότι η εκτιμώμενη ετερογένεια μπορεί να ποικίλει ανάλογα με την επιλογή της εκ των προτέρων κατανομής, και με τον τρόπο αυτό να επηρεάσει τη συνοχή του δικτύου και την εκτιμώμενη σχετική αποτελεσματικότητα των θεραπειών. Οι αναλύσεις ευαισθησίας είναι ζωτικής σημασίας ώστε να κατανοήσει κάποιος τον ενδεχόμενο αντίκτυπο τέτοιων υποθέσεων στη διαδικασία μοντελοποίησης. Ένας ελάχιστος αριθμός αναλύσεων ευαισθησίας πρέπει πάντα να προκαθορίζεται, ιδανικά σε δημοσιευμένο πρωτόκολλο.

Αν η δομή του δικτύου το επιτρέπει, αν δηλαδή υπάρχουν στο δίκτυο κλειστοί βρόχοι, μια στατιστική εκτίμηση της ασυνέπειας (*inconsistency*) θα πρέπει να λάβει χώρα. Η στατιστική ασυνέπεια αναφέρεται σε διαφορές ανάμεσα σε εκτιμήσεις που βασίζονται στην άμεση και την έμμεση πληροφορία στο δίκτυο. Η σωστή εκτίμηση της ασυνέπειας μπορεί να είναι δύσκολη, ειδικά όταν στην βάση δεδομένων υπάρχουν μελέτες που συγκρίνουν πολλαπλές θεραπείες. Στο Κεφάλαιο 2 αυτής της διατριβής παρουσιάσαμε μια ποικιλία μεθόδων και μοντέλων που μπορούν να χρησιμοποιηθούν για τον στατιστικό έλεγχο του δικτύου σε σχέση με την συνέπεια. Παράλληλα, συζητήσαμε τα πλεονεκτήματα και τους περιορισμούς της κάθε μεθόδου. Προτείνουμε την εφαρμογή όσο το δυνατό περισσότερων μοντέλων ή μεθόδων για την διερεύνηση της συνέπειας του δικτύου. Με αυτόν τον τρόπο διευκολύνεται ο εντοπισμός πιθανών πηγών ασυνέπειας. Οι ερευνητές θα πρέπει να έχουν κατά νου, ωστόσο, ότι η απουσία στατιστικά σημαντικών ευρημάτων για την ασυνέπεια δεν σημαίνει κατ' ανάγκη ότι η υπόθεση της μεταβατικότητας ισχύει: όλες τα στατιστικά τεστ για ασυνέπεια αναμένεται να έχουν χαμηλή ισχύ, ενώ μεγάλες τιμές της ετερογένειας μπορεί να κρύψουν τυχόν ασυνέπειες στα δεδομένα.

Εάν ανιχνευθεί στατιστικά σημαντική ασυνέπεια, οι ερευνητές καλούνται να διερευνήσουν πιθανές πηγές της και να προσπαθήσουν να τις εξηγήσουν. Τοπικές μέθοδοι

για την αξιολόγηση της ασυνέπειας μπορούν να επισημάνουν ενδεχομένως προβληματικές μελέτες, οι οποίες θα πρέπει στη συνέχεια να ελεγχθούν για πιθανά σφάλματα στην εξόρυξη δεδομένων, για τυχόν σημαντικές διαφορές στην κατανομή των τροποποιητών επίδρασης ή άλλα πιθανά σφάλματα. Επίσης, παρουσιάσαμε διάφορα μοντέλα που μπορούν να χρησιμοποιηθούν για την προσαρμογή της ΜΑΔ για μεροληψία, με την προσθήκη πληροφορίας σχετικής με τους τροποποιητές επίδρασης. Τέτοια μοντέλα μπορούν να εφαρμοστούν για να εξηγήσουν – και ενδεχομένως να εξαλείψουν – τυχόν ασυνέπειες στο δίκτυο. Αν η στατιστική ασυνέπεια δεν είναι δυνατόν να εξαλειφτεί με αυτόν τον τρόπο, οι ερευνητές μπορούν να εξετάσουν την κατάτμηση κόμβων (=θεραπειών) του δικτύου, π.χ. υψηλή δόση / χαμηλή δόση. Εναλλακτικά, μπορούν να παρουσιάσουν τα αποτελέσματα από το κατάλληλο μοντέλο ασυνέπειας (μοντέλο Lu&Ades όταν όλες οι μελέτες συγκρίνουν δύο θεραπείες και μοντέλο design-by-treatment όταν υπάρχουν μελέτες που συγκρίνουν πολλαπλές θεραπείες) σε συνδυασμό με την άμεση πληροφορία που προκύπτει από κλασσικές μετά-αναλύσεις.

Παρόλο που ένα στατιστικά σημαντικό τεστ για την ασυνέπεια συνεπάγεται ότι τα αποτελέσματα μιας ΜΑΔ πιθανώς δεν είναι έγκυρα, μια τέτοια παρατήρηση μπορεί να αποτελέσει κίνητρο για την διατύπωση πρόσθετων ερευνητικών υποθέσεων σχετικά με πιθανούς τροποποιητές επίδρασης. Μπορεί να παρακινήσει περαιτέρω αναλύσεις, όπως η περίληψη ατομικών δεδομένων των ασθενών ή πληροφορίες από μελέτες παρατήρησης. Έτσι, για τους σκοπούς της διαδικασίας λήψης αποφάσεων, η ΜΑΔ μπορεί να χρησιμοποιηθεί όχι μόνο ως μέθοδος για τη σύγκριση θεραπειών, αλλά μπορεί επίσης να χρησιμεύσει ως εργαλείο για την διερεύνηση των συνθηκών που επηρεάζουν την αποτελεσματικότητα των θεραπειών σε συνηθισμένες κλινικές συνθήκες.

Ο δεύτερος στόχος της παρούσας διατριβής ήταν η ανάπτυξη στατιστικής μεθοδολογίας για την από κοινού ανάλυση πολλαπλών συσχετισμένων εκβάσεων με ΜΑΔ. Στο Κεφάλαιο 3 εισαγάγαμε ένα μοντέλο το οποίο επικεντρώθηκε στην περίπτωση της ανάλυσης πολλαπλών διχότομων εκβάσεων το οποίο μοντελοποιεί τις συσχετίσεις ανάμεσά τους. Το μοντέλο συνθέτει πληροφορίες από τυχαιοποιημένες μελέτες, στις οποίες προστίθενται εξωτερικά δεδομένα που προέρχονται από εμπειρογνώμονες κλινικούς γιατρούς. Παρουσιάσαμε τις μαθηματικές λεπτομέρειες αυτού του μοντέλου και συζητήσαμε σε βάθος τη διαδικασία εκμαίευσης πληροφοριών από τους εμπειρογνώμονες. Στο Κεφάλαιο 4 παρουσιάσαμε δύο επιπλέον μοντέλα πολλαπλών συσχετισμένων εκβάσεων με ΜΑΔ. Και τα δύο μοντέλα μπορούν να χρησιμοποιηθούν για να συνθέσουν πολλαπλές διχότομες ή

συνεχείς εκβάσεις, ή και εκβάσεις που καθορίζονται από τον χρόνο μέχρι να εμφανιστεί το υπό εξέταση συμβάν. Το πρώτο από αυτά τα μοντέλα έχει πιο πλούσια δομή. Το δεύτερο μοντέλο είναι μια προσέγγιση του πρώτου, και μπορεί να χρησιμοποιηθεί σε περιπτώσεις περιορισμένης διαθεσιμότητας δεδομένων. Ο κώδικας λογισμικού που απαιτείται για να χρησιμοποιήσει κανείς όλα τα μοντέλα καθώς και πιθανές επεκτάσεις τους συζητούνται στο Παράρτημα της παρούσας διατριβής.

Για να δώσουμε ένα παράδειγμα πρακτικής εφαρμογής των μεθόδων μας εφαρμόσαμε όλα τα μοντέλα που παρουσιάστηκαν σε αυτήν την διατριβή σε ένα δίκτυο αντιμανιακών φαρμάκων. Σε αυτό το δίκτυο συγκρίνονται 15 φαρμακολογικές θεραπείες για την οξεία μανία καθώς και το εικονικό φάρμακο (placebo), ως προς την αποτελεσματικότητα (efficacy) και την δεκτικότητα (acceptability). Βρήκαμε ότι τα μοντέλα μας παρέχουν πιο ακριβείς εκτιμήσεις για τις περισσότερες συγκρίσεις ανάμεσα στις θεραπείες, και για τις δύο εκβάσεις. Αυτή η αύξηση της ακρίβειας ήταν πιο έντονη όταν υποθέσαμε μεγαλύτερη συσχέτιση μεταξύ των εκβάσεων. Επιπλέον, έχει δειχθεί ότι η πολύ-μεταβλητή μετα-ανάλυση μπορεί να μειώσει την μεροληψία των εκτιμήσεων όταν κάποιες από τις αρχικές μελέτες έκαναν επιλεκτική αναφορά των αποτελεσμάτων (178), π.χ. αποκρύβοντας αποτελέσματα που δεν έπιασαν το όριο στατιστικής σημαντικότητας ή δεν κρίθηκαν αρκετά ενδιαφέροντα.

Παρά το γεγονός ότι τα μοντέλα μας έδειξαν να αποδίδουν καλά και μπορεί να είναι προτιμότερα από μια σειρά ανεξάρτητων, μόνο-παραγοντικών αναλύσεων, έχουν κάποια μειονεκτήματα και περιορισμούς. Η πολυπλοκότητα των αναλύσεων αυξάνει με τον αριθμό των υπό ανάλυση εκβάσεων ή όταν στις συμπεριλαμβανόμενες μελέτες υπάρχουν κάποιες που συγκρίνουν μεγάλο αριθμό θεραπειών. Επίσης, το κέρδος στην ακρίβεια από την χρήση των μοντέλων μας μπορεί να είναι μικρό αν οι συντελεστές συσχέτισης ανάμεσα στις εκβάσεις είναι κοντά στο μηδέν. Σε τέτοιες περιπτώσεις, τα πλεονεκτήματα της κοινής από κοινού μοντελοποίησης των συσχετιζόμενων εκβάσεων μπορεί να είναι πολύ λίγα για να δικαιολογήσουν την αυξημένη πολυπλοκότητα της ανάλυσης (195). Το κέρδος σε ακρίβεια επίσης εξαρτάται από το ποσοστό των μελετών που δεν παρέχουν πληροφορίες για όλες τις εκβάσεις. Υπάρχει μια κρίσιμη ισορροπία στο να έχει κανείς αρκετές μελέτες που να δίνουν πληροφορίες και για τις δύο εκβάσεις ώστε να μπορεί να εκτιμήσει με ακρίβεια τον συσχετισμό τους, αλλά και αρκετές μελέτες που να μην παρέχουν πληροφορία για κάποια (ή κάποιες) από τις εκβάσεις, προκειμένου να επωφεληθεί κανείς από την αύξηση της ακρίβειας που μπορεί να προκύψει από το μοντέλο. Μια μελλοντική μελέτη προσομοίωσης

μπορεί να βοηθήσει στο να ερευνηθούν τα πιθανά οφέλη σε σχέση με την ακρίβεια των εκτιμήσεων από τα μοντέλα μας, για διαφορετικές τιμές των συντελεστών συσχέτισης, για διαφορετικούς αριθμούς μελετών και για διαφορετικά σενάρια επιλεκτικής αναφοράς των αποτελέσματα στις πρωτότυπες μελέτες, προκειμένου να εντοπιστούν οι περιπτώσεις κατά τις οποίες η χρήση πιο εξελιγμένων μοντέλων θα οδηγήσει σε πολύ πιο ακριβείς εκτιμήσεις.

Ένας περιορισμός των μοντέλων μας είναι ότι κάνουν την υπόθεση ότι δεν υπάρχουν ελλείπουσες τιμές στις μελέτες. Στο παράδειγμα της οξείας μανίας, δεδομένα για την αποτελεσματικότητα της θεραπείας δεν ήταν διαθέσιμα για ένα μικρό ποσοστό των ασθενών (λιγότερο από 10%). Σε αυτές τις περιπτώσεις, οι ελλείπουσες τιμές είχαν καταλογιστεί ως αποτυχίες. Αυτή η στρατηγική έχει δειχθεί ότι δεν επηρεάζει σημαντικά το αποτέλεσμα της ΜΑΔ, όταν το συνολικό ποσοστό ελλειπουσών τιμών είναι μικρό (134). Στην περίπτωση που υπάρχει μεγάλο ποσοστό ελλειπουσών τιμών για πολλαπλές συσχετισμένες εκβάσεις τότε ορισμένες μέθοδοι στρατηγική αντιμετώπισης των ελλειπουσών τιμών μπορεί να οδηγήσει σε σφάλματα. Για παράδειγμα, αν η ίδια στρατηγική να καταλογιστούν τα δεδομένα που λείπουν ως αποτυχίες ακολουθηθεί και για δύο εκβάσεις αποτελεσματικότητας, τότε αυτό αναμένεται να αυξήσει ψευδώς τις εκτιμώμενες συσχετίσεις ανάμεσά τους.

Γενικά, και με βάση τα ευρήματά της έρευνάς μας, συνιστούμε στους ερευνητές να χρησιμοποιούν όταν είναι δυνατόν και την μονοπαραγοντική και πολυπαραγοντική προσέγγιση, ώστε να διαπιστωθεί εάν τα κλινικά συμπεράσματα σχετικά με την κατάταξη των θεραπειών για κάθε έκβαση παραμένουν συνεπή κάτω από διαφορετικές επιλογές μοντέλου.

Δεδομένου ότι η ΜΑΔ πολλαπλών εκβάσεων αποτελεί μια νέα, σε μεγάλο βαθμό ανεξερεύνητη περιοχή, εξακολουθούν να υπάρχουν πολλά θέματα ανοικτά προς περεταίρω έρευνα. Μια πιθανή επέκταση των μοντέλων που παρουσιάστηκαν θα ήταν να συμπεριληφθούν προσωπικά δεδομένα στο επίπεδο των ασθενών, είτε αποκλειστικά είτε σε συνδυασμό με συγκεντρωτικά δεδομένα ανά μελέτη. Επιπλέον, θα ήταν σημαντικό για τα μοντέλα μας να εφαρμοστούν σε δημοφιλή λογισμικά ώστε να γίνει η ΜΑΔ πολλαπλών εκβάσεων πιο προσιτή στους ερευνητές.

Εν τέλει, πιστεύουμε ότι η έρευνα που παρουσιάστηκε στην παρούσα διατριβή είναι μια σημαντική εξέλιξη στον τομέα της ΜΑΔ. Πιστεύουμε επίσης ότι τα μοντέλα που παρουσιάστηκαν αποτελούν επί του παρόντος την καλύτερη διαθέσιμη μέθοδο για την μετα-ανάλυση δικτύου παρουσία πολλαπλών συσχετισμένων εκβάσεων.

Appendix

I. The acute mania dataset

In this Section we provide the data used in the analyses of Chapters 3 and 4. The data originates from a network of treatments for acute mania (206). It comprises 65 studies (47 two-arm and 18 three-arm) reporting on response to the treatment and dropout. Eighteen of the included studies did not provide data for response while one did not report data on dropout.

Table 7: The acute mania dataset

			Response			**Dropout**		
Study ID	**Treatment 1**	**Treatment 2**	**OR**	**95% C.I.** (lower)	**95% C.I.** (upper)	**OR**	**95% C.I.** (lower)	**95% C.I.** (upper)
1	Aripiprazole	Placebo	0.59	0.38	0.90	0.66	0.38	1.17
2	Aripiprazole	Placebo	0.41	0.25	0.67	1.12	0.70	1.81
3	Aripiprazole	Haloperidol	0.70	0.46	1.06	2.65	1.67	4.20
4	Aripiprazole	Placebo	0.35	0.20	0.62	2.64	1.53	4.55
5	Aripiprazole	Placebo	0.82	0.54	1.26	1.04	0.68	1.58
6	Quetiapine	Placebo	0.44	0.24	0.79	1.67	0.94	2.96
7	Quetiapine	Placebo	0.67	0.39	1.15	1.41	0.80	2.47
8	Quetiapine	Lithium	0.45	0.22	0.89	3.53	1.21	10.27
9	Quetiapine	Placebo	0.44	0.28	0.69	0.98	0.60	1.60
10	Ziprasidone	Placebo	0.56	0.31	1.03	1.41	0.79	2.51
11	Ziprasidone	Placebo	0.49	0.26	0.93	1.29	0.71	2.33
12	Ziprasidone	Placebo	0.91	0.52	1.57	0.86	0.47	1.56
13	Ziprasidone	Olanzapine	-	-	-	3.00	0.48	18.93
14	Ziprasidone	Placebo	0.96	0.69	1.33	0.61	0.41	0.91
15	Olanzapine	Lithium	-	-	-	3.50	0.32	38.23
16	Olanzapine	Placebo	0.32	0.15	0.66	2.99	1.50	5.96
17	Olanzapine	Placebo	0.38	0.18	0.81	2.27	1.07	4.79
18	Olanzapine	Divalproex	0.59	0.36	0.97	1.23	0.72	2.07
19	Olanzapine	Placebo	0.43	0.27	0.68	1.07	0.65	1.75
20	Olanzapine	Haloperidol	1.04	0.69	1.56	1.35	0.91	2.00
21	Olanzapine	Placebo	1.05	0.50	2.24	1.23	0.55	2.75
22	Risperidone	Placebo	0.21	0.13	0.35	3.43	1.82	6.43
23	Risperidone	Placebo	0.57	0.30	1.09	1.98	1.03	3.79
24	Risperidone	Placebo	0.43	0.25	0.74	1.78	1.09	2.92
25	Risperidone	Olanzapine	1.20	0.78	1.86	0.55	0.33	0.90
26	Divalproex	Lithium	6.67	0.66	67.46	-	-	-
27	Divalproex	Placebo	0.56	0.37	0.84	1.27	0.85	1.91
28	Divalproex	Placebo	0.12	0.02	0.67	0.86	0.25	2.98
29	Carbamazepine	Divalproex	2.41	0.52	11.10	1.00	0.17	5.98
30	Divalproex	Placebo	0.38	0.19	0.76	1.74	0.63	4.80
31	Haloperidol	Carbamazepine	-	-	-	1.07	0.06	18.62
32	Carbamazepine	Placebo	0.32	0.22	0.48	1.40	0.96	2.03
33	Lithium	Lamotrigine	0.76	0.18	3.24	0.62	0.09	4.34
34	Placebo	Topiramate	-	-	-	1.76	1.05	2.95
35	Placebo	Topiramate	-	-	-	2.19	1.23	3.89
36	Lithium	Olanzapine	2.44	1.01	5.85	0.36	0.13	0.98

37	Placebo	Paliperidone	1.14	0.75	1.73	0.88	0.58	1.35
38	Haloperidol	Carbamazepine	0.80	0.12	5.40	0.10	0.01	0.90
39	Haloperidol	Lithium	-	-	-	3.33	0.36	30.70
40	Haloperidol	Carbamazepine	-	-	-	6.00	0.53	67.65
41	Lithium	Carbamazepine	-	-	-	0.20	0.02	1.94
42	Olanzapine	Lithium	1.89	0.38	9.27	0.63	0.09	4.24
43	Divalproex	Placebo	-	-	-	0.94	0.46	1.93
44	Topiramate	Placebo	1.29	0.72	2.29	0.47	0.23	0.96
45	Gabapentin	Placebo	-	-	-	0.57	0.27	1.20
46	Lithium	Carbamazepine	-	-	-	2.16	0.71	6.57
47	Olanzapine	Placebo	0.64	0.36	1.16	0.92	0.53	1.61
48	Aripiprazole	Lithium	0.92	0.59	1.43	0.94	0.60	1.46
48	Aripiprazole	Placebo	0.59	0.38	0.93	0.99	0.64	1.54
48	Lithium	Placebo	0.64	0.41	1.01	1.06	0.69	1.64
49	Aripiprazole	Haloperidol	1.07	0.70	1.65	1.12	0.68	1.83
49	Aripiprazole	Placebo	0.70	0.45	1.09	1.24	0.75	2.04
49	Haloperidol	Placebo	0.65	0.41	1.01	1.11	0.68	1.81
50	Quetiapine	Lithium	0.99	0.57	1.72	1.62	0.68	3.83
50	Quetiapine	Placebo	0.32	0.18	0.58	4.34	1.99	9.48
50	Lithium	Placebo	0.32	0.18	0.59	2.69	1.32	5.47
51	Quetiapine	Haloperidol	1.72	0.98	3.00	0.52	0.28	0.98
51	Quetiapine	Placebo	0.73	0.41	1.28	1.20	0.68	2.12
51	Haloperidol	Placebo	0.42	0.24	0.75	2.30	1.24	4.26
52	Ziprasidone	Haloperidol	2.05	1.33	3.14	0.84	0.55	1.28
52	Ziprasidone	Placebo	0.45	0.25	0.82	1.75	1.01	3.04
52	Haloperidol	Placebo	0.22	0.12	0.40	2.09	1.20	3.63
53	Olanzapine	Divalproex	0.97	0.65	1.44	0.94	0.60	1.46
53	Olanzapine	Placebo	0.68	0.41	1.12	1.03	0.61	1.75
53	Divalproex	Placebo	0.70	0.42	1.17	1.10	0.64	1.88
54	Risperidone	Haloperidol	-	-	-	2.12	0.96	4.64
54	Risperidone	Placebo	-	-	-	1.82	0.82	4.01
54	Haloperidol	Placebo	-	-	-	0.86	0.40	1.85
55	Risperidone	Haloperidol	-	-	-	1.63	0.23	11.46
55	Risperidone	Lithium	-	-	-	0.46	0.04	5.75
55	Haloperidol	Lithium	-	-	-	0.29	0.03	3.12
56	Risperidone	Haloperidol	0.95	0.60	1.51	0.87	0.41	1.83
56	Risperidone	Placebo	0.53	0.32	0.86	1.42	0.72	2.82
56	Haloperidol	Placebo	0.56	0.34	0.91	1.64	0.80	3.37
57	Asenapine	Olanzapine	1.46	0.97	2.18	0.42	0.27	0.67
57	Asenapine	Placebo	0.49	0.29	0.83	1.09	0.67	1.77
57	Olanzapine	Placebo	0.34	0.20	0.57	2.56	1.51	4.35
58	Asenapine	Olanzapine	-	-	-	0.56	0.35	0.87
58	Asenapine	Placebo	-	-	-	1.46	0.88	2.42
58	Olanzapine	Placebo	-	-	-	2.63	1.56	4.43
59	Divalproex	Lithium	0.97	0.43	2.17	1.71	0.76	3.89
59	Divalproex	Placebo	0.31	0.15	0.64	1.90	0.97	3.71
59	Lithium	Placebo	0.32	0.14	0.75	1.11	0.49	2.52
60	Lamotrigine	Placebo	-	-	-	0.92	0.50	1.69
60	Lamotrigine	Lithium	-	-	-	2.07	0.94	4.56
60	Placebo	Lithium	-	-	-	2.24	1.03	4.89
61	Lamotrigine	Placebo	-	-	-	0.67	0.35	1.28
61	Lamotrigine	Lithium	-	-	-	0.37	0.19	0.72
61	Placebo	Lithium	-	-	-	0.55	0.28	1.08
62	Placebo	Topiramate	-	-	-	1.21	0.73	2.02

62	Placebo	Lithium	-	-	-	0.98	0.54	1.78
62	Topiramate	Lithium	-	-	-	0.81	0.48	1.34
63	Placebo	Topiramate	-	-	-	1.04	0.48	2.27
63	Placebo	Lithium	-	-	-	1.49	0.71	3.12
63	Topiramate	Lithium	-	-	-	1.43	0.69	2.96
64	Paliperidone	Quetiapine	0.80	0.54	1.19	1.05	0.64	1.71
64	Paliperidone	Placebo	0.44	0.27	0.72	2.48	1.47	4.19
64	Quetiapine	Placebo	0.55	0.34	0.90	2.38	1.41	4.00
65	Olanzapine	Haloperidol	1.82	0.67	4.93	3.27	1.22	8.76
65	Olanzapine	Placebo	0.75	0.43	1.31	1.97	1.11	3.49
65	Haloperidol	Placebo	0.41	0.15	1.13	0.60	0.23	1.60

II. Equivalence between different formulas for estimating the correlation of two log odds ratios.

In this section of the appendix we prove the equivalence between different formulas for estimating the correlation of two log odds ratios: equation (8) in Wei and Higgins (210), equation (10) in this dissertation and equation (4) in Bagos (208).

First we show the equivalence between equation (8) in (210) and equation (10) in this thesis. The covariance between two log odds ratios for a study comparing treatments A and B for outcomes R and D,following Wei and Higgins is (after dropping the study index):

$$cov(lnOR_R, lnOR_D) = \frac{\rho_W m_{A,RD}}{\sqrt{m_{A,R} m_{A,D}}} \sqrt{\frac{1}{e_{A,R}} + \frac{1}{f_{A,R}}} \sqrt{\frac{1}{e_{A,D}} + \frac{1}{f_{A,D}}} + \frac{\rho_W m_{B,RD}}{\sqrt{m_{B,R} m_{B,D}}} \sqrt{\frac{1}{e_{B,R}} + \frac{1}{f_{B,R}}} \sqrt{\frac{1}{e_{B,D}} + \frac{1}{f_{B,D}}} \quad (23)$$

In this equation:

- OR_R, OR_D are the log odds ratios of the comparison AB for outcomes R, D.
- $m_{A,R}$, $m_{A,D}$, $m_{A,RD}$ are the number of patients that reported on outcome R, D or both in group A; similarly for B.
- $e_{A,R}$, $f_{A,R}$ are the number of successes and failures for outcome R, arm A. Similarly for outcome D and treatment B.
- ρ_W is the correlation coefficient between the two outcomes.

Wei and Higgins used a fixed correlation coefficient independent of the treatment arm. Alternatively, correlation can be treatment-specific. Let us consider the data of Table 8. The correlation coefficient between the two binary outcomes R and D in arm A can be estimated as (217):

$$\hat{\rho}_{W,A} = \frac{n_{11} n_{00} - n_{01} n_{10}}{\sqrt{e_{A,R} f_{A,R} e_{A,D} f_{A,D}}} \quad (24)$$

Table 8: Full cross-classified table for a study reporting on treatment A for response (R) and dropout (D).

Treatment A	D^+	D^-	Total
R^+	n_{11}	n_{10}	$e_{A,R}$
R^-	n_{01}	n_{00}	$f_{A,R}$
Total	$e_{A,D}$	$f_{A,D}$	N

Following the methods for reconstructing the full cross tables when only the collapsed information is available (presented in Section 3.3.2 of this dissertation), we use the φ and ζ parameters to rewrite this coefficient as follows:

$$\hat{\rho}_{W,A} = \frac{e_{A,R}\hat{\varphi}_A\left(1-\hat{\zeta}_A\right)f_{A,R} - \left(1-\hat{\zeta}_A\right)e_{A,R}f_{A,R}\hat{\zeta}_A}{\sqrt{e_{A,R}f_{A,R}e_{A,D}f_{A,D}}} = \frac{e_{A,R}f_{A,R}(\hat{\varphi}_A - \hat{\zeta}_A)}{\sqrt{e_{A,R}f_{A,R}e_{A,D}f_{A,D}}} \tag{25}$$

If we substitute $\hat{\zeta}_A = \frac{1}{f_{A,R}}(e_{A,D} - e_{A,R}\hat{\varphi}_A)$ we get:

$$\hat{\rho}_{W,A} = \frac{e_{A,R}\hat{\varphi}_A f_{A,R} - e_{A,R}e_{A,D} + \hat{\varphi}_A e_{A,D}^2}{\sqrt{e_{A,R}f_{A,R}e_{A,D}f_{A,D}}} \tag{26}$$

We can now use Equation (23) to compute the covariance of the log odd ratios. In case that all patients report on both outcomes we have $m_{T,R} = m_{T,D} = m_{T,RD} = N_T$ for each treatment T. Thus, we get:

$$cov(lnOR_R, lnOR_D)$$
$$= \sum_{T=A,B} \frac{e_{T,R}\hat{\varphi}_T f_{T,R} - e_{T,R}e_{T,D} + \hat{\varphi}_T e_{T,D}^2}{\sqrt{e_{T,R}f_{T,R}e_{T,D}f_{T,D}}} \sqrt{\frac{1}{e_{T,R}} + \frac{1}{f_{T,R}}}\sqrt{\frac{1}{e_{T,D}} + \frac{1}{f_{T,D}}} \tag{27}$$

After some algebra, and substituting $e_{T,R} + f_{T,R} = e_{T,D} + f_{T,D} = N_T$ we get:

$$\hat{\rho}_\iota = \frac{1}{\hat{\sigma}_{i,R}\hat{\sigma}_{i,D}} \sum_{T=A,B} \frac{\hat{\varphi}_{i,T}\left(e_{i,T,R} + f_{i,T,R}\right)^2 - e_{i,T,D}(e_{i,T,R} + f_{i,T,R})}{e_{i,T,D}f_{i,T,R}f_{i,T,D}} \tag{28}$$

This is Equation (10) in our thesis.

We now show the equivalence between formula (8) of Wei and Higgins and formula (4) in Bagos. Using the notation of (208) for a study comparing treatment A with treatment B, we denote $e_{A,R} = n_{11+}$, $f_{A,D} = n_{1+0}$, $e_{B,R} = n_{01+}$, $f_{B,D} = n_{0+0}$ etc. In this notation Equation (23) by Wei and Higgins, after using Equation (24) for the correlation coefficient between the two outcomes for each treatment, can be written as follows:

$$cov(lnOR_R, lnOR_D) = \rho_{W,A}\sqrt{\frac{1}{e_{A,R}}+\frac{1}{f_{A,R}}}\sqrt{\frac{1}{e_{A,D}}+\frac{1}{f_{A,D}}} + \rho_{W,B}\sqrt{\frac{1}{e_{B,R}}+\frac{1}{f_{B,R}}}\sqrt{\frac{1}{e_{B,D}}+\frac{1}{f_{B,D}}} =$$

$$\sum_{i=0,1}\rho_{W,i}\sqrt{\frac{1}{n_{i1+}}+\frac{1}{n_{i0+}}}\sqrt{\frac{1}{n_{i+1}}+\frac{1}{n_{i+0}}} = \sum_{i=0,1}\frac{n_{i11}n_{i00}-n_{i01}n_{i10}}{\sqrt{n_{i1+}n_{i+1}n_{i0+}n_{i+0}}}\sqrt{\frac{N_i}{n_{i1+}n_{i0+}}}\sqrt{\frac{N_i}{n_{i+1}n_{i+0}}} =$$

$$\sum_{i=0,1}\frac{n_{i11}n_{i00}-n_{i01}n_{i10}}{n_{i1+}n_{i+1}n_{i0+}n_{i+0}}N_i = \sum_{i=0,1}\frac{n_{i11}n_{i00}-n_{i01}n_{i10}}{n_{i1+}n_{i+1}n_{i0+}n_{i+0}}(n_{i11}+n_{i00}+n_{i01}+n_{i10})$$

After some algebra this can be shown to be equivalent to Equation (4) of the paper by Bagos:

$$Cov(lnOR_R, lnOR_D) = \sum_{i=A,B}\sum_{j=0,1}\sum_{k=0,1}(-1)^{j-k}\left(\frac{n_{ijk}}{n_{ij+}n_{i+k}}\right) =$$

$$\sum_{i=A,B}\left(\frac{n_{i11}}{n_{i1+}n_{i+1}}-\frac{n_{i10}}{n_{i1+}n_{i+0}}-\frac{n_{i01}}{n_{i0+}n_{i+1}}+\frac{n_{i00}}{n_{i0+}n_{i+0}}\right)$$

Thus, we conclude that the three different formulas for the covariances between two log odd ratios (formula (10) in this thesis, formula 4 in Bagos (208) and formula 8 from Wei and Higgins (210)) are equivalent.

III. The variance-covariance matrix for heterogeneity

As we discuss in 3.3 of this thesis, for a three-arm study i, comparing treatments A, B and C the (4×4) the variance-covariance matrix is assumed to have the following structure:

$$\boldsymbol{\Delta}_{(4\times4)} = \begin{pmatrix} \tau_R^2 & \rho_\tau \tau_R \tau_D & \tau_R^2/2 & \chi_1 \tau_R \tau_D \\ \rho_\tau \tau_R \tau_D & \tau_D^2 & \chi_2 \tau_R \tau_D & \tau_D^2/2 \\ \tau_R^2/2 & \chi_2 \tau_R \tau_D & \tau_R^2 & \rho_\tau \tau_R \tau_D \\ \chi_1 \tau_R \tau_D & \tau_D^2/2 & \rho_\tau \tau_R \tau_D & \tau_D^2 \end{pmatrix} \tag{29}$$

We can now make the extra assumption that $\chi_1 = \chi_2 = \chi$. Let us pick a pair of the random effects $\delta_{i,AB,R}$, $\delta_{i,AC,R}$ and compute the variance of their difference:

$$\begin{aligned} Var\left(\delta_{i,AC,D} - \delta_{i,AB,R}\right) = Var\left(\delta_{i,AC,D}\right) + Var\left(\delta_{i,AB,R}\right) - 2Cov\left(\delta_{i,AC,D}, \delta_{i,AB,R}\right) = \\ = \tau_R^2 + \tau_D^2 - 2\,\chi\,\tau_R\,\tau_D \end{aligned} \tag{30}$$

The consistency assumption implies that:

$$\delta_{i,AC,D} - \delta_{i,AB,R} = \left(\delta_{i,BC,D} - \delta_{i,BA,D}\right) - \delta_{i,AB,R} = \delta_{i,BC,D} - \delta_{i,BA,D} + \delta_{i,BA,R}$$

and by taking the variances we obtain:

$$\begin{aligned} &Var\left(\delta_{i,AC,D} - \delta_{i,AB,R}\right) \\ &\quad = Var\left(\delta_{i,BC,D}\right) + Var\left(\delta_{i,BA,D}\right) - 2Cov\left(\delta_{i,BC,D}, \delta_{i,BA,D}\right) \\ &\quad + 2Cov\left(\delta_{i,BC,D}, \delta_{i,BA,R}\right) - 2Cov\left(\delta_{i,BA,D}, \delta_{i,BA,R}\right) \end{aligned}$$

After using Equation (30)we get:

$$\begin{aligned} \tau_R^2 + \tau_D^2 - 2\chi\tau_R\tau_D = \tau_R^2 + \tau_D^2 + \tau_R^2 - 2\frac{\tau_D^2}{2} + 2\chi\tau_R\tau_D - 2\rho_\tau\tau_R\tau_D \rightarrow \\ \chi = \frac{1}{2}\rho_\tau \end{aligned} \tag{31}$$

So the simplified form of the variance-covariance matrix $\Delta_{(4\times4)}$ is:

$$\boldsymbol{\Delta}_{(4\times4)} = \begin{pmatrix} \tau_R^2 & \rho_\tau \tau_R \tau_D & \tau_R^2/2 & \chi_1 \tau_R \tau_D \\ \rho_\tau \tau_R \tau_D & \tau_D^2 & \chi_2 \tau_R \tau_D & \tau_D^2/2 \\ \tau_R^2/2 & \chi_2 \tau_R \tau_D & \tau_R^2 & \rho_\tau \tau_R \tau_D \\ \chi_1 \tau_R \tau_D & \tau_D^2/2 & \rho_\tau \tau_R \tau_D & \tau_D^2 \end{pmatrix} \tag{32}$$

This can be conveniently decomposed as follows:

$$\boldsymbol{\Delta}_{(4\times4)} = \tau_R^2 \begin{pmatrix} 1 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \tau_D^2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1 \end{pmatrix} + \rho_\tau \tau_R \tau_D \begin{pmatrix} 0 & 1 & 0 & 1/2 \\ 1 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1 \\ 1/2 & 0 & 1 & 0 \end{pmatrix} \tag{33}$$

$$\boldsymbol{\Delta}_{(4\times4)} = \tau_R^2 \boldsymbol{\Delta_1} + \tau_D^2 \boldsymbol{\Delta_2} + \rho_\tau \tau_R \tau_D \boldsymbol{\Delta_3} \tag{34}$$

If we choose not to employ the $\chi_1 = \chi_2 = \chi$ assumption, then instead of (31) we get that $\chi_1 + \chi_2 = \rho^\tau$ and Equation (34) can be expressed in terms of two correlation parameters.

Note that the analysis in this section holds for all type of outcomes. R and D may be binary (in which case we analyze the log odds ratios, log risk ratios or log hazard ratios), continuous (and we can use mean difference or standardized mean difference) or a mixture of binary and continuous, e.g. R can be binary and D continuous.

IV. Computing within-study correlations from a full cross table

Consider the case of a three-arm study i for which we have the full cross-classified information presented in Table 9.

Table 9: Full cross table for a study i comparing treatments A,B and C for outcomes R and D.

Treatment A	$\boldsymbol{R^+}$	$\boldsymbol{R^-}$	**TOTAL**
$\boldsymbol{D^+}$	$n_{i,A(11)}$	$n_{i,A(01)}$	$n_{i,A(+1)}$
$\boldsymbol{D^-}$	$n_{i,A(10)}$	$n_{i,A(00)}$	$n_{i,A(+0)}$
TOTAL	$n_{i,A(1+)}$	$n_{i,A(0+)}$	$N_{i,A}$
Treatment B	$\boldsymbol{R^+}$	$\boldsymbol{R^-}$	**TOTAL**
$\boldsymbol{D^+}$	$n_{i,B(11)}$	$n_{i,B(01)}$	$n_{i,B(+1)}$
$\boldsymbol{D^-}$	$n_{i,B(10)}$	$n_{i,B(00)}$	$n_{i,B(+0)}$
TOTAL	$n_{i,B(1+)}$	$n_{i,B(0+)}$	$N_{i,B}$
Treatment C	$\boldsymbol{R^+}$	$\boldsymbol{R^-}$	**TOTAL**
$\boldsymbol{D^+}$	$n_{i,C(11)}$	$n_{i,C(01)}$	$n_{i,C(+1)}$
$\boldsymbol{D^-}$	$n_{iC(10)}$	$n_{i,C(00)}$	$n_{i,C(+0)}$
TOTAL	$n_{i,C(1+)}$	$n_{i,C(0+)}$	$N_{i,c}$

For the margins we have used the notation of Bagos (208), so that a plus sign in an index denotes a sum, e.g. $n_{i,B(+0)} = n_{i,B(00)} + n_{i,B(10)}$, $n_{i,A(0+)} = n_{i,A(00)} + n_{i,B(01)}$ etc; this notation significantly simplifies the expressions for the correlation coefficient.

In that paper it was shown that for the covariance of two log odds ratios of the same comparison it holds:

$$Cov\left(y_{AB_R}, y_{AB_D}\right) = \sum_{i=A,B} \sum_{j=0,1} \sum_{k=0,1} (-1)^{j-k} \left(\frac{n_{ijk}}{n_{ij+} n_{i+k}}\right)$$

So that the correlation coefficient is estimated as:

$$\rho_{i,AB_R,AB_D} = \frac{1}{\hat{\sigma}_{i,R}\hat{\sigma}_{i,D}} \sum_{i=A,B} \sum_{j=0,1} \sum_{k=0,1} (-1)^{j-k} \left(\frac{n_{ijk}}{n_{ij+} n_{i+k}} \right) \quad (35)$$

Following the same methods we show how to compute the covariance of different outcomes of different comparisons. Let us focus for example in the covariance of the log odds ratio of AB comparison for outcome R and BC comparison for outcome D. After dropping the study index i and setting $cov(\ln n_{Tjk}, \ln n_{T'lm}) = 0, \forall\, T \neq T'$ we obtain

$$cov(y_{AB,R}, y_{BC,D}) = cov(\ln n_{B0+}, \ln n_{B+1}) - cov(\ln n_{B0+}, \ln n_{B+0}) \\ -cov(\ln n_{B1+}, \ln n_{B+1}) + cov(\ln n_{B1+}, \ln n_{B+0})$$

and consequently:

$$Cov(y_{AB_R}, y_{AB_D}) = \sum_{j=0,1} \sum_{k=0,1} (-1)^{j-k} \left(\frac{n_{Bjk}}{n_{Bj+} n_{B+k}} \right)$$

Note that this method implies $cov(y_{AB,R}, y_{AC,D}) = cov(y_{AC,R}, y_{AB,D})$, etc. From this formula, after restoring the study index i, the correlation coefficient is easily computed as:

$$\rho_{i,AB_R,AB_D} = \frac{1}{\hat{\sigma}_{i,R}\hat{\sigma}_{i,D}} \sum_{j=0,1} \sum_{k=0,1} (-1)^{j-k} \left(\frac{n_{Bjk}}{n_{Bj+} n_{B+k}} \right) \quad (36)$$

When only the collapsed tables are available we can reconstruct the full cross tables in the way described in Section 3.3.2.1 of this thesis. With the full cross tables at our disposal we can use Equations (35) and (36) to estimate all the correlation coefficients needed.

Switching to the notation used in this dissertation, i.e. $e_{T,R} = n_{T(1+)}$, $f_{T,R} = n_{T(0+)}$, $e_{T,D} = n_{T(+1)}$ and $f_{T,D} = n_{T(+0)}$ and after some algebra we get:

$$\hat{\rho}_{i,AB_RAB_D} = \frac{1}{\hat{\sigma}_{i,AB,R}\hat{\sigma}_{i,AB,D}} \sum_{T=A,B} \frac{\hat{\varphi}_{i,T}\left(e_{i,T,R} + f_{i,T,R}\right)^2 - e_{i,T,D}(e_{i,T,R} + f_{i,T,R})}{e_{i,T,D} f_{i,T,R} f_{i,T,D}} \quad (37)$$

and

$$\hat{\rho}_{i,AB_RAB_D} = \frac{1}{\hat{\sigma}_{i,AB,R}\hat{\sigma}_{i,AC,D}} \frac{\hat{\varphi}_{i,A}\left(e_{i,A,R} + f_{i,A,R}\right)^2 - e_{i,A,D}(e_{i,A,R} + f_{i,A,R})}{e_{i,A,D} f_{i,A,R} f_{i,A,D}} \quad (38)$$

V. Eliciting prior distribution for the φ parameters from the experts

In this section we present the details of the elicitation process for the prior distributions for the φ parameters, using the method described in Section 3.3.2.2 of this thesis. The following program in R can be used to construct a beta distribution for the φ parameter based on expert opinions for the 95% confidence interval of φ:

```
betaparameters<-function(CIb,CIu){
ff=c(rep(0,1000000))
aaa=matrix(ff,1000)
for (i in 1:1000)
{a=i*0.1
for (k in 1:1000)
{b=k*0.1
aaa[i,k]=(abs(pbeta(CIb,a,b)-0.025)+abs(pbeta(CIu,a,b)-0.975))
}}
a1=which.min(aaa)
parameters=c(0,0)
parameters[1]=(a1 %% 1000)*0.1
parameters[2]=(a1%/%1000)*0.1
return(parameters)}
```

The inputs of this routine are CIu and Cib, the upper and lower limits of the 95% confidence interval. The outputs are a and b, the parameters of the beta distribution.

For the application of our methods to the acute mania example, we assigned a weight to each expert according to the years of his/her experience plus the number of randomized control trials he/she has participated in. For each expert k the two parameters α_k, β_k of the beta distribution were computed for each treatment that he/she gave information about. These parameters were then combined into a weighted average for α and β. Details are presented in Table 10.

Table 10: Parameters of the individual prior distribution for each expert and their weighted average

		Aripiprazole		**Placebo**		**Lithium**		**Haloperidol**		**Quetiapine**		**Ziprasidone**		**Olanzapine**	
ID	**weight**	***α***	***β***	***α***	***β***	***α***	***β***	***α***	***β***	***α***	***β***	***α***	***β***	***α***	***β***
1	0.08	16.8	52.3	23.1	55.1	17	52.3	19.2	53.5	16.8	52.3	23.1	55.1	14.2	49.8
2	0.04	6.7	22	12	116	6.6	13.1	6.6	13.1	10.9	46.4	10.9	46.4	10.9	46.4
3	0.06	8.3	12.9	10.1	17.4	14	15.6	15.7	14	14.1	15.6	15.6	23.8	12.4	15.3
4	0.10	4.8	12.5	11.3	11.3	8.3	12.9	8.3	12.9	6.6	13.1	6.6	13.1	3	11
5	0.08	10.9	46.4	8.3	12.9	23	55.1	36	54.4	25.5	90.1	47.6	99.8	71.5	215
6	0.02	23.1	55.1	4.8	12.5	11	46.4	4.8	12.5	10.9	46.4	–	–	23.1	55.1
7	0.04	3.9	18.6	4.9	19.6	9.6	23.6	4.8	12.5	10.9	46.4	–	–	10.9	46.4
8	0.01	4.8	12.5	10.9	46.4	4.5	4.5	8.3	12.9	8.3	12.9	–	–	8.3	12.9
9	0.02	0.9	7.7	5.3	3	0.4	7.98	0.68	10.6	3.9	18.6	4.1	36.2	0.42	7.98
10	0.03	12	116	12	116	12	116	10.9	46.4	27.8	161.9	12	115.6	27.8	162
11	0.11	11.3	11.3	6.9	6.9	11	11.3	6.9	6.8	9.9	12.2	6	7.5	4.9	7.8
12	0.07	4.8	12.5	10.9	46.4	11	46.4	6.6	13.1	4.8	12.5	4.8	12.5	4.8	12.5
13	0.15	3.4	5.1	2.2	1.5	23	55.1	11.3	11.3	3.1	5.2	8.3	12.9	96.3	225
14	0.03	18.3	22.5	23.1	55.1	4.8	12.5	11.3	11.3	23.1	55.1	11.3	11.2	23.1	55.1
15	0.04	13.2	32	9.8	71	345	982	46.9	87.8	35.4	176.4	42.2	143.4	80.4	157
16	0.05	24.5	73.3	79.5	362	57	85.6	45.4	49.2	29.8	76.1	54.6	131.7	42.4	43.2
17	0.04	11.2	27.4	9.8	71	8.3	12.9	5.5	8.7	13.3	27.9	11.6	31.3	21.8	31.7
18	0.04	31.6	43.9	6.9	34.8	8.7	9.4	15.7	11.5	24.3	45.9	6.9	11.8	15.6	23.8
19	0.02	42.1	51.6	10.9	46.4	7.6	5.9	13	8.2	9.9	26.9	14.4	24	11.9	27.6
Weighted average		**11.2**	**27.4**	**12.9**	**47.4**	**28.3**	**72.0**	**15.2**	**23.4**	**13.7**	**40.9**	**16.7**	**40.6**	**32.1**	**78.3**

		Lamotrigine		Divalproex		Risperidone		Asenapine		Carbamazepine		Topiramate		Gabapentin	
ID	**weight**	***α***	***β***	***α***	***β***	***α***	***β***	***α***	***β***	***α***	***β***	***α***	***β***	***α***	***β***
1	0.08	23.1	55.1	16.8	52	19.2	53.5	16.8	52.3	19.2	53.5	25.7	55.6	25.7	55.6
2	0.04	–	–	6.7	22	6.7	22	10.9	46.4	6.7	22	–	–	–	–
3	0.06	15.6	35	16.1	24	17.7	22.7	16.6	27.6	14.8	32.5	20.9	23.1	15.5	19.9
4	0.10	9.9	12.2	9.9	12	4.8	12.5	8.3	12.9	8.3	12.9	8.3	12.9	9.9	12.2
5	0.08	23.1	55.1	71.5	215	36.3	97.2	23.1	55.1	12.6	24.2	47.3	47.2	47.3	47.2
6	0.02	4.8	12.5	–	–	–	–	–	–	–	–	–	–	–	–
7	0.04	3.8	7.9	12.3	34	4.2	9.8	–	–	12.6	24.2	–	–	–	–
8	0.01	–	–	4.8	13	8.3	12.9	–	–	8.3	12.9	–	–	–	–
9	0.02	8.3	12.9	0.42	8	0.42	7.95	2.6	22	3.9	18.6	6.7	22	6.7	22
10	0.03	12	116	27.8	117	27.8	161.9	12	115.6	12	115.6	12	115.6	12	115.6
11	0.11	55.2	23	8.3	13	11.3	11.3	6.9	6.8	6.9	6.8	23.7	9.5	11.3	11.3
12	0.07	10.9	46.4	16.8	52	4.8	12.5	4.8	12.5	16.8	52.3	–	–	10.9	46.4
13	0.15	54.5	35.9	10.9	46	23.1	55.1	8.3	12.9	4.5	4.4	54.5	35.9	5.3	3
14	0.03	54.5	35.9	23.1	55	10.9	46.4	36	54.4	36	54.4	47.3	47.2	54.5	35.9
15	0.04	20.8	59.3	28.9	84	82.1	131.6	29.7	93.4	21.2	71.1	21.2	71.1	21.5	104.8
16	0.05	55.7	152	41.7	111	52.9	86.8	102	186.7	81	185.7	100	256.8	45.7	113.2
17	0.04	16.6	61.2	41.6	68	6.3	13.1	10.4	21	42.8	87.8	18	62.4	16.6	61.2
18	0.04	9.6	18.6	9.7	14	12.6	24.2	22.3	25.2	11.3	15.9	9.6	23.6	12.6	24.2
19	0.02	42.1	51.6	17.2	23	11.2	27.4	15.6	23.8	9.7	10.5	12.6	24.2	18.6	27.1
Weighted average		**29.5**	**43.5**	**20.4**	**55.2**	**19.2**	**43.4**	**17.8**	**38.8**	**16.3**	**37.6**	**32.6**	**48.8**	**20.6**	**39.9**

VI. Estimated values of the within-study correlation coefficients

In Table 11 we give the within-study correlation coefficients for the log odd ratios in two-arm studies reporting both response and dropout, as estimated using the elicited experts' opinion for the φ parameters and Equation (6) from the main paper.

Table 11: Within-study correlation coefficients estimated for two-arm studies reporting both outcomes.

Study ID	ρ	95% C.I. (lower)	95% C.I. (upper)
1	0.15	-0.05	0.36
2	-0.36	-0.51	-0.20
3	0.03	-0.17	0.27
4	-0.58	-0.69	-0.47
5	-0.58	-0.70	-0.44
6	-0.33	-0.45	-0.19
7	-0.28	-0.44	-0.11
8	0.18	0.09	0.23
9	-0.10	-0.24	0.06
10	-0.45	-0.56	-0.33
11	-0.28	-0.39	-0.17
12	-0.11	-0.26	0.05
14	0.02	-0.19	0.25
16	-0.37	-0.46	-0.27
17	-0.43	-0.57	-0.28
18	-0.10	-0.22	0.03
19	-0.63	-0.63	-0.63
20	0.07	-0.22	0.33
21	-0.09	-0.27	0.11
22	0.07	0.02	0.14
23	-0.32	-0.47	-0.17
24	-0.34	-0.44	-0.23
25	0.08	-0.06	0.23
27	-0.34	-0.45	-0.22
28	-0.15	-0.23	-0.07
29	0.24	0.03	0.38
30	0.17	0.07	0.28
32	-0.28	-0.40	-0.15
33	0.30	0.19	0.38
36	0.15	0.05	0.18
37	-0.30	-0.43	-0.16

38	-0.16	-0.30	-0.01
42	0.25	0.16	0.35
43	0.24	0.13	0.36
47	-0.25	-0.34	-0.15

VII. Extending the model presented in Chapter 3

In this section of the appendix we discuss how to extend the model discussed in Chapter 3, for the case of studies with more than three arms and for more than two outcomes of interest.

i. Handling studies with four arms or more

The models described so far can be extended when multi-arm studies with more than three arms are present. Suppose a four-arm study compares treatments A, B, C and E for efficacy (R) and dropout (D). If we choose the basic parameters to be the comparisons AB, AC and AE for both R and D, the random effects can be assumed to follow a multivariate normal distribution with variance-covariance matrix:

$$\boldsymbol{\Delta}_{(6\times6)} = \begin{pmatrix} \tau_R^2 & . & . & . & . & . \\ \rho_\tau\tau_R\tau_D & \tau_D^2 & . & . & . & . \\ \tau_R^2/2 & \chi_2\tau_R\tau_D & \tau_R^2 & . & . & . \\ \chi_1\tau_R\tau_D & \tau_D^2/2 & \rho_\tau\tau_R\tau_D & \tau_D^2 & . & . \\ \tau_R^2/2 & \rho_\tau\tau_R\tau_D & \tau_R^2/2 & \tau_D^2/2 & \tau_R^2 & . \\ \rho_\tau\tau_R\tau_D & \tau_D^2/2 & \rho_\tau\tau_R\tau_D & \tau_D^2/2 & \tau_D^2/2 & \tau_D^2 \end{pmatrix}$$

This matrix is analogous to the one in Equation (33) and the parameters P, τ_R, τ_D need to be estimated from the data. The random errors are also assumed to follow a multivariate normal distribution, the variance covariance-matrix is of the following form (the * represent the standard errors that multiply the ρ's, and we dropped the study index i for simplicity):

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{AB,R}^2 & . & . & . & . & . \\ \rho_{AB_RAB_D} * & \sigma_{AB,D}^2 & . & . & . & . \\ \kappa_{AB_RAC_R} & \rho_{AB_DAC_R*} & \sigma_{AC,R}^2 & . & . & . \\ \rho_{AB_RAC_D} * & \kappa_{AB_RAC_R} & \rho_{AC_RAC_D} * & \sigma_{AC,D}^2 & . & . \\ \kappa_{AB_RAE_R} & \rho_{AB_DAE_R} * & \kappa_{AC_RAE_R} & \rho_{AE_RAC_D} * & \sigma_{AE,R}^2 & . \\ \rho_{AB_RAE_D} * & \kappa_{AB_DAE_D} & \rho_{AC_RAE_D} * & \kappa_{AC_DAE_D} & \rho_{AE_RAE_D} * & \sigma_{AE,D}^2 \end{pmatrix}$$

The κ's can be readily estimated from the data and, as in the case of three-arm studies, we can use the φ parameters to reconstruct the full cross tables and then use formulas analogous to (37) and (38) to estimate all correlations needed. It is easy to see that this method can be applied without complications to multi-arm studies with any number of arms.

ii. Handling more than two correlated outcomes

Assume we have studies reporting on a specific comparison A versus B, for the correlated outcomes R, D and V. The random errors for every study can be assumed to follow a multivariate

normal distribution (after dropping the *i* study index):

$$\begin{pmatrix}\delta_{AB,R}\\ \delta_{AB,D}\\ \delta_{AB,V}\end{pmatrix}\sim N\left(0,\begin{pmatrix}\tau_R^2 & . & .\\ \rho_{\tau,RD}\tau_R\tau_D & \tau_D^2 & .\\ \rho_{\tau,RV}\tau_R\tau_V & \rho_{\tau,DV}\tau_V\tau_D & \tau_V^2\end{pmatrix}\right) \tag{39}$$

Note that there are in principal three heterogeneity variances and three different between-study correlation coefficients. The random errors follow a normal distribution:

$$\begin{pmatrix}\sigma_{AB,R}\\ \sigma_{AB,D}\\ \sigma_{AB,V}\end{pmatrix}\sim N\left(0,\begin{pmatrix}\sigma_R^2 & . & .\\ \rho_{RD}\sigma_R\sigma_D & \sigma_D^2 & .\\ \rho_{RV}\sigma_R\sigma_V & \rho_{DV}\sigma_V\sigma_D & \sigma_V^2\end{pmatrix}\right)$$

We can again reconstruct the full cross tables from the collapsed ones, but in this case we will need information on three different conditional probabilities, e.g. $P(D^+|R^+)$, $P(V^+|R^+)$ and $P(V^+|D^+)$ for every treatment. We can then use Equation (35) and compute every coefficient needed. For the case of a network meta-analysis with three correlated outcomes the between-study variance-covariance matrix for a three-arms trial is a (6×6) generalization of the matrix in Equation (34). The within-study correlations can again be estimated after eliciting the conditional probabilities $P(D^+|R^+)$, $P(V^+|R^+)$ and $P(V^+|D^+)$.

A generalization for more arms or more outcomes follows the same principles.

VIII. The variance-covariance matrix for random errors

In this section we describe how we can simplify the within-study variance-covariance matrix for multi-arm studies, by employing a set of assumptions. The results of this section are then employed in Section 4.2.3.1 of this dissertation.

For a three-arm study i that compares treatments A, B and C we assume that there are two different correlation coefficients, ρ_* that correlates same comparisons-different outcomes, and ρ_{**} for different comparisons different outcomes, i.e:

$$\rho_{i,AB_RAB_D} = \rho_{i,AC_RAC_D} \equiv \rho_i^*,\ \rho_{i,AC_RAB_D} = \rho_{i,AB_RAC_D} = \rho_{i,BC_RBA_D} \equiv \rho_i^{**}$$

The variance-covariance matrix for the random errors in this study is the following:

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma^2_{i,AB,R} & . & . & . \\ \rho_i^* \sigma_{i,AB,R}\sigma_{i,AB,D} & \sigma^2_{i,AB,D} & . & . \\ \kappa_{i,AB_RAC_R} & \rho_i^{**}\sigma_{i,AB,D}\sigma_{i,AC,R} & \sigma^2_{i,AC,R} & . \\ \rho_i^{**}\sigma_{i,AB,R}\sigma_{i,AC,D} & \kappa_{i,AB_DAC_D} & \rho_i^*\sigma_{i,AC,R}\sigma_{i,AC,D} & \sigma^2_{i,AC,D} \end{pmatrix} \quad (40)$$

The σ and κ parameters in the matrix above can be estimated from the data provided in the studies using well-known formulas.

Following a similar method to the one presented in Section III of Appendix 1, after assuming consistency we get:

$$y_{i,AC,R} - y_{i,BC,D} = y_{i,BC,R} - y_{i,BA,R} - y_{i,BC,D} \quad (41)$$

Taking the variance in both arms of Equation (41) we get:

$$\begin{aligned} Var(y_{i,AC,R}) + Var(y_{i,BC,D}) - 2\rho^{**}\sigma_{i,AC,R}\sigma_{i,BC,D} \\ = Var(y_{i,BC,R}) + Var(y_{i,BA,R}) + Var(y_{i,BC,D}) - 2\kappa_{i,BC_RBA_R} \\ - 2\rho^*\sigma_{i,BC,R}\sigma_{i,BC,D} + 2\rho^*\sigma_{i,BA,R}\sigma_{i,BC,D} \end{aligned} \quad (42)$$

If we use $e_{i,Y,W}$ to denote the number of successes and $f_{i,Y,W}$ the number of failures reported in every treatment arm Y, for outcome W of the study i, the left-hand side of Equation (42) can be written as follows:

$$\begin{aligned} LH = \left(\frac{1}{e_{i,A,R}} + \frac{1}{f_{i,A,R}} + \frac{1}{e_{i,C,R}} + \frac{1}{f_{i,C,R}}\right) + \left(\frac{1}{e_{i,B,D}} + \frac{1}{f_{i,B,D}} + \frac{1}{e_{i,C,D}} + \frac{1}{f_{i,C,D}}\right) \\ - 2\rho^{**}\sigma_{i,AC,R}\sigma_{i,BC,D} \end{aligned}$$

The right-hand side can be written as:

$$RH = \left(\frac{1}{e_{i,B,R}} + \frac{1}{f_{i,B,R}} + \frac{1}{e_{i,C,R}} + \frac{1}{f_{i,C,R}}\right) + \left(\frac{1}{e_{i,A,R}} + \frac{1}{f_{i,A,R}} + \frac{1}{e_{i,B,R}} + \frac{1}{f_{i,B,R}}\right) + \left(\frac{1}{e_{i,B,D}} + \frac{1}{f_{i,B,D}} + \frac{1}{e_{i,C,D}} + \frac{1}{f_{i,C,D}}\right) - 2\left(\frac{1}{e_{i,B,R}} + \frac{1}{f_{i,B,R}}\right) - 2\rho^{*}\sigma_{i,BC,R}\sigma_{i,BC,D} - 2\rho^{**}\sigma_{i,BA,R}\sigma_{i,BC,D}$$

By equating, we get:

$$\rho_i^{**} = \frac{\rho_i^{*}\sigma_{i,BC,R}}{\sigma_{i,BA,R} + \sigma_{i,AC,R}} \tag{43}$$

If we also assume that the standard deviations of different comparisons of the same outcome are equal within every study, i.e. $\sigma_{i,BC,R} = \sigma_{i,BA,R} = \sigma_{i,AC,R}$, we get that

$$\rho_i^{**} = \frac{\rho_i^{*}}{2} \tag{44}$$

Note that in order for this to be a consistent result we must also assume $\sigma_{i,BC,D} = \sigma_{i,BA,D} = \sigma_{i,AC,D}$.

Even though we have assumed equal variances to simplify the variance-covariance matrix of Equation (40), in the end of the day the σ and κ parameters are still left distinct and are estimated from the data. Equation (44) is just used to minimize the number of correlation parameters needed for the matrix in Equation (40).

The two assumptions we used (equal correlations, equal variances) are a justified approximation when all treatments in each study are comparable and the arms are balanced. This, however, may not always be the case. We can repeat the whole analysis without making any assumptions of equality in either the correlation coefficients or the variances. By taking analogous relations to the one in Equation (41) we get the following set of equations, after dropping the study index i for simplicity:

$$\begin{aligned}
\rho_{AC_RAC_D}\sigma_{AC_R} &= \rho_{AB_RAC_D}\sigma_{AC_D} + \rho_{BC_RAC_D}\sigma_{BC_R} \\
\rho_{AC_RAC_D}\sigma_{AC_D} &= \rho_{AC_RAB_D}\sigma_{AB_D} + \rho_{AC_RBC_D}\sigma_{BC_D} \\
\rho_{AB_RAB_D}\sigma_{AB_R} &= \rho_{AC_RAB_D}\sigma_{AC_R} - \rho_{BC_RAB_D}\sigma_{BC_R} \\
\rho_{AB_RAB_D}\sigma_{AB_D} &= \rho_{AB_RAC_D}\sigma_{AC_D} - \rho_{AB_RBC_D}\sigma_{BC_D} \\
\rho_{BC_RAB_D}\sigma_{BC_R} &= \rho_{AC_RBC_D}\sigma_{AC_R} - \rho_{AB_RBC_D}\sigma_{AC_R}
\end{aligned} \tag{45}$$

By cycling through the treatment and outcome indices we can produce more equations of this form, but it turns out they are linearly dependent to the ones above. Thus, out of the nine different correlation coefficients entering the five Equations (45) only four of them are independent. This set of equations is the most general solution to the problem of finding the correlation coefficients in a three-arm study.

Depending on the nature of the problem one can now make extra assumptions to simplify these equations. If for example out of the three treatments A, B and C being compared in a study, A is the placebo, while the other two are active treatments with similar results in both outcomes, it would be justifiable to assume $\rho_{AC_R AB_D} = \rho_{AB_R AC_D} = \rho^{**}$. If we also set $\rho_{AB_R AB_D} = \rho_{AC_R AC_D} = \rho_{BC_R BC_D} = \rho^{*}$, we find:

$$\rho^{**} = \rho^{**} \frac{\sigma_{AB_R}\sigma_{AB_D} + \sigma_{AC_R}\sigma_{AC_D} - \sigma_{BC_R}\sigma_{BC_D}}{\sigma_{AB_R}\sigma_{AC_D} + \sigma_{AB_D}\sigma_{AC_R}}$$

This equation allows a simplification of the variance-covariance matrix without the need of any assumption on the variances of the treatment effects. Also note that this equation further reduces to Equation (44) by employing the equal variance assumption.

IX. Ensuring the positive-definiteness of variance-covariance matrices

The correlation coefficient parameter ρ_i of the model described in Section 4.2.3.1 needs to be truncated separately for each three-arm study in order to ensure the positive-definiteness of the variance-covariance matrix of Equation (8) of the main paper. The following R (149) program can be used to compute the (study-specific) upper limits u_i for the correlation coefficient is the:

```
rho=c(rep(0,N))
ff=function(r,m){
  s=s1+r*s2
  ss=eigen(s[m,,],only.values = TRUE)
  mineg=min(ss[[1]])
  mineg}
  for(m in 1:N){
          for(i in 1:100){
                  if (ff(0.01*i,m)*ff(0.01*i+0.01,m)<0)
                          {rho[m]=0.01*i}}}
rho
```

The program utilizes the fact that a positive-definite matrix has only positive eigenvalues. The inputs needed are the number N of the three-arm studies and two arrays s1 and s2 which are $(N \times 4 \times 4)$-dimensional and contain the N in number (4×4)-dimensional matrices $\boldsymbol{\Sigma_{i,1}}$ and $\boldsymbol{\Sigma_{i,2}}$ of Equation (21) of the paper. These are estimated from the data. A similar program can be used to compute the lower values. This, however, is redundant, because the limits are symmetrical around zero. The results for the 18 three-arm studies of the acute mania dataset are given in Table 12.

Table 12: Upper limit for the correlation coefficient in the three-arm studies

Study	u_i
48	0.99
49	0.96
50	0.96
51	0.98
52	0.54
53	0.65
54	0.81
55	0.86
56	0.99
57	0.68
58	0.84
59	0.95
60	0.78
61	0.80
62	0.83
63	0.82
64	0.75
65	0.85

X. Generalizing the alternative model by Riley et al.

As we discuss in Section 4.2.2, Riley et al. (213) proposed a model for bivariate pairwise meta-analysis in which a single correlation coefficient models all correlations; this hybrid coefficient incorporates both within and between-study correlation.

For a two-arm study i reporting on outcomes R and D a bivariate normal distribution is assumed:

$$\begin{pmatrix} y_{i,R} \\ y_{i,D} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \boldsymbol{\Omega_i} \right)$$

With the variance-covariance matrix given by:

$$\boldsymbol{\Omega_i} = \begin{pmatrix} \psi_R^2 + \sigma_{i,R}^2 & \rho^h \sqrt{\psi_R^2 + \sigma_{i,R}^2}\sqrt{\psi_D^2 + \sigma_{i,D}^2} \\ \rho^h \sqrt{\psi_R^2 + \sigma_{i,R}^2}\sqrt{\psi_D^2 + \sigma_{i,D}^2} & \psi_D^2 + \sigma_{i,D}^2 \end{pmatrix} \quad (46)$$

The ψ parameters model for the additional variation apart from the sampling error that enters due to heterogeneity.

In this section of the Appendix we show how to extend the model for the case of a network of interventions. We restrict to the case of networks with two-arm and three-arm studies only. For a two-arm study i comparing treatments A and B for outcomes R and D, the variance-covariance matrix is again of the form of Equation (46). For a three-arm study comparing treatments A, B and C the variance-covariance matrix of Equation (46) can be generalized as follows:

$$\boldsymbol{\Omega_i} = \begin{pmatrix} \zeta_{i,AB,R} & . & . & . \\ \rho^h \sqrt{\zeta_{i,AB,R}\zeta_{i,AB,D}} & \zeta_{i,AB,D} & . & . \\ \rho_1 \sqrt{\zeta_{i,AB,R}\zeta_{i,AC,D}} & \rho_2 \sqrt{\zeta_{i,AB,D}\zeta_{i,AC,R}} & \zeta_{i,AC,R} & . \\ \rho_2 \sqrt{\zeta_{i,AB,R}\zeta_{i,AC,D}} & \rho_3 \sqrt{\zeta_{i,AB,D}\zeta_{i,AC,D}} & \rho^h \sqrt{\zeta_{i,AC,R}\zeta_{i,AC,D}} & \zeta_{i,AC,D} \end{pmatrix} \quad (47)$$

In the above we have set $\zeta_{i,AB,R} = \sigma_{i,AB,R}^2 + \psi_R^2$ and $\zeta_{i,AB,D} = \sigma_{i,AB,D}^2 + \psi_D^2$, similarly for the AC comparison. In Equations (46) and (47) we have assumed that the correlation coefficient ρ^h correlates treatment effects of the same treatment comparison but different outcomes (e.g. comparison AB for outcomes R and D), ρ_1 correlates different treatment comparisons of the R outcome, ρ_2 correlates different outcomes of different comparison and ρ_3 correlates different comparisons of the D outcome. In order to simplify this matrix we also assume that the variances of the treatment effects for comparisons of the same outcome

are equal within a study, irrespectively of the comparison being made:

$$\sigma^2_{i,AB,R} = \sigma^2_{i,AC,R} = \sigma^2_{i,BC,R} \text{ and } \sigma^2_{i,AB,D} = \sigma^2_{i,AC,D} = \sigma^2_{i,BC,D}.$$

This assumption, with the use of the consistency equations leads to $\rho_1 = \rho_3 = 1/2$, as it is easy to prove. For example, consistency states that $y_{i,AB,R} = y_{i,AC,R} - y_{i,BC,R}$. By taking the variance on both sides we get:

$$\zeta_{i,AB,R} = \zeta_{i,AC,R} + \zeta_{i,BC,R} - 2\rho_1\sqrt{\zeta_{i,AB,R}\zeta_{i,BC,R}} \rightarrow \rho_1 = 1/2$$

In the above we use $\zeta_{i,AB,R} = \zeta_{i,AC,R} = \zeta_{i,BC,R}$, which holds by virtue of the equal variance assumption we make. A similar proof holds for ρ_3.

Also, as we proved in Equation (42) the consistency equations give:

$$\begin{aligned} &Var(y_{i,AC,R}) + Var(y_{i,BC,D}) - 2Cov(y_{i,AC,R}, y_{i,BC,D}) \\ &\quad = Var(y_{i,BC,R}) + Var(y_{i,BA,R}) + Var(y_{i,BC,D}) - 2Cov(y_{i,BC,R}, y_{i,BA,R}) \\ &\quad - 2Cov(y_{i,BC,R}, y_{i,BC,D}) + 2Cov(y_{i,BA,R}, y_{i,BC,D}) \rightarrow \\ &\zeta_{i,R} + \zeta_{i,D} - 2\rho_2\sqrt{\zeta_{i,R}\zeta_{i,D}} \\ &\quad = \zeta_{i,R} + \zeta_{i,D} + \zeta_{i,D} - 2\rho_1\sqrt{\zeta_{i,R}\zeta_{i,D}} - 2\rho^h\sqrt{\zeta_{i,R}\zeta_{i,D}} + 2\rho_2\sqrt{\zeta_{i,R}\zeta_{i,D}} \end{aligned}$$

In the above we have set $\zeta_{i,AB,R} = \zeta_{i,AC,R} = \zeta_{i,BC,R} = \zeta_{i,R}$ and $\zeta_{i,AB,D} = \zeta_{i,AC,D} = \zeta_{i,BC,D} = \zeta_{i,D}$. By substituting $\rho_1 = 1/2$, and after some algebra we get:

$$\rho_2 = \frac{1}{2}\rho_h$$

The variance-covariance matrix takes the following form (after dropping the study index for simplicity):

$$\boldsymbol{\Omega} = \begin{pmatrix} \zeta_{AB,R} & . & . & . \\ \rho^h\sqrt{\zeta_{AB,R}\zeta_{AB,D}} & \zeta_{AB,D} & . & . \\ \frac{1}{2}\sqrt{\zeta_{AB,R}\zeta_{AC,D}} & \frac{\rho^h}{2}\sqrt{\zeta_{AB,D}\zeta_{AC,R}} & \zeta_{AC,R} & . \\ \frac{\rho^h}{2}\sqrt{\zeta_{AB,R}\zeta_{AC,D}} & \frac{1}{2}\sqrt{\zeta_{AB,D}\zeta_{AC,D}} & \rho^h\sqrt{\zeta_{AC,R}\zeta_{AC,D}} & \zeta_{AC,D} \end{pmatrix} \quad (48)$$

The σ parameters entering the ζ can be estimated from the data. This variance-covariance matrix is always positive-definite.

XI. Detailed results from fitting models of Chapter 4

In this Section we present the detailed results from all scenarios discussed Chapter 4, Section 4.3.1. As a reminder, the scenarios explored were the following:

I. Univariate (independent) NMA of response and dropout rate separately, assuming $\tau_R, \tau_D \sim U(0,1)$. This corresponds to setting all correlations equal to zero.

II. Multiple outcome network meta-analysis (MONMA) following the approach of Section 4.2.3.1. with minimally informative priors for the heterogeneity parameters: $\rho^{\tau} \sim U(-1,0)$, $\tau_R, \tau_D \sim U(0,1)$, and: (a) assuming a negative common $\rho_i = \rho$ with $\rho \sim U(-1,0)$, (b) assuming a strongly informative, negative, common $\rho \sim U(-0.7, -0.5)$, (c) assuming a common fixed $\rho_i = \rho$ with $\rho = -0.7$, (d) assuming two different within-studies correlation coefficients ρ_i: one for the studies comparing two active treatments, which we denote as $\rho_{Act-Act}$, and another for the studies comparing active treatments to placebo, ρ_{Act-Pl}. This distinction could be based on the assumption that the two relative treatment effects are differently correlated when one of the treatments compared is the placebo. For both parameters we used a uniform negative, $U(-1,0)$, prior distribution

III. MONMA following the approach in Section 4.2.3.2, assuming a common correlation coefficient and the following prior distributions for the parameters of the model: $\rho^h \sim U(-1,0)$, $\psi_R \sim U(0,1)$, $\psi_D \sim U(0,1)$.

In Figure 11 we present the odds ratios for the response outcome, for the comparison of active drugs vs. placebo. In Figure 12 we present the results for the acceptability outcome (dropout). In Table 13 we present the rankings for all treatments based on their SUCRA values (38) for each model.

Figure 11: Summary odds ratios for response, Treatment vs. Placebo for all scenarios presented in Section 4.3.1

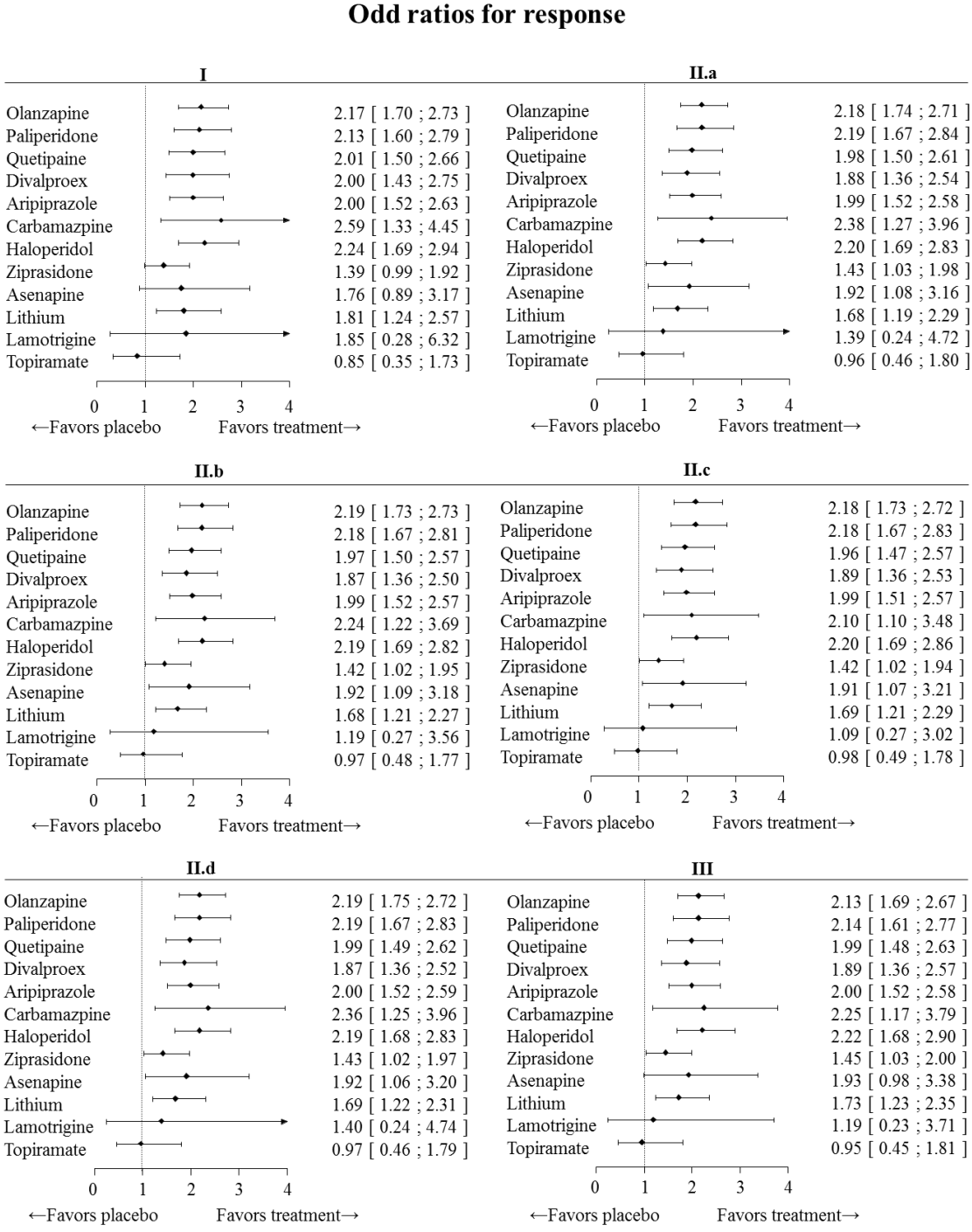


Figure 12: Summary odds ratios for dropout, Drug vs. Placebo for all scenarios presented in Section 4.3.1..

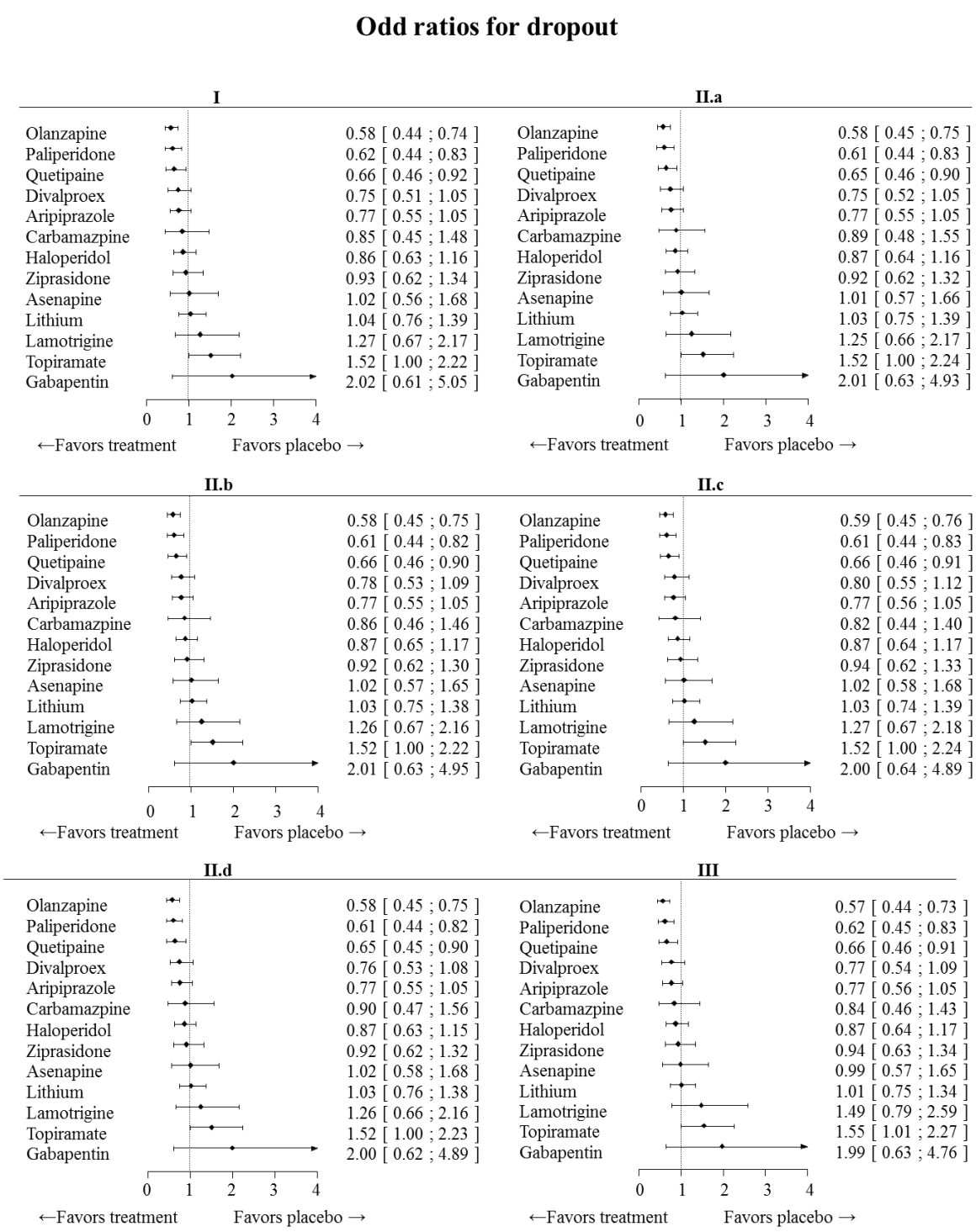


Table 13: Treatment ranking for all models in Section 4.3.1.based on the SUCRA values for response (R) and dropout (D).

I				**II.a**			
SUCRA – *R* (%)		**SUCRA – *D* (%)**		**SUCRA – *R* (%)**		**SUCRA – *D* (%)**	
Carbamazepine	79.6	Olanzapine	91.7	Haloperidol	76.6	Olanzapine	91.2
Haloperidol	75.8	Paliperidone	86.3	Olanzapine	76.5	Paliperidone	87.5
Olanzapine	72.1	Quetipaine	80.3	Carbamazepine	75.2	Quetipaine	81.4
Paliperidone	69.0	Divalproex	68.8	Paliperidone	74.9	Divalproex	68.1
Quetipaine	60.4	Aripiprazole	65.6	Aripiprazole	62.4	Aripiprazole	65.9
Aripiprazole	60.1	Carbamazepine	57.6	Quetipaine	61.5	Carbamazepine	53.3
Divalproex	59.6	Haloperidol	52.9	Asenapine	54.4	Haloperidol	52.5
Lithium	47.1	Ziprasidone	45.1	Divalproex	54.0	Ziprasidone	46.5
Asenapine	45.2	Asenapine	39.7	Lithium	40.8	Asenapine	40.1
Lamotrigine	39.1	Placebo	34.1	Ziprasidone	27.9	Placebo	34.1
Ziprasidone	25.5	Lithium	32.7	Lamotrigine	26.5	Lithium	33.5
Placebo	9.7	Lamotrigine	23.5	Topiramate	9.9	Lamotrigine	24.3
Topiramate	6.9	Gabapentin	12.7	Placebo	9.4	Gabapentin	12.6
		Topiramate	9.1			Topiramate	9.0
II.b				**II.c**			
SUCRA – *R* (%)		**SUCRA – *D* (%)**		**SUCRA – *R* (%)**		**SUCRA – *D* (%)**	
Olanzapine	77.9	Olanzapine	91.3	Haloperidol	78.5	Olanzapine	91.0
Haloperidol	77.0	Paliperidone	87.3	Olanzapine	78.4	Paliperidone	87.1
Paliperidone	76.4	Quetipaine	81.5	Paliperidone	77.1	Quetipaine	81.3
Carbamazepine	72.0	Aripiprazole	66.0	Carbamazepine	66.5	Aripiprazole	66.1
Aripiprazole	63.2	Divalproex	65.2	Aripiprazole	64.4	Divalproex	62.6
Quetipaine	61.8	Carbamazepine	57.1	Quetipaine	62.2	Carbamazepine	61.6
Asenapine	56.0	Haloperidol	51.7	Divalproex	56.6	Haloperidol	52.1
Divalproex	54.4	Ziprasidone	46.9	Asenapine	56.3	Ziprasidone	45.6
Lithium	42.2	Asenapine	39.9	Lithium	43.0	Asenapine	39.6
Ziprasidone	27.7	Placebo	34.1	Ziprasidone	28.3	Placebo	34.4
Lamotrigine	21.5	Lithium	33.7	Lamotrigine	17.8	Lithium	33.7
Topiramate	10.2	Lamotrigine	23.8	Topiramate	10.9	Lamotrigine	23.7
Placebo	9.7	Gabapentin	12.6	Placebo	10.0	Gabapentin	12.3
		Topiramate	9.0			Topiramate	9.1
II.d				**III**			
SUCRA – *R* (%)		**SUCRA – *D* (%)**		**SUCRA – *R* (%)**		**SUCRA – *D* (%)**	
Olanzapine	76.5	Olanzapine	91.2	Haloperidol	78.1	Olanzapine	92.8
Haloperidol	75.4	Paliperidone	87.6	Olanzapine	74.1	Paliperidone	86.6
Carbamazepine	75.3	Quetipaine	81.7	Paliperidone	73.0	Quetipaine	80.6
Paliperidone	75.1	Divalproex	67.3	Carbamazepine	71.4	Aripiprazole	66.1
Aripiprazole	62.4	Aripiprazole	66.2	Aripiprazole	63.4	Divalproex	65.8
Quetipaine	61.8	Carbamazepine	53.0	Quetipaine	63.2	Carbamazepine	58.7
Asenapine	55.1	Haloperidol	52.7	Asenapine	56.0	Haloperidol	52.2
Divalproex	53.6	Ziprasidone	46.8	Divalproex	55.4	Ziprasidone	45.4
Lithium	41.5	Asenapine	39.4	Lithium	44.5	Asenapine	42.9
Ziprasidone	27.4	Placebo	34.3	Ziprasidone	29.1	Lithium	36.1
Lamotrigine	26.7	Lithium	33.9	Lamotrigine	21.6	Placebo	34.6
Topiramate	9.9	Lamotrigine	23.9	Placebo	10.1	Lamotrigine	15.1
Placebo	9.4	Gabapentin	12.8	Topiramate	10.1	Gabapentin	13.3
		Topiramate	9.2			Topiramate	9.9

XII. Generalizing the models of Chapter 4

In this section of the Appendix we present methods for extending the two new models proposed in Chapter 4 for the case of studies with more than three arms, reporting on more than two correlated outcomes of interest.

a. Generalizing the first model

We start from the case of pairwise meta-analysis, when only two treatments are compared for three outcomes. Suppose there are studies reporting on a single comparison A versus B, for three correlated outcomes R, D and V. The random errors for every study are assumed to follow a multivariate normal distribution $(\delta_R \ \delta_D \ \delta_V)' \sim N(0, \boldsymbol{\Delta}_{(3\times3)})$, with variance-covariance matrix:

$$\boldsymbol{\Delta}_{(3\times3)} = \begin{pmatrix} \tau_R^2 & . & . \\ \rho_{\tau,RD}\tau_R\tau_D & \tau_D^2 & . \\ \rho_{\tau,RV}\tau_R\tau_V & \rho_{\tau,DV}\tau_V\tau_D & \tau_V^2 \end{pmatrix} \quad (49)$$

Note that there are in principle three heterogeneities and three different between-study correlation coefficients that need to be estimated.

The random errors of study i are also assumed to follow a multivariate normal distribution $(\varepsilon_{i,R} \ \varepsilon_{i,D} \ \varepsilon_{i,V})' \sim N(0, \boldsymbol{\Sigma_i})$. The within-study variance-covariance matrix is:

$$\boldsymbol{\Sigma_i} = \begin{pmatrix} \sigma_{i,R}^2 & . & . \\ \rho_{i,RD}\sigma_{i,R}\sigma_{i,D} & \sigma_{i,D}^2 & . \\ \rho_{i,RV}\sigma_{i,R}\sigma_{i,V} & \rho_{i,DV}\sigma_{i,V}\sigma_{i,D} & \sigma_{i,V}^2 \end{pmatrix} \quad (50)$$

Thus, there are also three different within-study correlation coefficients to estimate.

We now extend the method for the case of multi-arm studies reporting on three correlated outcomes for a multiplicity of treatments. If we focus on a three-arm study, the heterogeneity variance-covariance matrix will be a (6×6) generalization of the matrix of Equation (49):

$$\boldsymbol{\Delta}_{(6\times6)} = \begin{pmatrix} \tau_R^2 & . & . & . & . & . \\ \rho_{\tau,RD}\tau_R\tau_D & \tau_D^2 & . & . & . & . \\ \rho_{\tau,RV}\tau_R\tau_V & \rho_{\tau,DV}\tau_V\tau_D & \tau_V^2 & . & . & . \\ \tau_R^2/2 & \rho_{\tau,RD}\tau_R\tau_D & \rho_{\tau,RV}\tau_R\tau_V & \tau_R^2 & . & . \\ \rho_{\tau,RD}\tau_R\tau_D & \tau_D^2/2 & \rho_{\tau,DV}\tau_D\tau_V & \rho_{\tau,RD}\tau_R\tau_D & \tau_D^2 & . \\ \rho_{\tau,RV}\tau_R\tau_V & \rho_{\tau,DV}\tau_D\tau_V & \tau_V^2/2 & \rho_{\tau,RV}\tau_R\tau_V & \rho_{\tau,DV}\tau_D\tau_V & \tau_V^2 \end{pmatrix} \quad (51)$$

The within-study variance-covariance matrix for this study, after making the same simplifying assumptions as in Section III of this Appendix, can be estimated as follows:

$$\boldsymbol{\Sigma_i} = \begin{pmatrix} s_{i,AB,R}^2 & . & . & . & . & . \\ \rho_{i,RD} & s_{i,AB,D}^2 & . & . & . & . \\ \rho_{i,RV} & \rho_{i,DV} & s_{i,AB,V}^2 & . & . & . \\ \kappa_{i,AB_RAC_R} & 0.5\rho_{i,RD} & 0.5\rho_{i,RV} & s_{i,AC,R}^2 & . & . \\ 0.5\rho_{i,RD} & \kappa_{i,AB_DAC_D} & 0.5\rho_{i,VD} & \rho_{i,RD} & s_{i,AC,D}^2 & . \\ 0.5\rho_{i,RV} & 0.5\rho_{i,VD} & \kappa_{i,AB_VAC_V} & \rho_{i,RV} & \rho_{i,VD} & s_{i,AC,V}^2 \end{pmatrix} \quad (52)$$

In the above we have dropped the standard errors that multiply the correlation coefficients for simplicity. We now have three different within-study correlation coefficients to estimate for every three-arm study. As before, we can model these coefficients to be common across studies or among group of studies.

Extending for more arms or more outcomes is straightforward. For example, a four-arm study in the case of three outcomes of interest will require a 9×9 generalization of the above matrices.

b. Generalizing the second model

In this subsection we will show how to extend the second model presented in Section 3.2.2 of the main paper for the case of more than two correlated outcomes, or in the presence of studies than more than three arms. Let us start by assuming a network of studies reporting on three outcomes R, D and V. For a two-arm study the variance-covariance matrix can be estimated as:

$$\boldsymbol{\Omega_i} = \begin{pmatrix} \zeta_{i,AB,R} & . & . \\ \rho_{RD}^h\sqrt{\zeta_{i,AB,R}\zeta_{i,AB,D}} & \zeta_{i,AB,D} & . \\ \rho_{RV}^h\sqrt{\zeta_{i,AB,R}\zeta_{i,AB,V}} & \rho_{DV}^h\sqrt{\zeta_{i,AB,D}\zeta_{i,AB,V}} & \zeta_{i,AB,V} \end{pmatrix} \quad (53)$$

Note that we now need three different hybrid correlation coefficients to be estimated from the model. For a three-arm study comparing treatments A, B and C for three outcomes R, D and V a (6×6) variance-covariance matrix is needed instead:

$$\boldsymbol{\Omega_i} = \begin{pmatrix} \zeta_{i,AB,R} & . & . & . & . & . \\ \rho^h_{RD} & \zeta_{i,AB,D} & . & . & . & . \\ \rho^h_{RV} & \rho^h_{DV} & \zeta_{i,AB,V} & . & . & . \\ 1/2 & \rho^h_{VD}/2 & \rho^h_{RV}/2 & \zeta_{i,AC,R} & . & . \\ \rho^h_{RD}/2 & 1/2 & \rho^h_{VD}/2 & \rho^h_{RD}/2 & \zeta_{i,AC,D} & . \\ \rho^h_{RV}/2 & \rho^h_{VD}/2 & 1/2 & \rho^h_{RV}/2 & \rho^h_{VD}/2 & \zeta_{i,AC,V} \end{pmatrix} \quad (54)$$

In the above matrix we have dropped the z parameters in the elements off the diagonal, for simplicity. We can generalize for the case of studies with more arms, and for multiple outcomes by following the same pattern. Note that care should be taken so that all the variance-covariance matrices presented in this section remain definite-positive. In Section 4.2.3.1 and Section IX of this Appendix we discussed ways to ensure the positive-definiteness of these matrices.

XI. OpenBUGS code for fitting the model of Section 3.3.3

```
model{
## CONTROL for missing outcomes
for (i in 1:Ns){
      cR[i]<-step(e1r[i]-0.2)
      cD[i]<-step(e1d[i]-0.2)
      control[i]<-cR[i]*cD[i]
}
## truncate the φ's for every study
for (i in 1: Ns){
      p10[i]<-max((e1d[i]-f1r[i])/e1r[i],0)*control[i]
      p11[i]<-min(e1d[i]/(f1r[i]+e1r[i]),1)
      p20[i]<-max((e2d[i]-f2r[i])/e2r[i],0)*control[i]
      p21[i]<-min(e2d[i]/(f2r[i]+e2r[i]),1)
      zero1[i] <- 0
      phi1trunc[i] ~ dunif(p10[i],p11[i])
     cc1[i]<-step(abs(phi1trunc[i])-p10[i])*step(p11[i]-
     abs(phi1trunc[i]))+0.0001
     phi1[i] <- -(a[T1[i]]-1)*log(phi1trunc[i])-(b[T1[i]]-
     1)*log(1-phi1trunc[i])-log(cc1[i])+1000000
      zero1[i] ~ dpois(phi1[i])
      zero2[i] <- 0
      phi2trunc[i] ~ dunif(p20[i],p21[i])
     cc2[i]<-step(abs(phi2trunc[i])-p10[i])*step(p21[i]-
     abs(phi2trunc[i]))+0.0001
     phi2[i] <- -(a[T2[i]]-1)*log(phi2trunc[i])-(b[T2[i]]-
     1)*log(1-phi2trunc[i])-log(cc2[i])+1000000
    zero2[i] ~ dpois(phi2[i])}

for (i in (N2h+1):Ns){
      p30[i]<-max((e3d[i]-f1r[i])/e3r[i],0)*control[i]
      p31[i]<-min(e3d[i]/(f3r[i]+e3r[i]),1)
      zero3[i] <- 0
      phi3trunc[i] ~ dunif(p30[i],p31[i])
       cc3[i]<-step(abs(phi3trunc[i])-p30[i])*step(p31[i]-
     abs(phi3trunc[i]))+0.0001
     phi3[i] <- -(a[T3[i]]-1)*log(phi3trunc[i])-(b[T3[i]]-
     1)*log(1-phi3trunc[i])-log(cc3[i])+1000000
      zero3[i] ~ dpois(phi3[i])}

## two-arm studies
for (i in 1:N2h){
       test1[i]<-(phi1trunc[i]*(e1r[i]+f1r[i])*(e1r[i]+f1r[i])-
       e1d[i]*(e1r[i]+f1r[i]))/(e1d[i]*f1r[i]*f1d[i])
      +(phi2trunc[i]*(e2r[i]+f2r[i])*(e2r[i]+f2r[i])-

       e2d[i]*(e2r[i]+f2r[i]))/(e2d[i]*f2r[i]*f2d[i])
       rho2h[i]<-
     test1[i]/(sqrt(1/e1r[i]+1/f1r[i]+1/e2r[i]+1/f2r[i])*sqrt(1/e1
     d[i]+1/f1d[i]+1/e2d[i]+1/f2d[i]))
      rho22h[i]<-max(min(rho2h[i],0.98),-0.98)*control[i]
      s2[i,1,1]<-1/e1r[i]+1/f1r[i]+1/e2r[i]+1/f2r[i]
```

```
      s2[i,2,2]<-1/e1d[i]+1/f1d[i]+1/e2d[i]+1/f2d[i]
      s2[i,1,2]<-sqrt(s2[i,1,1]*s2[i,2,2])*rho22h[i]
      s2[i,2,1]<-s2[i,1,2]}

for( i in 1:N2h){prec2A[i,1:2,1:2]<-inverse(s2[i,,])}

for (i in 1:N2h)
      {y[(2*i-1):(2*i)]~dmnorm(theta[(2*i-1):(2*i)],prec2A[i,,])}

for(i in 1:2){
      for (j in 1:2){
     D2[i,j]<-
     tau1.sq*t1[i,j]+tau2.sq*t2[i,j]+sqrt(tau1.sq*tau2.sq)*(rhotau
     )*t3[i,j]}}

prec2B[1:2,1:2]<-inverse(D2[,])

for (i in 1:N2h)
      {theta[(2*i-1):2*i]~dmnorm(mean[(2*i-1):2*i],prec2B[,]) }

## three-arm studies
for (i in 1:(Ns-N2h))
      {s4[i,1,1]<-
     1/e1r[N2h+i]+1/f1r[N2h+i]+1/e2r[N2h+i]+1/f2r[N2h+i]
      s4[i,2,2]<-
     1/e1d[N2h+i]+1/f1d[N2h+i]+1/e2d[N2h+i]+1/f2d[N2h+i]
     s4[i,3,3]<-
     1/e1r[N2h+i]+1/f1r[N2h+i]+1/e3r[N2h+i]+1/f3r[N2h+i]
     s4[i,4,4]<-
     1/e1d[N2h+i]+1/f1d[N2h+i]+1/e3d[N2h+i]+1/f3d[N2h+i]
      s4[i,1,3]<-(1/e1r[N2h+i]+1/f1r[N2h+i])*cR[i]
      s4[i,3,1]<-(1/e1r[N2h+i]+1/f1r[N2h+i])*cR[i]
      s4[i,2,4]<-(1/e1d[N2h+i]+1/f1d[N2h+i])*cD[i]
      s4[i,4,2]<-(1/e1d[N2h+i]+1/f1d[N2h+i])*cD[i]

## rho(AB-R,AB-D)
     tests112[i]<-
     (phi1trunc[N2h+i]*(e1r[N2h+i]+f1r[N2h+i])*(e1r[N2h+i]+f1r[N2h
     +i])-
     e1d[N2h+i]*(e1r[N2h+i]+f1r[N2h+i]))/(e1d[N2h+i]*f1r[N2h+i]*f1
     d[N2h+i])
       +(phi2trunc[N2h+i]*(e2r[N2h+i]+f2r[N2h+i])*(e2r[N2h+i]+f2r[
     N2h+i])-
     e2d[N2h+i]*(e2r[N2h+i]+f2r[N2h+i]))/(e2d[N2h+i]*f2r[N2h+i]*f2
     d[N2h+i])
      tests212[i]<-tests112[i]/sqrt(s4[i,1,1]*s4[i,2,2])
      rhos12[i]<-max(-0.98,min(0.98,tests212[i]))*control[i]
      s4[i,1,2]<-rhos12[i]*sqrt(s4[i,1,1]*s4[i,2,2])
      s4[i,2,1]<-rhos12[i]*sqrt(s4[i,1,1]*s4[i,2,2])

## rho(AC-R,AC-D)
       test134[i]<-
     (phi1trunc[N2h+i]*(e1r[N2h+i]+f1r[N2h+i])*(e1r[N2h+i]+f1r[N2h
     +i])-
```

```
      e1d[N2h+i]*(e1r[N2h+i]+f1r[N2h+i]))/(e1d[N2h+i]*f1r[N2h+i]*f1
      d[N2h+i])
        +(phi3trunc[N2h+i]*(e3r[N2h+i]+f3r[N2h+i])*(e3r[N2h+i]+f3r[
      N2h+i])-
      e3d[N2h+i]*(e3r[N2h+i]+f3r[N2h+i]))/(e3d[N2h+i]*f3r[N2h+i]*f3
      d[N2h+i])
       test234[i]<-test134[i]/(sqrt(s4[i,1,1]*s4[i,3,3]))
       rhos34[i]<-max(-0.98,min(0.98,test234[i]))*control[i]
       s4[i,3,4]<-rhos34[i]*sqrt(s4[i,1,1]*s4[i,3,3])
       s4[i,4,3]<-rhos34[i]*sqrt(s4[i,1,1]*s4[i,3,3])

## rho(AB-R,AC,D), rho(AB-D,AC-R)
       test114[i]<-
      (phi1trunc[N2h+i]*(e1r[N2h+i]+f1r[N2h+i])*(e1r[N2h+i]+f1r[N2h
      +i])
       -
      e1d[N2h+i]*(e1r[N2h+i]+f1r[N2h+i]))/(e1d[N2h+i]*f1r[N2h+i]*f1
      d[N2h+i])
       test214[i]<-test114[i]/(sqrt(s4[i,1,1]*s4[i,4,4]))
       rhos14[i]<-min(max(test214[i],-0.98),0.98)*control[i]
       s4[i,1,4]<-rhos14[i]*sqrt(s4[i,1,1]*s4[i,4,4])
       s4[i,4,1]<-rhos14[i]*sqrt(s4[i,1,1]*s4[i,4,4])
       s4[i,3,2]<-rhos14[i]*sqrt(s4[i,1,1]*s4[i,4,4])
       s4[i,2,3]<-rhos14[i]*sqrt(s4[i,1,1]*s4[i,4,4])}

for (k in 1:(Ns-N2h)){
       prec3A[k,1:4,1:4]<-inverse(s4[k,,])
      y[2*N2h+4*k-3:2*N2h+4*k]~dmnorm(theta[2*N2h+4*k-
      3:2*N2h+4*k],prec3A[k,,])}

for (i in 1:4){
       for (j in 1:4){
      D3[i,j]<-
      tau1.sq*delta1[i,j]+tau2.sq*delta2[i,j]+sqrt(tau1.sq*tau2.sq)
      *(rhotau)*delta3[i,j]}}

for (k in 1:(Ns-N2h))
        {prec4A[k,1:4,1:4]<-inverse(D3[,])
        theta[2*N2h+4*k-3:2*N2h+4*k]~dmnorm(mean[2*N2h+4*k-
      3:2*N2h+4*k],prec4A[k,,])}

#Parameterization of the means#
for(i in 1:N2h) {
       mean[2*i-1] <- -dR[T2[i]] + dR[T1[i]]
       mean[2*i] <- -dD[T2[i]]+ dD[T1[i]]}

for(i in 1:(Ns-N2h)) {
       mean[2*N2h+4*i-3] <- -dR[T2[N2h+i]] + dR[T1[N2h+i]]
       mean[2*N2h+4*i-2] <- -dD[T2[N2h+i]] + dD[T1[N2h+i]]
      mean[2*N2h+4*i-1] <- -dR[T3[N2h+i]]+ dR[T1[N2h+i]]
      mean[2*N2h+4*i] <- -dD[T3[N2h+i]] + dD[T1[N2h+i]]}

#Priors#
for(k in 1:(ref-1)) {
```

```
       dR[k] ~ dnorm(0,.01)}
for(k in (ref+1):NT) {
       dR[k] ~ dnorm(0,.01)}
for(k in 1:(ref-1)) {
       dD[k] ~ dnorm(0,.01)}
for(k in (ref+1):NT) {
       dD[k] ~ dnorm(0,.01)}

tau1.sq<-tau1*tau1
tau1~dunif(0,1)
tau2.sq<-tau2*tau2
tau2~dunif(0,1)
rhotau~dunif(-0.99,0)
#Estimated  Effect Sizes#
dR[ref]<- 0
for (c in 1:(ref-1)) { Eff.refR[c]<- exp(dR[c] - dR[ref])}
for (c in (ref+1):NT) {Eff.refR[c]<- exp(dR[c] - dR[ref])}

for (c in 1:(NT-1)) {
      for (k in (c+1):NT) { EffR[c,k] <- exp(dR[k] - dR[c])}}
dD[ref]<- 0
for (c in 1:(ref-1)) {      Eff.refD[c]<- exp(dD[c] - dD[ref] )}
for (c in (ref+1):NT) {Eff.refD[c]<- exp(dD[c] - dD[ref] )}
for (c in 1:(NT-1)) {
       for (k in (c+1):NT) {EffD[c,k] <- exp(dD[k] - dD[c])}}

# Ranking of treatments - R
for (k in 1:13){ddR[k]<-dR[k]}
for(k in 1:13) {
      orderR[k]<-14- rank(ddR[],k)
       most.effectiveR[k]<-equals(orderR[k],1)
       for(j in 1: 13) {
      effectivenessR[k,j]<- equals(orderR[k],j)
              cumeffectivenessR[k,j]<-
sum(effectivenessR[k,1:j])}}
for(k in 1:13) {SUCRAR[k]<- sum(cumeffectivenessR[k,1:(13-1)])
/(13-1)}
#Ranking of treatments - D
for(k in 1:14) {
       orderD[k]<- rank(dD[],k)
      most.effectiveD[k]<-equals(orderD[k],1)
for(j in 1: 14) {
      effectivenessD[k,j]<- equals(orderD[k],j)
       cumeffectivenessD[k,j]<-
sum(effectivenessD[k,1:j])}}
for(k in 1:NT) {SUCRAD[k]<- sum(cumeffectivenessD[k,1:13]) /(13)}}
```

The inputs required for this program are the following:

N2h: the number of two-arm studies.

Ns: the total number of studies.

NT: the number of treatments.

ref: the treatment number for the reference treatment (e.g. placebo).

a, b: the parameters of the Beta prior distributions for the φ.

y: the 2(N2h+2Ns)- dimensional vector of observed effects (2 for every two-arm study, 4 for every three-arm). Odd positions correspond to R comparison, even to D.

varr: the (N2h+2Ns)- dimensional vector of the variance for every R comparison (one for each odd position in y). For studies with missing data impute a large variance (e.g. 100000).

vard: the (N2h+2Ns)- dimensional vector of the variance for every D comparison (one for each even position in y). For studies with missing data impute a large variance (e.g. 100000).

T1,T2,T3: the Ns – dimensional vector of treatments for every study. For two arm studies set T3=0.

e1r,f1r,e2r,f2r,e3r,f3r: the Ns – dimensional vectors containing the number of successes (e) and failures (f) of every arm for the R outcome. For studies with missing data impute small a number (e.g. 0.1).

e1d,f1d,e2d,f2d,e3d,f3d: the same for the D outcome.

t1,t2,t3: the (2?2) matrices needed for constructing the heterogeneity variance-covariance matrix for the two-arm studies, Ä_((2?2)) of Equation (2):

t1 = structure(.Data=c(1, 0, 0, 0),.Dim=c(2 , 2))

t2 = structure(.Data=c(0, 0, 0, 1),.Dim=c(2 , 2))

t3 = structure(.Data=c(0, 1, 1, 0),.Dim=c(2 , 2))

delta1,delta2,delta3: the (4x4) matrices needed for constructing the heterogeneity variance-covariance matrix for the three-arm studies.

delta1 = structure(.Data=c(1,0,0.5 0,0,0,0,0,0.5,0,1,0,0,0,0,0),.Dim=c(4,4))

delta2 = structure(.Data=c(0,0,0,0,0,1,0,0.5,0,0,0,0,0,0.5,0,1),.Dim=c(4 ,4))

delta3=structure(.Data=c(0,1,0,0.5,1,0,0.5,0,0,0.5,0,1,0.5,0,1,0),.Dim=c(4,4))

For the acute mania example the data are as follows:

```
list( N2h = 49, Ns=67, NT=14, ref=2,

y  =c(5.34831E-01, 4.09477E-01, 8.97209E-01, -1.16246E-01,
3.62793E-01, -9.74196E-01, 1.05322E+00, -9.70437E-01,
1.95066E-01, -3.72771E-02, -8.29426E-01, 5.10009E-01, -
3.99240E-01, 3.41398E-01, -8.09319E-01, 1.26194E+00, -
```

```
8.28099E-01, -2.15871E-02, -5.71553E-01, 3.43985E-01, -
7.05038E-01, 2.52997E-01, -9.69115E-02, -1.54010E-01,     NA,
-1.09861E+00, -4.37973E-02, -4.89940E-01,     NA,
1.25276E+00, -1.14054E+00, 1.09397E+00, -9.65081E-01,
8.18310E-01, 5.29278E-01, -2.02999E-01, -8.48977E-01,
6.91444E-02, 3.84200E-02, 3.00429E-01,     NA,     NA,
5.26437E-02, 2.05852E-01, -1.55186E+00, 1.23117E+00, -
5.38997E-01, 6.80725E-01, -8.34938E-01, 5.79166E-01,
1.84498E-01, -6.00690E-01, 1.89712E+00,     NA, -5.87877E-01,
2.39369E-01, -2.10191E+00, -1.54151E-01, 8.78070E-01,
0.0E+00, -9.68826E-01, 5.53768E-01,     NA, -6.45385E-02, -
1.13197E+00, 3.34935E-01, 2.71934E-01, 4.85508E-01,     NA, -
5.66183E-01,     NA, -7.84831E-01, -8.90315E-01, 1.03407E+00,
-1.28437E-01, 1.23193E-01,     NA,     NA, 2.23144E-01,
2.35138E+00,     NA, 1.20397E+00,     NA, -1.79176E+00,
NA, 1.60944E+00, 6.35989E-01, -4.62624E-01,     NA, -
6.53195E-02, 2.51314E-01, -7.58439E-01,     NA, -5.59162E-01,
NA, -7.71790E-01, -4.39488E-01, -8.15296E-02, -5.23822E-01, -
7.06051E-03, -4.40040E-01, 5.91889E-02, -3.61506E-01,
2.15552E-01, -4.32809E-01, 1.04443E-01, -1.12719E+00,
9.88264E-01, -1.13561E+00, 1.46863E+00, -8.57450E-01,
8.30769E-01, -3.17969E-01, 1.84141E-01, -1.52128E+00,
7.37673E-01, -8.05123E-01, 5.60758E-01, -3.86448E-01,
3.19516E-02, -3.51284E-01, 9.36559E-02,     NA, -1.52549E-01,
NA, 5.96768E-01,     NA, -1.25276E+00,     NA, -7.67255E-01,
-5.86445E-01, 4.93876E-01, -6.37310E-01, 3.52167E-01, -
1.09030E+00, 9.40983E-01, -7.14470E-01, 8.20439E-02,     NA,
9.67736E-01,     NA, 3.79929E-01, -1.13498E+00, 1.02326E-01,
-1.16397E+00, 6.41322E-01,     NA, -8.07657E-01,     NA, -
7.99573E-02,     NA, 6.03875E-01,     NA, -3.94654E-01,
NA, 2.40976E-02,     NA, -1.92126E-01,     NA, -3.98348E-01,
NA, -3.88398E-02, -5.98762E-01, 8.64997E-01, -8.25390E-01,
9.09235E-01, -8.83191E-01, -5.06561E-01, -2.83200E-01,
6.79062E-01)
,
a=c(11.2,12.9,28.3,15.2,13.7,16.7,32.1,29.5,20.4,19.2,17.8,16
.3,32.6,20.6),
b=c(27.4,47.4,72,23.4,40.9,40.6,78.3,43.5,55.2,43.4,38.8,37.6
,48.8,39.9),
T1=c( 1, 1, 1, 1, 1, 2, 2, 3, 2, 2, 2, 2, 6, 2, 3, 2, 2, 7,
2, 4, 7, 2, 2, 2, 2, 7,3, 2, 2, 9 ,2, 4 ,2, 3, 2, 2, 3, 2, 2,
4 ,3, 4, 3, 3, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 2,
2, 2, 2, 2, 2, 2, 2, 2, 2),
T2=c(2,2,4,2,2,5,5,5,5,6,6,6,7,6,7,7,7,9,7,7,9,7, 10 ,10 ,10,
10,9,9,9 ,12,9, 12, 12,8 ,13 ,13,7,10, 10, 12,4 ,12 ,12,7,9
,13 ,14 ,12,7,1,1,3,4,4,7,4,4,4,7,7,3,3,3,3,3,5,4),
T3=c( 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
```

```
0,  0,  0,  0,  0,  3 , 4,  5,  5,  6,  9, 10, 10, 10, 11 ,11
, 9 , 8 , 8 ,13 ,13 ,10,  7),
e1r=c(1.55000E+02, 7.20E+01, 8.90E+01, 4.90E+01, 1.10E+02,
2.90E+01, 4.80E+01, 4.60E+01, 5.30E+01, 2.30E+01, 1.90E+01,
4.80E+01,     0.1, 1.29000E+02,     0.1, 1.60E+01, 2.40E+01,
6.80E+01, 5.10E+01, 1.58000E+02,     0.1, 3.90E+01, 5.10E+01,
3.00E+01, 2.90E+01, 8.00E+01, 1.20E+01, 6.00E+01, 2.00E+00,
1.10E+01, 3.00E+01,     0.1, 5.40E+01, 9.00E+00,     0.1,
0.1, 5.20E+01, 5.10E+01,     0.1, 5.00E+00,     0.1,     0.1,
0.1, 5.00E+00,     0.1, 3.20E+01,     0.1,     0.1, 3.00E+01,
5.60E+01, 5.80E+01, 2.60E+01, 3.50E+01, 18, 3.10E+01,
0.1,     0.1, 3.90E+01, 2.60E+01,     0.1, 1.80E+01,     0.1,
0.1,     0.1,     0.1, 3.60E+01, 4.30E+01),
e2r=c(6.3E+01, 4.2E+01, 7.2E+01, 2.3E+01, 4.9E+01, 4.4E+01,
5.9E+01, 6.0E+01, 8.2E+01, 6.5E+01, 6.3E+01, 5.0E+01,
0.1, 2.71000E+02,     0.1, 3.4E+01, 3.5E+01, 5.2E+01,
1.49000E+02, 1.67000E+02,     0.1, 3.7E+01, 1.05000E+02,
4.0E+01, 5.5E+01, 7.2E+01, 9.0E+00, 8.9E+01, 9.0E+00,
8.0E+00, 4.7E+01,     0.1, 1.12000E+02, 8.0E+00,     0.1,
0.1, 6.0E+01, 1.56000E+02,     0.1, 4.0E+00,     0.1,
0.1,     0.1, 3.0E+00,     0.1, 2.6E+01,     0.1,     0.1,
4.0E+01, 7.2E+01, 7.8E+01, 5.2E+01, 5.5E+01, 9.3E+01,
8.2E+01,     0.1,     0.1, 5.9E+01, 9.4E+01,     0.1,
1.8E+01,     0.1,     0.1,     0.1,     0.1, 9.4E+01,
1.3E+01),
e3r=c(    0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1, 7.10000E+01, 8.00000E+01, 5.70000E+01, 4.30000E+01,
6.50000E+01, 7.50000E+01,     0.1,     0.1, 6.50000E+01,
7.80000E+01,     0.1, 3.50000E+01,     0.1,     0.1,     0.1,
0.1, 1.06000E+02, 5.30000E+01),
f1r=c(9.8E+01, 6.5E+01, 8.6E+01, 8.1E+01, 1.57000E+02,
7.1E+01, 5.7E+01, 3.1E+01, 1.08000E+02, 4.7E+01, 4.7E+01,
5.5E+01,     0.1, 9.3E+01,     0.1, 5.3E+01, 3.6E+01,
5.7E+01, 6.4E+01, 6.1E+01,     0.1, 2.1E+01, 9.4E+01,
4.5E+01, 9.6E+01, 8.5E+01, 1.0E+00, 1.25000E+02, 2.0E+01,
4.0E+00, 3.7E+01,     0.1, 1.66000E+02, 6.0E+00,     0.1,
0.1, 1.9E+01, 7.1E+01,     0.1, 4.0E+00,     0.1,     0.1,
0.1, 1.5E+01,     0.1, 1.12000E+02,     0.1,     0.1,
7.1E+01, 1.09000E+02, 9.5E+01, 7.1E+01, 6.6E+01, 7.0E+01,
7.4E+01,     0.1,     0.1, 1.01000E+02, 7.9E+01,     0.1,
5.6E+01,     0.1,     0.1,     0.1,     0.1, 6.9E+01,
5.6E+01),
f2r=c(6.80000E+01, 9.30000E+01, 1.00000E+02, 1.09000E+02,
8.50000E+01, 4.70000E+01, 4.70000E+01, 1.80000E+01,
```

```
7.30000E+01, 7.50000E+01, 7.70000E+01, 5.20000E+01,     0.1,
1.87000E+02,     0.1, 3.60000E+01, 2.00000E+01, 7.40000E+01,
8.00000E+01, 6.70000E+01,     0.1, 2.10000E+01, 4.10000E+01,
3.50000E+01, 7.90000E+01, 9.20000E+01, 5.00000E+00,
1.03000E+02, 1.10000E+01, 7.00000E+00, 2.20000E+01,     0.1,
1.11000E+02, 7.00000E+00,     0.1,     0.1, 9.00000E+00,
1.91000E+02,     0.1, 4.00000E+00,     0.1,     0.1,     0.1,
1.70000E+01,     0.1, 1.17000E+02,     0.1,     0.1,
6.10000E+01, 8.30000E+01, 8.90000E+01, 4.60000E+01,
4.40000E+01, 7.90000E+01, 1.33000E+02,     0.1,     0.1,
8.50000E+01, 9.60000E+01,     0.1, 1.80000E+01,     0.1,
0.1,     0.1,     0.1, 9.90000E+01, 7.00000E+00),
f3r=c(    0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1, 8.90000E+01, 8.50000E+01, 5.00000E+01, 5.90000E+01,
1.13000E+02, 1.26000E+02,     0.1,     0.1, 8.90000E+01,
1.16000E+02,     0.1, 3.40000E+01,     0.1,     0.1,     0.1,
0.1, 8.90000E+01, 5.20000E+01),
e1d=c(5.4E+01, 6.2E+01, 4.1E+01, 7.6E+01, 1.57000E+02,
5.1E+01, 4.3E+01, 1.5E+01, 4.5E+01, 3.9E+01, 3.0E+01,
2.9E+01, 1.0E+01, 4.3E+01, 3.0E+00, 4.5E+01, 3.5E+01,
3.9E+01, 8.2E+01, 7.8E+01,     0.1, 1.8E+01, 4.3E+01,
4.0E+01, 7.3E+01, 3.5E+01,     0.1, 8.9E+01, 8.0E+00,
3.0E+00, 1.1E+01, 1.5E+01, 1.1E+02, 3.0E+00, 2.8E+01,
2.8E+01, 1.5E+01, 5.0E+01,     0.1, 7.0E+00, 4.0E+00,
1.0E+00, 5.0E+00, 2.0E+00, 6.4E+01, 1.3E+01, 2.1E+01,
1.0E+01, 4.1E+01, 8.7E+01, 4.4E+01, 3.0E+01, 4.0E+01,
6.3E+01, 2.8E+01, 2.5E+01, 1.0E+00, 2.1E+01, 4.1E+01,
4.1E+01, 4.7E+01, 3.4E+01, 3.1E+01, 2.9E+01, 1.4E+01,
4.1E+01, 4.7E+01),
e2d=c(2.0E+01, 6.5E+01, 7.7E+01, 1.04000E+02, 8.0E+01,
3.5E+01, 3.5E+01, 5.0E+00, 4.4E+01, 6.6E+01, 5.5E+01,
3.2E+01, 1.2E+01, 1.29000E+02, 1.0E+00, 2.7E+01, 2.1E+01,
4.5E+01, 1.6E+02, 6.8E+01,     0.1, 1.5E+01, 1.6E+01,
2.7E+01, 5.9E+01, 5.4E+01,     0.1, 8.1E+01, 8.0E+00,
3.0E+00, 7.0E+00, 1.6E+01, 9.3E+01, 2.0E+00, 8.7E+01,
4.8E+01, 6.0E+00, 1.32000E+02,     0.1, 2.0E+00, 2.0E+00,
4.0E+00, 1.0E+00, 3.0E+00, 1.22000E+02, 2.5E+01, 2.9E+01,
1.4E+01, 4.3E+01, 8.2E+01, 4.1E+01, 1.4E+01, 2.2E+01,
9.4E+01, 5.6E+01, 2.8E+01, 3.0E+00, 1.4E+01, 3.8E+01,
4.4E+01, 2.2E+01, 2.0E+01, 2.1E+01, 2.9E+01, 2.0E+01,
4.1E+01, 1.2E+01),
e3d=c(    0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
```

```
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1, 8.20000E+01, 4.40000E+01, 1.00000E+01, 3.60000E+01,
1.05000E+02, 5.00000E+01, 1.80000E+01, 2.00000E+00,
1.70000E+01, 7.20000E+01, 6.10000E+01, 3.30000E+01,
3.20000E+01, 3.70000E+01, 6.60000E+01, 1.50000E+01,
4.00000E+01, 3.30000E+01),
f1d=c(1.99000E+02, 7.50000E+01, 1.34000E+02, 5.40000E+01,
1.10000E+02, 4.90000E+01, 6.20000E+01, 6.20000E+01,
1.16000E+02, 3.10000E+01, 3.60000E+01, 7.40000E+01,
5.00000E+00, 1.79000E+02, 1.20000E+01, 2.40000E+01,
2.50000E+01, 8.60000E+01, 3.30000E+01, 1.41000E+02,     0.1,
4.20000E+01, 1.02000E+02, 3.60000E+01, 5.20000E+01,
1.30000E+02,     0.1, 9.60000E+01, 1.40000E+01, 1.20000E+01,
5.60000E+01, 1.00000E+00, 1.10000E+02, 1.20000E+01,
7.20000E+01, 7.80000E+01, 5.60000E+01, 7.20000E+01,     0.1,
2.00000E+00, 3.00000E+00, 9.00000E+00, 1.40000E+01,
1.80000E+01, 1.40000E+01, 1.31000E+02, 3.80000E+01,
1.70000E+01, 6.00000E+01, 7.80000E+01, 1.09000E+02,
6.70000E+01, 6.10000E+01, 2.50000E+01, 7.70000E+01,
2.60000E+01, 1.40000E+01, 1.19000E+02, 6.40000E+01,
5.70000E+01, 2.70000E+01, 6.10000E+01, 4.60000E+01,
8.20000E+01, 9.80000E+01, 6.40000E+01, 5.20000E+01),
f2d=c(1.11000E+02, 7.00000E+01, 9.50000E+01, 2.80000E+01,
5.40000E+01, 5.60000E+01, 7.10000E+01, 7.30000E+01,
1.11000E+02, 7.40000E+01, 8.50000E+01, 7.00000E+01,
2.00000E+00, 3.29000E+02, 1.40000E+01, 4.30000E+01,
3.40000E+01, 8.10000E+01, 6.90000E+01, 1.66000E+02,     0.1,
4.30000E+01, 1.30000E+02, 4.80000E+01, 7.50000E+01,
1.10000E+02,     0.1, 1.11000E+02, 1.20000E+01, 1.20000E+01,
6.20000E+01, 1.00000E+00, 1.30000E+02, 1.30000E+01,
1.27000E+02, 6.10000E+01, 6.30000E+01, 2.15000E+02,     0.1,
6.00000E+00, 5.00000E+00, 6.00000E+00, 1.40000E+01,
1.70000E+01, 2.50000E+01, 1.18000E+02, 3.00000E+01,
1.10000E+01, 5.80000E+01, 7.30000E+01, 1.26000E+02,
8.40000E+01, 7.70000E+01, 7.80000E+01, 1.59000E+02,
2.50000E+01, 1.20000E+01, 1.30000E+02, 1.52000E+02,
1.61000E+02, 1.40000E+01, 1.60000E+01, 5.70000E+01,
8.40000E+01, 9.40000E+01, 1.52000E+02, 8.00000E+00),
f3d=c(    0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1,     0.1,     0.1,     0.1,     0.1,     0.1,     0.1,
0.1, 7.80000E+01, 1.21000E+02, 9.70000E+01, 6.60000E+01,
7.30000E+01, 1.51000E+02, 3.40000E+01, 1.30000E+01,
```

```
1.37000E+02, 1.22000E+02, 1.24000E+02, 3.60000E+01,
5.30000E+01, 3.70000E+01, 1.54000E+02, 1.01000E+02,
1.55000E+02, 7.20000E+01),
t1 = structure(.Data=c(1, 0, 0, 0),.Dim=c( 2 , 2 )),
t2 = structure(.Data=c(0, 0, 0, 1),.Dim=c( 2 , 2 )),
t3 = structure(.Data=c(0, 1, 1, 0),.Dim=c( 2 , 2 )),
delta1 = structure(.Data=c(1, 0, 0.5,
0,0,0,0,0,0.5,0,1,0,0,0,0,0),.Dim=c( 4 , 4 )),
delta2 =
structure(.Data=c(0,0,0,0,0,1,0,0.5,0,0,0,0,0,0.5,0,1),.Dim=c
( 4 , 4 )),
delta3
=structure(.Data=c(0,1,0,0.5,1,0,0.5,0,0,0.5,0,1,0.5,0,1,0),.
Dim=c( 4 , 4 )))
```

XII. OpenBUGS code for the model in Section 4.3.2.1

```
model{
# this controls for studies with one outcome not reported by
setting the correlation equal to zero

for (k in 1:(2*Ns-N2h))
     {control[k]<-step(9999-varr[k])*step(9999-vard[k])}

# two-arm studies
for( k in 1:N2h){
     s[k,1,1]<-varr[k]
     s[k,2,2]<-vard[k]
     s[k,1,2]<-control[k]*rhosigma*sqrt(varr[k]*vard[k])
     s[k,2,1]<-control[k]*rhosigma*sqrt(varr[k]*vard[k])

     prec2A[k,1:2,1:2]<-inverse(s[k,,])
     y[(2*k-1):(2*k)]~dmnorm(theta[(2*k-
1):(2*k)],prec2A[k,,])
          for(i in 1:2){
               for (j in 1:2){
                    D2[k,i,j]<-
                    tau1.sq*t1[i,j]+tau2.sq*t2[i,j]+sqrt(tau1
                    .sq*tau2.sq)*control[k]*(rhotau)*t3[i,j]}
                    }
     prec2B[k,1:2,1:2]<-inverse(D2[k,,])
     theta[(2*k-1):2*k]~dmnorm(mean[(2*k-
1):2*k],prec2B[k,,]) }

# three-arm studies
for (k in 1:(Ns-N2h)){
     rhosigma1T[k]<-max(rhosigma,-ul[k])
     rhosigmaT[k]<-min(rhosigma1T[k],ul[k])
     for (i in 1:4){
          for (j in 1:4){
               S[k,i,j]<-
               sigma1[k,i,j]+control[k]*rhosigmaT[k]*sigma2[k
               ,i,j]}}
     prec3A[k,1:4,1:4]<-inverse(S[k,,])
     y[2*N2h+4*k-3:2*N2h+4*k]~dmnorm(theta[2*N2h+4*k-
     3:2*N2h+4*k],prec3A[k,,])
     for (i in 1:4){
          for (j in 1:4){
               D3[k,i,j]<-
               tau1.sq*delta1[i,j]+tau2.sq*delta2[i,j]+sqrt(t
               au1.sq*tau2.sq)*(control[k]*rhotau)*delta3[i,j
               ]}}
     prec4A[k,1:4,1:4]<-inverse(D3[k,,])
     theta[2*N2h+4*k-3:2*N2h+4*k]~dmnorm(mean[2*N2h+4*k-
```

```
     3:2*N2h+4*k],prec4A[k,,])}

# Parameterization of the means
for(i in 1:N2h) {
     mean[2*i-1] <- -dR[T2[i]] + dR[T1[i]]
     mean[2*i] <- -dD[T2[i]]+ dD[T1[i]]}
for(i in 1:(Ns-N2h)) {
     mean[2*N2h+4*i-3] <- -dR[T2[N2h+i]] + dR[T1[N2h+i]]
     mean[2*N2h+4*i-2] <- -dD[T2[N2h+i]] + dD[T1[N2h+i]]
     mean[2*N2h+4*i-1] <- -dR[T3[N2h+i]]+ dR[T1[N2h+i]]
     mean[2*N2h+4*i] <- -dD[T3[N2h+i]] + dD[T1[N2h+i]] }

# Priors

for(k in 1:(ref-1)) {dR[k] ~ dnorm(0,.01)}
for(k in (ref+1):NT) {dR[k] ~ dnorm(0,.01)}
for(k in 1:(ref-1)) {dD[k] ~ dnorm(0,.01)}
for(k in (ref+1):NT) {dD[k] ~ dnorm(0,.01)}
tau1.sq<-tau1*tau1
tau1~dunif(0,1)
tau2.sq<-tau2*tau2
tau2~dunif(0,1)
rhosigma<-0
rhotau<-0

# Estimated  Effect Sizes
dR[ref]<- 0
for (c in 1:(ref-1)) {Eff.refR[c]<- exp(dR[c] - dR[ref] )}
for (c in (ref+1):NT) {Eff.refR[c]<- exp(dR[c] - dR[ref] )}
for (c in 1:(NT-1)) {
     for (k in (c+1):NT) {EffR[c,k] <- exp(dR[k] - dR[c])}}
dD[ref]<- 0
for (c in 1:(ref-1)) {Eff.refD[c]<- exp(dD[c] - dD[ref] )}
for (c in (ref+1):NT) {  Eff.refD[c]<- exp(dD[c] - dD[ref] )}
for (c in 1:(NT-1)) {
     for (k in (c+1):NT) {EffD[c,k] <- exp(dD[k] - dD[c])}}

# SUCRA rankings
# Ranking of treatments for response. This part is customized
# for the acute mania dataset,
# where one of the treatments was not compared for response.

for(k in 1:13) {ddR[k]<-dR[k]}
for(k in 1:13) {
     orderR[k]<-14- rank(ddR[],k)
     most.effectiveR[k]<-equals(orderR[k],1)
     for(j in 1: 13) {
          effectivenessR[k,j]<- equals(orderR[k],j)
          cumeffectivenessR[k,j]<-
sum(effectivenessR[k,1:j])}}
```

```
for(k in 1:13) {
     SUCRAR[k]<- sum(cumeffectivenessR[k,1:(13-1)]) /(13-1)}

#Ranking of treatments for dropout
for(k in 1:NT) {
     orderD[k]<- rank(dD[],k)
     most.effectiveD[k]<-equals(orderD[k],1)
for(j in 1: NT) {
     effectivenessD[k,j]<- equals(orderD[k],j)
     cumeffectivenessD[k,j]<- sum(effectivenessD[k,1:j])}}
     for(k in 1:NT) {
     SUCRAD[k]<- sum(cumeffectivenessD[k,1:(NT-1)]) /(NT-
1))}}
```

The inputs required for this program are the following:

N2h: the number of two-arm studies.

Ns: the total number of studies.

NT: the number of treatments.

ref: the treatment number for the reference treatment (e.g. placebo).

y: the $2(N2h + 2Ns)$-dimensional vector of observed effects (two for every two-arm study, four for every three-arm). Odd positions correspond to R comparison, even to D. Impute NA when in a study an outcome is missing.

varr: the $(N2h + 2Ns)$-dimensional vector of the variance for every R comparison (one for each odd position in y). For studies with missing data impute a large variance (10,000).

vard: the $(N2h + 2Ns)$- dimensional vector of the variance for every D comparison (one for each even position in y). For studies with missing data impute a large variance (10,000).

T1, T2, T3: these are Ns – dimensional vector of treatments for every study. T1 refers to the first treatment of every study (chosen arbitrarily), T2 to the second. For two arm-studies set T3 = 0.

t1, t2, t3: the (2×2) matrices needed for constructing the heterogeneity variance-covariance matrix for the two-arm studies, $\Delta_{(2\times2)}$ in Equation (2) of the paper:

t1 = structure(.Data=c(1, 0, 0, 0),.Dim=c(2 , 2))

t2 = structure(.Data=c(0, 0, 0, 1),.Dim=c(2 , 2))

t3 = structure(.Data=c(0, 1, 1, 0),.Dim=c(2 , 2))

delta1, delta2, delta3: the (4×4) matrices needed for constructing the heterogeneity variance-covariance matrix $\Delta_{(4\times4)}$ for the three-arm studies, Equation (5) of the paper.

delta1 = structure(.Data=c(1,0,0.5 0,0,0,0,0,0.5,0,1,0,0,0,0,0),.Dim=c(4,4))

delta2 = structure(.Data=c(0,0,0,0,0,1,0,0.5,0,0,0,0,0,0.5,0,1),.Dim=c(4 ,4))

delta3=structure(.Data=c(0,1,0,0.5,1,0,0.5,0,0,0.5,0,1,0.5,0,1,0),.Dim=c(4,4))

sigma1, sigma2: the $(Ns - N2h) \times 4 \times 4$ – dimensional arrays entering Equation (8) for every three-arm study, as computed from the data.

XIII. OpenBUGS code for the model in Section 4.2.3.2

```
model{

# this controls for studies with one outcome not reported by
setting the correlation equal to zero

for (k in 1:(2*Ns-N2h)){control[k]<-step(9999-
varr[k])*step(9999-vard[k])}

# two-arm studies
for( i in 1:N2h){
     s[i,1,1]<-varr[i]+psiR.sq
     s[i,2,2]<-vard[i]+psiD.sq
     s[i,1,2]<-
     control[i]*rho1*sqrt(varr[i]+psiR.sq)*sqrt(vard[i]+psiD.
     sq)
     s[i,2,1]<-s[i,1,2]
     prec2A[i,1:2,1:2]<-inverse(s[i,,])
     y[(2*i-1):(2*i)]~dmnorm(mean[(2*i-
1):(2*i)],prec2A[i,,])}

# three-arm studies
for (i in 1:(Ns-N2h)){
     S[i,1,1]<- varr[N2h+2*i-1]+psiR.sq
     S[i,2,2]<-vard[N2h+2*i-1]+psiD.sq
     S[i,3,3]<- varr[N2h+2*i]+psiR.sq
     S[i,4,4]<-vard[N2h+2*i]+psiD.sq

     S[i,1,2]<- control[i]*rho1*sqrt(S[i,1,1])*sqrt(S[i,2,2])
     S[i,2,1]<- S[i,1,2]
     S[i,1,3]<- control[i]*sqrt(S[i,1,1])*sqrt(S[i,3,3])/2
     S[i,3,1]<- S[i,1,3]
     S[i,1,4]<-
     control[i]*rho1*sqrt(S[i,1,1])*sqrt(S[i,4,4])/2
     S[i,4,1]<- S[i,1,4]
     S[i,2,3]<-
     control[i]*rho1*sqrt(S[i,2,2])*sqrt(S[i,3,3])/2
     S[i,3,2]<- S[i,2,3]
     S[i,2,4]<- control[i]*sqrt(S[i,2,2])*sqrt(S[i,4,4])/2
     S[i,4,2]<- S[i,2,4]
     S[i,4,3]<- control[i]*rho1*sqrt(S[i,4,4])*sqrt(S[i,3,3])
     S[i,3,4]<- S[i,4,3] }
for (k in 1:(Ns-N2h)){
     prec3A[k,1:4,1:4]<-inverse(S[k,,])
     y[2*N2h+4*k-3:2*N2h+4*k]~dmnorm(mean[2*N2h+4*k-
     3:2*N2h+4*k],prec3A[k,,])}

# Parameterization of the means
```

```
for(i in 1:N2h) {
     mean[2*i-1] <- -dR[T2[i]] + dR[T1[i]]
     mean[2*i] <- -dD[T2[i]]+ dD[T1[i]]}
for(i in 1:(Ns-N2h)) {
     mean[2*N2h+4*i-3] <- -dR[T2[N2h+i]] + dR[T1[N2h+i]]
     mean[2*N2h+4*i-2] <- -dD[T2[N2h+i]] + dD[T1[N2h+i]]
     mean[2*N2h+4*i-1] <- -dR[T3[N2h+i]]+ dR[T1[N2h+i]]
     mean[2*N2h+4*i] <- -dD[T3[N2h+i]] + dD[T1[N2h+i]] }

# Priors
for(k in 1:(ref-1)) {    dR[k] ~ dnorm(0,.01)}
for(k in (ref+1):NT) {dR[k] ~ dnorm(0,.01)}
for(k in 1:(ref-1)) {    dD[k] ~ dnorm(0,.01)}
for(k in (ref+1):NT) {dD[k] ~ dnorm(0,.01)}
psiR.sq<-psi1*psi1
psi1~dunif(0,1)
psiD.sq<-psi2*psi2
psi2~dunif(0,1)
rho1~dunif(-1,0)

#Estimated  Effect Sizes
dR[ref]<- 0
for (c in 1:(ref-1)) { Eff.refR[c]<- exp(dR[c] - dR[ref] )}
for (c in (ref+1):NT) { Eff.refR[c]<- exp(dR[c] - dR[ref] )
     }
for (c in 1:(NT-1)) {
     for (k in (c+1):NT) {
          EffR[c,k] <- exp(dR[k] - dR[c])}}
dD[ref]<- 0
for (c in 1:(ref-1)) { Eff.refD[c]<- exp(dD[c] - dD[ref] )}
for (c in (ref+1):NT) { Eff.refD[c]<- exp(dD[c] - dD[ref] )
     }
for (c in 1:(NT-1)) {
     for (k in (c+1):NT) {
          EffD[c,k] <- exp(dD[k] - dD[c]) }}

# SUCRA rankings
# Ranking of treatments for response. This part of the code is
# adjusted for the acute mania dataset
for (k in 1:13){ddR[k]<-dR[k]}
for(k in 1:13) {
     orderR[k]<-13- rank(ddR[],k)
     most.effectiveR[k]<-equals(orderR[k],1)
for(j in 1: 13) {
     effectivenessR[k,j]<- equals(orderR[k],j)
     cumeffectivenessR[k,j]<- sum(effectivenessR[k,1:j])}}

for(k in 1:13) {SUCRAR[k]<- sum(cumeffectivenessR[k,1:(13-
1)]) /(13-1)}
```

```
#Ranking of treatments for dropout

for(k in 1:NT) {
     orderD[k]<- rank(dD[],k)
     most.effectiveD[k]<-equals(orderD[k],1)
     for(j in 1: NT) {
          effectivenessD[k,j]<- equals(orderD[k],j)
          cumeffectivenessD[k,j]<-
sum(effectivenessD[k,1:j])}}
for(k in 1:NT) {
     SUCRAD[k]<- sum(cumeffectivenessD[k,1:( NT-1)]) /(NT-
1)}}
```

The data needed as inputs for this program are y, varr, vard and T1, T2, T3, described in the end of the previous Section.

Bibliography

1. Institute of Medicine (US) Roundtable on Evidence-Based Medicine. Leadership Commitments to Improve Value in Healthcare: Finding Common Ground: Workshop Summary [Internet]. Washington (DC): National Academies Press (US); 2009 [cited 2017 Mar 1]. (The National Academies Collection: Reports funded by National Institutes of Health). Available from: http://www.ncbi.nlm.nih.gov/books/NBK52851/

2. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. Stat Med. 2008;27(5):625–650.

3. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005 Oct 15;331(7521):897–900.

4. Higgins JPT, Whitehead A. Borrowing Strength from External Trials in a Meta-Analysis. Stat Med. 1996;15(24):2733–2749.

5. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct 30;23(20):3105–24.

6. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. Stat Methods Med Res. 2008 Jun;17(3):279–301.

7. Coleman CI, Phung OJ, Cappelleri JC, Baker WL, Kluger J, White CM, et al. Use of mixed treatment comparison in systematic reviews. (Prepared by the university of Connecticut/Harford Hospital Evidence-Based Practice Center under Contract No. 290-2007-10067-I). AHRQ. 2012;Publication No. 12-EHC119-EF. Rockville, MD: Agency for Healthcare Research and Quality.

8. Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. Characteristics of networks of interventions: a description of a database of 186 published networks. PloS One. 2014 Jan 22;9(1):e86754.

9. Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. Health TechnolAssess. 2005 Jul;9(1366–5278 (Print)):1–iv.

10. D'Ascenzo F, Biondi-Zoccai G. Network Meta-Analyses: The "White Whale" for Cardiovascular Specialists. J Cardiothorac Vasc Anesth. 2014 Feb;28(1):169–73.

11. Donegan S, Williamson P, Gamble C, Tudur Smith C. Indirect comparisons: a review of reporting and methodological quality. PLoSOne. 2010;5(1932–6203 (Electronic)):e11054.

12. Sobieraj DM, Cappelleri JC, Baker WL, Phung OJ, White CM, Coleman CI. Methods used to conduct and report Bayesian mixed treatment comparisons published in the medical literature: a systematic review. BMJ Open. 2013;3(2044–6055 (Electronic)).

13. NICE DSU Decision Support Unit - Evidence Synthesis TSD series [Internet]. [cited 2017 Mar 1]. Available from: http://www.nicedsu.org.uk/Evidence-Synthesis-TSD-series%282391675%29.htm

14. Cooper NJ, Peters J, Lai MC, Juni P, Wandel S, Palmer S, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? ValueHealth. 2011 Mar;14(1524–4733 (Electronic)):371–80.

15. Cucherat M, Izard V. Summary report: Indirect comparisons: Methods and validity. HAS. 2009;

16. EUnetHTA – European network for Health Technology Assessment (last). Comparators and comparisons: Direct and indirect comparisons. Available Httpwwweunethtaeusites5026fedimbobelgiumbefilesDirect20and20indirect20comparisonspdf. 2013;

17. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2011 Jun;14(4):417–28.

18. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ. 2009 Oct 13;181(1488–2329 (Electronic)):488–93.

19. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. ValueHealth. 2008 Sep;11(1524–4733 (Electronic)):956–64.

20. Malone DC. Using indirect comparisons in pharmacoeconomic studies--time for implementation. ClinTher. 2007 Nov;29(0149–2918 (Print)):2454–5.

21. Mills EJ, Bansback N, Ghement I, Thorlund K, Kelly S, Puhan MA, et al. Multiple treatment comparison meta-analyses: a step forward into complexity. ClinEpidemiol. 2011;3(1179–1349 (Electronic)):193–202.

22. Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. BMJ. 2013;346(1756–1833 (Electronic)):f2914.

23. Naci H, Fleurence R. Using indirect evidence to determine the comparative effectiveness of prescription drugs: Do benefits outweight risks? HealthOutcResMed. 2011;2(4):241–9.

24. 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 Methods. 2012;3(2):80–97.

25. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics. 2008;26(1170–7690

(Print)):753–67.

26. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect evidence: Indirect treatment comparisons in meta-analysis. Ott Can Agency Drugs Technol Health. 2009;

27. Welton NJ, Sutton AJ, Cooper N, Abrams KR, Ades AE. Mixed and indirect treatment comparisons. In: Evidence synthesis for decision making in healthcare. New York: John Wiley & Sons; 2012.

28. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. ValueHealth. 2011 Jun;14(1524–4733 (Electronic)):429–37.

29. Cipriani A, Barbui C, Rizzo C, Salanti G. What is a multiple treatments meta-analysis? EpidemiolPsychiatrSci. 2012 Jun;21(2045–7960 (Print)):151–3.

30. Sutton AJ, Cooper NJ, Jones DR. Evidence synthesis as the key to more coherent and efficient research. BMCMedResMethodol. 2009;9(1471–2288 (Electronic)):29.

31. O'Regan C, Ghement I, Eyawo O, Guyatt GH, Mills EJ. Incorporating multiple interventions in meta-analysis: an evaluation of the mixed treatment comparison with the adjusted indirect comparison. Trials. 2009;10(1745–6215 (Electronic)):86.

32. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ. 2003 Mar 1;326(1756–1833 (Electronic)):472.

33. Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. J Clin Epidemiol. 2008 May;61(5):455–63.

34. Jonas DE, Wilkins TM, Bangdiwala S, Bann CM, Morgan LC, Thaler KJ, et al. Findings of Bayesian mixed treatment comparison meta-analysis: comparison and exploration using real-word trial data and simulation. (Prepared by RTI-UNC Evidence-Based Practice Center under Contract No. 290-2007-10056-I). AHRQ. 2013;Publication No. 13-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality.

35. Madan J, Stevenson MD, Cooper KL, Ades AE, Whyte S, Akehurst R. Consistency between direct and indirect trial evidence: is direct evidence always more reliable? ValueHealth. 2011 Sep;14(1524–4733 (Electronic)):953–60.

36. 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 Sep 12;12(1):138.

37. Falissard B, Zylberman M, Cucherat M, Izard V, Meyer F. Real medical benefit assessed by indirect comparison. Therapie. 2009 May;64(0040–5957 (Print)):225–

32.

38. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011 Feb;64(2):163–71.

39. Thorlund K, Zafari Z, Druyts E, Mills EJ, Sadatsafavi M. The impact of incorporating Bayesian network meta-analysis in cost-effectiveness analysis - a case study of pharmacotherapies for moderate to severe COPD. Cost Eff Resour Alloc CE. 2014 Mar 13;12(1):8.

40. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org. [Internet]. 2011. Available from: Available from www.cochrane-handbook.org.

41. Wiley: Introduction to Meta-Analysis - Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, et al [Internet]. [cited 2017 Mar 16]. Available from: http://www.wiley.com/WileyCDA/WileyTitle/productCd-EHEP002313.html

42. Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. Evid Based Ment Health. 2014 May;17(2):53–7.

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

44. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986 Sep;7(3):177–88.

45. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods. 2016 Mar 1;7(1):55–79.

46. Dias S, Sutton AJ, Ades AE, Welton NJ. 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 Jul;33(5):607–17.

47. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. Res Synth Methods. 2013 Jul 1;n/a-n/a.

48. Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, et al. Issues in methodological research: perspectives from researchers and commissioners. Health Technol Assess Winch Engl. 2001;5(8):1–57.

49. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. JClinEpidemiol. 1997 Jun;50(0895–4356 (Print)):683–91.

50. Gartlehner G, Moore CG. Direct versus indirect comparisons: a summary of the evidence. IntJTechnolAssessHealth Care. 2008;24(0266–4623 (Print)):170–7.

51. Edwards SJ, Clarke MJ, Wordsworth S, Borrill J. Indirect comparisons of treatments based on systematic reviews of randomised controlled trials. Int J Clin Pract. 2009;63(6):841–854.

52. Ioannidis JP. Indirect comparisons: the mesh and mess of clinical trials. Lancet. 2006 Oct 28;368(1474–547X (Electronic)):1470–2.

53. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. AnnInternMed. 2008 Apr 1;148(1539–3704 (Electronic)):544–53.

54. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010 Mar 30;29(1097–0258 (Electronic)):932–44.

55. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ. 2009;338(1756–1833 (Electronic)):b1147.

56. Baker SG, Kramer BS. The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? BMCMedResMethodol. 2002 Nov 13;2(1471–2288 (Electronic)):13.

57. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. MedDecisMaking. 2013 Jul;33(1552–681X (Electronic)):641–56.

58. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMCMed. 2013;11(1741–7015 (Electronic)):159.

59. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. Res Synth Methods. 2013 Dec 1;4(4):291–323.

60. Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med. 2013 Jul 16;159(2):130–7.

61. Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. JClinEpidemiol. 2009 Aug;62(1878–5921 (Electronic)):857–64.

62. Julious SA, Wang SJ. How biased are indirect comparisons, particularly when comparisons are made over time in controlled trials. DrugInformJ. 2008;42(6):625.

63. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics. 2009 Oct;10(1468–4357 (Electronic)):792–805.

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

65. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med. 2002 Aug 30;21(16):2313–24.

66. Coory M, Jordan S. Frequency of treatment-effect modification affecting indirect comparisons: a systematic review. Pharmacoeconomics. 2010;28(1179–2027 (Electronic)):723–32.

67. Dewilde S, Hawkins N. Investigating incoherence gives insight: clopidogrel is equivalent to extended-release dipyridamole plus aspirin in secondary stroke prevention. JClinEpidemiol. 2012 Aug;65(1878–5921 (Electronic)):835–45.

68. Griffin S, Bojke L, Main C, Palmer S. Incorporating direct and indirect evidence using bayesian methods: an applied case study in ovarian cancer. ValueHealth. 2006 Mar;9(1098–3015 (Print)):123–31.

69. van Valkenhoef G, Tervonen T, Zhao J, de BB, Hillege HL, Postmus D. Multicriteria benefit-risk assessment using network meta-analysis. JClinEpidemiol. 2012 Apr;65(1878–5921 (Electronic)):394–403.

70. Xiong T, Parekh-Bhurke S, Loke YK, Abdelhamid A, Sutton AJ, Eastwood AJ, et al. Overall similarity and consistency assessment scores are not sufficiently accurate for predicting discrepancy between direct and indirect comparison estimates. JClinEpidemiol. 2013 Feb;66(1878–5921 (Electronic)):184–91.

71. Franchini AJ, Dias S, Ades AE, Jansen JP, Welton NJ. Accounting for correlation in network meta-analysis with multi-arm trials. ResSynthMeth. 2012;3(2):142–60.

72. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. Stat Med. 1998;17(22):2537–2550.

73. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. Stat Med. 2010;29(12):1282–1297.

74. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med. 2002 Feb 28;21(4):589–624.

75. Thorlund K, Thabane L, Mills EJ. Modelling heterogeneity variances in multiple treatment comparison meta-analysis--are informative priors the better solution? BMCMedResMethodol. 2013;13(1471–2288 (Electronic)):2.

76. Lu G, Welton NJ, Higgins JPT, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. ResSynthMeth. 2011;2(1):43–60.

77. White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in

network meta-analysis: model estimation using multivariate meta-regression. ResSynthMeth. 2012;3(2):111–25.

78. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. Stat Methods Med Res. 2013 Apr 1;22(2):133–58.

79. Rücker G. Network meta-analysis, electrical networks and graph theory. ResSynthMeth. 2012;3(4):312–24.

80. Yang G, Liu D, Liu RY, Xie M, Hoaglin DC. Efficient network meta-analysis: A confidence distribution approach. Stat Methodol [Internet]. 2014 Mar 14 [cited 2014 Mar 14]; Available from: http://www.sciencedirect.com/science/article/pii/S1572312714000136

81. Greco T, Landoni G, Biondi-Zoccai G, D'Ascenzo F, Zangrillo A. A Bayesian network meta-analysis for binary outcome: how to do it. StatMethods MedRes. 2013 Aug 22;(1477–0334 (Electronic)).

82. Ohlssen D, Price KL, Amy XH, Hong H, Kerman J, Fu H, et al. Guidance on the implementation and reporting of a drug safety Bayesian network meta-analysis. PharmStat. 2013 Aug 30;(1539–1612 (Electronic)).

83. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates. StatMed. 2012 Dec 20;31(1097–0258 (Electronic)):3840–57.

84. 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. JClinEpidemiol. 2010 Aug;63(1878–5921 (Electronic)):875–82.

85. van Valkenhoef G, Tervonen T, de Brock B, Hillege H. Algorithm parameterization of mixed treatment comparisons. StatComput. 2012;22(5):1099–111.

86. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. ResSynthMeth. 2012;3(2):98–110.

87. Piepho HP, Williams ER, Madden LV. The use of two-way linear mixed models in multitreatment meta-analysis. Biometrics. 2012;68(4):1269–77.

88. Caldwell DM, Welton NJ, Dias S, Ades AE. Selecting the best scale for measuring treatment effect in a network meta-analysis: a case study in childhood nocturnal enuresis. ResSynthMeth. 2012;3(2):126–41.

89. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Ser B Stat Methodol. 2002 Oct 1;64(4):583–639.

90. Krahn U, Binder H, Konig J. A graphical tool for locating inconsistency in network meta-analyses. BMCMedResMethodol. 2013;13(1471–2288 (Electronic)):35.

91. Chung H, Lumley T. Graphical exploration of network meta-analysis data: the use of multidimensional scaling. ClinTrials. 2008;5(1740–7745 (Print)):301–7.

92. Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. BMJ. 2011;343(1756–1833 (Electronic)):d4909.

93. Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G. Evaluation of inconsistency in networks of interventions. IntJEpidemiol. 2013 Feb;42(1464–3685 (Electronic)):332–45.

94. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique and future directions. Stat Med. 2009;28(25):3049–3067.

95. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian Modelling Framework: Concepts, Structure, and Extensibility. Stat Comput. 2000 Oct;10(4):325–337.

96. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. 2003.

97. Achana FA, Cooper NJ, Dias S, Lu G, Rice SJ, Kendrick D, et al. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. StatMed. 2013 Feb 28;32(1097–0258 (Electronic)):752–71.

98. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. StatMed. 2003 Oct 15;22(0277–6715 (Print)):2995–3016.

99. Ades AE, Mavranezouli I, Dias S, Welton NJ, Whittington C, Kendall T. Network meta-analysis with competing risk outcomes. ValueHealth. 2010 Dec;13(1524–4733 (Electronic)):976–83.

100. Carlin BP, Hong H, Shamliyan TA, Sainfort F, Kane RL. Case study comparing bayesian and frequentist approaches for multiple treatment comparisons. (Prepared by the Minnesota Evidence-Based Practice Center under Contract No. 290-2007-10064-I2). AHRQ. 2013;Publication No. 12(13)-EHC103-EF. Rockville, MD: Agency for Healthcare Research and Quality.

101. Chaimani A, Vasiliadis HS, Pandis N, Schmid CH, Welton NJ, Salanti G. Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. Int J Epidemiol. 2013 Aug;42(4):1120–31.

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

103. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med. 2009 Jun 30;28(1097–0258 (Electronic)):1861–81.

104. Cope S, Jansen JP. Quantitative summaries of treatment effect estimates obtained with network meta-analysis of survival curves to inform decision-making. BMC Med Res Methodol. 2013 Dec 1;13(1):147.

105. Dakin HA, Welton NJ, Ades AE, Collins S, Orme M, Kelly S. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. StatMed. 2011 Jul 5;(1097–0258 (Electronic)).

106. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. MedDecisMaking. 2013 Jul;33(1552–681X (Electronic)):618–40.

107. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. MedDecisMaking. 2013 Jul;33(1552–681X (Electronic)):657–70.

108. Dias S, Welton NJ, Ades AE. Study designs to detect sponsorship and other biases in systematic reviews. JClinEpidemiol. 2010 Jun;63(1878–5921 (Electronic)):587–8.

109. Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. JRoyStatistSocA. 2010;173(3):613–29.

110. Ding Y, Fu H. Bayesian indirect and mixed treatment comparisons across longitudinal time points. StatMed. 2012 Dec 11;(1097–0258 (Electronic)).

111. Donegan S, Williamson P, D'Alessandro U, Garner P, Tudur Smith C. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. StatMed. 2013 Mar 15;32(1097–0258 (Electronic)):914–30.

112. Del Giovane C, Vacchi L, Mavridis D, Filippini G, Salanti G. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. Stat Med. 2013 Jan 15;32(1097–0258 (Electronic)):25–39.

113. Govan L, Ades AE, Weir CJ, Welton NJ, Langhorne P. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. StatMed. 2010 May 30;29(1097–0258 (Electronic)):1340–56.

114. Hong H, Carlin BP, Chu H, Shamliyan TA, Wang S, Kane RL. A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 [cited 2013 Dec 3]. (AHRQ Methods for Effective Health Care). Available from:

http://www.ncbi.nlm.nih.gov/books/NBK116689/

115. Hong H, Carlin BP, Shamliyan TA, Wyman JF, Ramakrishnan R, Sainfort F, et al. Comparing Bayesian and Frequentist Approaches for Multiple Outcome Mixed Treatment Comparisons. MedDecisMaking. 2013 Apr 2;(1552–681X (Electronic)).

116. Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMCMedResMethodol. 2011;11(1471–2288 (Electronic)):61.

117. Jansen JP, Cope S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. BMCMedResMethodol. 2012;12(1471–2288 (Electronic)):152.

118. Jones B, Roger J, Lane PW, Lawton A, Fletcher C, Cappelleri JC, et al. Statistical approaches for conducting network meta-analysis in drug development. PharmStat. 2011 Nov;10(1539–1612 (Electronic)):523–31.

119. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. StatMed. 2007 Sep 10;26(0277–6715 (Print)):3681–99.

120. Madan J, Chen Y-F, Aveyard P, Wang D, Yahaya I, Munafo M, 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 Stat Soc. 2014;177(1):295–314.

121. Mavridis D, Sutton A, Cipriani A, Salanti G. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. StatMed. 2013 Jan 15;32(1097–0258 (Electronic)):51–66.

122. Mills EJ, Kanters S, Thorlund K, Chaimani A, Veroniki AA, Ioannidis JP. The effects of excluding treatments from network meta-analyses: survey. BMJ. 2013;347(1756–1833 (Electronic)):f5195.

123. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. StatMed. 2007 Mar 15;26(0277–6715 (Print)):1237–54.

124. Ouwens MJNM, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. ResSynthMeth. 2010;1(3–4):258–71.

125. Price MJ, Welton NJ, Ades AE. Parameterization of treatment effects for meta-analysis in multi-state Markov models. Stat Med. 2011 Jan 30;30(1097–0258 (Electronic)):140–51.

126. Price MJ, Welton NJ, Briggs AH, Ades AE. Model averaging in the presence of structural uncertainty about treatment effects: influence on treatment decision and expected value of information. ValueHealth. 2011 Mar;14(1524–4733 (Electronic)):205–18.

127. Salanti G, Dias S, Welton NJ, Ades AE, Golfinopoulos V, Kyrgiou M, et al. Evaluating novel agent effects in multiple-treatments meta-regression. StatMed. 2010 Oct 15;29(1097–0258 (Electronic)):2369–83.

128. Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. StatMed. 2012 Dec 10;31(1097–0258 (Electronic)):3516–36.

129. Schmid CH, Trikalinos TA, Olkin I. Bayesian network meta-analysis for unordered categorical outcomes with incomplete data. Res Synth Methods. 2013 Dec 1;n/a-n/a.

130. Schmidli H, Wandel S, Neuenschwander B. The network meta-analytic-predictive approach to non-inferiority trials. Stat Methods Med Res. 2013 Apr 1;22(2):219–40.

131. Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. StatMed. 2013 Feb 25;(1097–0258 (Electronic)).

132. Schmitz S, Adams R, Walsh C. The use of continuous data versus binary data in MTC models: a case study in rheumatoid arthritis. BMCMedResMethodol. 2012;12(1471–2288 (Electronic)):167.

133. Soares MO, Dumville JC, Ades AE, Welton NJ. Treatment comparisons for decision making: facing the problems of sparse and few data. JRoyStatistSocA. 2013;

134. Spineli LM, Higgins JP, Cipriani A, Leucht S, Salanti G. Evaluating the impact of imputations for missing participant outcome data in a network meta-analysis. Clin Trials Lond Engl. 2013;10(3):378–88.

135. Sturtz S, Bender R. Unsolved issues of mixed treatment comparison meta-analysis: network size and inconsistency. ResSynthMeth. 2012;3(4):300–11.

136. Thorlund K, Zafari Z, Druyts E, Mills EJ, Sadatsafavi M. The impact of incorporating Bayesian network meta-analysis in cost-effectiveness analysis - a case study of pharmacotherapies for moderate to severe COPD. Cost Eff Resour Alloc CE. 2014 Mar 13;12(1):8.

137. Thorlund K, Mills E. Stability of additive treatment effects in multiple treatment comparison meta-analysis: a simulation study. ClinEpidemiol. 2012;4(1179–1349 (Electronic)):75–85.

138. Trinquart L, Abbe A, Ravaud P. Impact of reporting bias in network meta-analysis of antidepressant placebo-controlled trials. PLoSOne. 2012;7(1932–6203 (Electronic)):e35219.

139. Trinquart L, Chatellier G, Ravaud P. Adjustment for reporting bias in network meta-analysis of antidepressant trials. BMCMedResMethodol. 2012;12(1471–2288 (Electronic)):150.

140. Tu YK, Faggion CM. A primer on network meta-analysis for dental research.

ISRNDent. 2012;2012(2090–438X (Electronic)):276520.

141. van Valkenhoef G, Lu G, de Brock B, Hillege H, Ades AE, Welton NJ. Automating network meta-analysis. ResSynthMeth. 2012;3(4):285–99.

142. Warren FC, Abrams KR, Sutton AJ. Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. Stat Med. 2014 Mar 13;

143. Welton NJ, Willis SR, Ades AE. Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. ResSynthMeth. 2010;1(3–4):239–57.

144. Welton NJ, Caldwell DM, Adamopoulos E, Vedhara K. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. AmJEpidemiol. 2009 May 1;169(1476–6256 (Electronic)):1158–65.

145. Welton NJ, Cooper NJ, Ades AE, Lu G, Sutton AJ. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. StatMed. 2008 Nov 29;27(0277–6715 (Print)):5620–39.

146. Woods BS, Hawkins N, Scott DA. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: a tutorial. BMCMedResMethodol. 2010;10(1471–2288 (Electronic)):54.

147. Zhang J, Carlin BP, Neaton JD, Soon GG, Nie L, Kane R, et al. Network meta-analysis of randomized clinical trials: Reporting the proper summaries. ClinTrials. 2013 Oct 3;(1740–7753 (Electronic)).

148. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0. 2014.

149. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org; 2014. Available from: http://www.R-project.org

150. Jansen JP. Network meta-analysis of individual and aggregate level data. ResSynthMeth. 2012;3(2):177–90.

151. Mills EJ, Ghement I, O'Regan C, Thorlund K. Estimating the power of indirect comparisons: a simulation study. PLoSOne. 2011;6(1932–6203 (Electronic)):e16237.

152. Naudet F, Millet B, Charlier P, Reymann JM, Maria AS, Falissard B. Which placebo to cure depression? A thought-provoking network meta-analysis. BMCMed. 2013 Oct 25;11(1741–7015 (Electronic)):230.

153. Rücker G, Schwarzer G. Reduce dimension or reduce weights? Comparing two

approaches to multi-arm studies in network meta-analysis. Stat Med. 2014 Nov 10;33(25):4353–69.

154. Tan SH, Cooper NJ, Bujkiewicz S, Welton NJ, Caldwell DM, Sutton AJ. Novel presentational approaches were developed for reporting network meta-analysis. J Clin Epidemiol [Internet]. 2014 Feb 28 [cited 2014 Feb 28]; Available from: http://www.sciencedirect.com/science/article/pii/S0895435613004800

155. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP. 2013.

156. 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. JClinPsychopharmacol. 2005 Apr;25(0271–0749 (Print)):127–31.

157. Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical Tools for Network Meta-Analysis in STATA. PloS One. 2013;8(10):e76654.

158. Norton EC, Miller MM, Wang JJ, Coyne K, Kleinman LC. Rank reversal in indirect comparisons. ValueHealth. 2012 Dec;15(1524–4733 (Electronic)):1137–40.

159. SAS Institute. The SAS system for Windows. Release 9.2. SAS Inst.., Cary, NC. 2011.

160. Piepho H-P. Network-meta analysis made easy: Detection of inconsistency using factorial analysis-of-variance models. BMC Med Res Methodol. 2014 May 10;14(1):61.

161. Piepho H-P, Madden LV, Williams ER. Multiplicative interaction in network meta-analysis. Stat Med. 2015 Feb 20;34(4):582–94.

162. Signorovitch J, Erder MH, Xie J, Sikirica V, Lu M, Hodgkins PS, et al. Comparative effectiveness research using matching-adjusted indirect comparison: an application to treatment with guanfacine extended release or atomoxetine in children with attention-deficit/hyperactivity disorder and comorbid oppositional defiant disorder. Pharmacoepidemiol Saf. 2012 May;21 Suppl 2(1099–1557 (Electronic)):130–7.

163. Neupane B, Richer D, Bonner AJ, Kibret T, Beyene J. Network Meta-Analysis Using R: A Review of Currently Available Automated Packages. PLoS ONE. 2014 Dec 26;9(12):e115065.

164. Gleser L, Olkin LV. Stochastically dependent effect sizes. In The Handbook of Research Synthesis and Meta-Analysis, (Cooper H, Hedges LV (eds). Russell Sage Foundation; 2009. 633 p.

165. Eckermann S, Coory M, Willan AR. Indirect comparison: relative risk fallacies and odds solution. JClinEpidemiol. 2009 Oct;62(1878–5921 (Electronic)):1031–6.

166. van Valkenhoef G, Ades AE. Evidence synthesis assumes additivity on the scale of

measurement: response to “Rank reversal in indirect comparisons” by Norton et al. ValueHealth. 2013 Mar;16(1524–4733 (Electronic)):449–51.

167. Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc. 2009 Jan 1;172(1):21–47.

168. Copas J. What works?: selectivity models and meta-analysis. J R Stat Soc Ser A Stat Soc. 1999 Jan 1;162(1):95–109.

169. Jansen JP, Schmid CH, Salanti G. Directed acyclic graphs can help understand bias in indirect and mixed treatment comparisons. JClinEpidemiol. 2012 Jul;65(1878–5921 (Electronic)):798–807.

170. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. JClinEpidemiol. 2013 Sep 30;(1878–5921 (Electronic)).

171. Bafeta A, Trinquart L, Seror R, Ravaud P. Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review. BMJ. 2013;347(1756–1833 (Electronic)):f3675.

172. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer’s checklist. MedDecisMaking. 2013 Jul;33(1552–681X (Electronic)):679–91.

173. Mills EJ, Ioannidis JP, Thorlund K, Schunemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. JAMA. 2012 Sep 26;308(1538–3598 (Electronic)):1246–53.

174. Senn S, Gavini F, Magrez D, Scheen A. Issues in performing a network meta-analysis. StatMethods MedRes. 2013 Apr;22(1477–0334 (Electronic)):169–89.

175. Fadda V, Maratea D, Trippoli S, Messori A. Results can be summarised in a simple figure. BMJ. 2011 Mar 23;342(mar23 2):d1555–d1555.

176. König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Stat Med. 2013 Oct 1;n/a-n/a.

177. Tan SH, Bujkiewicz S, Sutton A, Dequen P, Cooper N. Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons. JHealth ServResPolicy. 2013 Oct;18(1758–1060 (Electronic)):224–32.

178. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Stat Med. 2012;31(20):2179–2195.

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

180. Li T, Puhan MA, Vedula SS, Singh S, Dickersin K, The Ad Hoc Network Meta-

analysis Methods Meeting Working Group. Network meta-analysis-highly attractive but more methodological research is needed. BMC Med. 2011 Jun 27;9(1):79.

181. Hawkins N, Scott DA, Woods BS, Thatcher N. No study left behind: a network meta-analysis in non-small-cell lung cancer demonstrating the importance of considering all relevant data. ValueHealth. 2009 Sep;12(1524–4733 (Electronic)):996–1003.

182. Hawkins N, Scott DA, Woods B. How far do you go? Efficient searching for indirect evidence. MedDecisMaking. 2009 May;29(0272–989X (Print)):273–81.

183. Mills EJ, Thorlund K, Ioannidis JP. Calculating additive treatment effects from multiple randomized trials provides useful estimates of combination therapies. JClinEpidemiol. 2012 Dec;65(1878–5921 (Electronic)):1282–8.

184. Debray TPA, Moons KGM, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RHH, et al. Methods for Individual Participant Data meta-analysis of relative treatment effects: a systematic review. Res Synth Methods. 2015;Submitted.

185. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. ValueHealth. 2011 Jun;14(1524–4733 (Electronic)):417–28.

186. Ali S, Mealing S, Hawkins N, Lescrauwaet B, Bjork S, Mantovani L, et al. The use of individual patient-level data (IPD) to quantify the impact of pretreatment predictors of response to treatment in chronic hepatitis B patients. BMJ Open. 2013;3(2044–6055 (Electronic)).

187. Ades AE, Sutton AJ. Multiparameter evidence synthesis in epidemiology and medical-decision making: current approaches. JRoyStatistSocA. 2006;169(1):5–35.

188. Thorlund K, Mills EJ. Sample size and power considerations in network meta-analysis. SystRev. 2012;1(2046–4053 (Electronic)):41.

189. Snapinn S, Jiang Q. Indirect comparisons in the comparative efficacy and non-inferiority settings. PharmStat. 2011 Sep;10(1539–1612 (Electronic)):420–6.

190. Naci H, O'Connor AB. Assessing comparative effectiveness of new drugs before approval using prospective network meta-analyses. J Clin Epidemiol. 2013 Aug;66(8):812–6.

191. Petropoulou M, Nikolakopoulou A, Veroniki A-A, Rios P, Vafaei A, Zarin W, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. J Clin Epidemiol. 2017 Feb;82:20–8.

192. Efthimiou O, Debray TPA, van Valkenhoef G, Trelle S, Panayidou K, Moons KGM, et al. Get real in network meta-analysis: a review of the methodology. Submitt Res Synth Methods.

193. Ishak KJ, Platt RW, Joseph L, Hanley JA. Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. Stat Med. 2008;27(5):670–686.

194. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. J R Stat Soc Ser A Stat Soc. 2009;172(4):789–811.

195. Trikalinos TA, Hoaglin DC, Schmid CH. Empirical and Simulation-Based Comparison of Univariate and Multivariate Meta-Analysis for Binary Outcomes [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 [cited 2013 Dec 3]. (AHRQ Methods for Effective Health Care). Available from: http://www.ncbi.nlm.nih.gov/books/NBK132562/

196. Chen H, Manning AK, Dupuis J. A Method of Moments Estimator for Random Effect Multivariate Meta-Analysis. Biometrics. 2012;68(4):1278–1284.

197. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Res Synth Methods. 2010;1(1):39–65.

198. Jackson D, White IR, Riley RD. A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. Biom J Biom Z. 2013 Mar;55(2):231–45.

199. Ma Y, Mazumdar M. Multivariate meta-analysis: a robust approach based on the theory of U-statistic. Stat Med. 2011;30(24):2911–2929.

200. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Stat Med. 2007 Jan 15;26(1):78–97.

201. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. Biostatistics. 2008 Jan 1;9(1):172–86.

202. Wei Y, Higgins JPT. Bayesian multivariate meta-analysis with multiple outcomes. Stat Med. 2013 Jul 30;32(17):2911–34.

203. Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DPM, et al. Multivariate meta-analysis of mixed outcomes: a Bayesian approach. Stat Med. 2013 Sep 30;32(22):3926–43.

204. Garthwaite PH, Kadane JB, O'Hagan A. Statistical Methods for Eliciting Probability Distributions. J Am Stat Assoc. 2005;100(470):680–701.

205. Gokhale DV, Press SJ. Assessment of a Prior Distribution for the Correlation Coefficient in a Bivariate Normal Distribution. J R Stat Soc Ser Gen. 1982 Jan 1;145(2):237–49.

206. Cipriani A, Barbui C, Salanti G, Rendell J, Brown R, Stockton S, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments

meta-analysis. Lancet. 2011 Oct 8;378(9799):1306–15.

207. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. Stat Med. 1993;12(24):2273–2284.

208. Bagos PG. On the covariance of two correlated log-odds ratios. Stat Med. 2012;31(14):1418–1431.

209. Vieta E, T'joen C, McQuade RD, Carson WH Jr, Marcus RN, Sanchez R, et al. Efficacy of adjunctive aripiprazole to either valproate or lithium in bipolar mania patients partially nonresponsive to valproate/lithium monotherapy: a placebo-controlled study. Am J Psychiatry. 2008 Oct;165(10):1316–25.

210. Wei Y, Higgins JP. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. Stat Med. 2013;32(7):1191–1205.

211. Daniels MJ. A prior for the variance in hierarchical models. Can J Stat. 1999;27(3):567–578.

212. Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 2006 Sep;1(3):515–34.

213. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. Biostatistics. 2008 Jan 1;9(1):172–86.

214. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. BMC Med Res Methodol. 2007 Jan 12;7:3.

215. Efthimiou O, Mavridis D, Cipriani A, Leucht S, Bagos P, Salanti G. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. Stat Med. 2014 Jun 15;33(13):2275–87.

216. Efthimiou O, Mavridis D, Riley RD, Cipriani A, Salanti G. Joint synthesis of multiple correlated outcomes in networks of interventions. Biostatistics. 2015 Jan;16(1):84–97.

217. Cramer H. Mathematical Methods of Statistics. Princeton: Princeton University Press; 1946.