A Bayesian Network Meta-analysis for Binary Outcome: A Simulation Study

A BAYESIAN NETWORK META-ANALYSIS FOR BINARY OUTCOME: A SIMULATION STUDY

BY

TADDELE CHERINET KIBRET

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2013) McMaster University

(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE:	A Bayesian Network Meta-analysis for Binary Outcome: A Simulation Study
AUTHOR:	Taddele Cherinet Kibret M.Sc., (Statistics) McMaster University, Canada
SUPERVISOR:	Dr. Joseph Beyene
NUMBER OF PAGES:	xi, 69

Abstract

Meta-analysis is a method of synthesizing results of different studies conducted to answer a specific question. Meta-analysis applications have been published in a wide range of disciplines including medicine, education, psychology and many others. However, for many years, only pair-wise and direct comparisons have been made using standard meta-analysis methods. It is only recently that network meta-analysis emerged enabling the comparison of multiple treatments based on estimates from different studies. With network meta-analysis, the relative efficacy (or safety) of a particular intervention versus competing interventions can be obtained even in the absence of head-to-head evidence via a common comparator.

An increasing number of methodologies related to network meta-analysis, assessments of underlying assumptions, and strategies for presentation of results have been proposed by several researchers. But only few simulation studies have been done to investigate different characteristics of this emerging statistical method. Hierarchical Bayesian meta-analysis model is commonly used in network meta-analysis to estimate effect of each intervention relative to every other. This model facilitates the

calculation of the rank probabilities of a set of alternative treatments. However, various factors can determine the performance of the model which needs to be considered before using results for decision.

This project aimed to investigate how the Bayesian hierarchical model estimates the rank probability of the best overall most effective treatment (i.e., the treatment ranked first) under different scenarios for modelling a binary outcome. Different network geometries, numbers of studies per comparison, sets of probabilities of success for treatments and sample sizes were investigated in our simulation study for binary outcome.

Our simulation study showed that the estimates of treatments under consideration can be affected by network structures. Similar geometries affect the estimate in similar ways. Unbalanced number of studies per comparison influenced estimates of treatments in the geometries we considered. When a superior treatment is involved in the network, the hierarchical Bayesian mixed treatment model correctly identified it regardless of network patterns, number of studies and individual study sample size.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Dr. Joseph Beyene for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me throughout my time of research and the writing of this thesis. I could not have imagined having a better advisor and mentor for my MSc study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Roman Viveros-Aguillera and Dr. Gregory Pond for their encouragement, insightful comments, and useful questions. My special thanks go to Dr. Jemila Hamid, who has been encouraging me during the challenging times I had experienced.

I thank my fellow lab mates in the Statistics for Integrative Genomics and Methods Advancement (SIGMA) research Group: Ahmed Hossain, Russel de Souza, Binod Neupane, Mateen Shaikh, Ashley Bonner, Shofiqul Islam, Huda Alsulami, Sathish Pichika, Lydia Makoroka and Danielle Richer for the stimulating discussions in our weekly meetings and for all the fun we have had during our studies. I would like to thank once more Ashley Bonner, Danielle Richer, and Binod Neupane, for unreserved support and guidance through coding issues and for our lovely talks.

I also want to thank my professors, Prof. Narayanaswamy Balakrishnan, Prof.

Roman Viveros-Aguillera, and Prof. Angelo Canty, as they greatly supported me during my graduate studies.

I would also like to thank my wife Meseret Samuel and my very young daughter Nardos Taddele, for shouldering all responsibilities I could have shared during my study. Lastly, my Mom (Kelemwa Werku), Dad (Cherinet Kibret), my Brothers Fikru, Ashebir and my Sister Ayelech.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

In the last three decades, there has been an exponential growth in the production and application of systematic reviews in the social and medical science literature. This stimulates researchers to find optimal ways on how to organize and summarize research findings in order to identify what is known about a specific scientific question, for example a relative treatment effect comparing two treatments for a given condition, and focus research on areas that are less well known. A number of researchers have developed systematic techniques of combining effect sizes of empirical researches across samples of related studies (Cook et al., 1997). Meta-analysis is the most popular quantitative method that is used for synthesizing effect sizes from different studies (Hanka, 1994).

The statistical basis of meta-analysis goes back to the 17th century when it was

suggested that combining data might be better than attempts to select among them (Egger et al., 2002). It was in 1904 that the famous statistician, Karl Pearson addressed questions of combining clinical trial results using formal technique that are referred today as meta-analysis (O'Rourke, 2007). However, the method did not receive a wide recognition in medicine for several years. It was rather well demonstrated as a method of combining findings of different studies in the social sciences, particularly in psychology and educational research. The founding of The Cochrane Collaboration in early 1990s as an international network of health care professionals who prepare and regularly update systematic reviews facilitated the popularity of meta-analysis in all areas of health care (Egger et al., 2002).

Meta-analysis is the use of statistical methods to obtain a pooled estimate of treatment effect relative to some comparator. It involves combining of either summary results or individual data or both of several studies in order to obtain a summary treatment effect estimate (Hedges et al., 1992). Meta analysis is mainly focused on contrasting and combining results from randomized control trials to identify patterns and disagreement between individual studies. Meta-analysis has gained increasing recognition in the statistical, medical, and social science research areas (Hedges and Vevea, 1998; Nelson and Kennedy, 2009).

Traditionally systematic reviews compare only two interventions by using pairwise meta-analysis methods while others examine the comparative effectiveness of several or all available interventions for a given condition. Effect size refers to a value that reflects the magnitude of the treatment effect, which includes correlation coefficient, odds ratio, relative risk, and standardized mean difference. When studies

have consistent treatment effect, meta-analysis can be used to estimate their common effect size. But sometimes the generality of intervention effects could be of interest which focuses on variability of measures. When the effect size differs across studies, meta-analysis estimate the average effect size as accurately as possible. It may also be used to explore sources of heterogeneity and its implications (Borenstein et al., 2010).

When studies possess different precision, we should assign relative importance in such a way that more weight is assigned to studies that carried more information. There are two popular models in meta analysis, namely fixed effects model and random effects model (Borenstein et al., 2010). Each of the models makes different assumptions about the nature of the studies which leads to different definitions for the combined effect and different mechanisms for assigning weights (Hedges and Vevea, 1998). The fixed effect model assumes that there is one true effect size which underlies all the studies in the analysis implying that any variation of effect size is the result of random error inherent in each study (Borenstein et al., 2010).

Unlike fixed effects model, the random effects model allows the true effects underlying the studies to vary. Random effects model is suitable for unexplained heterogeneity. Heterogeneity refers to differences between effect sizes across studies. Random effects model is mainly useful to estimate the mean of the distribution of the true effects in a specific population using the true effects of individual studies. This implies that both the number of subjects within studies and total number of studies determine the precision of estimate of the combined effect (Borenstein et al., 2010).

Inpractice, there are multiple health-care interventions that can be used to treat diseases(Song et al., 2009). As a result, traditional method of comparison of two therapies is not always optimal. Health intervention decision-making bodies including clinicians, drug manufacturers, and regulatory agencies require comparisons of all available competing interventions. The challenge is how to choose the most effective intervention from a multiple potentially competing interventions thereby complicating decision making. The standard meta-analysis which focus on comparing only two therapies is not useful in such cases (Salanti et al., 2008b). Decision makers would also like to know how all the different options rank against each other and how big the differences are between all the available options. This requires an appropriate statistical methodology for analysis when the comparative effectiveness of a range of interventions is important.

Network Meta Analyses (NMA), an extension of conventional pairwise meta-analysis, is a valuable alternative to synthesize evidence when the interest is to compare multiple interventions (Hoaglin et al., 2011). With network meta-analysis, the relative efficacy of a particular intervention versus competing interventions can be obtained even in the absence of head-to-head evidence (i.e., an indirect comparison of two interventions is made via a common comparator). Furthermore, network meta-analysis is believed to improve inference by increasing precision (Hoaglin et al., 2011) since it uses more evidence.

In practice, there may exist both direct and indirect evidence. However, plenty of competing treatments may lack direct comparisons for many reasons. For instance,

due to commercial interests and the regulatory approval process, head-to-head comparison of two active treatments is often not available. Decision making cannot wait for the direct evidence which may never happen and hence using indirect evidence is important when possible (Psaty et al., 2003; Wells et al., 2012). That is, we derive direct comparisons between therapies and common comparator (say placebo), to obtain an estimate of evidence for indirect comparison. Therefore, both generators of evidence for health technology assessment and reimbursement agencies need to develop greater understanding of this methodology. This thesis contributes in creating awareness about network meta-analysis and providing some guidance about performance of these methods in a wide range of scenarios.

The use of all potentially relevant available evidence is an attractive feature of network meta-analysis for analysts, as opposed to other methods that rely solely on one type of evidence (Li et al., 2011; Jansen et al., 2011). Some argue that frequentist methods for network meta-analysis become increasingly difficult to fit for less constrained models where as, a Bayesian method can easily construct such complicated models with less assumptions (Hoaglin et al., 2011). However, the Bayesian approach is not without difficulties. Selection of appropriate prior distributions for unknown model parameters and checking of consistency assumptions are the main problem. Nevertheless, the Bayesian approach in network meta-analysis helps to arrive at a single, integrated, estimate of the effect of all included treatments based on all included studies. Moreover, it allows for a probabilistic interpretation and it has ability to select the most effective treatment which is vital for decision making (Hoaglin et al., 2011). The Bayesian method also requires greater statistical expertise

because of its complexities. Multiple treatment comparisons rely on important assumptions such as homogeneity, consistency and transitivity (Cipriani et al., 2013). A considerable amount of studies have been dedicated to investigate these properties and assumptions of network meta-analysis.

1.2 Scope of the Project

In recent years, several papers have been published on network meta-analysis on various topics including statistical methods for network meta-analysis, test of assumptions of network meta-analysis, methods of result presentation in network meta-analysis and so on. However, there is limited work done on simulation studies. There are a number of issues that need comprehensive simulation studies to understand performance of the statistical methods for network meta-analysis Jonas et al. (2013).

A few previous simulation studies includes Song et al. (2012) where they evaluated the properties and the performance of commonly used indirect treatment comparison (ITC) and mixed treatment comparison (MTC) methods. They showed the strength and weakness of the existing indirect treatment comparison and mixed treatment comparison models. (Mills et al., 2011) also conducted a simulation study to investigate the fragility of confidence interval estimation and hypothesis testing. For the settings they considered in their study, indirect confidence interval estimation suffers from under-coverage while indirect hypothesis testing suffers from low power in the presence of moderate to large between-study heterogeneity.

The simulation report by Jonas et al. (2013) pinpointed that the number of

studies, the geometry of networks and effect sizes can affect the identification of best treatment in mixed treatment model; and suggested that detailed investigation about MTC model through simulation study is required for understanding how the methods perform in the real life scenarios. In their simulation scenarios, they assumed equal number of studies in each comparison. However, varying the number of studies per comparison may change the results. They also assumed that the number of patients in each of the individual studies was fixed to be 100 per treatment arm, which is again unrealistic.

This project is a simulation study that aimed to investigate how differences in the number of studies per comparison, the individual study sample sizes and the network geometry influence the rank of the most effective treatment using mixed treatment comparison model. We used a Bayesian framework and restricted our simulations to binary outcomes. The simulations were implemented using the R statistical package with the Bayesian network- meta-analysis carried out by using the "gemtc" (generate mixed treatment comparison) library.

The remainder of the thesis is organized as follows. Chapter 2 provides brief descriptions of meta-analysis methods. Chapter 3 presents simulation study results assessing the effect of different number of studies per-comparison, and the effect of number of patients per individual studies on the ranking of treatments and we conclude with discussion and future direction in Chapter 4.

Chapter 2

Methods

2.1 Standard Meta-analysis

2.1.1 Frequentist Approach

Meta-analysis is an increasingly popular method of evidence synthesis that has been used with greater frequency in the past three decades in medical research. Meta-analysis can be accomplished using either summary data or individual participant data approach. An individual participant data approach might be considered the ideal because it allows for standardizing statistical analyses from each study, obtaining summary results directly, checking the assumptions of models, assessing participant-level effects and examining interactions . However, since the meta-analyst is often most interested in the overall result, both the individual patient data and summary data approaches should yield similar findings (Olkin and Sampson, 1998).

Some of the benefits of employing meta-analysis include, quantifying treatment

effects and their uncertainties, higher power, better precision (i.e, more information is used to estimate pooled effect size), the ability to explore differences between studies, settle controversies between conflicting studies, generate new hypothesis Haidich (2010), and to put knowledge gained in to future application (prediction in random effect model). The exact selection from these for a specific meta-analysis will depend on its aim and context.

The first step in meta-analysis involves describing the results of each study via numerical indicators such as a standardized mean difference, a correlation coefficient, or an odds ratio. These effect size estimates reflect the magnitude of the association of interest in each study. The second step involves pulling the effect size estimates from individual studies to produce a single indicator that summarizes the relationship of interest across the sample of studies. Then, meta-analytic procedures produce summary statistics, which are then tested to determine their statistical significance and importance. One of the foremost decisions during meta-analysis is whether to use a fixed-effect or a random effects model. One may then be interested in first testing for heterogeneity.

Exploring heterogeneity is indispensable for any meta-analysis before pooling the results of primary studies into a summary estimate. Then based on the outcome, one can use either fixed effect models or random effects models. Graphical presentation or eyeballing of the individual study results can aid in the investigation of possible heterogeneity. The forest plot is the most common graph in meta-analysis reports. An example of which represented in Figure 2.1, indicates that the estimate and confidence interval for each study and the corresponding summary estimate. The

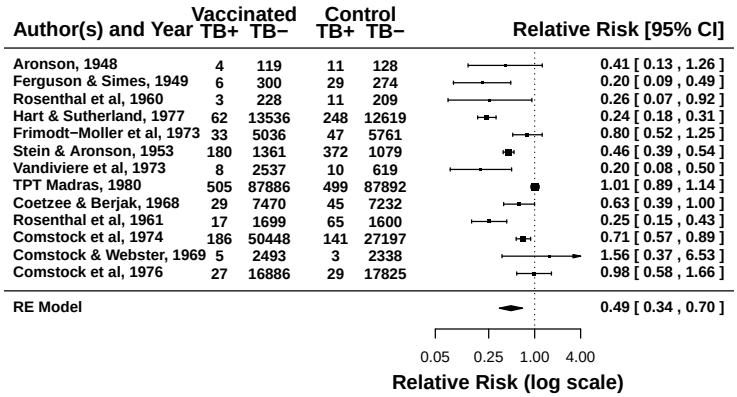

Figure 2.1: Forest plot showing the results of 13 studies examining the effectiveness of the BCG vaccine for preventing tuberculosis (taken from "metafor package" package)

plot represent each study by a line. There is a box in the line for each study and the mid-point of the box represents the point effect estimate for each study. The size of the box represents the weight given to the study. The overall effect of studies is usually denoted by the diamond at the bottom. The width of the line shows the confidence intervals of the effect estimate of individual studies. The width of the diamond shows the confidence intervals for the overall effect estimate. Absence of overlap in confidence intervals may indicate heterogeneity.

The funnel plot is another common plot in meta-analysis. A funnel plot has a measure for effect size on the x-axis and a measure related to the within-study variance (e.g., inverse standard error) on the y-axis. Each study is represented by a single equal-sized marker. Usually, there are more small studies (with larger variance) than big studies in a meta-analytical data set, and the smaller studies have estimates

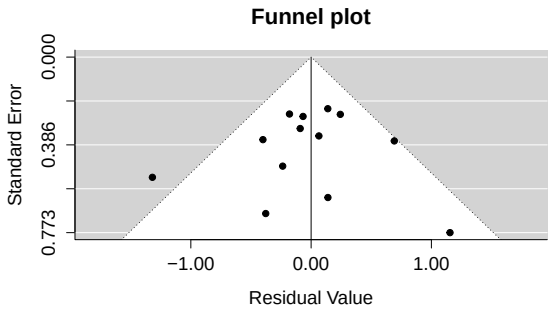

Figure 2.2: Funnel plot for a meta analysis model, taken from "metafor package"

that are more scattered and further removed from the summary estimate. This creates a funnel-like, as indicated by Figure 2.2, distribution of the markers in the plot that can be accompanied with confidence intervals around the fixed effect summary estimate or p-value limits around the null-hypothesis value of no effect. Plots with confidence intervals can be used to check for markers outside of the intervals and can be referred to as heterogeneity funnel plots. The plot shows study markers outside the confidence bands, indicating possible heterogeneity.

Alternatively, the Galbraith plot is designed to assess the extent of heterogeneity between studies in a meta-analysis. It has an effect size divided by its standard error (resulting in a z score) on the y-axis, and the inverse of the standard error on the x-axis. Each study is indicated by a single marker and a regression line runs centrally through the plot. Parallel to the regression line, as shown on figure:2.3 at a 2-standard deviation distance, two lines create an interval in which most markers would fall if the studies were estimating a single fixed parameter.

More formal statistical tests for heterogeneity are available but depends on the

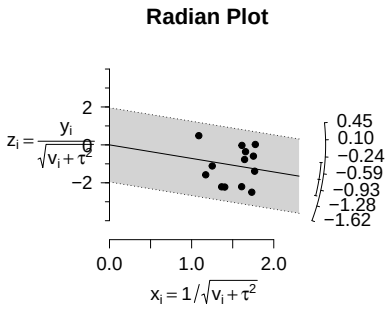

Figure 2.3: Radian plot for a meta analysis model, taken from "metafor package"

number of studies in the meta-analysis (Higgins and Thompson, 2002). Let the observed effect of the i^{th} study be denoted by Y_i, where $\{i = 1, 2, ...k\}$. The classical measure of heterogeneity is Cochrans Q-test calculated as

$$Q = \sum_{i=1}^{k} w_i (Y_i - \hat{\mu})^2 \tag{2.1}$$

where w_i denotes weights of the i^{th} study, $\hat{\mu} = \frac{\sum_{i=1}^{k} w_i Y_i}{\sum_{i=1}^{k} w_i}$ represents the pooled estimate of effect size and k denotes number of studies. A p-value is frequently quoted as an indication of the extent of between-study variability. However, care must be taken in the interpretation of the Q test, since it has low power when studies have small sample size. This means that while a statistically significant result may indicate a problem with heterogeneity, a non-significant result must not be taken as evidence of no heterogeneity. Conversely, Q has too much power as a test of heterogeneity if the number of studies is very large (Higgins et al., 2003). Moreover, the Q test only

informs us about the presence versus the absence of heterogeneity, but it does not report on the extent of such heterogeneity.

Some researchers argue that, clinical and methodological diversity always exists in a meta-analysis and statistical heterogeneity is inevitable (Higgins et al., 2003). Heterogeneity will always exist whether or not we happen to be able to detect it using a statistical test. Thus the test for heterogeneity is irrelevant to the choice of analysis. Consequently, methods have been developed for quantifying inconsistency across studies and assess impact of heterogeneity on the meta-analysis. A useful statistic to quantify the degree of heterogeneity in a meta-analys that is proposed by Higgins and Thompson (2002) is

$$I^2 = \begin{cases} 1, & \text{if } Q - (k-1 > 0. \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

where Q is the commonly used for assessing heterogeneity and $k-1$ is its degrees of freedom (Higgins and Thompson, 2002). The statistic I^2 describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error. When it is proved that there is clear heterogeneity and we are unable to explain it, we have to reflect it in the method of estimation. This can be done using a random effects model where we regard each study as estimating a different effect. We should look for possible explanations of this variation in study characteristics. For instance, there may be subsets of studies within which there is little heterogeneity.

Fixed Effect models

Fixed effect models need two assumptions. One is assuming each study measures the same parameter. The other is assuming that there is no variation in source population ,i.e., only within the study variation (Hedges and Vevea, 1998). Consequently, Borenstein et al. (2009) adopt the wording of fixed effect (no s) to emphasize the effect sizes in our meta-analysis differ only because of sampling error and they all share a common mean. The goal in analyzing fixed-effect models will be making conditional inference only for the k studies included in the meta-analysis (Viechtbauer, 2010). This means, fixed-effects models provide valid inferences under heterogeneity, if the inferences, say, estimating common effect and its standard error (i.e., the conclusions about the size of the average effect) are confined to the set of studies included in the meta-analysis and not generalize to the rest of the population. Let the observed effect of the i^{th} study be denoted by Y_i, where $\{i = 1, 2, ...k\}$, and μ represents the population mean effect size. The customarily used fixed effect model can be written as:

$$Y_i = \mu + \epsilon_i \tag{2.3}$$

where ϵ_i is the error term for study i and its distribution is assumed to be normal, that is, $\epsilon_i \sim N(0, \sigma_i^2)$ and σ_i^2 denotes the within-study variances for study i. The most precise effect size can be obtained by weighting the average of each study by their corresponding inverse variance in such a way that more weight is assigned to the studies that carry more information. Thus, the weight is $w_i = \frac{1}{\sigma_i^2}$. Therefore,

the weighted least squares estimate of common mean effect size is given by:

$$M = \frac{\sum_{i=1}^{k} w_i Y_i}{\sum_{j=1}^{k} w_i} \tag{2.4}$$

with the variance given by: $var(M) = \frac{1}{\sum_{i=1}^{k} w_i}$. The square root of this variance, $\sqrt{var(M)}$ is the standard error that can be used in estimation and hypothesis testing process.

Random Effects Models

In order to combine a group of studies in a meta-analysis, the studies are required to have enough in common so that it makes sense to synthesize the information. However, there is generally no reason to assume that they are identical. This is because studies aiming to address a common question might differ clinically, methodologically or statistically (Higgins and Thompson, 2002). Random effects models have an underlying assumption that a distribution of effects, resulting in heterogeneity among study results, called τ^2.

Heterogeneity refers to the variation among effect parameters across studies. Basically, there can be two sources of variability in a set of studies in a meta-analysis.

1. The variability due to sampling error, also known as within-study variability. The sampling error variability is always present in a meta-analysis, because every single study uses different samples.

2. The other source of variability is heterogeneity, the between-studies variability,

which can occur in a meta-analysis when there is true heterogeneity among the population effect sizes estimated by the individual studies. The causes for between study variation include:

- Difference in characteristics of participants, eligibility criteria for trials, conditions under investigation, dose variations and so on.
- Clinical variation: differences in outcome such as definition of an event, follow-up duration, ways of measuring outcomes, cut-off points on scales and others.
- Methodological heterogeneity: diversity in design,(randomized versus non-randomized, cross over versus parallel, group versus cluster randomized etc);

As a result, when the interest is generalizing results across a variety of situations and the included studies are unlikely to be functionally equal, the random-effects model is usually the more appropriate model (Borenstein et al., 2009).

The random effects models provide an unconditional inference about the average effect in the entire population of studies from which k studies are included in the meta-analysis (Hedges and Vevea, 1998). The simplest and the most popular method is to use the normal random effects model, where a treatment in each study is assumed to be randomly selected from a normal distribution. The normal random effects model is given by

$$Y_i = \theta_i + \epsilon_i = \mu + \eta_i + \epsilon_i \tag{2.5}$$

where $\eta_i \sim N(0, \tau^2)$, $\epsilon_i \sim N(0, \sigma_i^2)$ for $i = 1, 2, , ..., k$. The aim is then, to estimate μ, the mean true effect and τ^2, the total amount of heterogeneity among the true

effects. The within study variance, σ_i^2, represents the extent of estimation error of θ_i, usually estimated by individual study sample variance. Any measure of true effect, Y_i, that satisfies the normality assumption (at least asymptotically normal) can be used. Consequently,

$$Y_i \sim N(\mu, \sigma_i^2 + \tau^2) \tag{2.6}$$

The overall weighted mean effect size, μ can be obtained similar to fixed effect model but now the weight is different due to between study variance. That is,

$$\mu = \frac{\sum_{i=1}^{k} w_i^* Y_i}{\sum_{j=1}^{k} w_i^*} \tag{2.7}$$

where, $w_i^* = \frac{1}{\sigma_i^2 + \tau_i^2}$. In practice the value of τ^2 is unknown and its estimate $\hat{\tau}^2$ from data is obtained by using the DerSimonian-Liard moment-based method of estimation,

$$\hat{\tau}^2 = max\left(0, \frac{Q - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}\right) \tag{2.8}$$

where, $w_i = 1/\sigma_i^2$ for $i = 1, 2, ..., k$.

Alternatively, we may also include one or more study-level variables that account for some part of variability in the true effect sizes in random effects models and use the mixed effects modeling approach to measure how much the moderators included in the model influence the size of the average true effect.

Along with quantifying the amount of heterogeneity present, exploring the root causes of the heterogeneity gain high priorities in meta-analyses. Common practices

for dealing with heterogeneity are

- Stratify the studies into homogeneous subgroups and then fit a separate fixed effects model,
- Construct a random effects model,
- Fit a meta-regression model that explains the heterogeneity in terms of study level covariates.

Meta Regression

Meta regression seeks to explain the observed heterogeneity of the findings in empirical literature by means of moderator variables. Random effects meta-regression has been recommended for use in summary data meta-analysis. The difference is mainly in the calculation of the between study variance (Thompson and Sharp, 1999). Regardless of the model chosen, random-effects meta-regression models will initially estimate τ^2, the between study variance, followed by β, the regression coefficients.

In most meta-regression approaches, the unit of analysis, that is each observation in the regression model, is a study. Predictors in the regression are at the study-level. It may be used to investigate the treatment effect controlling for differences across studies, and determining which study-level covariates account for the heterogeneity. The major difficulties faced in a meta-regression are the degrees-of-freedom available can be small due to the fact most meta-analyses do not include a large number of studies, and high collinearity of covariates, for example all studies in rural areas may administer the medicine in a particular way, while urban hospitals use a different

protocol.

Assume that study y_i is an estimate of effect size of interest, such as a log odds-ratio, log risk-ratio, or difference in means in the i^{th} study, and n denotes the total number of studies. Suppose that each study also reported a standard error for this estimate, σ_i, which we assume is known, as is common in meta-analysis (in practice, it will have been estimated from the data in that study). Then the simplest model, an extension of fixed effect model with individual level covariates, is:

$$y_i = \theta_i + \epsilon_i, i = 1, 2, ..., n \tag{2.9}$$

where $\epsilon_i \sim N(0, \sigma_i^2)$ and $\theta_i = x_i\beta$. Here, β is a $k \times 1$ vector of coefficients (including a constant if fitted) and x_i is a vector of covariate values in the i^{th} study (including 1 if constant is fitted).

Similarly, random effect meta-regression which takes into account the variability across studies assuming that the true effect sizes follow a normal distribution around the linear predictor is given by:

$$y_i = \theta_i + \eta_i + \epsilon_i, i = 1, 2, ..., n \tag{2.10}$$

where $\epsilon_i \sim N(0, \sigma_i^2)$, $\theta_i = x_i\beta$ and $\eta_i \sim N(\theta_i, \sigma_i^2 + \tau^2)$. In the above equation, τ^2 is the between-study variance and must be estimated from the data.

2.1.2 Bayesian Approach

In order to carry out a hierarchical Bayesian meta-analysis, the data available from each study must have a certain commonality or comparability. Each of the estimates is also assumed to be accompanied by its standard error which is denoted by s_i, for $i = 1, 2, ..., k$.

We have seen above that fixed effect model assumes common mean μ for the observed treatment effect Y_i, where $i = 1, , 2, ..., k$ independent studies. That is $E(Y_i) = \mu$ with known variance, $var(Y_i) = s_i^2$. Conventionally, Y_i is assumed to come from a normal distribution, $Y_i \sim N(\mu, s_i^2)$. However, in random effects model rather than a common mean effect size, we assume a study specific mean effect size θ_i. Thus, there are two levels to be considered in modeling. The first level describes the variability of Y_i given θ_i and s_i^2. That is,

$$Y_i/\theta_i, s_i^2 \sim N(\theta_i, s_i^2) \tag{2.11}$$

The second level of the hierarchy describes the variability of the study level parameters θ_i. Given the mean μ and the between variance τ^2 we assume the study specific mean θ_i is drawn from normal distribution:

$$\theta_i/\mu, \tau^2 \sim N(\mu, \tau^2) \implies Y_i \sim N(\theta_i, \sigma_i^2 + \tau^2) \tag{2.12}$$

Placing priors on hyper-parameters (μ and τ^2) makes the Bayesian hierarchical model. In the absence of more specific information, we use non-informative normal priors for μ and non-informative inverse gamma or uniform prior for τ^2 is frequently

used.

2.2 Data Formats

In meta-analysis, input data are usually the summary statistics collected from the published literature rather than the primary data directly collected from trial units. In addition, the aggregate input data might be available in two formats:

1. Arm-level summaries: refers to effect measures that are reported for each arm (such as odds, absolute risk, hazard or mean) or

2. Contrast-level summaries: refers to the relative effect measures which include odds ratio, risk ratio, hazard ratios or mean difference that are presented as the difference in effect between arms.

One advantage of the arm-level summaries over that of contrast-level is that it is possible to adopt the exact likelihood for the data (i.e. binomial for binary data) rather than its normal approximation, as for the contrast-level summary (Hoaglin et al., 2011). Both frequentist and Bayesian approaches can be used to specify models based on either of the two data formats (Franchini et al., 2012). In the next sections we discuss the analysis of data by means of arm-level summaries using Bayesian approach, which enable more flexible and precise analyses.

The existing randomized controlled trials comparing multiple treatment options are not feasible for comparison of interventions of complex structure since a series of individual meta-analysis cannot provide all information but answering the question about pairs of treatments (Salanti et al., 2008b; Sutton et al., 2008). This does

not support optimal clinical decision-making, because each meta-analysis is only one part of the whole picture. Therefore, a more general method is required to tackle the problem.

2.3 Network Meta-analysis

2.3.1 Overview

Medical decision making requires the evaluation of the relative value of a particular intervention relative to all other relevant interventions of choice. The value of an intervention is dependent on the available evidence (i.e. quality of study, effect size and associated uncertainty in estimates) for that intervention. Ideally, evidence should be available from a randomized controlled trial (RCT) of sufficient size that compares all interventions of interest simultaneously. However, such a study is often not available because randomized control trials are usually designed for registration purposes and generally only include placebo or one active comparator. Network meta-analyses(NMA) are invaluable alternative to harmonize evidence when the interest is to compare multiple interventions.

Network meta-analysis (NMA), which is also referred to as multiple treatment meta-analysis (MTM) or mixed treatment comparison (MTC) is an extension of traditional meta-analysis (that include only studies that compare the same intervention with the same comparator) by including multiple different pair-wise comparisons across a range of interventions (Lumley, 2002; Salanti, 2012). Consequently, complex analysis and comparison of several treatments by network meta-analysis under

the same condition is quickly gaining popularity among clinicians, guideline decision makers and health technology agencies (Caldwell et al., 2005; Li et al., 2012; Mills et al., 2013).

An important benefit gained from network meta-analysis over that of conventional pairwise meta-analysis is that it borrows strength from indirect evidence to gain certainty about all treatment comparisons and allows for estimation of comparative effects that have not been investigated head-to-head in randomized clinical trials (Lumley, 2002). The indirect estimation is done by preserving the within-trial randomization which Bucher et al. (1997) called it an adjusted indirect comparison. For instance, if two particular treatments have never been compared against each other head-to-head, but these two treatments have been compared to a common comparator, then an indirect treatment comparison (ITC) can use the relative effects of the two treatments versus the common comparator.

2.3.2 Assumptions

The basic assumptions of network meta-analysis are:

- Homogeneity: different trials in every pair-wise comparison in the network are assumed to be sufficiently homogeneous for which there is direct evidence. The homogeneous assumption is satisfied when, for particular treatment pairing, the true treatment effect is the same throughout the trials that assigns the true treatments of interest (Donegan et al., 2013)

- Similarity (Transitivity), is an assumption for indirect comparison: the relative effects of treatment should be held similar across the entire set of trials. Transitivity should hold for any indirect comparison so that generalization can be made.

- Consistency (or coherence), assumption for the combination of direct and indirect evidence. The direct and indirect estimates are assumed to be the same(i.e, $d_{AB}^{dir} = d_{AB}^{indir}$ where, $d_{AB}^{indir} = d_{AC}^{dir} - d_{BC}^{dir}$). Consistency is a property of loops of evidence. That is, it can be evaluated only when there is both direct and indirect evidence for a particular comparison of interventions (Salanti et al., 2008b; Dias et al., 2010; Lu and Ades, 2006; Cipriani et al., 2013).

In complex network structures, the transitivity assumption should hold for all cases where indirect or mixed estimates are derived. Salanti (2012) gave five possible ways to look at transitivity across the network:

- Every treatment in the network has a fixed definition irrespective of the comparator meaning, Treatment C is similar when it appears in AC and BC trials;

- The missing treatments in each trial may be viewed as missing at random;

- There are no differences between observed and unobserved relative effects of AC and BC beyond what can be explained by heterogeneity;

- There are no differences between observed and unobserved effects for every comparison in the network beyond those attributed to heterogeneity;

- Participants included in the network could in principle be randomized to any of the treatments.

Inconsistency models or node-splitting models are the most popular approaches used to assess the consistency of the trials in the network. In the inconsistency models inconsistency factors are added to the closed loops and consistency is assessed by comparing the models with and without these inconsistency factors. The consistency and inconsistency models are respectively,

$$d_{BC} = d_{AC} - d_{AB} \tag{2.13}$$

$$d_{BC} = d_{AC} - d_{AB} + \omega \tag{2.14}$$

where ω represents inconsistency factor (i.e, the discrepancy between the direct and indirect evidence).

Each inconsistency factors assumed to have the same variations (Dias et al., 2010). It has been proposed that larger consistency variation as compared to the random effects variance shows the presence of the discrepancy between direct and indirect estimates of effect size. However, except to check for all potential inconsistencies in a single model, it is difficult to interpret the result. By contrast, the node splitting method compares all direct and indirect evidences for each nodes. It is the general method of assessing inconsistencies and the results are easy to interpret. However, it is computationally intensive and time consuming (Dias et al., 2010).

Network meta-analysis is believed to improve the precision of the direct estimate by reducing the width of the confidence intervals compared with the direct evidence

alone (Bucher et al., 1997; Salanti et al., 2008b). However, several methodological researchers have raised many concerns about the validity of multiple treatment meta-analysis methods. The main criticism is associated with the difficulty in evaluating the assumptions underlying the statistical synthesis of direct and indirect evidence (Ioannidis, 2006).

2.3.3 Fixed Effects Models in Network Meta-Analysis

A statistical model typically depends on unknown parameter(s) and the objective may either be to estimate the parameters or some functions of them or test hypotheses about them. While parameters are assumed to be fixed in frequentist approach, both the data and parameters are treated as random variable, hence having a joint distribution, in a Bayesian framework. Bayesian methods use probability to describe uncertainty about parameters, uncertainty about future predicted values, and to represent one's beliefs about parameters using a prior distribution. Moreover, the advantage of Bayesian methods in comparison to frequentist approaches is that inference is exact for any sample size, assuming the prior assumptions are valid.

With the recent development in mixed treatment comparisons, there is an increasing interest in a hierarchical Bayesian approach to network meta-analysis modeling. The model is estimated using a Markov Chain Monte Carlo (MCMC) simulation. Availability of packages such as the WinBUGS language have greatly helped tackling the computational difficulties that arises from the complex structure of the model and contributed to the growing use of Bayesian hierarchical models (Lu and Ades, 2004, 2006).

The Bayesian setting in network meta-analysis also uses knowledge from the observed data to modify the understanding of how different treatments perform against each other and how treatments should be ranked. Lu and Ades (2004, 2006) proposed the general model for mixed treatment comparison for K treatments. A common generalized linear regression model is used which can be implemented with normal, binomial, poisson, and multinomial data. In this study we focus on the model that uses arm-level binary outcome.

Suppose that N randomized controlled trials make mixed comparisons among K treatments. Let r_{ik} denotes the number of binary events that occurred for every clinical trial i for each treatment k. Let n_{ik} denotes the number of observations and p_{ik} represents the corresponding probability of the successful outcome on treatment k. Then the distribution of the number of events is given by:

$$r_{ik} \sim Bin(p_{ik}, n_{ik}), i = 1, ..., N; k = 1, 2, ..., K \tag{2.15}$$

Where p_{ik} represents the corresponding probability of the successful outcome on treatment k and is modelled on the logit scale as:

$$logit(p_{ib}) = log\left(\frac{p_{ib}}{1 - p_{ib}}\right) = \mu_i, i = 1, 2, ..., N; b = k = 1, ...K \tag{2.16}$$

$$logit(p_{ik}) = log\left(\frac{p_{ik}}{1 - p_{ik}}\right) = \mu_i + d_{i,1k} \Leftrightarrow p_{ik} = logit^{-1}(\mu_i + d_{i,1k}), i = 1, 2, ..., N; k = 2, ...K; b < k \tag{2.17}$$

where, μ_i are the trial specific baselines representing the log odds of event in the

referent treatment (k=b), and $d_{i,1k}$ are the fixed trial-specific $ln(OR)$ of event occurrence of the group k compared with referent treatment and hence the model is called fixed effect model. The logit link function maps the probability of treatments response on to the real number.

2.3.4 Random Effects Models in Network Meta-Analysis

In a random effects model, each study i provides an estimate of the study-specific log odds, $\delta_{i,1k}$, which are assumed not to be equal but instead similar in the sense that assumes the information that the trials provide is independent of the order in which they were carried out (exchangeable), over the population of interest. Hence, the random effects model is obtained by replacing $d_{i,1k}$ in the fixed effects model by $\delta_{i,1k}$. That is,

$$logit(p_{ik}) = log\left(\frac{p_{ik}}{1-p_{ik}}\right) = \mu_i + \delta_{i,bk}, i = 1, 2, ..., N; k = 2, ...K; b < k \qquad (2.18)$$

Equivalently, we can say that the trial-specific treatment effects come from a the same family distribution. That is we usually assume that, $\delta_{i,bk} \sim N(d_{bk} = d_{1k} - d_{1b}, \sigma^2)$. In the above model the notation, $k > b$ stands for k is $after$ b and $d_{11} = 0$. Prior distributions for basic parameters, $d_{12}, d_{13}, d_{14}, ...$, are assumed in Bayesian framework while the remaining contrasts (functional parameters) are defined in terms of those treatments compared with the baseline treatment directly

assuming consistency. That means we can write the functional parameters as

$$d_{s,t} = d_{b,t} - d_{b,s}, b = 1, 2, ..., K, s = 2, 3, ..., K, t = 3, 4, ..., K; s < t. \tag{2.19}$$

It is not rare to see network meta-analysis that involves trials with more than two arms (multi-arms) in pooling data across a network of treatments. It has been pointed out that, in the analysis of multi-arm data, any assumptions about heterogeneity have implications on the relative efficacy of parameters ((Higgins and Whitehead, 1996; Franchini et al., 2012)). It is also characterized by induced correlation between data-points due to the use of a common comparator in the comparations . This suggests that there is a need for an adjustment in the likelihood to account for this induced correlation. Otherwise, results from the contrast-level format will be incorrect (Higgins and Whitehead, 1996; Franchini et al., 2012). One multi-arm trial i which compares α_i will create a correlated vector of random effects of $\alpha_i - 1$ given by $\delta_i = (\delta_{i,12}, ..., \delta_{i,ka_i})^T$. Assuming consistency, the functional parameters obtained from the $K - 1$ direct treatment effects using $\delta_{s,t} = \delta_{b,t} - \delta_{b,s}, b = 1, 2, ..., K, s = 2, 3, ..., K, t = 3, 4, ..., K$; that is $s < t$. When there is at least one multi-arm in the network, assuming homogeneity between trial variance, the univariate normal distribution of a single random effects discussed earlier become multivariate normal distribution for a vector of the random effects indicated above Salanti et al. (2008a). That is,

$$\begin{pmatrix} \delta_{i,12} \\ \delta_{i,13} \\ \vdots \\ \delta_{i,ba_i} \end{pmatrix} \sim N_{a_i} \left(\begin{pmatrix} d_{t_{i1},t_{i2}} \\ d_{t_{i1},t_{i3}} \\ \vdots \\ d_{t_{i1},t_{ia_i}} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & ... & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 & \sigma^2/2 & ... & \sigma^2/2 \\ \vdots & & & & \\ \sigma^2/2 & \sigma^2/2 & ... & \sigma^2 \end{pmatrix} \right)$$

where, δ_i is the vector of random effects which follows a multivariate normal distribution, a_i denotes the number of arms in the i^{th} trial and $d_{t_{i1},t_{ik}} = d_{1,t_{ik}} - d_{1,t_{i1}}$. In order to understand how this model works, let us suppose that one study is a multi-arm trial in a network of K=4 trials comparing treatments A, B and C. Imagine also that two studies are two-arm trials that compare B with A, and D with A, respectively. For this network we have three basic parameters (d_{AB}, d_{AC}, d_{AD}) and three functional parameters (d_{BC}, d_{BD}, d_{CD}), assuming consistency. Then we can write the random effects model as:

$$logit \begin{pmatrix} p_{1B} \\ p_{2D} \\ p_{3B} \\ p_{3C} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \delta_{1,AB} \\ \delta_{2,AD} \\ \delta_{3,AB} \\ \delta_{3,AC} \end{pmatrix} \tag{2.20}$$

and

$$\begin{pmatrix} \delta_{1,AB} \\ \delta_{2,AD} \\ \delta_{3,AB} \\ \delta_{3,AC} \end{pmatrix} \sim N_4 \left(\begin{pmatrix} \delta_{1,AB} \\ \delta_{2,AD} \\ \delta_{3,AB} \\ \delta_{3,AC} \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2/2 & 0 & 0 \\ 0 & 0 & \sigma^2 & \sigma^2/2 \\ 0 & 0 & \sigma^2/2 & \sigma^2 \end{pmatrix} \right) \tag{2.21}$$

2.3.5 Network Meta-Regression

Network meta-regression can serve to explore inconsistency due to differences across studies and comparisons by regression of aggregate (study-level) covariates or individual patient data, when available. It can also be used to examine causes of inconsistency or to adjust for different effect modifiers across studies making different comparisons (Cipriani et al., 2013). For binary data, meta-regression models are fitted specifying fixed or random effects models and adjusting the $lnOR$ for study-level prognostic factors. The fixed treatment effect meta-regression model is obtained adjusting fixed effects model that we have discussed above for the effect of covariates:

$$logit(p_{ib}) = \mu_i + \beta x_i, i = 1, 2, ..., N; b = 1, 2, ..., K \tag{2.22}$$

$$logit(p_{ik}) = \mu_i + d_{i,1k} + \beta x_i, i = 1, 2, ..., N; b = 1, 2, ..., K; b < k \tag{2.23}$$

where x_i is covariate for the i^{th} trial which can represent a subgroup or continuous variable.

In the same way, the meta-regression with random effects is obtained adjusting

random effects model as follows:

$$logit(p_{ib}) = \mu_i + \beta x_i, i = 1, 2, ..., N; b = 1, 2, ..., K \tag{2.24}$$

$$logit(p_{ik}) = \mu_i + \delta_{i,bk} + \beta x_i, i = 1, 2, ..., N; b = 1, 2, ..., K; b < k \tag{2.25}$$

where x_i is covariate for the i^{th} trial which can represent a subgroup or continue variable and $\delta_{i,bk} \sim N(d_{bk} = d_{1k} - d_{1b}, \sigma^2)$.

2.3.6 Choice of Priors

With regard to choice of priors, it is common to set weakly-informative prior distributions (unless we have strong evidence to select informative ones), usually $\mu_i, d_{bk}, \beta \sim N(0, 10^4)$ (Lu and Ades, 2004, 2006; Salanti et al., 2008b). For the likelihood of binomial outcomes with logit link models, people usually set a uniform flat prior for the standard deviation, $(i.e, \sigma \sim Uniform(0, 5))$(Stettler et al., 2007). An other option, which was once popular but has fallen out of interest is to place a flat Gamma prior on the precision $(say, 1/\sigma^2 \sim gamma(0.001, 0.001)$. This prior assigns small weights to large standard deviation (i.e., it assigns high weights to values of standard deviation near zero),usually not needed. However this feature may be useful particularly, when data is sparse to improve numerical stability and speed convergence of MCMC sampling.

The same linear regression framework can be applied to continuous outcomes. That is, the core models remain the same but the likelihood and the link functions differ to indicate the nature of the data (continuous, categorical, rate and so on) and

the sampling process that generated these data (normal, binomial, Poisson).

2.3.7 Estimation of Rank Probabilities

Hierarchical Bayesian approach to mixed treatment comparison allows the straightforward calculation of rank probabilities of a set of alternative treatments. In each MCMC run, each treatment in the study is ranked according to its estimated magnitude. Then, the proportion of MCMC cycles in which the treatment K ranks first gives the probability that the specific treatment is best among all competing treatments in the study. Similarly, other probabilities are calculated for being the second best, the third best, and so on for every treatment.

Different statistical tools such as (rankograms) or the cumulative rank curves can be used to present the rank probabilities of the treatments. Rankograms represent the distribution and probability of rankings of each treatment compared to all other interventions in the network. Cumulative ranking probabilities obtained by plotting against ranks of all competing treatments facilitates the estimation of the surface under the cumulative rank curve (SUCRA) line for each treatments. For each treatment j out of the a treatments in the network, we can obtain a vector of cumulative probabilities $Cum_{j,b}$ to be best among b total treatments, where $b = 1, ., a$, then the SUCRA index is defined as:

$$SUCRA_j = \frac{\sum_{j=1}^{a-1} Cum_{j,b}}{a-1} \tag{2.26}$$

A SUCRA value 1 suggests that the corresponding treatment is always ranked as first

best among other treatments in the network while a SUCRA value of 0 indicates that the treatment is ranked the last all the time.

2.3.8 Evaluation of Goodness of Fit

The goodness of fit of statistical model describes how well the model fits the observed set of observations. Goodness of fit indices summarize the difference between the observed values and expected under a statistical model. When more than one model exist, one may also be interested in a relative model fit (i.e., the discrepancy between two models). Let $l(\hat{\theta}/y)$ denotes the log likelihood, $\hat{\theta}$ denotes the maximum likelihood estimates for parameters in the model and k denotes the number of parameters in the model. Two popular indices are given as follows:

$$\text{Akaike Information Criteria(AIC)} = -2l(\hat{\theta}/y) + 2k \tag{2.27}$$

$$\text{Schwarz Bayesian Information Criteria(BIC)} = -2l(\hat{\theta}/y) + kln(N) \tag{2.28}$$

where N denotes the total number of observations included in the model. If we aimed to select the best fitting model, say selection between fixed and the random effect models, the smaller the value of the indices, the better the fit of the model. In Bayesian posterior computation using Markov Chain Monte Carlo method determining the best fit model is much important. Deviance information criteria (DIC) is a hierarchical model selection criteria that can be viewed as a generalization of the AIC and BIC (Spiegelhalter et al., 2002). It is particularly useful in Bayesian model selection problems where the posterior distributions of parameters have been

obtained by Markov Chain Monte Carlo simulation. The classical deviance $D(\theta)$ is given by:

$$D(\theta) = -2ln(f(y/\theta)) + 2ln(g(y)) \tag{2.29}$$

where, $f(y/\theta)$ represents the likelihood function and $g(y)$ denotes a fully specified standardizing term that is completely determined by the observed data. This has been extended to deviance information criteria (DIC) defined as:

$$DIC = \bar{D} + P_D \tag{2.30}$$

where

$$\bar{D} = E_{\theta/y}(D(\theta)) = E_{\theta/y}(-2ln(f(y/\theta))) \tag{2.31}$$

Smaller values of DIC indicate a better fitting model. The P_D measures the complexity of model and defined as the difference between the posterior mean of deviance and the deviance evaluated at the posterior mean of the parameters.That is,

$$\begin{aligned} P_D &= \bar{D} - D(\bar{\theta}) \\ &= E_{\theta/y}(D(\theta)) - D(E_{\theta/y}(\theta)) \\ &= E_{\theta/y}(-2ln(f(y/\theta))) + 2ln(f(y/\bar{\theta}) \end{aligned} \tag{2.32}$$

Consequently, $\bar{D} = P_D + D(\bar{\theta}))$ and hence, $DIC = 2P_D + D(\bar{\theta}))$. For binomial likelihoods, each trial arm contributes 1 independent data point, and the residual

deviance is calculated (for each iteration of the MCMC simulation) as:

$$D_{res} = \sum_i \sum_k 2\left(r_{ik} log\left(\frac{r_{ik}}{\hat{r}_{ik}}\right) + (n_{ik} - r_{ik}) log\left(\frac{n_{ik} - r_{ik}}{n_{ik} - \hat{r}_{ik}}\right)\right) = \sum_i \sum_k dev_{ik}; \tag{2.33}$$

where $\hat{r}_{ik} = n_{ik} p_{ik}$ is the expected number of events in each trial and dev_{ik} is the residual deviance for each data point. In order to assess the influence of individual data points on the model parameters, the leverage is calculated by subtracting the mean residual deviance at the posterior mean of the fitted value from the posterior mean of the residual deviance.

$$P_D = \sum_i \sum_k [\overline{dev}_{ik} - \tilde{dev}_{ik}] \tag{2.34}$$

where, $\tilde{r}_{ik}$ denotes posterior mean of $\hat{r}_{ik}$ and $\tilde{dev}_{ik}$ represent the posterior mean of the deviance residuals. The lower the residual deviance, the better the fit.

2.3.9 Geometry of Networks

The first task in network meta-analysis is to view the nature and structure of studies that are involved in the analysis. In order to view the distribution of the studies involved in the analysis one can use network diagram. The plot of a network geometry is a visual representation of the evidence base and offers a concise description of its characteristics. It offers an intuitive approach to symbolically represent all the direct relations among interventions.

The basic assumption underlying mixed treatment comparison methods is that

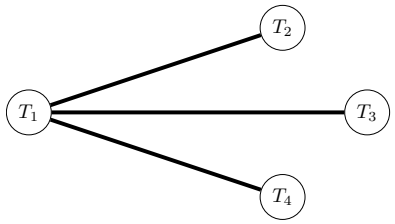

Figure 2.4: Star geometry

the network must be a connected one. Nodes are usually used to represent the treatments or interventions and edges that connects nodes represent the direct comparison of interventions. Geometry of network refers to the overall structure of treatment contrasts. Network configuration can dictate the reference treatment. The treatment that appear more frequently than others or the one which is most popularly used in clinical practice may be taken as reference. The amount of available information can be reflected by weighting the nodes and edges using different node sizes and line thicknesses and hence networks may also be considered as symmetric or asymmetric.

The geometry of network can be imposed by investigators preference for specific treatments, comparator preference bias, sponsorship bias, selective reporting bias, and time lag bias (Salanti et al., 2008b). Networks may ranges from simple (a network in which only 3 treatment options exist) to the very complex (a network involving large number of treatments). In this study, we consider four network patterns namely, star, loop, one-closed loop and ladder (linear). In our simulation study, we considered the networks given in figures 2.4 to 2.7.

Each network, figures 2.4 to 2.7 comprises four nodes representing four interventions (denoted by, treatment 1 or T_1, treatment 2 (T_2) ,treatment 3 (T_3), treatment

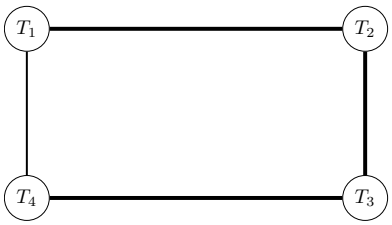

Figure 2.5: Loop geometry

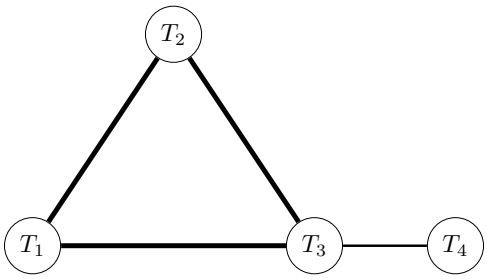

Figure 2.6: One closed loop geometry

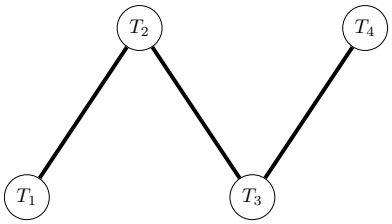

Figure 2.7: Ladder geometry

4 (T_4) and three edges (for star and ladder) and four edges for loop and one closed loop. Except for star pattern, each contrast has both direct and indirect evidence (closed loops).

Chapter 3

Simulation Studies

3.1 Simulation Design

In this simulation study, we aimed to investigate the impact of number of studies per-comparison and individual study sample size on the rank probability of treatments. We also investigated how hierarchical Bayesian model identify the most effective treatment under different network geometries. We compared four interventions assuming one of them as placebo (reference treatment) where the rest were assumed to be competing treatments.

In the first scenario, we considered three equally effective treatments relative to one less effective reference treatment. In another scenario we considered three treatments that are equivalently and modestly effective and one treatment which is superior. We created pseudo populations of binary data under each scenario from which we generated samples of different network patterns. We limited our study to

two-arm studies with binary outcome and fitted mixed treatment comparison models.

In network meta-analysis, when existing randomized control trials fail to compare all interventions but each trial compares only a subset of the interventions of interest, then it is possible to represent the evidence base as a network where all trials have at least one intermediate treatment in common. Therefore, four different network patterns of trials involving treatments compared directly or indirectly were synthesized by means of a network meta-analysis. As introduced in Chapter 2, these patterns of interest consist of star, one closed loop, loop and ladder.

Let us denote the treatments under investigation by treatment 1 (T_1) which we considered to be placebo, treatment 2 (T_2), treatment 3 (T_3), and treatment 4 (T_4). The parameters of interest in this study include probability of success for the binary outcome, sample size of individual studies, number of studies and network pattern as indicated in Table 3.1. Using the probability sets (0.1, 0.5, 0.5, 0.5), (0.5, 0.5,

Table 3.1: Parameters varied during simulations

Parameters		**Values**	
Probability of Success for T1, T2, T3, T4		$(.1, .5, .5, .5), (.5, .5, .1, .5), (.2, .2, .2, .8)$	
Study Sample Size		$50, 100, 200$	
Number of Studies per Comparison	Network:	Equal:	Unequal:
	Star and Ladder	$1, 2, 3, 5, 10, 15$	$(1, 5, 15)$
	Loop and One Closed Loop	$1, 2, 3, 5, 10, 15$	$(1, 3, 5, 15)$
Network Pattern		Star, Ladder, Loop, One Closed Loop	

0.1, 0.5), (0.2, 0.2, 0.2, 0.8) indicated in Table 3.1, we created populations of 80,000 binary data at random from a binomial distribution. Data sets generated by the first and second set of probabilities represent three equivalent treatments compared with placebo. The third set of probabilities is the condition when one superior treatment

is compared with other treatments. In this simulation study, the specified study sample sizes are assumed to be constant within each study for the sake of simplicity.

For the unequal number of studies for the ladder and star geometries, $(1, 5, 15)$ denotes the number of studies for the comparisons between treatments $(T_1\&T_2), (T_1, \&T_3)$ and $(T_1\&T_4)$, respectively. Likewise, the unbalanced number of studies $(1, 3, 5, 15)$ for loop and one closed loop geometries denote the numbers of studies compairing the pairs of treatments $(T_1\&T_2), (T_2, \&T_3), (T_3\&T_4)$ and $(T_1\&T_4)$, respectively. Moreover, the number of studies between $(T_1\&T_2), (T_1\&T_3), (T_1\&T_4), (T_2\&T_3), (T_2\&T_4)$, and $(T_3\&T_4)$ comparisons for the completely connected loop geometry were set at $(1, 2, 3, 5, 10, 15)$, respectively.

Next we randomly sampled 1000 data sets for each of the four network patterns considering both equal and unequal number of studies from the simulated data set. Then we fitted hierarchical Bayesian mixed treatment comparison model on each data set and performed statistical inference.

It is common practice to discard the initial iterations of iterative simulation as they are highly influenced by starting values and do not provide good information about the target distribution. These early MCMC samples are not used for inference. In our simulation, we discarded 5000 iterations as a burn-in. This practice of discarding an initial portion of a Markov chain sample minimizes the effect of initial values on inference based on the posterior distribution.

Moreover, the models might experience poor mixing (or slow convergence) of the Markov chain if there is significant auto-correlation between samples. A common strategy is to *thin* the Markov chain in order to reduce sample auto-correlations. We

thinned a chain by keeping every 10^{th} simulated draw from each MCMC sequence. Then $20,000$ iterations after a burn-in phase of $5,000$ iterations were used to make inference about parameters of interest. Convergence was assessed using the Brooks-Rubin diagnostic test where a potential scale reduction factor of 1.05 or lower was considered sufficient for the convergence (Brooks and Gelman, 1998).

3.2 Performance Evaluation

The bias of estimates of rank probabilities were computed as the difference between the average percentage of times a given treatment was considered most effective based on the estimates from models fitted to the 1000 data sets and the expected percentage based on the assumptions used to generate the data. The average standard deviation of the 1000 samples was also calculated.

3.3 Simulation Results

In this section, the simulation results are presented. Each table displays results for a specific scenario, determined by the set of parameters. Values in each table are estimates of the Mean, Bias and Standard deviation of the rank probability that each treatment is the best, as estimated from the network meta-analysis models over the $1,000$ sample data sets. First, simulation results from scenarios with three treatments of equivalent and all superior effectiveness compared to placebo are presented, followed by simulation results from the scenario where one treatment is superior compared to the other three.

For the scenario with $p = (0.1, 0.5, 0.5, 0.5)$, the competing treatments are assumed to be equally effective, the expected value of the event of interest for T_1 (placebo) is 0.10 and the rest of the treatments, T_2, T_3, and T_4, is 0.50. For this scenario, if the network meta-analysis model is appropriate, we expect no difference between the average rank probabilities of T_2, T_3, and T_4. Table 3.2 shows results for this scenario with a star geometry for a range of equal and unequal number of studies per comparison. With an equal number of studies per comparison, the star geometry identified treatments with rank probabilities very close to the corresponding expected probability values, resulting in low bias. As the equal number of studies per comparison increases from 1 to 15, the rank probabilities approach the true rank probabilities. However, when the number of studies per comparison is only 1 or 2 the rank probabilities were slightly deflated or inflated, though these estimates are very close to their expected values when the number of studies per comparison is increased to 15. With an unequal number of studies per comparison, the star geometry resulted in maximum bias and standard deviation.

Table 3.2: Star Network Pattern with with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 50$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.00	0.34	0.32	0.34	0.00	0.01	-0.01	0.00	0.00	0.17	0.16	0.17
(2,2,2)	0.00	0.34	0.33	0.34	0.00	0.00	-0.01	0.00	0.00	0.20	0.20	0.20
(3,3,3)	0.00	0.34	0.33	0.33	0.00	0.00	-0.00	0.00	0.00	0.23	0.23	0.23
(5,5,5)	0.00	0.34	0.33	0.33	0.00	0.00	0.00	-0.00	0.00	0.24	0.25	0.24
(10,10,10)	0.00	0.32	0.34	0.33	0.00	-0.01	0.01	0.00	0.00	0.25	0.26	0.24
(15,15,15)	0.00	0.33	0.33	0.34	0.00	-0.00	-0.00	0.00	0.00	0.25	0.25	0.25
(1,5,15)	0.00	0.46	0.30	0.24	0.00	0.12	-0.04	-0.09	0.00	0.28	0.23	0.21
(15,5,1)	0.00	0.26	0.30	0.44	0.00	-0.07	-0.04	0.11	0.00	0.22	0.23	0.27
(15,1,5)	0.00	0.26	0.45	0.29	0.00	-0.07	0.11	-0.04	0.00	0.22	0.28	0.24
(1,15,5)	0.00	0.46	0.26	0.28	0.00	0.13	-0.07	-0.05	0.00	0.28	0.22	0.24
(5,15,1)	0.00	0.30	0.26	0.44	0.00	-0.03	-0.08	0.11	0.00	0.24	0.21	0.27
(5,1,15)	0.00	0.30	0.46	0.24	0.00	-0.03	0.13	-0.09	0.00	0.24	0.28	0.21

As we increase sample size per study from $n = 50$ to $n = 100$ or $n = 200$, as

Table 3.3: Star Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 100$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.00	0.34	0.33	0.33	0.00	0.01	-0.01	-0.00	0.00	0.13	0.13	0.13
(2,2,2)	0.00	0.33	0.34	0.33	0.00	0.00	0.00	-0.01	0.00	0.19	0.21	0.20
(3,3,3)	0.00	0.33	0.33	0.33	0.00	0.00	-0.00	0.00	0.00	0.22	0.23	0.22
(5,5,5)	0.00	0.33	0.33	0.34	0.00	-0.01	0.00	0.01	0.00	0.24	0.23	0.24
(10,10,10)	0.00	0.34	0.34	0.32	0.00	0.01	0.00	-0.01	0.00	0.25	0.25	0.24
(15,15,15)	0.00	0.33	0.33	0.34	0.00	0.00	-0.00	0.00	0.00	0.25	0.25	0.25
(1,5,15)	0.00	0.43	0.30	0.26	0.00	0.10	-0.03	-0.07	0.00	0.28	0.24	0.22
(15,5,1)	0.00	0.25	0.30	0.44	0.00	-0.08	-0.03	0.11	0.00	0.20	0.23	0.27
(15,1,5)	0.00	0.25	0.45	0.30	0.00	-0.08	0.11	-0.03	0.00	0.21	0.27	0.23
(1,15,5)	0.00	0.43	0.27	0.30	0.00	0.10	-0.07	-0.03	0.00	0.28	0.22	0.24
(5,15,1)	0.00	0.29	0.27	0.45	0.00	-0.05	-0.07	0.11	0.00	0.23	0.22	0.27
(5,1,15)	0.00	0.30	0.43	0.27	0.00	-0.04	0.10	-0.06	0.00	0.23	0.27	0.22

Table 3.4: Star Network Pattern with p=(0.1,0.5,0.5,0.5) and n=200

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.00	0.33	0.33	0.34	0.00	-0.00	-0.00	0.00	0.00	0.10	0.10	0.10
(2,2,2)	0.00	0.34	0.32	0.34	0.00	0.00	-0.01	0.00	0.00	0.19	0.20	0.20
(3,3,3)	0.00	0.34	0.32	0.33	0.00	0.01	-0.01	-0.00	0.00	0.23	0.22	0.23
(5,5,5)	0.00	0.33	0.34	0.33	0.00	-0.00	0.01	-0.00	0.00	0.23	0.24	0.24
(10,10,10)	0.00	0.35	0.33	0.33	0.00	0.01	-0.01	-0.01	0.00	0.25	0.24	0.24
(15,15,15)	0.00	0.34	0.32	0.34	0.00	0.01	-0.01	0.01	0.00	0.25	0.24	0.25
(1,5,15)	0.00	0.43	0.31	0.26	0.00	0.09	-0.02	-0.07	0.00	0.27	0.23	0.21
(15,5,1)	0.00	0.26	0.30	0.44	0.00	-0.07	-0.03	0.10	0.00	0.22	0.23	0.27
(15,1,5)	0.00	0.27	0.43	0.31	0.00	-0.07	0.09	-0.02	0.00	0.21	0.27	0.24
(1,15,5)	0.00	0.43	0.26	0.31	0.00	0.10	-0.07	-0.02	0.00	0.27	0.21	0.24
(5,15,1)	0.00	0.30	0.26	0.44	0.00	-0.03	-0.08	0.11	0.00	0.23	0.21	0.27
(5,1,15)	0.00	0.30	0.44	0.25	0.00	-0.03	0.11	-0.08	0.00	0.23	0.27	0.21

shown in Tables 3.3 and 3.4, the rank probabilities changed only slightly. For an unequal number of studies per treatment comparison within the star geometry, the rank probability of the treatment that corresponds to the least number of studies was increased by as much as 0.13. As shown in the lower rows of Table 3.4, for example, all possible permutations of the unbalanced number of studies have been checked and we can confirm that the amplification of rank probability moves with the treatment that corresponds to the smaller number of studies. Higher bias and standard deviations resulted from this overestimation.

We also investigated if the position of the reference treatment within the network

has any effect on the rank probability. Table 3.5 shows results from a star geometry with $n = 200$, but this time commuting the lower efficacy from the base position of T_1 to the outer position of T_3.

Table 3.5: Star Network Pattern with p=(0.5,0.5,0.1,0.5) and n=200

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.23	0.37	0.02	0.38	-0.10	0.04	0.02	0.04	0.09	0.12	0.00	0.12
(2,2,2)	0.24	0.37	0.00	0.39	-0.09	0.04	0.00	0.05	0.16	0.20	0.00	0.20
(3,3,3)	0.25	0.39	0.00	0.36	-0.08	0.06	0.00	0.03	0.18	0.23	0.00	0.22
(1,5,15)	0.25	0.43	0.00	0.32	-0.08	0.10	0.00	-0.02	0.19	0.26	0.00	0.22
(15, 5,1)	0.25	0.31	0.00	0.44	-0.09	-0.02	0.00	0.11	0.21	0.23	0.00	0.27
(15,1,5)	0.24	0.35	0.00	0.41	-0.09	0.01	0.00	0.07	0.20	0.25	0.00	0.26
(5,15,1)	0.25	0.33	0.00	0.42	-0.08	-0.00	0.00	0.08	0.20	0.23	0.00	0.26
(5,1,15)	0.24	0.43	0.00	0.33	-0.09	0.09	0.00	-0.00	0.20	0.27	0.00	0.24

In contrast to the equal rank probabilities observed with $p = (0.1, 0.5, 0.5, 0.5)$ and an equal number of studies per comparison, the treatment structure of $p = (0.5, 0.5, 0.1, 0.5)$ for the star geometry returns increased rank probabilities for T_2 and T_4, resulting in underestimation for that of T_1 at the base of the star. For an unequal number of studies per comparison, the gain in rank probability again follows the smaller number of studies.

For instance, if we look at the results given in Tables 3.4 and 3.5 for $n = 200$, in both cases higher rank probability is held by smaller number of studies. It can be seen that the position of the inferior treatment altered the magnitudes of rank probabilities. For instance, the rank probability of T_3 was increased by 10% while the rank probability of T_4 reduced by the same amount from their corresponding expected rank probabilities.

Tables 3.6,3.7, and 3.8 show results from the loop geometry for $p = (0.1, 0.5, 0.5, 0.5)$

with $n = 50$, $n = 100$, and $n = 200$, respectively. With an equal number of studies per comparison, T_3 (the only treatment without direct comparison with T_1) was consistently underestimated although it was expected to be equivalent in efficacy to T_2 and T_4. With an unequal number of studies per comparison, the treatment with the least number of studies directly connected to T_1 received higher rank probability, but T_3 remained to have the lowest and underestimated rank probability.

Table 3.6: Loop Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 50$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.00	0.35	0.30	0.34	0.00	0.02	-0.03	0.01	0.00	0.17	0.14	0.17
(2,2,2,2)	0.00	0.35	0.29	0.36	0.00	0.02	-0.04	0.02	0.00	0.21	0.18	0.21
(3,3,3,3)	0.00	0.36	0.28	0.36	0.00	0.02	-0.05	0.03	0.00	0.24	0.21	0.25
(5,5,5,5)	0.00	0.35	0.28	0.37	0.00	0.02	-0.05	0.04	0.00	0.24	0.21	0.25
(10,10,10,10)	0.00	0.35	0.28	0.37	0.00	0.02	-0.05	0.03	0.00	0.25	0.22	0.26
(15,15,15,15)	0.00	0.37	0.28	0.35	0.00	0.04	-0.05	0.02	0.00	0.26	0.23	0.25
(1,3,5,15)	0.00	0.39	0.27	0.34	0.00	0.05	-0.06	0.01	0.00	0.26	0.22	0.25
(15,3,5,1)	0.00	0.38	0.28	0.34	0.00	0.05	-0.06	0.01	0.00	0.25	0.21	0.24

Table 3.7: Loop Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 100$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.00	0.35	0.29	0.36	0.00	0.02	-0.05	0.02	0.00	0.15	0.12	0.15
(2,2,2,2)	0.00	0.36	0.29	0.35	0.00	0.03	-0.05	0.02	0.00	0.22	0.19	0.23
(3,3,3,3)	0.00	0.36	0.28	0.35	0.00	0.03	-0.05	0.02	0.00	0.24	0.20	0.24
(5,5,5,5)	0.00	0.34	0.29	0.37	0.00	0.01	-0.04	0.03	0.00	0.24	0.22	0.25
(10,10,10,10)	0.00	0.37	0.27	0.36	0.00	0.04	-0.06	0.02	0.00	0.26	0.23	0.26
(15,15,15,15)	0.00	0.36	0.28	0.36	0.00	0.03	-0.06	0.03	0.00	0.26	0.23	0.26
(1,3,5,15)	0.00	0.39	0.26	0.35	0.00	0.06	-0.07	0.01	0.00	0.26	0.21	0.25
(15,3,5,1)	0.00	0.31	0.24	0.45	0.00	-0.02	-0.09	0.11	0.00	0.23	0.19	0.26

Tables 3.9, 3.10, and 3.11 show results from changing the position of the inferior treatment from T_1 to T_3 in the loop geometry. Results are comparable to those when $p = (0.1, 0.5, 0.5, 0.5)$, except with T_1 having the underestimated rank probabilities instead of T_3.

We also investigated if estimates from a loop geometry with every treatment directly compared to one another (complete, or connected network polygon) would be

Table 3.8: Loop Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 200$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.00	0.35	0.29	0.35	0.00	0.02	-0.04	0.02	0.00	0.13	0.12	0.13
(2,2,2,2)	0.00	0.34	0.31	0.35	0.00	0.01	-0.03	0.02	0.00	0.22	0.20	0.21
(3,3,3,3)	0.00	0.36	0.29	0.35	0.00	0.03	-0.05	0.02	0.00	0.24	0.21	0.24
(5,5,5,5)	0.00	0.36	0.28	0.36	0.00	0.02	-0.05	0.03	0.00	0.25	0.22	0.25
(10,10,10,10)	0.00	0.36	0.28	0.36	0.00	0.02	-0.06	0.03	0.00	0.27	0.23	0.26
(15,15,15,15)	0.00	0.35	0.30	0.36	0.00	0.02	-0.04	0.02	0.00	0.25	0.23	0.25
(1,3,5,15)	0.00	0.38	0.27	0.35	0.00	0.04	-0.06	0.02	0.00	0.25	0.21	0.25
(15,3,5,1)	0.00	0.37	0.28	0.35	0.00	0.04	-0.05	0.02	0.00	0.25	0.21	0.24

Table 3.9: Loop Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 50$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.29	0.35	0.00	0.35	-0.04	0.02	0.00	0.02	0.15	0.17	0.00	0.17
(2,2,2,2)	0.29	0.36	0.00	0.36	-0.05	0.02	0.00	0.02	0.19	0.21	0.00	0.21
(3,3,3,3)	0.28	0.36	0.00	0.36	-0.05	0.02	0.00	0.03	0.21	0.24	0.00	0.25
(5,5,5,5)	0.28	0.35	0.00	0.37	-0.05	0.02	0.00	0.03	0.22	0.25	0.00	0.24
(10,10,10,10)	0.29	0.35	0.00	0.37	-0.05	0.01	0.00	0.03	0.23	0.26	0.00	0.26
(15,15,15,15)	0.29	0.36	0.00	0.35	-0.04	0.03	0.00	0.01	0.23	0.26	0.00	0.25
(1,3,5,15)	0.28	0.45	0.00	0.27	-0.05	0.12	0.00	-0.06	0.22	0.27	0.00	0.22
(15,3,5,1)	0.29	0.27	0.00	0.44	-0.05	-0.06	0.00	0.11	0.22	0.21	0.00	0.27

different. The results are shown in Tables 3.12, 3.13 and 3.14. Under this condition and an equal number of studies per comparison, the loop network estimates were improved, moving very close to the expected values. Again, however, differing the number of studies per comparison resulted in the highest rank probability following the lowest number of studies.

Results from the one closed loop geometry with $p = (0.1, 0.5, 0.5, 0.5)$ are presented in Tables 3.15 to 3.17. Under an equal number of studies per comparison in this scenario, the rank probability of T_3 is underestimated while that of T_2 and T_4 is overestimated, as was also found in the case of the loop geometry. No improvement in rank probability is attained by increasing the study sample size and number of studies per comparison.

For an unbalanced number of studies per comparison, higher rank probability

Table 3.10: Loop Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 100$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.30	0.34	0.00	0.35	-0.04	0.01	0.00	0.02	0.13	0.15	0.00	0.15
(2,2,2,2)	0.30	0.35	0.00	0.35	-0.03	0.01	0.00	0.02	0.20	0.22	0.00	0.22
(3,3,3,3)	0.29	0.36	0.00	0.35	-0.05	0.03	0.00	0.02	0.21	0.24	0.00	0.23
(5,5,5,5)	0.29	0.34	0.00	0.37	-0.04	0.01	0.00	0.03	0.22	0.25	0.00	0.25
(10,10,10,10)	0.28	0.36	0.00	0.36	-0.06	0.03	0.00	0.02	0.22	0.26	0.00	0.25
(15,15,15,15)	0.29	0.34	0.00	0.37	-0.04	0.01	0.00	0.03	0.23	0.26	0.00	0.26
(1,3,5,15)	0.29	0.43	0.00	0.28	-0.04	0.10	0.00	-0.05	0.23	0.28	0.00	0.22
(15,3,5,1)	0.29	0.28	0.00	0.43	-0.04	-0.05	0.00	0.10	0.23	0.23	0.00	0.27

Table 3.11: Loop Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 200$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.30	0.35	0.00	0.35	-0.04	0.02	0.00	0.01	0.12	0.13	0.00	0.14
(2,2,2,2)	0.28	0.37	0.00	0.35	-0.06	0.04	0.00	0.02	0.19	0.22	0.00	0.21
(3,3,3,3)	0.28	0.35	0.00	0.37	-0.05	0.02	0.00	0.03	0.21	0.24	0.00	0.24
(5,5,5,5)	0.29	0.34	0.00	0.37	-0.04	0.01	0.00	0.03	0.22	0.24	0.00	0.24
(10,10,10,10)	0.28	0.35	0.00	0.36	-0.05	0.02	0.00	0.03	0.23	0.25	0.00	0.25
(15,15,15,15)	0.28	0.36	0.00	0.36	-0.06	0.03	0.00	0.03	0.23	0.26	0.00	0.26
(1,3,5,15)	0.28	0.45	0.00	0.27	-0.06	0.12	0.00	-0.06	0.22	0.27	0.00	0.22
(15,3,5,1)	0.27	0.28	0.00	0.45	-0.06	-0.05	0.00	0.12	0.21	0.22	0.00	0.27

were assigned to treatments with the fewer number of studies by the one closed geometry but unlike in the case of loop, a higher number of studies slightly adjusted the overestimated rank probabilities of T_2 and T_4 down to its expected value. Bias of this pattern was relatively large as compared to other networks patterns.

Tables 3.18 to 3.20 show the results when the inferior efficacy is placed on the T_3 position for the one closed loop geometry. Under these scenarios with a balanced number of studies, T_4 was identified as the most efficacious with a rank probability of 0.41 or more. Changing the study size n and the number of studies per comparison has negligible effect on the estimates.

With an unbalanced number of studies per comparison, rank probability estimates for treatments associated to a lesser number of studies were pulled close to their expected values whereas previously overestimated associated to a higher number of

Table 3.12: Loop (complete) Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 50$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1,1,1)	0.00	0.33	0.33	0.34	0.00	-0.01	-0.00	0.00	0.00	0.19	0.19	0.19
(2,2,2,2,2,2)	0.00	0.33	0.33	0.34	0.00	-0.00	0.00	0.00	0.00	0.23	0.23	0.23
(3,3,3,3,3,3)	0.00	0.33	0.34	0.33	0.00	-0.00	0.01	-0.00	0.00	0.23	0.24	0.23
(5,5,5,5,5,5)	0.00	0.33	0.34	0.33	0.00	0.00	0.01	-0.01	0.00	0.25	0.25	0.24
(10,10,10,10,10,10)	0.00	0.33	0.35	0.33	0.00	-0.01	0.01	-0.01	0.00	0.25	0.26	0.24
(15,15,15,15,15,15)	0.00	0.34	0.33	0.33	0.00	0.01	-0.00	-0.01	0.00	0.25	0.25	0.25
(1,2,3,5,10,15)	0.00	0.35	0.34	0.31	0.00	0.02	0.01	-0.03	0.00	0.26	0.25	0.23
(15,10,5,3,2,1)	0.00	0.29	0.33	0.38	0.00	-0.04	-0.01	0.05	0.00	0.23	0.24	0.26

Table 3.13: Loop (complete) Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 100$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1,1,1)	0.00	0.34	0.33	0.33	0.00	0.01	-0.01	-0.00	0.00	0.19	0.19	0.19
(2,2,2,2,2,2)	0.00	0.34	0.33	0.33	0.00	0.01	-0.00	-0.00	0.00	0.23	0.23	0.23
(3,3,3,3,3,3)	0.00	0.34	0.31	0.35	0.00	0.00	-0.02	0.02	0.00	0.24	0.24	0.24
(5,5,5,5,5,5)	0.00	0.32	0.35	0.33	0.00	-0.02	0.02	-0.00	0.00	0.24	0.25	0.24
(10,10,10,10,10,10)	0.00	0.34	0.33	0.33	0.00	0.01	-0.00	-0.00	0.00	0.25	0.25	0.25
(15,15,15,15,15,15)	0.00	0.32	0.34	0.33	0.00	-0.01	0.01	-0.00	0.00	0.24	0.26	0.25
(1,2,3,5,10,15)	0.00	0.36	0.35	0.30	0.00	0.02	0.01	-0.04	0.00	0.26	0.25	0.23
(15,10,5,3,2,1)	0.00	0.29	0.33	0.38	0.00	-0.04	-0.00	0.04	0.00	0.22	0.24	0.25

studies were adjusted down to their expected values. We observe that the treatment with a lesser number of studies always gains rank probability whiles the one with a higher number of studies loses.

Tables 3.21 to 3.26 show comparable results for the ladder geometry. Despite the difference in geometries, rank probabilities for treatments in a ladder geometry are similar to those in one closed loop. However, as apparent from Tables 3.24 to 3.26, the underestimation of the rank probability for T_2 when the set of probabilities is $p = (0.5, 0.5, 0.1, 0.5)$ (i.e, when inferior efficacy moves from T_1 to T_3) is more pronounced in the ladder geometry.

Table 3.14: Loop (complete) Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 200$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1,1,1)	0.00	0.33	0.33	0.34	0.00	-0.00	0.00	0.00	0.00	0.19	0.19	0.19
(2,2,2,2,2,2)	0.00	0.32	0.33	0.35	0.00	-0.01	0.00	0.01	0.00	0.23	0.23	0.23
(3,3,3,3,3,3)	0.00	0.33	0.33	0.34	0.00	0.00	-0.01	0.00	0.00	0.23	0.23	0.24
(5,5,5,5,5,5)	0.00	0.35	0.32	0.32	0.00	0.02	-0.01	-0.01	0.00	0.25	0.24	0.24
(10,10,10,10,10,10)	0.00	0.32	0.33	0.35	0.00	-0.01	-0.00	0.01	0.00	0.24	0.25	0.25
(15,15,15,15,15,15)	0.00	0.34	0.33	0.33	0.00	0.01	-0.01	-0.00	0.00	0.26	0.25	0.25
(1,2,3,5,10,15)	0.00	0.37	0.33	0.29	0.00	0.04	-0.00	-0.04	0.00	0.26	0.25	0.23
(15,10,5,3,2,1)	0.00	0.29	0.34	0.38	0.00	-0.05	0.00	0.04	0.00	0.23	0.25	0.26

Table 3.15: One Closed Loop Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 50$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.00	0.36	0.25	0.39	0.00	0.02	-0.08	0.05	0.00	0.16	0.13	0.16
(2,2,2,2)	0.00	0.36	0.25	0.39	0.00	0.03	-0.08	0.05	0.00	0.22	0.18	0.22
(3,3,3,3)	0.00	0.36	0.24	0.39	0.00	0.03	-0.09	0.06	0.00	0.24	0.18	0.24
(5,5,5,5)	0.00	0.36	0.24	0.40	0.00	0.03	-0.09	0.06	0.00	0.25	0.20	0.26
(10,10,10,10)	0.00	0.37	0.24	0.38	0.00	0.04	-0.09	0.05	0.00	0.26	0.21	0.26
(15,15,15,15)	0.00	0.36	0.26	0.38	0.00	0.03	-0.07	0.04	0.00	0.26	0.22	0.26
(1,3,5,15)	0.00	0.42	0.26	0.32	0.00	0.08	-0.07	-0.01	0.00	0.26	0.21	0.24
(15,3,5,1)	0.00	0.31	0.25	0.44	0.00	-0.03	-0.08	0.11	0.00	0.24	0.20	0.27

Table 3.16: One Closed Loop Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 100$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.00	0.36	0.25	0.38	0.00	0.03	-0.09	0.05	0.00	0.15	0.12	0.15
(2,2,2,2)	0.00	0.36	0.26	0.39	0.00	0.02	-0.08	0.05	0.00	0.22	0.18	0.23
(3,3,3,3)	0.00	0.36	0.25	0.40	0.00	0.02	-0.09	0.06	0.00	0.23	0.19	0.24
(5,5,5,5)	0.00	0.37	0.26	0.38	0.00	0.03	-0.08	0.04	0.00	0.25	0.21	0.25
(10,10,10,10)	0.00	0.37	0.26	0.37	0.00	0.04	-0.08	0.04	0.00	0.25	0.21	0.26
(15,15,15,15)	0.00	0.37	0.25	0.38	0.00	0.04	-0.09	0.05	0.00	0.26	0.21	0.26
(1,3,5,15)	0.00	0.41	0.25	0.34	0.00	0.08	-0.08	0.00	0.00	0.26	0.21	0.24
(15,3,5,1)	0.00	0.31	0.26	0.43	0.00	-0.02	-0.07	0.09	0.00	0.23	0.21	0.27

Table 3.17: One Closed Loop Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 200$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.00	0.36	0.24	0.39	0.00	0.03	-0.09	0.06	0.00	0.14	0.10	0.14
(2,2,2,2)	0.00	0.35	0.26	0.38	0.00	0.02	-0.07	0.05	0.00	0.22	0.19	0.22
(3,3,3,3)	0.00	0.35	0.26	0.39	0.00	0.02	-0.08	0.06	0.00	0.23	0.19	0.24
(5,5,5,5)	0.00	0.37	0.25	0.37	0.00	0.04	-0.08	0.04	0.00	0.25	0.20	0.25
(10,10,10,10)	0.00	0.36	0.25	0.39	0.00	0.03	-0.09	0.06	0.00	0.26	0.21	0.27
(15,15,15,15)	0.00	0.36	0.24	0.40	0.00	0.03	-0.09	0.07	0.00	0.26	0.21	0.27
(1,3,5,15)	0.00	0.42	0.25	0.33	0.00	0.08	-0.08	-0.00	0.00	0.27	0.21	0.24
(15,3,5,1)	0.00	0.31	0.25	0.44	0.00	-0.02	-0.08	0.11	0.00	0.23	0.20	0.26

Table 3.18: One Closed Loop Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 50$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.29	0.29	0.00	0.41	-0.04	-0.04	0.00	0.08	0.15	0.15	0.00	0.19
(2,2,2,2)	0.29	0.29	0.00	0.43	-0.05	-0.05	0.00	0.09	0.21	0.20	0.00	0.25
(3,3,3,3)	0.29	0.29	0.00	0.42	-0.04	-0.04	0.00	0.09	0.22	0.21	0.00	0.26
(5,5,5,5)	0.28	0.29	0.00	0.43	-0.06	-0.04	0.00	0.10	0.22	0.23	0.00	0.27
(10,10,10,10)	0.30	0.28	0.00	0.41	-0.03	-0.05	0.00	0.08	0.23	0.22	0.00	0.27
(15,15,15,15)	0.30	0.29	0.00	0.42	-0.04	-0.05	0.00	0.08	0.23	0.22	0.00	0.27
(1,3,5,15)	0.36	0.32	0.00	0.31	0.03	-0.01	0.00	-0.02	0.25	0.23	0.00	0.24
(15,3,5,1)	0.26	0.24	0.00	0.50	-0.07	-0.09	0.00	0.16	0.23	0.21	0.00	0.30

Table 3.19: One Closed Loop Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 100$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.30	0.30	0.00	0.41	-0.04	-0.04	0.00	0.07	0.13	0.14	0.00	0.17
(2,2,2,2)	0.29	0.28	0.00	0.42	-0.04	-0.05	0.00	0.09	0.21	0.20	0.00	0.24
c(3,3,3,3)	0.28	0.29	0.00	0.42	-0.05	-0.04	0.00	0.09	0.22	0.22	0.00	0.26
(5,5,5,5)	0.29	0.29	0.00	0.41	-0.04	-0.04	0.00	0.08	0.23	0.22	0.00	0.27
(10,10,10,10)	0.28	0.29	0.00	0.42	-0.05	-0.04	0.00	0.09	0.23	0.24	0.00	0.28
(15,15,15,15)	0.30	0.27	0.00	0.43	-0.04	-0.06	0.00	0.10	0.24	0.22	0.00	0.27
(1,3,5,15)	0.36	0.31	0.00	0.33	0.03	-0.02	0.00	-0.01	0.25	0.23	0.00	0.25
(15,3,5,1)	0.27	0.25	0.00	0.47	-0.06	-0.08	0.00	0.14	0.22	0.21	0.00	0.28

Table 3.20: One Closed Loop Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 200$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.30	0.29	0.00	0.41	-0.04	-0.04	0.00	0.08	0.12	0.12	0.00	0.15
(2,2,2,2)	0.28	0.28	0.00	0.44	-0.06	-0.05	0.00	0.10	0.20	0.20	0.00	0.24
(3,3,3,3)	0.29	0.29	0.00	0.42	-0.04	-0.05	0.00	0.09	0.22	0.22	0.00	0.26
(5,5,5,5)	0.30	0.29	0.00	0.41	-0.03	-0.05	0.00	0.08	0.23	0.22	0.00	0.27
(10,10,10,10)	0.29	0.30	0.00	0.41	-0.05	-0.03	0.00	0.08	0.22	0.23	0.00	0.27
(15,15,15,15)	0.28	0.30	0.00	0.42	-0.06	-0.03	0.00	0.09	0.22	0.24	0.00	0.28
(1,3,5,15)	0.35	0.33	0.00	0.32	0.02	-0.00	0.00	-0.01	0.25	0.24	0.00	0.24
(15,3,5,1)	0.25	0.26	0.00	0.49	-0.08	-0.08	0.00	0.16	0.21	0.22	0.00	0.28

Table 3.21: Ladder Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 50$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.02	0.36	0.24	0.38	0.02	0.02	-0.09	0.04	0.01	0.15	0.11	0.15
(2,2,2)	0.00	0.38	0.24	0.38	0.00	0.04	-0.09	0.05	0.00	0.20	0.15	0.20
(3,3,3)	0.00	0.37	0.25	0.39	0.00	0.03	-0.09	0.06	0.00	0.23	0.19	0.24
(5,5,5)	0.00	0.38	0.25	0.37	0.00	0.05	-0.08	0.03	0.00	0.25	0.20	0.24
(10,10,10)	0.00	0.38	0.25	0.37	0.00	0.04	-0.08	0.04	0.00	0.26	0.21	0.26
(15,15,15)	0.00	0.38	0.26	0.37	0.00	0.04	-0.07	0.03	0.00	0.26	0.21	0.26
(1,5,15)	0.00	0.41	0.25	0.33	0.00	0.08	-0.08	0.00	0.00	0.27	0.21	0.25
(15,5,1)	0.00	0.32	0.24	0.43	0.00	-0.01	-0.09	0.10	0.00	0.23	0.19	0.25

Table 3.22: Ladder Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 100$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.02	0.36	0.25	0.37	0.02	0.03	-0.09	0.04	0.00	0.12	0.09	0.12
(2,2,2)	0.00	0.38	0.24	0.38	0.00	0.04	-0.09	0.05	0.00	0.20	0.16	0.21
(3,3,3)	0.00	0.38	0.24	0.38	0.00	0.04	-0.09	0.05	0.00	0.24	0.18	0.23
(5,5,5)	0.00	0.38	0.26	0.36	0.00	0.05	-0.08	0.03	0.00	0.25	0.20	0.24
(10,10,10)	0.00	0.35	0.26	0.38	0.00	0.02	-0.07	0.05	0.00	0.25	0.21	0.26
(15,15,15)	0.00	0.37	0.26	0.37	0.00	0.04	-0.08	0.04	0.00	0.26	0.21	0.26
(1,5,15)	0.00	0.41	0.26	0.33	0.00	0.08	-0.08	0.00	0.00	0.26	0.20	0.24
(15,5,1)	0.00	0.31	0.24	0.45	0.00	-0.02	-0.09	0.11	0.00	0.23	0.19	0.26

Table 3.23: Ladder Network Pattern with $p = (0.1, 0.5, 0.5, 0.5)$ and $n = 200$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.02	0.36	0.24	0.37	0.02	0.03	-0.09	0.04	0.00	0.09	0.07	0.09
(2,2,2)	0.00	0.37	0.25	0.38	0.00	0.04	-0.09	0.05	0.00	0.20	0.16	0.20
(3,3,3)	0.00	0.38	0.25	0.38	0.00	0.04	-0.09	0.05	0.00	0.23	0.18	0.23
(5,5,5)	0.00	0.37	0.25	0.38	0.00	0.04	-0.08	0.04	0.00	0.24	0.20	0.24
(10,10,10)	0.00	0.39	0.24	0.37	0.00	0.05	-0.09	0.04	0.00	0.26	0.21	0.26
(15,15,15)	0.00	0.37	0.25	0.38	0.00	0.04	-0.09	0.05	0.00	0.26	0.21	0.26
(1,5,15)	0.00	0.41	0.25	0.34	0.00	0.08	-0.08	0.00	0.00	0.26	0.21	0.25
(15,5,1)	0.00	0.32	0.25	0.43	0.00	-0.01	-0.09	0.09	0.00	0.22	0.19	0.25

Table 3.24: Ladder Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 50$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.33	0.25	0.00	0.42	-0.00	-0.08	0.00	0.08	0.14	0.12	0.00	0.18
(2,2,2)	0.33	0.25	0.00	0.42	-0.01	-0.08	0.00	0.09	0.18	0.16	0.00	0.23
(3,3,3)	0.32	0.24	0.00	0.44	-0.01	-0.09	0.00	0.10	0.21	0.19	0.00	0.25
(5,5,5)	0.34	0.25	0.00	0.42	0.00	-0.08	0.00	0.08	0.24	0.20	0.00	0.27
(10,10,10)	0.34	0.25	0.00	0.42	0.00	-0.09	0.00	0.08	0.25	0.22	0.00	0.27
(15,15,15)	0.33	0.25	0.00	0.42	-0.00	-0.08	0.00	0.08	0.24	0.21	0.00	0.27
(1,5,15)	0.40	0.25	0.00	0.35	0.07	-0.08	0.00	0.01	0.25	0.20	0.00	0.25
(15,5,1)	0.27	0.23	0.00	0.49	-0.06	-0.10	0.00	0.16	0.23	0.21	0.00	0.29

Table 3.25: Ladder Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 100$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.34	0.25	0.00	0.41	0.00	-0.09	0.00	0.08	0.11	0.09	0.00	0.14
(2,2,2)	0.33	0.25	0.00	0.42	-0.00	-0.09	0.00	0.09	0.19	0.16	0.00	0.22
(3,3,3)	0.32	0.25	0.00	0.42	-0.01	-0.08	0.00	0.09	0.22	0.18	0.00	0.25
(5,5,5)	0.32	0.26	0.00	0.42	-0.01	-0.08	0.00	0.09	0.23	0.20	0.00	0.27
(10,10,10)	0.32	0.24	0.00	0.43	-0.01	-0.09	0.00	0.10	0.24	0.20	0.00	0.27
(15,15,15)	0.33	0.26	0.00	0.41	-0.00	-0.08	0.00	0.08	0.25	0.22	0.00	0.27
(1,5,15)	0.41	0.25	0.00	0.34	0.08	-0.08	0.00	0.01	0.25	0.19	0.00	0.23
(15,5,1)	0.28	0.25	0.00	0.48	-0.06	-0.09	0.00	0.14	0.22	0.20	0.00	0.28

Table 3.26: Ladder Network Pattern with $p = (0.5, 0.5, 0.1, 0.5)$ and $n = 200$

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.34	0.25	0.00	0.41	0.01	-0.09	0.00	0.07	0.09	0.08	0.00	0.11
(2,2,2)	0.34	0.26	0.00	0.41	0.00	-0.08	0.00	0.07	0.18	0.16	0.00	0.22
(3,3,3)	0.33	0.25	0.00	0.42	-0.00	-0.08	0.00	0.08	0.22	0.19	0.00	0.25
(5,5,5)	0.32	0.26	0.00	0.43	-0.02	-0.08	0.00	0.10	0.23	0.20	0.00	0.26
(10,10,10)	0.33	0.25	0.00	0.42	0.00	-0.09	0.00	0.09	0.25	0.21	0.00	0.27
(15,15,15)	0.31	0.26	0.00	0.42	-0.02	-0.07	0.00	0.09	0.24	0.22	0.00	0.27
(1,5,15)	0.41	0.24	0.00	0.35	0.07	-0.09	0.00	0.02	0.26	0.19	0.00	0.24
(15,5,1)	0.26	0.26	0.00	0.48	-0.07	-0.08	0.00	0.15	0.22	0.22	0.00	0.28

We now present results from the next scenario, where $p = (0.2, 0.2, 0.2, 0.8)$, meaning one treatment (T_4) is superior in efficacy compared to three equally less effective treatments (T_1, T_2, and T_3). The expected value of the event for T_4 is 0.8, and for each of T_1, T_2 and T_3, the mean is 0.2. Here, we expect T_4 to be identified by the model under each condition as the first effective treatment among others.

Under this scenario, all network shapes we considered in our study identified the most powerful treatment with highest rank probability as expected. In each patterns, both number of studies per comparison and individual study sample size slightly contributes to the valid detection of the superior treatment (See tables 3.27 to 3.30). The impact of the number of studies on the rank probability was at most about 0.01.

In this scenario, bias and standard deviations were generally very small in all cases. Since the results from every scenarios is not really different, we chose to present only the results with sample size 50 for each network patterns. The results of the scenarios for which we did not include here are very similar even when there is only a single treatment between comparison.

Table 3.27: Star for n=50,p=(0.2,0.2,0.2,0.8)

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.01	0.08	0.08	0.83	0.01	0.08	0.08	-0.17	0.00	0.04	0.04	0.04
(2,2,2)	0.00	0.01	0.01	0.98	0.00	0.01	0.01	-0.02	0.00	0.01	0.01	0.02
(3,3,3)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(5,5,5)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(10,10,10)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(15,15,15)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(1,5,15)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.01	0.00	0.01
(15,5,1)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00

Table 3.28: Loop for n=50,p=(0.2,0.2,0.2,0.8)

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.02	0.03	0.02	0.94	0.02	0.03	0.02	-0.06	0.01	0.02	0.01	0.02
(2,2,2,2)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(3,3,3,3)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(5,5,5,5)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(10,10,10,10)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(15,15,15,15)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(1,3,5,15)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(15,3,5,1)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00

Table 3.29: One closed Loop for n=50,p=(0.2,0.2,0.2,0.8)

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1,1)	0.04	0.04	0.01	0.91	0.04	0.04	0.01	-0.09	0.02	0.02	0.00	0.03
(2,2,2,2)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.01
(3,3,3,3)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(5,5,5,5)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(10,10,10,10)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(15,15,15,15)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(1,3,5,15)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(15,3,5,1)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00

Table 3.30: Ladder for n=50,p=(0.2,0.2,0.2,0.8)

	Rank Probability				Bias				Standard deviation			
No Studies	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
(1,1,1)	0.11	0.05	0.02	0.83	0.11	0.05	0.02	-0.17	0.06	0.02	0.00	0.06
(2,2,2)	0.02	0.01	0.00	0.97	0.02	0.01	0.00	-0.03	0.02	0.01	0.00	0.03
(3,3,3)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.01	0.00	0.00	0.01
(5,5,5)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
(10,10,10)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(15,15,15)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(1,5,15)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.01	0.00	0.00	0.01
(15,5,1)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00

Chapter 4

Summary, Discussion and Future Directions

4.1 Summary and Discussion

Meta-analysis is the statistical process of combining the results from separate studies concerned with the same intervention. Standard meta-analyses involve only pair-wise comparisons of two treatments but network meta-analysis, which can include multiple treatments in the same meta-analysis, has attracted much attention in recent years. Network meta-analysis offers the advantage of being able to compare any treatments included in the network, including those that have not been compared directly. Applications that compare multiple treatments and network meta-analysis modeling are increasingly becoming common, but research on this method is still in its infancy.

In Chapter 2, we have noted in our methodology section that network meta-analysis can be done in both pure frequentist framework or in Bayesian framework. However, the literature is in favor of Bayesian network meta-analysis because of many advantages conferred by this approach. One benefits from Bayesian approach is that there is an allowance to incorporate prior belief about parameters of interest in the analysis. The other important advantage of Bayesian method in network meta-analysis is that the posterior probability distribution allows calculating the probability of each of the competing interventions which decision makers might want to know.

In Chapter 3 we have described the design of our simulation. In the Bayesian network meta-analysis, it is important to determine the priors of the basic parameters and variances. The default values in the "gemtc" R package were used to estimate parameters in the network meta-analysis models using MCMC approach. The default parameters included weakly informative priors, number of burn-in samples, number of iterations for inference, and variance scaling factor. We also allowed the package to set its initial values to all parameters as it can assign initial values if not specified. However, as we remarked in Chapter 3,the thinning number is used to reduce the impact of auto-correlation on the convergence of the posterior distribution which we have changed from the default value 1 in "gemtc" to 10 and the convergence of posterior distributions were checked by using Brooks-Rubin test statistic. Therefore, in the real data network meta-analysis, there is a room to change all these default values when needed. With this quick summary, we discuss our simulation study findings as follows.

In this simulation study we have observed that, under balanced treatment comparison scenario with evenly distributed number of studies in the network geometry, the star pattern resulted in almost correct conclusion as compared to other network patterns. However, with unevenly distributed number of studies, star can lead to wrong conclusion as it over estimated the rank probability of the treatment associated with smaller number of studies.

The network patterns, Loop, One closed loop and Ladder were found to deflate the estimates of the treatment that compared with larger number of other treatments (T_3) even though the treatment was as effective as other treatments in competition (T_2 and T_4) by its nature. These geometries, favor treatments (T_2 and T_4) that are compared less number of times with others. Moreover, increasing number of studies per comparison does not help to solve the problem of deflating the rank probability of the treatment that compared with more treatments and inflating the estimates of those that are compared with others less number of times. Increasing the sample size of individual studies does not avoid the problem.

We have investigated the effect of the position of the reference treatment in the network by switching the reference treatment (T_1) and the treatment that compared with more number of treatments (T_3). Under this experiment, Ladder and One closed loop network patterns assigned significant rank probability(over 41%) to T_4 that was compared only with the inferior reference treatment. Other treatments, T_2 and T_3 which are equivalent in efficacy took relatively equally deflated rank probability. Similar result were obtained for Loop geometry. This shows that the extent to which specific treatments or specific comparisons represented in the network can dictate

the estimates of the treatment. This is termed as the effect of asymmetry.

When all of the treatments included in the network have been compared against each other (a connected polygonal network configuration), all possible closed loops are represented. To investigate how treatments can be identified for symmetric or all connected polygonal geometry, we network meta-analyzed the data for all connected loop geometry. The results showed that,the problem of deflating estimates of treatments that connected to larger number of treatments and inflating estimates of those connected to less number of treatments has solved. This shows that, a fully connected or symmetric geometry helps in drawing valid conclusion. Unlike in the case of balanced treatments scenario, when there is super effective treatment among treatments included in the network, all the network geometries we considered supported the correct conclusion.

In our simulation, we found that under equivalent scenario with unequal number of studies per comparison, all geometries evolved in the study can lead to wrong conclusion, (i.e., the estimate of treatment associated with smaller number of studies was inflated by the Bayesian Hierarchical model). The source of this bias and possible way of adjustment should be identified.

A treatment connected to large number of competent treatments was seen to lose its power of competence, (i.e., always underestimated) in all closed connected networks. Except star geometry, all other patterns identified treatments that are linked to less number of other treatments as the first best treatment even though originally there was no difference among the treatments in efficacy. Therefore, due consideration should be given to the number of treatments compared to a particular

treatment before making conclusion particularly when there is a closed connection among treatments.

When there is a superior treatment among the competing treatments, the network meta-analysis model was able to identify the most effective treatment regardless of the network geometry, individual study sample sizes and number of studies per comparisons. We have also investigated if the study sample sizes can drastically change the decision and no results have significantly changed. However, the degree of superiority might need further investigation, (i.e., for what degree of the difference in efficacy can the model correctly identify treatments correctly). When every treatment is symmetrically connected to one another, then the model would be much likely to approximate the underlying truth.

4.2 Future Directions

In this study we have evaluated through simulations the performance of hierarchical Bayesian network meta-analysis model in identifying the first effective treatment under different scenarios using binary outcome. In reality, the primary outcome can be continuous, count or time-to-event type. We would like to extend our simulation study to some critical issues related to network meta-analysis with all these different types of outcomes. We also limited our study to only two arm studies while in practice there are cases where multi-arm studies are available in network meta-analysis. We plan to carry out further investigation for contrast-based network meta-analysis, especially when there exist multi-arm studies that induce correlation

between contrasts that involve common comparator.

Bibliography

Borenstein, M., Hedges, L. V., Higgins, J., and Rothstein, H. R. (2009). Fixed-effect versus random-effects models. *Introduction to Meta-analysis*, pages 77–86.

Borenstein, M., Hedges, L. V., Higgins, J., and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2):97–111.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455.

Bucher, H. C., Guyatt, G. H., Griffith, L. E., and Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of clinical epidemiology*, 50(6):683–691.

Caldwell, D. M., Ades, A., and Higgins, J. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ: British Medical Journal*, 331(7521):897.

Cipriani, A., Higgins, J. P., Geddes, J. R., and Salanti, G. (2013). Conceptual

and technical challenges in network meta-analysis. *Annals of internal medicine*, 159(2):130–137.

Cook, D. J., Mulrow, C. D., and Haynes, R. B. (1997). Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of internal medicine*, 126(5):376–380.

Dias, S., Welton, N., Caldwell, D., and Ades, A. (2010). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in medicine*, 29(7-8):932–944.

Donegan, S., Williamson, P., D'Alessandro, U., and Tudur Smith, C. (2013). Assessing key assumptions of network meta-analysis: a review of methods. *Research Synthesis Methods*.

Egger, M., Ebrahim, S., and Smith, G. D. (2002). Where now for meta-analysis? *International Journal of Epidemiology*, 31(1):1–5.

Franchini, A., Dias, S., Ades, A., Jansen, J., and Welton, N. (2012). Accounting for correlation in network meta-analysis with multi-arm trials. *Research Synthesis Methods*, 3(2):142–160.

Haidich, A. (2010). Meta-analysis in medical research. *Hippokratia*, 14(Suppl 1):29.

Hanka, R. (1994). The handbook of research synthesis. *BMJ*, 309(6952):488–489.

Hedges, L. V., Cooper, H., and Bushman, B. J. (1992). Testing the null hypothesis in meta-analysis: A comparison of combined probability and confidence interval procedures. *Psychological Bulletin*, 111(1):188.

Hedges, L. V. and Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods*, 3(4):486.

Higgins, J. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11):1539–1558.

Higgins, J. and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in medicine*, 15(24):2733–2749.

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, 327(7414):557.

Hoaglin, D. C., Hawkins, N., Jansen, J. P., Scott, D. A., Itzler, R., Cappelleri, J. C., Boersma, C., Thompson, D., Larholt, K. M., Diaz, M., et al. (2011). Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ispor task force on indirect treatment comparisons good research practices: part 2. *Value in Health*, 14(4):429–437.

Ioannidis, J. (2006). Indirect comparisons: the mesh and mess of clinical trials. *The Lancet*, 368(9546):1470–1472.

Jansen, J. P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., Lee, K., Boersma, C., Annemans, L., and Cappelleri, J. C. (2011). 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 in Health*, 14(4):417–428.

Jonas, D. E., Wilkins, T. M., Bangdiwala, S., Bann, C. M., Morgan, L. C., Thaler, K. J., Amick, H. R., and Gartlehner, G. (2013). Findings of bayesian mixed treatment comparison meta-analyses.

Li, T., Puhan, M., Vedula, S., Singh, S., Dickersin, K., et al. (2011). Network meta-analysis-highly attractive but more methodological research is needed. *BMC medicine*, 9(1):79.

Li, T., Vedula, S. S., Scherer, R., and Dickersin, K. (2012). What comparative effectiveness research is needed? a framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. *Annals of internal medicine*, 156(5):367–377.

Lu, G. and Ades, A. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in medicine*, 23(20):3105–3124.

Lu, G. and Ades, A. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, 101(474).

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in medicine*, 21(16):2313–2324.

Mills, E. J., Ghement, I., O'Regan, C., and Thorlund, K. (2011). Estimating the power of indirect comparisons: a simulation study. *PLoS One*, 6(1):e16237.

Mills, E. J., Thorlund, K., and Ioannidis, J. (2013). Demystifying trial networks and network meta-analysis. *BMJ: British Medical Journal*, 346.

Nelson, J. P. and Kennedy, P. E. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment. *Environmental and Resource Economics*, 42(3):345–377.

Olkin, I. and Sampson, A. (1998). Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics*, pages 317–322.

O'Rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*, 100(12):579–582.

Psaty, B. M., Lumley, T., Furberg, C. D., Schellenbaum, G., Pahor, M., Alderman, M. H., and Weiss, N. S. (2003). Health outcomes associated with various antihypertensive therapies used as first-line agents. *JAMA: the journal of the American Medical Association*, 289(19):2534–2544.

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

Salanti, G., Higgins, J. P., Ades, A., and Ioannidis, J. P. (2008a). Evaluation of networks of randomized trials. *Statistical methods in medical research*, 17(3):279–301.

Salanti, G., Kavvoura, F. K., and Ioannidis, J. P. (2008b). Exploring the geometry of treatment networks. *Annals of internal medicine*, 148(7):544–553.

Song, F., Clark, A., Bachmann, M. O., and Maas, J. (2012). Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC medical research methodology*, 12(1):138.

Song, F., Loke, Y. K., Walsh, T., Glenny, A.-M., Eastwood, A. J., and Altman, D. G. (2009). Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ: British Medical Journal*, 338.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stettler, C., Wandel, S., Allemann, S., Kastrati, A., Morice, M. C., Schömig, A., Pfisterer, M. E., Stone, G. W., Leon, M. B., de Lezo, J. S., et al. (2007). Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *The Lancet*, 370(9591):937–948.

Sutton, A., Ades, A., Cooper, N., and Abrams, K. (2008). Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics*, 26(9):753–767.

Thompson, S. G. and Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in medicine*, 18(20):2693–2708.

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, 36(3):1–48.

Wells, G., Sultan, S., Chen, L., Khan, M., and Coyle, D. (2012). Indirect evidence: indirect treatment comparisons in meta-analysis. ottawa: Canadian agency for drugs and technologies in health; 2009.