Network Meta-Analysis Using Bayesian Methods and Some Diagnostics

by

MOHAMMED MUJAAB KAMSO

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg

Abstract

Network meta-analysis (NMA), also known as mixed treatment or multiple treatment comparisons, is commonly used to incorporate direct and indirect evidences comparing treatments. This is an extension to meta-analysis which seeks to estimate the combined estimate of treatment comparisons from multiple studies making use of just direct evidence from the treatment comparison.

With recent advances in methods and software, Bayesian approaches to NMA have become quite popular and allow models of previously unanticipated complexity. However, when direct and indirect evidence differ in an NMA, the model is said to suffer from inconsistency which is a critical assumption. If it is violated, interpretation and conclusion of results will be affected. Current inconsistency detection in NMA is usually based on contrast-based (CB) models; however, this approach has certain limitations. In this work, we look at an arm-based random effects model, where we detect discrepancy of direct and indirect evidence for comparing two treatments using the fixed effects in the model while flagging extreme trials using the random effects. We define discrepancy factors to characterize evidence of inconsistency for particular treatment comparisons, which is novel in NMA research. Our approaches permit users to address issues previously tackled via CB models. We compare sources of inconsistency identified by our approach and existing loop-based CB methods using real

and simulated datasets and demonstrate that our methods can offer powerful inconsistency detection.

After the detection of inconsistency, we try to perform some diagnostics to network meta-analysis to see if the trials that are causing the inconsistencies are just outliers or influential. Specifically, we address the question: if these trials are removed, will they affect the conclusion of results?

Acknowledgment Page

I am most grateful to my Heavenly Father for making a way for me to come to Canada for my higher studies and for his provision. I thank my thesis supervisors Dr. Saman Muthukumarana and Dr. Saumen Mandal for providing me with the funding for my graduate studies. Thanks to my supervisors again for their guidance on my thesis. I am thankful to my committee members Dr. Po Yang and Dr. Darren Gillis for their time, comments and corrections.

Dedication Page

I dedicate this page to my Father, Mother and family for their constant prayers and support throughout my journey to academia. All I can say is a very big thank you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Meta-Analysis

Meta-analysis is a statistical technique for combining the results of independent but similar studies to obtain an overall estimate of treatment effect size (Borenstein et al., 2015). This treatment effect size can be measured on some scale, for example, odd-ratios, log-odds ratio and mean effects. The motivation for meta-analysis lies in the fact that, since we are interested in the combination of independent studies (Sutton et al., 2000), treatment effects or differences are estimated precisely. So there is a need to assign weights to the studies to see how precise a particular study is. For example, if a particular study has more information, it should be assigned more weight (Borenstein et al., 2015). Meta-analysis is also known as the extension of the traditional pairwise comparison (Whitehead, 2002). It has proven to have an enormous validity of its use. But with all its benefits, it also has some demerits which every researcher

should look at when using this statistical method (Shuttleworth, 2009). The benefits that one gets from the use of this statistical method are

1. It allows data to be collected from a field which will be impossible for a single research group, for example a rare medical condition.

2. A meta-analysis study can reduce the need for long, expensive and potentially intrusive repeated research studies.

3. It throws more light to studies that may have some level of correlation and relationship between them that may not be obvious to see initially (Whitehead, 2002).

Also with all these benefits alluding to the use of meta-analysis, it is prudent to state its demerits. Such demerits are

1. One of the most important disadvantage is that there is the tendency for publication bias which must be accessed thoroughly.

2. It is a very sensitive statistical method. Erroneous or poorly conducted studies leads to invalid results and conclusions. Hence setting criteria for inclusion of a study may lead to ending up with few studies to be analyzed, this results in poor conclusions.

3. The researcher must ensure all the studies are quantitative rather than qualitative for valid comparisons (Whitehead, 2002).

In meta-analysis, we wish to estimate the combined effects. Hence it begs the question how best we can estimate the "combined effects". It turns out that there are two models that suggest how to estimate the "combined effects". These are fixed effects model and random effects model. Under the fixed effects model, we assume there is a common effect that is shared among all the studies. Hence a study with more information is assigned more weight (Borenstein et al., 2015). The combined effects is an estimate of the common effect size. In fixed effects model, the only source of error in our estimate of the combined effects is the random error within studies (Borenstein et al., 2010). Therefore, with a large enough sample size the error will tend toward zero. This holds true whether the large sample size is confined to one study or distributed across many studies. Hence they have a common effect size. The model is given below

$$y_i = \mu + \epsilon_i, \tag{1.1}$$

where $\epsilon_i \sim N(0, \sigma_i^2)$ and $y_i \sim N(\mu, \sigma_i^2)$. The model (1.1) is the fixed effects model. Steps are shown below on how the combined effects are computed under the fixed effects model. The effect size is shared and only one source of error is expected. The process of computing the combined effects denoted by $\bar{Y}_.$, is as follows:

1. Identify the variance of each effect within each study, denoted as σ_i^2 and the effect size of each study, denoted as y_i .

2. Find the weight for each study, W_i, where $W_i = \frac{1}{\sigma_i^2}$. Thus, a weight is

the reciprocal of the variance.

3. Now, assuming that we have n studies, the combined effect is computed as

$$\bar{Y}_{\cdot} = \frac{\sum_{i=1}^{n} W_i y_i}{\sum_{i=1}^{n} W_i}.$$

4. Compute the weighted variance,

$$V_{\cdot} = \frac{1}{\sum_{i=1}^{n} W_i}.$$

5. Finally compute the standard error, $S.E. = \sqrt{V_{\cdot}}$.

With the random effects model, here it is assumed that the true effect could vary from study to study (Borenstein et al., 2015). The studies included in the meta-analysis are assumed to be a random sample of the relevant distribution of effects, and the combined effect estimates, be the mean effect in this distribution. In random effects model, there are two levels of sampling and two levels of error. First, each study is used to estimate the true effect in a specific population. Second, all of the true effects are used to estimate the mean of the true effects (Lumley, 2002). Therefore, our ability to estimate the combined effect precisely will depend on both the number of subjects within studies (which addresses the first source of error) and also the total number of studies (which addresses the second) (Borenstein et al., 2015). The mathematical model representation is given below

$$y_i = \mu + \theta_i + \epsilon_i \tag{1.2}$$

where $\theta_i \sim N(0, \tau^2)$, $\epsilon_i \sim N(0, \sigma_i^2)$ and $y_i \sim N(\mu, \sigma_i^2 + \tau^2)$. The model (1.2) is the random effects model. In this model, between study variation exist, which is represented as the τ^2 and the within study variation denoted as σ_i^2. These two form the total variation in the combined study analysis. In the fixed effects model, the between study variation is zero, thus $\tau^2 = 0$. But in the random effects model we need to estimate this between study variation. A Generalized Method of Moments is used to estimate this τ^2 (Borenstein et al., 2010), this method is shown below. We let M which represents the total variation, given by

$$M = \sum_{i=1}^{n} a_i (y_i - \bar{y})^2,$$

where a_i are some constants and $\bar{y} = \frac{\sum_{i=1}^{n} a_i y_i}{\sum_{i=1}^{n} a_i}$. Now we find the expected value of M, that is $E(M)$. Then

$$\begin{aligned}
E(M) &= E\left(\sum_{i=1}^{n} a_i (y_i - \overline{y})^2\right) \\
&= E\left(\sum_{i=1}^{n} a_i y_i^2 - 2\overline{y}\sum_{i=1}^{n} a_i y_i + \overline{y}^2 \sum_{i=1}^{n} a_i\right) \\
&= E\left(\sum_{i=1}^{n} a_i y_i^2 - 2\overline{y}^2\sum_{i=1}^{n} a_i + \overline{y}^2 \sum_{i=1}^{n} a_i\right) \\
&= E\left(\sum_{i=1}^{n} a_i y_i^2 - \overline{y}^2 \sum_{i=1}^{n} a_i\right)
\end{aligned}$$

$$= E\left(\sum_{i=1}^{n} a_i y_i^2 - \frac{(\sum_{i=1}^{n} a_i y_i)^2}{\sum_{i=1}^{n} a_i}\right).$$

We know $E(y_i) = \mu$ and $E(y_i^2) = Var(y_i) + [E(y_i)]^2 = \sigma_i^2 + \tau^2 + \mu^2$. Hence

$$E(M) = \sum_{i=1}^{n} a_i E(y_i^2) - \frac{E[(\sum_{i=1}^{n} a_i y_i)^2]}{\sum_{i=1}^{n} a_i}$$

$$= \sum_{i=1}^{n} a_i(\sigma_i^2 + \tau^2 + \mu^2) - \frac{E[(\sum_{i=1}^{n} a_i y_i)^2]}{\sum_{i=1}^{n} a_i}.$$

$E[(\sum_{i=1}^{n} a_i y_i)^2] = Var(\sum_{i=1}^{n} a_i y_i) + (E[(\sum_{i=1}^{n} a_i y_i)])^2 = \sum_{i=1}^{n} a_i^2(\sigma_i^2 + \tau^2) + \mu^2(\sum_{i=1}^{n} a_i)^2,$

$$= \sum_{i=1}^{n} a_i(\sigma_i^2 + \tau^2) + \mu^2 \sum_{i=1}^{n} a_i - \frac{\sum_{i=1}^{n} a_i^2(\sigma_i^2 + \tau^2)}{\sum_{i=1}^{n} a_i} - \mu^2 \sum_{i=1}^{n} a_i$$

$$= \left(\sum_{i=1}^{n} a_i \sigma_i^2 - \frac{\sum_{i=1}^{n} a_i^2 \sigma_i^2}{\sum_{i=1}^{n} a_i}\right) + \left(\sum_{i=1}^{n} a_i - \frac{\sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} a_i}\right) \tau^2$$

Thus, $$\tau^2 = \frac{E(M) - (\sum_{i=1}^{n} a_i \sigma_i^2 - \frac{\sum_{i=1}^{n} a_i^2 \sigma_i^2}{\sum_{i=1}^{n} a_i})}{\sum_{i=1}^{n} a_i - \frac{\sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} a_i}}.$$

Now the above generalized method of moments reduces to different forms of estimating τ^2 when the constants $a_i's$ are changed accordingly. Let $a_i = \frac{1}{s_i^2}$ and the population variances, σ^2 are replaced by the sample variances, s_i^2. Then, the method of estimating τ^2 is called the *DerSimonian-Laird* (DerSimonian and Laird, 1986) method. Thus the estmte of τ^2 is

$$\widehat{\tau}^2 = \frac{Q - (n-1)}{\sum_{i=1}^{n} w_i - \frac{\sum_{i=1}^{n} w_i^2}{\sum_{i=1}^{n} w_i}}, \tag{1.3}$$

where $Q = \sum_{i=1}^{n} w_i(y_i - \overline{y})^2$ and $w_i = \frac{1}{s_i^2}$. Since the effect size are not shared and two sources of errors are expected, the process of computing the combined effect, denoted as $\bar{Y}_.$, is as follows:

1. Identify the variance of effect within and between each study, denoted as s_i^2 and $\widehat{\tau}^2$ and the effect size of each study, y_i .

2. Find the weight at each study, W_i^*, where $W_i^* = \frac{1}{s_i^2 + \widehat{\tau}^2}$.

3. Now the combined effect is computed as

$$\bar{Y}_.^* = \frac{\sum_{i=1}^{n} W_i^* y_i}{\sum_{i=1}^{n} W_i^*}.$$

4. Compute the weighted variance

$$Var(\bar{Y}_.^*) = \frac{1}{\sum_{i=1}^{n} W_i^*}.$$

5. Finally compute the standard error, $S.E. = \sqrt{Var(\bar{Y}_.^*)}$.

1.2 Heterogeneity in Meta-Analysis.

Heterogeneity is the differences in studies that are not due to chance. The presence of heterogeneity helps in the selection of the type of model to use, either fixed effects model or random effects model (Sutton et al., 2000). When the presence of heterogeneity is significant, one should settle for the random

effects model. If there is no heterogeneity, one should just settle for the fixed effects model. This strategy is quite dicey because it is sensitive to the number of studies to consider. If one considers a small number of studies, the test for the presence of heterogeneity will not be effective. In this situation, we need large number of studies for analysis (Schwarzer, 2015). There are two types of heterogeneity present in such analysis. These are

1. Clinical heterogeneity: This is always present. This is as a result of the design from one study to the next, the study setting and how the interventions or drugs are administered.

2. Statistical heterogeneity: This may not be always present. It is the variation between the results of the studies that leads to differences in the results and unlike clinical heterogeneity.

Hence, in studies like this, one has to test for its presence, quantify it and investigate. Therefore we can use the Chi-square (χ^2) test for testing the presence of heterogeneity. This is usually accurate when the number of studies is large. Let k be the number of studies. We take the degree of freedom to be the number of studies minus 1, thus $k-1$, under the null hypothesis (Sutton et al., 2000). Where the null hypothesis has the statement of homogeneity among studies and the alternative hypothesis suggesting there is the presence of heterogeneity, this will have a corresponding p-value quantifying evidence for or against the alternative hypothesis. This might not be enough and we

need to quantify this with some indices. Thus, for testing heterogeneity, we try to quantify the presence of heterogeneity in our analysis. Some of these indices are Higgins' (I^2) and H-index, H^2 (Schwarzer, 2015), which are used to quantify heterogeneity and are discussed as follows. Higgins (I^2) measures the percentage of variation across studies that is due to heterogeneity and not due to chance. The representation of Higgin's index is given by

$$I^2 = \frac{(Q - df)}{Q} \times 100.$$

where the $df = k - 1$ and Q is the overall heterogeneity in the study. Also the threshold in taking a decision on the presence of heterogeniety using I^2 is shown below

1. 0% to 25% indicates that a low amount of heterogeniety is present.

2. 25% to 50% indicates that a moderate amount of heterogeniety is present.

3. 50% to 100% shows a considerable high amount of heterogeniety is present (Borenstein et al., 2010).

H-Index (H^2) is the ratio of the overall heterogeneity to the degree of freedom, where the degree of freedom is the number of studies minus one. Thus, the index is given by

$$H^2 = \frac{Q}{df}.$$

Higgins and Thompson (2002) used it to develope a confidence interval for I^2. The interval is formulated by calculating the H^2 index of their proposed measures of heterogeneity (Higgins and Thompson, 2002), which is known as Birge's ratio (Birge, 1932). They also derieved the I^2 in terms of the H^2 index

$$I^2 = \frac{H^2 - 1}{H^2} \times 100.$$

This allows us to express inferences of H^2 in terms of I^2. For practical application, Higgins and Thompson (2002) recommends a confidence interval for the natural logarithm of H, that is $ln(H)$, assuming a standard normal distribution. The interval is given by

$$exp\{ln(H) \pm Z_{\alpha/2}\, SE(ln(H))\},$$

where $Z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution, $SE(ln(H))$ is the standard error of the $ln(H)$, estimated by

$$SE(ln(H)) = \begin{cases} \frac{1}{2}\frac{ln(Q)-ln(k-1)}{\sqrt{2Q}-\sqrt{2k-3}} & \text{if } Q > k, \\ \sqrt{\frac{1}{2(k-2)}[1 - \frac{1}{3(k-2)^2}]} & \text{if } Q \leq k. \end{cases}$$

1.3 Network Meta-Analysis

Usually after the collection of studies to estimate the combined effect, the information obtained can have a direct comparison, indirect comparison and mixed treatment comparison. A combination of the direct and indirect comparisons

results in a mixed treatments comparison or multiple treatments comparison meta-analysis. Hence it is called network meta-analysis (Lumley, 2002). Network meta-analysis expands the scope of a conventional pair-wise meta-analysis by analyzing simultaneously both direct comparisons of interventions within randomized controlled trials (RCTs) and indirect comparisons across trials based on a common comparator (e.g., placebo or some standard treatment). In the simplest case, one may be interested in comparing two interventions, say A and C. Indirect evidence can be obtained from RCTs of either A or C versus a common comparator, say B, keeping intact the randomized comparisons within the RCTs. When both direct and indirect evidences are available, the two sources of information can be combined as a weighted average when appropriate. Data structure of this type can be extended to k-comparisons to facilitate simultaneous inference for all available treatments, and to provide evidence for selecting the best of several treatment options (Bucher et al., 1997).

Many assumptions behind network meta-analysis methods appear to be similar to those made in standard pair-wise meta-analysis (Jeroen et al., 2011). However, for a conventional pair-wise meta-analysis, the methodology for network meta-analysis must be carefully developed and rigorously evaluated before the technique is applied widely. Direct treatment comparison is an estimation of the relative treatment effect or the other relative characteristic of one technology compared to another informed only by head-to-head comparison (Stegenga et al., 2008). Also, indirect treatment comparison is an analysis for comparing interventions that have not been compared directly within a head-to-head ran-

domized trial (Stegenga et al., 2008). Finally the mixed treatment comparison is also known as network meta-analysis. It is an analysis that compares two or more interventions using a combination of direct evidence (from head-to-head trials of the intervention of interest) and indirect evidence (trials that do not compare the interventions of interest directly in head-to-head trials) (Stegenga et al., 2008).

Now we take a look at an example. Assume that there are four medical interventions, drug A(control), B, C and D. There is direct evidence for drug A and C, direct evidence for drug A and B and direct evidence for drug A and D. Direct estimates of the differences of the treatment effects of A and C, A and B and A and D are denoted by $\hat{\theta}_{AC}^{direct}$, $\hat{\theta}_{AB}^{direct}$ and $\hat{\theta}_{AD}^{direct}$ respectively. These evidences are measured on some scales such as odds ratio, log odds ratio or mean difference effect. Here we realize there is no direct evidence for drug B and C as well as drug C and D. However, it will suffice if we use the existing evidences to estimate the evidence we do not have at this time since there is a common comparator, A. This method of estimating this new evidence is called the indirect comparison (Bucher et al., 1997). Thus, indirect evidence for comparing B and C, and C and D are estimated as

$$\hat{\theta}_{BC}^{Indirect} = \hat{\theta}_{AC}^{direct} - \hat{\theta}_{AB}^{direct}$$

$$\hat{\theta}_{CD}^{Indirect} = \hat{\theta}_{AD}^{direct} - \hat{\theta}_{AC}^{direct}.$$

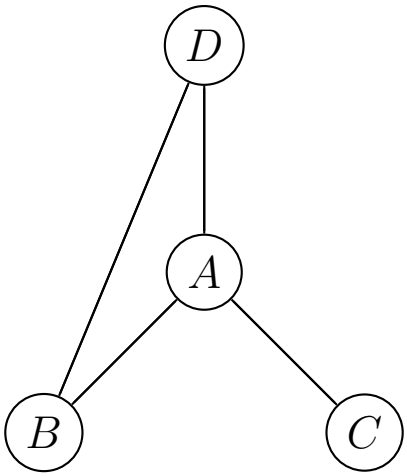

Figure 1.1: The comparison of treatment A, B, C and D in a network.

Hence its corresponding variances are

$$Var(\hat{\theta}_{BC}^{Indirect}) = Var(\hat{\theta}_{AC}^{direct}) + Var(\hat{\theta}_{AB}^{direct})$$

$$Var(\hat{\theta}_{CD}^{Indirect}) = Var(\hat{\theta}_{AD}^{direct}) + Var(\hat{\theta}_{AC}^{direct}).$$

The variances of the indirect estimates have no co-variances because we assume that the direct comparison studies are independent. This combination of evidences, through direct and indirect methods brings the rise of *network meta-analysis*. This is illustrated in Figure 1.1.

1.3.1 Assumptions for Network Meta-Analysis

For every statistical model to be valid so as to apply some statistical techniques, we need some assumptions to hold. For network meta-analysis there are three main assumptions that must be met. These are consistency, homogeneity and

similarity (exchangeability) (Kiefer et al., 2015).

For the similarity assumption, all trials included in the analysis should be comparable in terms of potential effect modifiers like patient characteristics and geographical area of the patients. The similarity of the individual trials should be examined on the basis of their essential characteristics. However, in this case, this must be done for all the investigated interventions. There is a well-known approach called the PICOS approach, population (P), intervention (I), comparator (C), outcome (O), and study design (S) (Kiefer et al., 2015). Important information can be obtained from comparisons of trials regarding relevant patient characteristics as well as comparisons of trial arms representing reasonable reference interventions regarding relevant endpoints.

For the homogeneity assumption, there should be no relevant heterogeneity between trials of the pairwise comparisons. Homogeneity should be examined using standard procedures such as forest plots and measures of heterogeneity like the I^2 and the Cochran Q statistic. Depending on the size of the network, however, this may be a very lengthy process, as all possible combinations of two interventions must be included (Jeroen et al., 2011).

For the consistency assumption, there should be some agreement of the indirect estimates and the direct estimates. Thus there should be no discrepancies in their comparison (Ades, 2003). It can usually be examined if the direct and indirect comparisons are available. This gives rise to a consistency equation, $d_{st}^{indirect} = d_{bt}^{direct} - d_{bs}^{direct}$. Now assuming we have the direct comparison of

d_{st}^{direct}, then we can assess the level of inconsistency $ICF = d_{st}^{direct} - d_{st}^{indirect}$. Where ICF is the inconsistency factor (Jeroen et al., 2011).

1.3.2 Evidence of a Network

Usually before applying any statistical techniques, it is important for the analyst to visualize how the network is. There are two key concepts one has to have in mind when visualizing the network. These two concepts are Geometry and Asymmetry (Greco et al., 2013). Geometry of a network shows the shape of the network and how the nodes or vertices are connected to one another. Thus it is the overall structure of the interventions or treatments comparisons. Asymmetry of a network derives from the amount of data for a particular comparison of interventions or treatments. This helps us to know the weights of each comparison in a network.

Visualization of the network can also enable us know how to call the available evidence comparisons. Thus some call network meta-analysis as indirect treatment comparisons and also mixed treatment comparisons. A network may have one or more loops in the network, where a number of nodes connected together form a loop. We called this a mixed treatment comparison. When the network does not have loops in the network, we then call it an indirect treatment comparison (Jeroen et al., 2011).

A star-shape can be thought of using an indirect treatment comparison analysis. For Figure 1.2 we see a clear Star-Shaped network. Here it has one node (A) which has been connected to all the other nodes. Hence it shows a direct

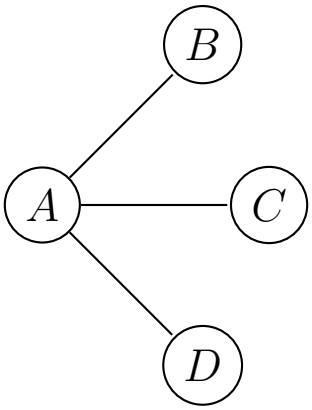

Figure 1.2: The Star-Shaped network of treatments A, B, C and D.

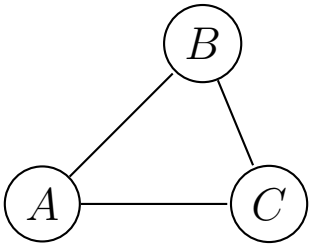

Figure 1.3: The Triangular-Shaped network of treatments A, B, C and D.

evidence from AB, AC and AD. Now BC, BD and CD do not have direct evidence. Hence they can be estimated indirectly from the above direct evidences. A triangular-shaped network from Figure 1.3 which is clearly a closed-loop, and can be thought of by using a mixed treatment comparison analysis. Here we see that there are direct evidences for all the nodes in this network. Thus we have direct evidence for AB, AC and BC. Also we can estimate its indirect evidence from the direct evidence as well. Now the combination of these available evidences makes the mixed treatment comparison analysis possible and a good approach since it makes use of more information.

1.3.3 Matrix Representation of Network Meta-Analysis

Most network graphs can be represented in terms of matrices to describe its pattern and structure. There are two types of matrix representation of network graphs that are important. These are adjacency matrices and incidence matrices. An adjacency matrix shows the existence of an edge between two immediate nodes. In the adjacency matrix, both the rows and columns represent the vertices of the graph (Molitierno, 2012). Let M be a network graph on n nodes, and we have nodes running from $1, ..., n$ (Molitierno, 2012). Then the adjacency matrix of M on n vertices is the $n \times n$ matrix $\boldsymbol{A} = [a_{ij}]$, where

$$a_{i,j} = \begin{cases} 1 & \text{if } i \neq j \text{ and } i \text{ and } j \text{ are adjacent,} \\ 0 & \text{if } i = j \text{ and } i \text{ and } j \text{ are not adjacent.} \end{cases}$$

Hence we take an example of a network graph and try to find out how to formulate the adjacency matrix. Let the graph 1.4 be the network graph ω and the adjacency matrix be as follows

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

Also the incidence matrix is a matrix which is based on which two vertices, say v_a and v_b share an edge (Molitierno, 2012). It is conventional to have the vertex with lower-number labelled as v_a and the high-number labelled as v_b. For each

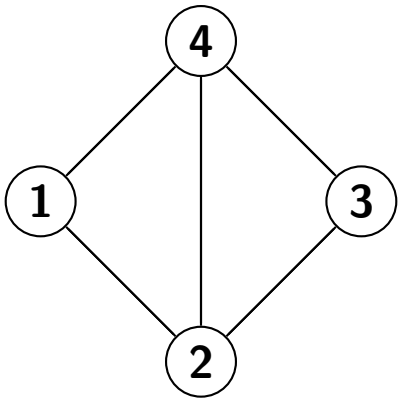

Figure 1.4: Network plot, ω for the adjacency matrix of treatments 1, 2, 3 and 4.

vertex-edge pair (v, e) in which the vertex v is incident to the edge e, we define a function $p(v, e)$ where $p(v_a, e) = 1$ and $p(v_b, e) = -1$. Mathematically, we can define an incidence matrix to be as follows. Let ω be a network graph on n nodes, thus we have nodes running from $v_1, ..., v_n$ and m edges labeled as $e_1, ..., e_n$. Then the incidence matrix on n vertices and m edges is the $n \times m$ matrix $\boldsymbol{G} = [g_{ij}]$, where

$$g_{i,j} = \begin{cases} p(v_i, e_j) & \text{if } v_i \text{ and } e_j \text{ are incident,} \\ 0 & \text{if } v_i \text{ and } e_j \text{ are not incident.} \end{cases}$$

Also, let us take a look at an example of how to get the incidence matrix from the network graph. Let the Figure 1.5 be the network graph ω. Then the incidence matrix is as follows

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \end{bmatrix}.$$

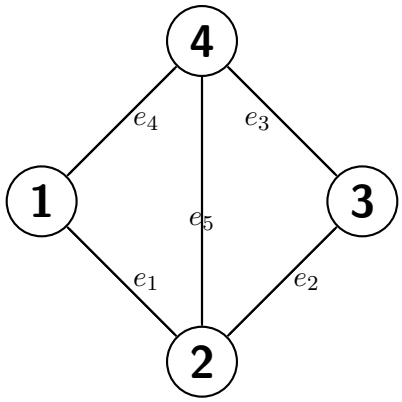

Figure 1.5: Network plot for the incidence matrix of treatments 1, 2, 3 and 4.

The Laplacian matrix is a matrix in graph theory, which is used to describe the pattern in a network (Molitierno, 2012). Mathematically, let ω be a graph on n vertices labeled $1, ..., n$ with d_{ij} to be the number of out/in degree directions coming out of a particular node of an undirected network graph. The Laplacian matrix of ω is the $n \times n$ matrix $\boldsymbol{L} = [l_{i,j}]$, where

$$l_{i,j} = \begin{cases} -1 & \text{if } i \neq j \text{ and } i \text{ and } j \text{ are adjacent,} \\ 0 & \text{if } i \neq j \text{ and } i \text{ is not adjacent to } j, \\ d_{i,j} & \text{if } i = j. \end{cases}$$

Hence from the graph ω above Figure 1.5, the Laplacian matrix can be derived as

$$\boldsymbol{L} = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}.$$

We shall see that the Laplacian matrix is related to the other matrices

discussed earlier.

1. $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{D}$ is the diagonal matrix consisting of the degrees of the vertices of a graph ω and $\boldsymbol{A}$ is the adjacency matrix for ω (Molitierno, 2012).

2. $\boldsymbol{L} = \boldsymbol{G}\boldsymbol{G}^{\boldsymbol{T}}$, where $\boldsymbol{G}$ is the incidence matrix of a graph ω (Molitierno, 2012).

Matrix Representation of a Weighted Graph Network

For a weighted graph network, adjacency matrix, incidence matrix and laplacian matrix can be generalized (Molitierno, 2012). The procedure to deal with the weighted graph of a network is shown below.

- For adjacency and laplacian matrices, we replace the off-diagonal entries with w and $-w$, respectively, where w denotes the weight of the corresponding edge.

- Also for the Laplacian matrix of a weighted graph, each diagonal entry is the sum of the weights of the edges incident to the corresponding vertex.

- For the incidence matrix of a weighted graph, we change the definition of $p(v_i, e_j)$ by defining $p(v_a, e_j) = \sqrt{w}$ and $p(v_b, e_j) = -\sqrt{w}$, where w is the weight at each edge.

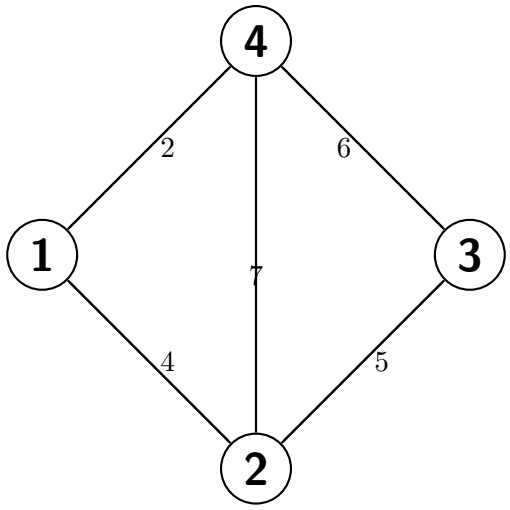

Figure 1.6: Network plot for a weighted graph of treatments 1, 2, 3 and 4.

Let the Figure 1.6 be a weighed network graph denoted as, ω. Then below is the matrix representaton of the above matrices.

- The adjacency matrix of the the graph, ω, will be

$$\boldsymbol{A} = \begin{bmatrix} 0 & 4 & 0 & 2 \\ 4 & 0 & 5 & 7 \\ 0 & 5 & 0 & 6 \\ 2 & 7 & 6 & 0 \end{bmatrix}.$$

- The incidence matrix of the the graph, ω, will be

$$\boldsymbol{G} = \begin{bmatrix} \sqrt{4} & 0 & 0 & \sqrt{2} & 0 \\ -\sqrt{4} & \sqrt{5} & 0 & 0 & \sqrt{7} \\ 0 & -\sqrt{5} & \sqrt{6} & 0 & 0 \\ 0 & 0 & -\sqrt{6} & -\sqrt{2} & -\sqrt{7} \end{bmatrix}.$$

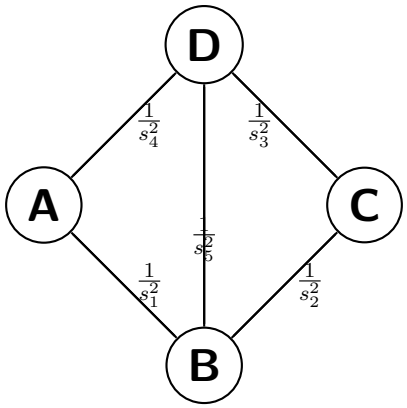

Figure 1.7: Network plot with weights of treatments A, B, C and D.

- The laplacian matrix of the the graph, ω, will be

$$\boldsymbol{L} = \begin{bmatrix} 6 & -4 & 0 & -2 \\ -4 & 16 & -5 & -7 \\ 0 & -5 & 11 & -6 \\ -2 & -7 & -6 & 15 \end{bmatrix}.$$

The weighted Laplacian matrix can also be represented as given in equation 1.5 where the W is a diagonal matrix where diagonal entries are the respective weights of each study, $(1/s_1^2, ..., 1/s_m^2)$, where s_i^2 are the sample variances. Let the Figure 1.7 be the network graph ω. Then we have the model

$$\begin{bmatrix} \hat{\theta}_{AB} \\ \hat{\theta}_{BC} \\ \hat{\theta}_{CD} \\ \hat{\theta}_{AD} \\ \hat{\theta}_{BD} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \theta_A \\ \theta_B \\ \theta_C \\ \theta_D \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}.$$

The design matrix $\boldsymbol{X}$ is not of full rank, thereby $\boldsymbol{X}^\top \boldsymbol{X}$ is not of full rank

too. Hence we use the Moore-Penrose pseudo-inverse. We first derive the weighted Laplcian matrix, $\boldsymbol{L} = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$, as given by

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{s_1^2} & & & & \\ & \frac{1}{s_2^2} & & 0 & \\ & & \frac{1}{s_3^2} & & \\ & 0 & & \frac{1}{s_4^2} & \\ & & & & \frac{1}{s_5^2} \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

$$\boldsymbol{L} = \begin{bmatrix} \frac{1}{s_1^2} + \frac{1}{s_4^2} & -\frac{1}{s_1^2} & 0 & -\frac{1}{s_4^2} \\ -\frac{1}{s_1^2} & \frac{1}{s_1^2} + \frac{1}{s_2^2} + \frac{1}{s_5^2} & -\frac{1}{s_2^2} & -\frac{1}{s_5^2} \\ 0 & -\frac{1}{s_2^2} & \frac{1}{s_2^2} + \frac{1}{s_3^2} & -\frac{1}{s_3^2} \\ -\frac{1}{s_4^2} & -\frac{1}{s_5^2} & -\frac{1}{s_3^2} & \frac{1}{s_3^2} + \frac{1}{s_4^2} + \frac{1}{s_5^2} \end{bmatrix}.$$

We see that this matrix notation of Laplacian matrix, $\boldsymbol{L} = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$ can be equally derived from the various approaches of the weighted graph.

1.3.4 Model for Network Meta-Analysis

Let $\hat{\boldsymbol{\theta}}$ be a vector of $k \times 1$, where k is the number of studies if all the pairwise comparisons are from a two arm trial, else it should be a vector of $m \times 1$, where m is the number of pairwise comparisons of the treatments. Also, let $\boldsymbol{\theta^{treat}}$ be a vector of $n \times 1$, where n is the number of treatments or interventions under consideration. Let $\boldsymbol{X}$ be a matrix which is of dimension $k \times n$ or $m \times n$ depending whether the pairwise comparisons are all from a two-arm trial or

from a multi-arm trial. The matrix $\boldsymbol{X}$ is called the design matrix (Schwarzer, 2015). The model for a network meta-analysis is then given by

$$\hat{\boldsymbol{\theta}} = \boldsymbol{X}\boldsymbol{\theta}^{\boldsymbol{treat}} + \boldsymbol{\epsilon}, \tag{1.4}$$

where $\boldsymbol{\epsilon} \sim N(0, \sum)$, $\sum$ is the variance-covariance matrix. So, for the above example of Figure 1.1, our model will be

$$\begin{bmatrix} \hat{\theta}_{AB} \\ \hat{\theta}_{AC} \\ \hat{\theta}_{AD} \\ \hat{\theta}_{BD} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \theta_A \\ \theta_B \\ \theta_C \\ \theta_D \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}.$$

As we mentioned earlier, the design matrix $\boldsymbol{X}$ is usually not of full rank. Hence its $\boldsymbol{X^T X}$ is not invertible. Hence we use the Moore-Penrose pseudo inverse to estimate the treatment effects. Here we assume an $n \times n$ Laplacian-matrix $\boldsymbol{L}$. This is denoted as

$$\boldsymbol{L} = \boldsymbol{X^T W X}, \tag{1.5}$$

where $\boldsymbol{W}$ is an $m \times m$ diagonal matrix where the diagonal entries are the inverses of the variances of the respective studies, that is, $(1/s_1^2, ..., 1/s_m^2)$. The Laplacian-matrix $\boldsymbol{L}$ is of rank $n-1$, so it is not of full rank. Therefore we need the Moore-Penrose pseudo-inverse as given by

$$\boldsymbol{L}^+ = (\boldsymbol{L} - \boldsymbol{J}/n)^{-1} + \boldsymbol{J}/n, \tag{1.6}$$

where $\boldsymbol{J}$ is an $n \times n$ matrix with all entries being 1's.

1.3.5 Estimation of the Treatment Effects

Let $\hat{\boldsymbol{\theta}}^{\boldsymbol{nma}}$ be the network estimates (estimates of the fitted values) of order $n \times 1$ vector. Here we use the frequentist method of estimation

$$\begin{aligned} \hat{\boldsymbol{\theta}}^{\boldsymbol{nma}} &= \boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W}\hat{\boldsymbol{\theta}} \\ &= \boldsymbol{H}\hat{\boldsymbol{\theta}}. \end{aligned} \tag{1.7}$$

Now $\boldsymbol{H}$ can be seen as the Hat matrix with dimension $m \times m$. We realize that elements of $\hat{\boldsymbol{\theta}}^{\boldsymbol{nma}}$ are linear combinations of the elements of $\hat{\boldsymbol{\theta}}$. Hence the rows of the matrix $\boldsymbol{H}$ are seen as the generalized weights (Schwarzer, 2015). Therefore, the variance- covariances matrix of $\hat{\boldsymbol{\theta}}^{\boldsymbol{nma}}$ is given by

$$var(\hat{\boldsymbol{\theta}}^{\boldsymbol{nma}}) = \boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}. \tag{1.8}$$

As we know that $\hat{\boldsymbol{\theta}}^{\boldsymbol{nma}} = \boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W}\hat{\boldsymbol{\theta}}$, $var(\hat{\boldsymbol{\theta}}^{\boldsymbol{nma}})$ can be derieved as

$$\begin{aligned} var(\hat{\boldsymbol{\theta}}^{\boldsymbol{nma}}) &= var(\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W}\hat{\boldsymbol{\theta}}) \\ &= (\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W})var(\hat{\boldsymbol{\theta}})(\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W})^{\top} \\ &= (\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W})\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W})^{\top} \\ &= (\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W})\boldsymbol{\Sigma}\boldsymbol{W}(\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}) \\ &= (\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W})\boldsymbol{I}(\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}) \\ &= \boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}\boldsymbol{W}\boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top} \\ &= \boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{L}\boldsymbol{L}^{+}\boldsymbol{X}^{\top} \quad \text{(since Moore-Penrose inverse is } L^{+}LL^{+} = L^{+}) \\ &= \boldsymbol{X}\boldsymbol{L}^{+}\boldsymbol{X}^{\top}. \end{aligned}$$

1.4 Organization of the Thesis

The motivation for this thesis is based on the fact that for a given disease, there is likely to be many other substitute drugs or new drugs that can be used to treat the patients. Also, there might be a comparison of treatments (head to head) and comparison of treatments that have never been compared before (not head to head). Pooling these sources of evidences together makes a better estimate of the available treatments in the study. But these drugs may not all be at the same cost, some may possibly have adverse side effects and the method of application could be complex for others. On grounds of these information, we do equivalence testing to see if two different drugs can be regarded as equivalent in terms of their treatment effect as well as treatments that have never been compared before to see if they are equivalent. If there are some significant differences, we try to rank the treatments to see which treatment is most effective. Network meta-analysis using Bayesian methods seeks to answer these questions. The remaining section of this thesis is organized as follows. In Chapter 2, some statistical models in network meta-analysis are discussed in detail.

Chapter 3 presents some diagnostics to network meta-analysis to test that its major assumptions for its use are met. In Chapter 4, we perform some data analysis using the frequentist approach on a dataset comprising of diabetes treatments. We then look at the anaesthetic drug comparison using the Bayesian approach to network meta-analysis. Chapter 5 presents a discussion of the

results and some concluding remarks. As future work, we will be interested in exploring network meta-analysis using Dirichlet process. This is a meta-analysis in which multiple treatments are compared in multivariate analysis thus using a distribution over probability distributions. We assume that there is a probability distribution over a measurable space. Then a Dirichlet process is a probability distribution over all the distributions of the subsets of the measurable space (Sethuraman, 1994).

Chapter 2

Statistical Models

2.1 Types of Data for Analysis

In the network meta-analysis, the data we use for analysis are from published results. Hence we use synthesized input data in our analysis instead of the individual data from individual trials. The synthesized data can be classified into two formats, normally the Arm-Level summaries and the Contrast-Level summaries. In arm-level summaries, the effect measures are reported for each arm, and they are in the form of mean effect, odds effects or absolute risk. In contrast-level summaries, the effect measures are reported as the difference in effects between arms, as such they are in the form of mean difference, odds-ratio, risk-ratio or hazard ratio (Spielgelhalter and Myles, 2004).

There is one advantage of arm-level summaries over the contrast-level summaries. With the arm-level summaries, it is possible to obtain an exact likelihood for the data like binomial for binary data instead of the normal approximation

for the contrast-level summary data. Here we try to do some analysis in the Bayesian setting. We look at the arm-level summaries analysis.

2.1.1 Arm-Level Summaries

As mentioned earlier, the arm-level summaries usually provide exact likelihood. In the setting of binary data, it provides a binomial likelihood. Assume there are N randomized controlled trials which have a mixed comparison with K treatments. The number of events on treatment k in trial i is denoted by r_{ik}, and the total number of observations on treatment k in trial i is denoted by n_{ik}. The probability of the event of occurrence on treatment k in trial i is denoted by p_{ik}. Given this fixed probability and total number of observations on treatment k in trial i, r_{ik} follows a binomial distribution

$$r_{ik} \sim Bin(p_{ik}, n_{ik}), i = 1, 2, 3, ...N; k = 1, 2, ...K. \tag{2.1}$$

With this binary data available, we can have the probability, p_{ik}, b as the baseline treatment, μ_i as the log-odds of the baseline treatment and $\delta_{i,bk}$ as the log-odds of treatment k compared to the baseline treatment in study i, modeled on a *Logit* scale as

$$logit(p_{ib}) = log\left(\frac{p_{ib}}{1-p_{ib}}\right) = \mu_i, i = 1, 2, 3, ...N; k = b = 1, 2, ...K.$$

$$logit(p_{ik}) = log\left(\frac{p_{ik}}{1-p_{ik}}\right) = \mu_i + \delta_{i,bk}, i = 1, 2, 3, ...N; k = 1, 2, ...K; b < k.$$

$$p_{ik} = \frac{e^{\mu_i+\delta_{i,bk}}}{1+e^{\mu_i+\delta_{i,bk}}}, i = 1, 2, 3, ...N; k = 1, 2, ...K; b < k,$$

2.2 Fixed Effects Model

We can represent the model based on the assumption of a fixed effects model. Here b is the baseline treatment, μ_i's are the trial specific baselines representing the log odds of the event in the reference treatment (k = b), and instead of $\delta_{i,bk}$ which is the trial-specific log odds ratio of event occurrence of the treatment group k compared with the reference treatment, we use $d_{i,bk}$ under the fixed effects model. The logit link function maps the probability of treatment response on the real number and the model is shown below (Higgins and Whitehead, 1996).

$$logit(p_{ik}) = log\left(\frac{p_{ik}}{1-p_{ik}}\right) = \mu_i + d_{i,bk}, i = 1, 2, 3, ...N; k = 2, ...K; b < k.$$

2.3 Random Effects Model

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

$$logit(p_{ik}) = log\left(\frac{p_{ik}}{1-p_{ik}}\right) = \mu_i + \delta_{i,bk}, i = 1, 2, 3, ...N; k = 2, ...K; b < k.$$

Equivalently, we can say that the trial-specific treatment effects come from the same family of distributions. 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. Thus, for example, $d_{23} = d_{13} - d_{12}$ and so on assuming consistency. Thus, here $1 = b$ the baseline treatment. This means that we can generalize how to estimate the functional parameters as $d_{st} = d_{bt} - d_{bs}$. 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. Studies have shown that, in the analysis of multi-arm data, any assumptions about heterogeneity have implications on the relative efficacy of parameters (Higgins and Whitehead, 1996). 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 (summaries) will be incorrect (Higgins and Whitehead, 1996). One multi-arm trial i which compares α_i, which is the number of arms in study i, will create a correlated vector of random effects of $\alpha_i - 1$ given by $\delta_i = (\delta_{i,12}, ..., \delta_{i,k\alpha_i})^\top$. Assuming consistency, the functional parameters are obtained from the $K - 1$ direct treatment effects using $\delta_{st} = \delta_{bt} - \delta_{bs}$, $b = 1, 2, ..., K$, $s = 2, 3, ...K$,

$t = 3, 4, ..., K$; 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 effect discussed earlier become multivariate normal distribution for a vector of the random effects as shown below (Salanti et al., 2008). Also the assumption of homogeneity between-trial variance means that all σ_{bk}^2 are the same and equal to σ^2. This implies that the covariance between two contrasts in the multi-arm trial is $\sigma^2/2$. Thus we have the vector of random effects that follows a multivariate normal distribution given by

$$\begin{pmatrix} \delta_{i,12} \\ \delta_{i,13} \\ \vdots \\ \delta_{i,ba_i} \end{pmatrix} \sim N_{\sum_{i=1}^{k} a_i - k} \left(\begin{pmatrix} \delta_{i,12} \\ \delta_{i,13} \\ \vdots \\ \delta_{i,ba_i} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \cdots & \frac{\sigma^2}{2} \\ \vdots & \ddots & \vdots \\ \frac{\sigma^2}{2} & \cdots & \sigma^2 \end{pmatrix} \right).$$

Let us assume there are four interventions or treatments (say A, B, C and D) to bed-wetting. Also assume that we have 4 studies or randomized controlled trials, making a comparison among the 4 treatments. Now 2 studies are two arm trials consisting of AB and AD and 2 studies are three arms consisting of A, B and C for the first three arm study and A, C and D for the second three arm study. We take treatment A as the base treatment. Hence, from the above set-up, we have the basic parameters as δ_{AB}, δ_{AC} and δ_{AD}, and the functional parameters derived from the basic parameters as δ_{BC}, δ_{BD} and δ_{CD}. The basic parameters will be estimated by the network and the functional parameters will be estimated by the consistency criteria. As such the random effects model

will be

$$logit \begin{pmatrix} p_{1B} \\ p_{2D} \\ p_{3B} \\ p_{3C} \\ p_{4C} \\ p_{4D} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_4 \\ \mu_4 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \delta_{AB} \\ \delta_{AD} \\ \delta_{AB} \\ \delta_{AC} \\ \delta_{AC} \\ \delta_{AD} \end{pmatrix}$$

and

$$\begin{pmatrix} \delta_{AB} \\ \delta_{AD} \\ \delta_{AB} \\ \delta_{AC} \\ \delta_{AC} \\ \delta_{AD} \end{pmatrix} \sim N_6 \left(\begin{pmatrix} \delta_{AB} \\ \delta_{AD} \\ \delta_{AB} \\ \delta_{AC} \\ \delta_{AC} \\ \delta_{AD} \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & \frac{\sigma^2}{2} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{2} & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & \frac{\sigma^2}{2} \\ 0 & 0 & 0 & 0 & \frac{\sigma^2}{2} & \sigma^2 \end{pmatrix} \right).$$

Here we have $i = 1, 2, 3, 4$, and $a_1 = 2$, $a_2 = 2$, $a_3 = 3$ and $a_4 = 3$. Hence $\sum_{i=1}^{k} a_i - k = \sum_{i=1}^{4} a_i - 4 = 2 + 2 + 3 + 3 - 4 = 6$. Hence, in order to compute the *logit* of a particular treatment comparison for the basic parameters and functional parameters, we first look at what the basic and functional parameters are. Basic parameters are actually the direct evidence obtained from the study. Hence from above, we have δ_{AB}, δ_{AC} and δ_{AD} as the basic parameters of our assumed example. Then the *logit* for the comparison of δ_{AB} is

$$\begin{aligned} logit[p_{AB}] &= logit[p_{AB}] - logit[p_{AA}] \\ &= \mu_i + \delta_{AB} - (\mu_i + \delta_{AA}) \\ &= \delta_{AB} - \delta_{AA}. \end{aligned}$$

But we know that the comparison between the baseline treatment and itself is zero, thus $\delta_{AA} = 0$. Hence we have

$$\begin{aligned} logit[p_{AB}] &= \delta_{AB} \\ &= log\left(\frac{p_{AB}}{1 - p_{AB}}\right) = \delta_{AB}. \end{aligned}$$

Therefore the odds ratio of AB is

$$\frac{p_{AB}}{1 - p_{AB}} = e^{\delta_{AB}}.$$

Hence the other basic parameters δ_{AC} and δ_{AD} follow as above. Also the functional parameters use the concept of consistency to estimate the other functional parameters. Thus, for example, the logit for the comparison of δ_{AB} is given by

$$\begin{aligned} logit[p_{BC}] &= logit[p_{AC}] - logit[p_{AB}] \\ &= \mu_i + \delta_{AC} - (\mu_i + \delta_{AB}) \\ &= \delta_{AC} - \delta_{AB} \\ &= log\left(\frac{p_{BC}}{1 - p_{BC}}\right) = \delta_{AC} - \delta_{AB}. \end{aligned}$$

Therefore the odds ratio of BC is

$$\frac{p_{BC}}{1 - p_{BC}} = e^{\delta_{AC} - \delta_{AB}}.$$

Hence, the other basic parameters, δ_{BD} and δ_{CD}, follow as above.

2.4 Assessing the Goodness of Fit of a Model (Model Selection)

After modeling data every statistician wants to make inferences from the model. They will like to know how best their data fits the model. This gives rise to the concept of goodness of fit. Goodness of fit is a statistical concept that shows how well the model fits the observed set of observations. It allows one to describe the difference between the observed and expected values and has the ability to help discriminate between alternative models. This enables a statistician explore and compare differing models that could be used in analyzing the data and to give the most precise inference. There is a classic test for the goodness of fit called the *likelihood ratio test*. However, here we are going to concentrate more on the following two methods used in the assessment of goodness of fit. These two are

1. Akaike Information Criteria (AIC).

2. Bayesian Information Criteria (BIC).

Akaike Information Criteria (AIC) for a given model is a function of its maximized log-likelihood and the number of estimable parameters, say s. Mathematically AIC is represented by

$$AIC = -2\log(L(y|\hat{\theta})) + 2s.$$

For the Bayesian Information Criteria (BIC), it is also called the Schwarz Bayesian Information Criteria which consist of s parameters and n observations. Mathematically BIC is represented as

$$BIC = -2\log(L(y|\hat{\theta})) + (2s \times \log(n)).$$

Now the question is how best one can come to a conclusion based on the results of these model assessment statistics. It has been shown that smaller values for these statistics (that is, AIC and BIC) gives a direction of which model is relatively better. For example, to compare which model to use between fixed effects model and random effects model, one can use the above model assessment statistics. The model that has the smallest statistics is adopted to be the relatively efficient one. One of the major advantages of the AIC and BIC statistics is that they can be used for non-nested or non-heirarchical models. Consequently, a generalization and a Bayesian version of AIC was proposed and it has some level of relationship with the BIC, which is called the Deviance Information Criteria (DIC). Spielgelhalter et al. (1998), indicated this in their paper, where they proposed the DIC. The DIC comes in handy and is very useful in Bayesian model selection issues, where the posterior distributions of parameters have been obtained by the Markov Chain Monte Carlo (MCMC) simulation (Spielgelhalter and Myles, 2004). They suggested a classical deviance, say $D(\theta)$, to examine the posterior distribution, which is given by

$$D(\theta) = -2ln(f(y/\theta)) + 2ln(f(y)),$$

where $f(y/\theta)$ denotes the likelihood function and $f(y)$ denotes a fully specified standardizing term that is completely determined by the observed data. The DIC is composed of two components and they sum to form the DIC. The first component measures the goodness of fit of the model by the posterior expectation of the overall residual deviance and the second component measures the complexity of the model by the effective number of parameters. This is defined by the difference between the posterior mean of the overall residual deviance and the deviance evaluated at the posterior mean of the parameters of interest. Hence, mathematically, the DIC is represented by

$$DIC = \bar{D} + p_D,$$

where $\bar{D}$ is the first component and p_D is the second component. Also the two components can be further broken down to,

$$\bar{D} = E_{\theta|y}(D(\hat{\theta})) = E_{\theta|y}(-2ln(f(y|\theta))),$$

$$p_D = E_{\theta|y}(D(\hat{\theta})) - D(E_{\theta|y}(\theta)) = \bar{D} - D(\hat{\theta}).$$

Therefore, DIC is given by

$$DIC = \bar{D} + p_D = 2\bar{D} - D(\hat{\theta}) = 2\bar{D} - 2D(\hat{\theta}) + 2D(\hat{\theta}) - D(\hat{\theta})$$

$$= D(\hat{\theta}) + 2p_D.$$

As mentioned earlier, for the interpretation of the AIC and BIC statistics, small values of DIC suggest a better fit of a particular model over the other, hence, it efficiently predicts the observed data. The question naturally arises:

how "small is small"? Hence there is a rule of thumb, that guides in the choice of which "small" DIC to choose. A difference of more than 10 can definitely rule out the model with higher DIC, differences between 5 and 10 are considerable and differences in DIC less than 5 may provide very different inferences, and caution should be taken when referring to the model with the lowest DIC.

Another way for model selection is by using the Posterior distribution of the Sum of Residual Deviance, $\bar{D}$. From our previous method, where we used the contrast based method, we had a likelihood distribution as binomial. Hence we shall have to develope a binomial likelihood function for the sum of residual deviance. This is given by

$$\bar{D} = \sum_{i=1}^{N} Dev_i = \sum_{i=1}^{N}\sum_{k=1}^{K} 2\left[r_{ik}\log\left(\frac{r_{ik}}{n_{ik}p_{ik}}\right) + (n_{ik}-r_{ik})log\left(\frac{n_{ik}-r_{ik}}{n_{ik}-n_{ik}p_{ik}}\right)\right],$$

where r_{ik} denotes the number of events of the treatment k in trial i, n_{ik} is the total number of observations and p_{ik} is the probability of event occurence. The posterior distribution of the model deviance difference is denoted by

$$\bar{D}_{1,2} = \bar{D}_1 - \bar{D}_2.$$

The posterior probability obtained from the posterior distribution of the model deviance difference is

$$P[\bar{D}_{1,2} > \beta(\bar{D})],$$

where β is defined as follows. To allow inference from this Bayesian approach and to have its association with the frequentist approach, the posterior distribution of the model difference follows approximately a chisquare distribution

Table by Kass and Raftery	
β	Evidence in favour of model 2
0-2	Not worth more than a bare mention
2-6	Positive
6-10	Strong
>10	Very strong

Table 2.1: Kass and Raftery Table of β values in model selection.

with degrees of freedom, $d.f = p_2 - p_1$, this is, the difference between the number of parameters estimated from the models involved. Hence the above β can be approximately, $\beta = \chi^2_{p_2-p_1}$. The higher the probability, the stronger is the evidence in favour of model 2 against model 1. Also we can use the value β to determine the best fitted model between model 1 and model 2. A yardstick is to consider which value of β will have

$$P[\bar{D}_{1,2} > \beta(\bar{D})] = 0.5.$$

The question now boils down to how to interpret that β value. This is because that β value can be used to quantify the evidence against model 1. A table postulated, which is cited in (Kass and A.E., 1995) can be used for the interpretation on the β values. This is given in Table 2.1.

2.5 Ranking of Treatments/Interventions using Baseyian Probability

The advantage of Bayesian approach in network meta-analysis is the ranking of the treatments so that we know which treatment is the best. In each of the

MCMC run, the treatments are ranked according to their magnitude. The proportion of the MCMC cycles in which the treatment ranks first give the probability that the treatment is the best among the other treatment. Salanti et al. (2011), suggested a simple method called Surface Under the Cummulative Rank Curve (SUCRA). This ranks the cummulative probabilities for each treatment by computing an index of max 1(100%). It can be witten as

$$SUCRA_k = \frac{\sum_{z=1}^{k-1} Cum_{k,z}}{k-1}, \tag{2.2}$$

where k is the number of treatments and cum is the cummulative probabilities at rank z. A graphical representation of this index is called Rankograms.

2.6 Inconsistency Detection in Network Meta-Analysis Using Arm-Based Model

Consistency is one of the most important assumption of network meta-analysis which must be met before its use because if inconsistency exist, it affects the accuracy and conclusion of the study. Consistency arises when the direct and indirect evidences agree. Hence, if they disagree it brings about inconsistency (Zhao et al., 2016). Inconsistency arises from non-comparability of trials and different control groups or differences in patient characteristics. Hence, there is the need to identify the presence of inconsistency in network meta-analysis as well as to identify the source of inconsistency. The contrast based model and arm based model have been used in the detection of inconsistency in

network meta-analysis. The contrast based method has been widely studied by Lu and Ades (2006), where they modelled the inconsistency by introducing some inconsistency factors called w. Also they assesed this inconsistency by having loops in the treatment comparisons. Hence, this method made use of the inconsistency degree of freedom (ICDF). This ICDF is the number of independent loops of evidence in the network (Dias et al., 2011), where $ICDF = T - K + 1$ and T is the number of direct comparisons and K is the number of treatments. For example, if we have four treatments, say A, B, C and D, then $T = 6$ and $K = 4$. Therefore the $ICDF = 6 - 4 + 1 = 3$, thus we will have 3 independent loops to model the inconsistency, in addition to the inconsistency factor. Here the direct evidence consist of d_{AB}, d_{AC} and d_{AD}. This is denoted by

$$d_{BC} = d_{AC} - d_{AB} + w_{ABC}$$

$$d_{BD} = d_{AD} - d_{AB} + w_{ABD}$$

$$d_{CD} = d_{AD} - d_{AC} + w_{ACD}.$$

Here we look at how to use the arm based model to identify inconsistency and investigate the sources of the inconsistency. The arm based (AB) model is denoted by

$$logit(p_{ik}) = \mu_k + \eta_{ik}, \tag{2.3}$$

where μ_k is the (fixed-effect) mean outcome for treatment k and η_{ik} is the random effect for treatment k in study i. Then the random effects $\boldsymbol{\eta_i}$ for study

i are modeled as

$$\boldsymbol{\eta_i} = (\eta_{i1}, \eta_{i2}, ...\eta_{ik})^\top \sim MVN(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is a $K \times K$ unstructured covariance matrix to allow correlation between treatment arms in each trial. Compared with the CB model framework, AB models are more straightforward to interpret, especially when implemented in a missing-data framework that imputes values for any treatment arms missing in a given study, thus allowing use of a common baseline across all trials.

Under the AB model, there are two possible ways to determine inconsistency. The first one is by using estimates of the fixed effects in the AB model to test the discrepancy of direct and indirect evidence for comparing two treatment either loop-based or not. The second one is by using estimates of specific AB random effects to detect inconsistency at certain trial-by-arm combinations, once inconsistency has been detected through the AB model fixed effects Zhao et al. (2016). The AB model does not use loops to study inconsistency since it implicitly performs inconsistency. But we can use it to investigate inconsistency in a loop-based manner by defining discrepancy factors using a different subsetting method for the groups. There are four groups that can be formed when trying to investigate inconsistency. Assume that we have a network meta-analysis study, where two treatments are suspected to have some level of inconsistency say treatment A and B. Hence, for the first group (i), we look at trials that have compared both treatment A and B. For the second group (ii), we group all trials of treatment comparison with treatment A excluding treatment B. The third group (iii) will consist of all trials of

treatment comparison with treatment B excluding treatment A. The fourth group (iv) will consist of trials of treatment comparisons that exclude treatment A and B. This can be computed as follows for the discrepancy factor

$$\Delta_{AB} = (\mu_A^{(i)} - \mu_B^{(i)}) - (\mu_A^{(ii)} - \mu_B^{(iii)}), \tag{2.4}$$

which is the difference in treatment effects in trials including both arms (the direct evidence) minus the difference in trials including just one arm (the indirect evidence). If zero is found to be in the far tail of this posterior distribution, we can conclude that the two sources of evidence for comparing A and B thus the direct and indirect evidence are discrepant, and thus, inconsistency exists. The above approach, uses the fixed effects in the arm based model to investigate the presence of inconsistencies. Using just this method has some short commings since it only investigates the presence of inconsistencies but does not identify the potential sources of these inconsistencies in the study. Hence we need to consider the second step where the random effects of the arm based model is used. Zhao et al. (2016), considered the top 5% of the absolute values of the random components in each trial with its corresponding treatment. Here, we try to find the random components of each treatment in each trial. We then rank these absolute values of the random components. Then the top 5% absolute random components are considered to be the ones as the sources of these inconsistencies.

Chapter 3

Diagnostics in Network Meta-Analysis

3.1 Diagnostics to Linear Hierarchical Model in Network Meta Analysis

Before a model is used for predictive or inference purposes, we need to test the robustness or check to see if all its assumptions are met. One approach to appraise a model's robustness is a diagnostic approach. Here we look at how to perform some diagnostics on linear hierarchical model. A classic example of a linear model for which diagnostics can be applied is the linear regression model, which is of the form $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{Y}$ is the vector of observations or response variable, $\boldsymbol{X}$ is the design matrix, $\boldsymbol{\beta}$ is the vector of unknown parameters and $\boldsymbol{\epsilon}$ is the vector of random errors. The regression model is a linear model in which we made some assumptions that need to be met before

inference is made. Such assumptions are

1. There should be a linear relationship and $\boldsymbol{X}$ is fixed.
2. The random errors should follow a normal distribution with mean, zero and a common variance, σ^2.
3. There should be equality of variances, thus satisfying homoscedasticity assumption.
4. There should be a little or no multicollinearity.

These assumptions are tested using some procedures. Some of these procedures are

1. Check the residuals and the fitted values. If there is a non-linear pattern between them, then we have a good indication that the linearity assumption is met.
2. Check the normality of the residuals. This is done by using the normal $Q-Q$ plot.
3. Check the scale-location or spread-location plot. This is to check if the homoscedasticity assumption is met by plotting the squared root of the standardized residuals against the fitted.
4. Check the residuals versus the leverage. This plot helps to find influential cases or subjects.

3.2 Generalized Linear Hierarchical Model-Network Meta-Analysis as a Form of Hierarchical Model

Hierarchical linear models are statistical models of parameters that vary at more than one level. They are also known as multilevel models, nested data models, mixed models, random coefficient, random-effects models, random parameter models, or split-plot designs (Bryk et al., 2002). Network meta-analysis are modeled in a hierarchical format because of the different levels to which the parameters of interests are assigned. Network meta-analysis is a classic example of hierarchical model. We consider the models for binary outcomes. Assume that the outcome y_{ik} follows a binomial distribution for the study i and the treatment arm k, as given below

$$y_{ik}|p_{ik} \sim Bin(n_{ik}, p_{ik}), i = 1, ..., I, k \in S_i, S_i \subset \{1, ..., K\}, \tag{3.1}$$

where S_i denotes the set of treatments compared in trial i, n_{ik} is the total of observations on treatment k in trial i and p_{ik} is the probability of occurence of the event on treatment k in trial i.

Contrast Based Model and Arm Based Model

In the contrast based (CB) model, we try to model the random effects to capture the heterogeneity between studies. The model is given by

$$logit(p_{ik}) = \alpha_{iB} + \delta_{iBk}, \tag{3.2}$$

where B is a base treatment chosen for and specific to study i, α_{iB} is the log odds of an event on the base treatment in study i, and δ_{iBk} is the log odds ratio of an event for the k treatments compared with the base treatment in study i. We further assume indepedent normal distributions for the random effect δ_{iBk}

$$\delta_{iBk} \sim N(d_k - d_B, \sigma^2). \tag{3.3}$$

Diagnostics for the Arm Based Model

To simplify matters, the key is to express a hierarchical model in the form of a linear model by adding artificial "cases" to the dataset. As discussed earlier, we realised network meta-analysis is a classic example of hierarchical model. We need to do some transformation of it to a linear hierarcical model for applying the diagnostic methods.

Also, the errors of the arm-based models are not normally distributed since they are in a binary format. Hence we need to transform by the normal approximation to the likelihood. The binary model is of the form

$$logit(p_{ik}) = \mu_k + \eta_{ik}, \tag{3.4}$$

$$\boldsymbol{\eta_i}^\top = (\eta_{i1}, ..., \eta_{ik})^\top \sim MVN(0, \Sigma). \tag{3.5}$$

Steps to transform the Network Meta-Analysis binary errors to Normal

1. We begin by transforming each data point, y_i to $\widetilde{y}_i$.

2. This is done by using $\widetilde{y}_i = logit(\frac{y_{ik}}{n_{ik}})$, where $logit(\frac{y_{ik}}{n_{ik}}) = log\left(\frac{\frac{y_{ik}}{n_{ik}}}{1-\frac{y_{ik}}{n_{ik}}}\right) = log\left(\frac{y_{ik}}{n_{ik}-y_{ik}}\right)$.

3. If the data point $y_{ik} = 0$ and $y_{ik} = n_{ik}$, then $\widetilde{y}_i$ is undefined. Hence there is the need to add 0.5 to y_{ik} and $n_{ik} - y_{ik}$ when the undefined case happens.

4. Using the delta method, $var(\widetilde{y}_{ik}) \approx \frac{1}{n_{ik}p_{ik}(1-p_{ik})}$.

5. We finally try to approximate the binary model by applying Hodges's method to generalized linear hierarchical models in NMA. The above transformation method was proposed by Zhao et al. (2017) using the delta method.

3.3 Formulation of Generalized Linear Hierarchical Model

In the formulation of a generalized linear hierarchical model, it was shown from previous literature that the following equations 3.6 are used to break down a

hierarchical model to a linear model (Zhao et al., 2017),

$$\begin{aligned} \widetilde{\boldsymbol{y}}_{\boldsymbol{ik}} &= \boldsymbol{\xi}_{\boldsymbol{ik}} + \boldsymbol{\epsilon}_{\boldsymbol{ik}}, \\ \boldsymbol{\xi}_{\boldsymbol{ik}} &= \boldsymbol{\mu}_{\boldsymbol{k}} + \boldsymbol{\eta}_{\boldsymbol{ik}}, \\ \boldsymbol{\mu}_{\boldsymbol{k}} &= \boldsymbol{M}_{\boldsymbol{k}} + \boldsymbol{\upsilon}_{\boldsymbol{k}}. \end{aligned} \tag{3.6}$$

From equations 3.6 (Hodges, 1998), we have $\boldsymbol{\epsilon_{ik}} \sim N(0, \frac{1}{n_{ik}p_{ik}(1-p_{ik})}$, $\boldsymbol{\eta_i} = (\eta_{i1}, ..., \eta_{ik}) \sim MVN(0, \Sigma)$. Also, $\boldsymbol{\mu_k}$ represents the priors with mean $\boldsymbol{M_k}$ and $\boldsymbol{\upsilon_k} \sim N(0, 1000)$. And $\boldsymbol{\upsilon_k}$ has variance to be 1000 because we need a vague prior so that the analysis is driven solely by the data. Hodges (1998) showed that we can move the known terms to the left hand side and the unknown terms to the right handside of the equations

$$\begin{aligned} \widetilde{\boldsymbol{y}}_{\boldsymbol{ik}} &= \boldsymbol{\xi}_{\boldsymbol{ik}} + \boldsymbol{\epsilon}_{\boldsymbol{ik}}, \\ \boldsymbol{0} &= -\boldsymbol{\xi}_{\boldsymbol{ik}} + \boldsymbol{\mu}_{\boldsymbol{k}} + \boldsymbol{\eta}_{\boldsymbol{ik}}, \\ -\boldsymbol{M}_{\boldsymbol{k}} &= -\boldsymbol{\mu}_{\boldsymbol{k}} + \boldsymbol{\upsilon}_{\boldsymbol{k}}. \end{aligned} \tag{3.7}$$

Let $\#S_i$ denote the number of arms in study i. Hence the total number of arms is $N = \sum_{i=1}^{I} \#S_i$. Then the above set of equations from 3.7 can be written in a matrix form as given below

$$\begin{bmatrix} \widetilde{\boldsymbol{y}}_{\boldsymbol{ikN}} \\ \boldsymbol{0_N} \\ -\boldsymbol{M_k} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I_{N\times N}} & \boldsymbol{0_{N\times K}} \\ -\boldsymbol{I_{N\times N}} & \boldsymbol{H_{N\times K}} \\ \boldsymbol{0_{K\times N}} & -\boldsymbol{I_{K\times K}} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_N \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\eta} \\ \boldsymbol{\upsilon} \end{bmatrix},$$

$$\boldsymbol{Y} = \boldsymbol{X\Theta} + \boldsymbol{E}, \tag{3.8}$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)^\top$, $\boldsymbol{\xi}$, $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta}$ are vectors of length N and $\boldsymbol{\upsilon} = (\upsilon_1, ..., \upsilon_k)^\top$. The $\boldsymbol{I}$ matrix is an identity matrix, the $\mathbf{0}$ matrix is the zero matrix and $\boldsymbol{H}$ is matrix which shows how treatment arms are arranged in a particular study.

Example of Linear Hierarchical Model Diagnostics

Assume we have three studies or trials where the first trial consists of 3-arms and the other studies consist of 2-arms with three treatments to be considered. Hence, we have $N = 3 + 2 + 2 = 7$, $K = 3$ which is the number of treatments. Also let us asssume $M_k = 0$. Therefore we have our matrix representation as shown below. From equation 3.8, $\boldsymbol{Y}$, the transformed data, and $\boldsymbol{X}$, the design matrix, are known. But $\boldsymbol{\Theta}$ is an unknown parameter vector and $\boldsymbol{E}$ is an error term with mean $\mathbf{0}$ and block diagonal covariance matrix, $\boldsymbol{\Gamma}$. The block of $\boldsymbol{\Gamma}$ corresponds to the covariance matrices of $\boldsymbol{\epsilon}$, $\boldsymbol{\eta}$ and $\boldsymbol{\upsilon}$, where the upper left 7×7 block of $\boldsymbol{\Gamma}$ contains the transformed variances for the $\boldsymbol{\epsilon_{ik}}$, the next 7 rows and columns are composed of 3 block covariance matrices for the random effects of the 3 trails and the last 3×3 diagonal block contains the variances of the $\boldsymbol{\upsilon_k}$ (the prior variances), all set to 1000. Our interest is in the estimates of the $\boldsymbol{\mu_k}$'s. This format is referred to as the constraint case

formulation of a hierarchical model.

$$\begin{bmatrix} \widetilde{\boldsymbol{y}}_{\boldsymbol{ik7}} \\ \mathbf{0_7} \\ \mathbf{0_3} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I_{7\times 7}} & \mathbf{0_{7\times 3}} \\ -\boldsymbol{I_{7\times 7}} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{7\times 3} \\ \mathbf{0_{3\times 7}} & -\boldsymbol{I_{3\times 3}} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_7 \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_7 \\ \eta_1 \\ \vdots \\ \eta_7 \\ \upsilon_1 \\ \upsilon_2 \\ \upsilon_3 \end{bmatrix}.$$

- Rows of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{E}$ in equation 3.8 corresponding to $\widetilde{\boldsymbol{y}}$ and also the $\boldsymbol{y_{ik}}$ are called data cases.
- Rows of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{E}$ corresponding to $\boldsymbol{H}$ are the constrained cases. This imposes constraints on the parameters, $\boldsymbol{\Theta}$.
- Rows of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{E}$ in equation 3.8 corresponding to $\boldsymbol{\upsilon_k}$ are denoted as prior cases. These are generated by the hyperprior.

3.4 Presteps Before Diagnostics

From equation 3.8, we realized the error terms do not have equal variances, that is from $\boldsymbol{\Gamma}$. Hence we have to transform it such that we obtain equal variances. In doing so we have to premultiply equation 3.8 by $\boldsymbol{\Gamma}^{-\frac{1}{2}}$. Hence we get

$$\boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{Y} = \boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{X}\boldsymbol{\Theta} + \boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{E}, \tag{3.9}$$

where $\boldsymbol{E} \sim N(0, \boldsymbol{\Gamma})$, with $\boldsymbol{\Gamma}$, corresponding to the covariance matrices of $\boldsymbol{\epsilon}$, $\boldsymbol{\eta}$ and $\boldsymbol{\upsilon}$, suggesting unequal variances. Hence we have $\boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{E} \sim N(0, \boldsymbol{I}_{(2N+k)\times(2N+k)})$, we then obtain equal varinces. From equation 3.9, we can rewrite it as

$$\boldsymbol{Y}^* = \boldsymbol{X}^*\boldsymbol{\Theta} + \boldsymbol{E}^*. \tag{3.10}$$

We estimate the residuals as: $e_i = y_i - \hat{y}_i$, where $\hat{y}_i$ is the ith row of $\boldsymbol{X}\hat{\boldsymbol{\beta}}$. Therefore from this idea, if we define the hat matrix as $\boldsymbol{V} = \boldsymbol{X}^*(\boldsymbol{X}^{*\top}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\top}$, then the vector of residuals, $\hat{\boldsymbol{E}}^*$ will be

$$\begin{aligned}\hat{\boldsymbol{E}}^* &= \boldsymbol{Y}^* - \hat{\boldsymbol{Y}}^* = \boldsymbol{Y}^* - \boldsymbol{X}^*\hat{\boldsymbol{\Theta}} \\ &= \boldsymbol{Y}^* - \boldsymbol{X}^*(\boldsymbol{X}^{*\top}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\top}\boldsymbol{Y}^* = \boldsymbol{Y}^* - \boldsymbol{V}\boldsymbol{Y}^* \\ &= (\boldsymbol{I} - \boldsymbol{V})\boldsymbol{Y}^*.\end{aligned} \tag{3.11}$$

3.5 Diagnostics for case Influence

Case influence diagnostics show how estimates of parameters change when cases are deleted. In a Bayesian MCMC framework of which we are considering (Zhao et al., 2017), the most accurate way of detecting influential cases is to rerun the MCMC algorithm for each deleted cases but this is computationally cumbersome. Hence there is a an approach to approximate this result. This approach is the linear approximation method. This uses a classic result in linear model theory, where if $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, then $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$. Also

if we have some cases deleted from $\boldsymbol{X}$, then the new $\boldsymbol{X}$ will be $\boldsymbol{X_R}$. Hence a standard result of the inverse of $\boldsymbol{X_R^\top X_R}$ will be

$$(\boldsymbol{X_R^\top X_R})^{-1} = (\boldsymbol{X^\top X})^{-1} + (\boldsymbol{X^\top X})^{-1}\boldsymbol{X_R^\top}(\boldsymbol{I - V_R})^{-1}\boldsymbol{X_R}(\boldsymbol{X^\top X})^{-1}, \quad (3.12)$$

where $\boldsymbol{V_R}$ is the new hat matrix after the deleted cases. This representation is given when we have more than one deleted cases at the same time. The updated $\hat{\boldsymbol{\beta}}$ (after deletion of cases) is given by $\hat{\boldsymbol{\beta}}_{\boldsymbol{R}} = (\boldsymbol{X_R^\top X_R})^{-1}\boldsymbol{X_R^\top Y_R} = (\boldsymbol{X_R^\top X_R})^{-1}(\boldsymbol{X^\top Y} - \boldsymbol{X_R^\top Y_R})$, where $\boldsymbol{V_R} = \boldsymbol{X_R}(\boldsymbol{X^\top X})^{-1}\boldsymbol{X_R^\top}$.

Let us consider the case where we have just one deleted case. Equation 3.12 cn be simplified further. This can be expressed as

$$(\boldsymbol{X_r^\top X_r})^{-1} = (\boldsymbol{X^\top X})^{-1} + (\boldsymbol{X^\top X})^{-1}\boldsymbol{x_r^\top}(\boldsymbol{1 - v_r})^{-1}\boldsymbol{x_r}(\boldsymbol{X^\top X})^{-1}, \quad (3.13)$$

where $v_r = \boldsymbol{x_r}(\boldsymbol{X^\top X})^{-1}\boldsymbol{x_r^\top}$. Now we post multiply Equation 3.13 by $(\boldsymbol{X^\top Y} - \boldsymbol{X_i^\top Y_i}$. Hence we have

$$\begin{aligned}(\boldsymbol{X_r^\top X_r})^{-1}(\boldsymbol{X^\top Y} - \boldsymbol{x_r^\top y_r}) = (\boldsymbol{X^\top X})^{-1}\boldsymbol{X^\top Y} - (\boldsymbol{X^\top X})^{-1}\boldsymbol{x_r^\top y_r} \\ + \frac{(\boldsymbol{X^\top X})^{-1}\boldsymbol{x_r^\top x_r}(\boldsymbol{X^\top X})^{-1}\boldsymbol{X^\top Y}}{1 - v_r} \\ - \frac{(\boldsymbol{X^\top X})^{-1}\boldsymbol{x_r^\top x_r}(\boldsymbol{X^\top X})^{-1}\boldsymbol{x_r^\top y_r}}{1 - v_r}\end{aligned}$$

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{r}} = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top \boldsymbol{y}_{\boldsymbol{r}} + \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top \boldsymbol{x}_{\boldsymbol{r}} \hat{\boldsymbol{\beta}}}{1 - v_r}$$

$$- \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top \boldsymbol{x}_{\boldsymbol{r}} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top \boldsymbol{y}_{\boldsymbol{r}}}{1 - v_r}$$

$$= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top \boldsymbol{y}_{\boldsymbol{r}} + \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top \boldsymbol{x}_{\boldsymbol{r}} \hat{\boldsymbol{\beta}}}{1 - v_r}$$

$$- \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top v_r \boldsymbol{y}_{\boldsymbol{r}}}{1 - v_r}$$

$$= \hat{\boldsymbol{\beta}} - \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top}{1 - v_r} [(1 - v_r) \boldsymbol{y}_{\boldsymbol{r}} - \boldsymbol{x}_{\boldsymbol{r}} \hat{\boldsymbol{\beta}} + v_r \boldsymbol{y}_{\boldsymbol{r}}]$$

$$= \hat{\boldsymbol{\beta}} - \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top}{1 - v_r} [\boldsymbol{y}_{\boldsymbol{r}} - \boldsymbol{x}_{\boldsymbol{r}} \hat{\boldsymbol{\beta}}]$$

$$= \hat{\boldsymbol{\beta}} - \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top}{1 - v_r} \hat{E}_r$$

Thus, $\quad \hat{\boldsymbol{\beta}}_{\boldsymbol{r}} - \hat{\boldsymbol{\beta}} = - \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top}{1 - v_r} \hat{E}_r.$

Therefore we have a linear approxiamtion for the change in $\hat{\boldsymbol{\beta}}$ arising from the deletion of the r^{th} case, as giiven by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{r}} - \hat{\boldsymbol{\beta}} \approx - \frac{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{\boldsymbol{r}}^\top}{1 - v_r} \hat{E}_r. \tag{3.14}$$

Now relating equation 3.14 to our equation 3.9, we can approximate the change in $\hat{\Theta}$ arising by deleting the r^{th} case, as given by

$$\hat{\boldsymbol{\Theta}}_{\boldsymbol{r}} - \hat{\boldsymbol{\Theta}} \approx - \frac{(\boldsymbol{X}^{*\top} \boldsymbol{X}^*)^{-1} \boldsymbol{x}_{\boldsymbol{r}}^{*\top}}{1 - v_r} \hat{E^*}_r. \tag{3.15}$$

where $\boldsymbol{X}^*$ is transformed using the posterior mean of $\boldsymbol{\Gamma}$ as in equation 3.9, $\hat{\boldsymbol{\Theta}}$ is the posterior mean of $\boldsymbol{\Theta}$ using the full dataset, $\hat{\boldsymbol{\Theta}}_{\boldsymbol{r}}$ is the posterior mean of Θ after deleting the r^{th} case, $\hat{E}^*{}_r$ is the r^{th} row of 3.11 which is a scalar. Also v_r is the r^{th} diagonal element of $\boldsymbol{V}$ and also a scalar.

Now to give a decision rule about the change in the size after deleting a case, we try to standardize this change in size and call it a Relative Change (RC). Let θ be an element of $\boldsymbol{\Theta}$, then the RC arising from deleting the r^{th} observation is given as

$$RC\{\theta; r\} = \frac{\hat{\theta}_{(r)} - \hat{\theta}}{psd\{\theta|Y\}}, \tag{3.16}$$

where *psd* stands for Posterior Standard Deviation computed using the full dataset. By convention, if $|RC| \geq 2$, it suggests the deleted observation is an influential case (Zhao et al., 2017).

3.5.1 Outlier Detection Using Residuals

We can also check to see if observations that are suspected to be inconsistent are actually outliers. Since we are using the linear model to approximate the generalized linear hierarchical model, we can obtain the residuals from the linear model. Residuals are generally used in the detection of outliers. We can obtain the residuals from equation 3.11. The rth case standardized residual is calculated by

$$\frac{\hat{E}^*{}_r}{\sqrt{var(\hat{E}^*{}_r)}}. \tag{3.17}$$

We use the above standardized residuals to conclude whether a case is an outlier or not. Note that the mean of the residuals is zero in ordinary linear models. However, the mean is not zero in general for hierarchical models.

Chapter 4

Data Analysis

4.1 Example 1: Network Meta-Analysis Using Frequentist Approach.

Here we try to do an example with *R* software with a diabetes drug dataset by Senn et al. (2013). This data is involved with the comparison of the effectiveness of varieties of 10 diabetes drugs from 26 independent studies. The data is given in Table 4.1 which is a classic example of a continuous outcome in the analyis of network meta-analysis. From Table 4.1 on page 59 , we see there are 8 columns, namely the laboratory in which the study was conducted, the treatment effect (TE), the standard error of the treatment effect (seTE), Treatment 1 (treat1), Treatment 2 (treat1), the adjusted standard error (seTEadj), the number of arms in each study (narms) and multi-arm studies (multiarm). Looking at the first column we realize the number of rows are 28 suggesting this will be the number of studies considered for this analysis. However, in actual sense, there are 26 studies for this analysis.

This is because of the presence of multi-arm study in this example. Rows 26, 27 and 28 of the Senn et al. (2013) diabetes data, with the study label, *Willms1999* is a multi-arm study which has been partitioned into series of two-arm studies. The 10 drugs (treatments) are acar = Acarbose, benf = Benfluorex, metf = Metformin, migl = Miglitol, piog = Pioglitazone, plac = Placebo, rosi = Rosiglitazone, sita = Sitagliptin, sulf = Sulfonylurea and vild = Vildagliptin. Now we use R and a package in R, *netmeta* (Rcker et al., 2018), to run a network meta-analysis first using the assumption of a fixed effects model. Table 4.1 on page 59 shows the Senn diabetes drug dataset showing which study is a multi-arm. This is depicted with an asterisk, $*$. Also, a reverse of a comparison of a treatment results in the same effect size, only the sign of the effect size will be changed accordingly. Also Table 4.2 on page 60 shows the direct and indirect comparions of the various treatments in a matrix-like table, having each pairwise comparison present. Here, we realize from the Figure 4.1 that it has a link between meft-acar, hence there is direct evidence from this comparison and can be estimated by the network model. This estimate is shown in Table 4.2 on page 60 as -0.29. Also, for example, we realize from Figure 4.1 that there is no link between migl-acar, hence we can have an indirect estimate of this comparison by

$$\hat{\theta}^{indirect}_{migl-acar} = \hat{\theta}^{direct}_{plac-acar} - \hat{\theta}^{direct}_{plac-migl}. \tag{4.1}$$

where *plac* is the common comparator for *migl* and *acar*. Hence the corresponding estimate is $\hat{\theta}^{indirect}_{migl-acar} = 0.83 - 0.94 = -0.12$. This estimate is also given in Table 4.2. This sort of computations is carried out through out Table 4.2 on page 60, for the indirect estimate, hence using the assumption of consistency.

	treat1	treat2	TE	seTE	seTE.adj	narms	multiarm
DeFronzo1995	metf	plac	-1.90	0.1414	0.1414	2	
Lewin2007	metf	plac	-0.82	0.0992	0.0992	2	
Davidson2007	plac	rosi	1.34	0.1435	0.1435	2	
Wolffenbuttel1999	plac	rosi	1.10	0.1141	0.1141	2	
Kipnes2001	piog	plac	-1.30	0.1268	0.1268	2	
Kerenyi2004	plac	rosi	0.77	0.1078	0.1078	2	
Hanefeld2004	metf	piog	-0.16	0.0849	0.0849	2	
Derosa2004	piog	rosi	0.10	0.1831	0.1831	2	
Baksi2004	plac	rosi	1.30	0.1014	0.1014	2	
Rosenstock2008	plac	rosi	1.09	0.2263	0.2263	2	
Zhu2003	plac	rosi	1.50	0.1624	0.1624	2	
Yang2003	metf	rosi	0.14	0.2239	0.2239	2	
Vongthavaravat2002	rosi	sulf	-1.20	0.1436	0.1436	2	
Oyama2008	acar	sulf	-0.40	0.1549	0.1549	2	
Costa1997	acar	plac	-0.80	0.1432	0.1432	2	
Hermansen2007	plac	sita	0.57	0.1291	0.1291	2	
Garber2008	plac	vild	0.70	0.1273	0.1273	2	
Alex1998	metf	sulf	-0.37	0.1184	0.1184	2	
Johnston1994	migl	plac	-0.74	0.1839	0.1839	2	
Johnston1998a	migl	plac	-1.41	0.2235	0.2235	2	
Kim2007	metf	rosi	0.00	0.2339	0.2339	2	
Johnston1998b	migl	plac	-0.68	0.2828	0.2828	2	
Gonzalez-Ortiz2004	metf	plac	-0.40	0.4356	0.4356	2	
Stucci1996	benf	plac	-0.23	0.3467	0.3467	2	
Moulin2006	benf	plac	-1.01	0.1366	0.1366	2	
Willms1999	acar	metf	0.20	0.3579	0.3884	3	*
Willms1999	metf	plac	-1.20	0.3758	0.4125	3	*
Willms1999	acar	plac	-1.00	0.4669	0.8242	3	*

Table 4.1: Senn dataset showing the number of arms.

	acar	benf	metf	migl	piog	plac	rosi	sita	sulf	vild
acar	0.00	0.08	0.29	0.12	0.24	-0.83	0.37	-0.26	-0.39	-0.13
benf	-0.08	0.00	0.21	0.04	0.16	-0.91	0.30	-0.34	-0.47	-0.21
metf	-0.29	-0.21	0.00	-0.17	-0.05	-1.11	0.09	-0.54	-0.67	-0.41
migl	-0.12	-0.04	0.17	0.00	0.12	-0.94	0.26	-0.37	-0.50	-0.24
piog	-0.24	-0.16	0.05	-0.12	0.00	-1.07	0.14	-0.50	-0.63	-0.37
plac	0.83	0.91	1.11	0.94	1.07	0.00	1.20	0.57	0.44	0.70
rosi	-0.37	-0.30	-0.09	-0.26	-0.14	-1.20	0.00	-0.63	-0.76	-0.50
sita	0.26	0.34	0.54	0.37	0.50	-0.57	0.63	0.00	-0.13	0.13
sulf	0.39	0.47	0.67	0.50	0.63	-0.44	0.76	0.13	0.00	0.26
vild	0.13	0.21	0.41	0.24	0.37	-0.70	0.50	-0.13	-0.26	0.00

Table 4.2: Output sowing the Direct and indirect estimates of the treatments.

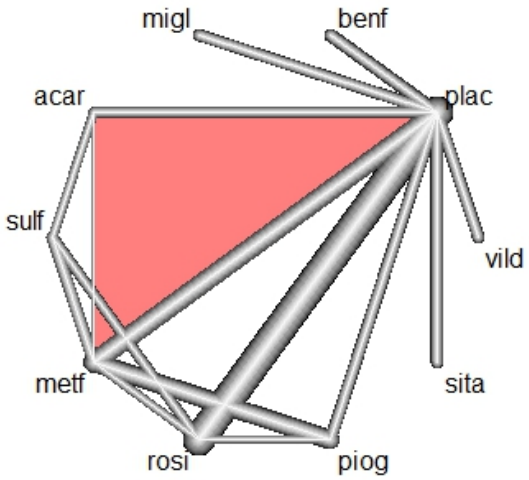

Figure 4.1: Network Plot from the Senn 2013 data of Drugs for Diabetes.

	Q	df	pval
Total	96.99	18.00	0.0001
Within designs	74.45	11.00	0.000
Between designs	22.53	7.00	0.002

Table 4.3: Test of heterogeneity / inconsistency.

Figure 4.1 shows the network plot of the Senn et al. (2013) data. Also this figure shows a colored triangular shape. This colored part denotes the three-arm trial. From the Senn diabetes drug dataset, we now conduct test for heterogeneity and provide the value of test statistics in the corresponding p-value in Table 4.3. We can see that $Q = 96.99$ with a $d.f = 18$ and a corresponding $p - value = 0.0001$ which is less than 0.05 if we are taking a 5% level of significance. Hence we reject the notion of homogeneity and can say there is some level of heterogeneity. This Q is the total heterogeneity which can be decomposed into heterogeneity due to "within designs" and "between designs". With the "within designs" it assesses the heterogeneity between studies with the same design and the "between designs" assesses the design inconsistency. Also this is supported by the Higgins' index, $I^2 = 81.4\%$ showing a very high level of heterogeneity in this analysis with $\tau^2 = 0.1087$. This τ^2 value suggests that heterogeneity is present since its value was not zero. Now we look at studies of the same design, thus studies with the same treatment comparison and the same number of treatment arm-comparison. Hence this is the breakdown of the "Within design" to see the contribution of each design to the heterogeneity. This is shown in Table 4.4 on page 62. In the table we see that some treatment comparisons have 0.00 degrees of freedom. Suggesting that these treatment comparisons were compared in just one study. Hence it is obvious that they will have no heterogeneity

	design	Q	df	pval
1	acar:plac	0.00	0.00	NA
2	acar:sulf	0.00	0.00	NA
3	benf:plac	4.38	1.00	0.04
4	metf:piog	0.00	0.00	NA
5	metf:plac	42.16	2.00	0.00
6	metf:rosi	0.19	1.00	0.67
7	metf:sulf	0.00	0.00	NA
8	migl:plac	6.45	2.00	0.04
9	piog:plac	0.00	0.00	NA
10	piog:rosi	0.00	0.00	NA
11	plac:rosi	21.27	5.00	0.00
12	plac:sita	0.00	0.00	NA
13	plac:vild	0.00	0.00	NA
14	rosi:sulf	0.00	0.00	NA
15	acar:metf:plac	0.00	0.00	NA

Table 4.4: Heterogeneity analysis within design.

in themselves suggesting $Q = 0.00$ and the corresponding p-value will be unavailable. For design "metf:plac" we have a heterogeneity quantity of $Q = 42.16$ and degrees of freedom of 2 suggesting that there were 3 studies that implemented this design with a p-value of 0.00. Hence there is heterogeneity present in this design. But with design "metf:rosi", $Q = 0.19$ which is small with a p-value of 0.67 suggesting heterogeneity is not present in this design. Therefore all designs with $Q \neq 0$ have some heterogeneity in them with the exception of design "metf:rosi", since it has a large p-value suggesting homogeneity between the two studies that analysed this treatment combination.

4.2 Example 2: Network Meta-Analysis Using Bayesian Approach.

Here we give a numerical output from Winbugs, firstly using the fixed effects model, where we use the Anaesthetic drug agent example (Greco et al., 2013). With the Anaesthetic drug dataset, which is about the beneficial impact of volatile agents on a 30-day mortality of patients. Here there are four treatments for the comparison labeled as desflurane (A), isoflurane (B), sevoflurane (C) and total i.v. anaesthesia (D) and made up of 30 trials or studies. Using Markov Chains Monte Carlo (MCMC) we calculate an estimate for the effects of interest. Thus d_{AB}, d_{AC} and d_{AD} for the basic parameters and d_{BC}, d_{BD} and d_{CD} for the functional parameters obtained by the consistency equation explained earlier. In Table 4.5, we provide the output from the MCMC process from Winbugs. Here $d_{AB} = d[2]$, $d_{AC} = d[3]$ and $d_{AD} = d[4]$ corresponds to the estimate of the basic parameters. Now these estimates are used to estimate the functional parameters using the consistency equation, thus $d_{BC} = d[3] - d[2]$, $d_{BD} = d[4] - d[2]$ and $d_{CD} = d[4] - d[3]$. Now using the logit function from above, the various odds ratios can be computed as follows and the final results are shown in table 4.7

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
d[2]	-0.7743	0.6802	0.005549	-2.142	-0.7681	0.5405	20001	60000
d[3]	-0.9500	0.5963	0.004083	-2.162	-0.9343	0.1781	20001	60000
d[4]	-0.8658	0.5909	0.004556	-2.067	-0.8526	0.2525	20001	60000

Table 4.5: Estimates of the basic parameters from the Anaesthetic drug dataset using the fixed effects model.

$$
\begin{aligned}
logit[p_{AB}] &= logit[p_{AB}] - logit[p_{AA}] \\
&= \mu_i + d_{AB} - (\mu_i + d_{AA}) \\
&= \mu_i - 0.7743 - (\mu_i + 0) \\
&= \mu_i - 0.7743 - \mu_i \\
&= -0.7743
\end{aligned}
$$

$$
logit[p_{AB}] = log\left(\frac{p_{AB}}{1 - p_{AB}}\right) = -0.7743
$$

That is $\quad \dfrac{p_{AB}}{1 - p_{AB}} = e^{-0.7743}$

That is $\quad \dfrac{p_{AB}}{1 - p_{AB}} = 0.46.$

Hence the odds ratio for treatment comparison AB is 0.46 and the other basic parameters odds ratios are computed as above. With the functional parameters we can estimate the treatment comparison BC as

$$
\begin{aligned}
logit[p_{BC}] &= logit[p_{AC}] - logit[p_{AB}] \\
&= \mu_i - 0.95 - (\mu_i - 0.7743) = \mu_i - 0.95 - \mu_i + 0.7743 \\
&= -0.95 + 0.7743 = -0.1757
\end{aligned}
$$

Thus, $\quad log\left(\dfrac{p_{BC}}{1 - p_{BC}}\right) = -0.1757.$

Therefore the odds ratio of BC is

$$
\frac{p_{BC}}{1 - p_{BC}} = e^{-0.1757}
$$

$$
\frac{p_{BC}}{1 - p_{BC}} = 0.84.
$$

Hence, the other treatment comparisons BD and CD are computed as above. Also the 95% credible intervals are computed as follows. From Table 4.5, the column for the standard deviation, we compute the credible intervals for the basic parameters AB, AC and AD by the following approach. For AB we have

$$e^{lnOR_{AB} \pm (1.96 * SE(lnOR_{AB}))}.$$

Here $lnOR_{AB} = d[2]$, where $d[2]$ is in table 4.5. Hence we have,

$$e^{-0.7743 \pm (1.96 * 0.6802)}.$$

So the credible interval for d_{AB} is $(0.12, 1.75)$. The remaining basic parameters are computed like above. For the functional parameters, we have to make a slight adjustment to the standard error because of the combination of two basic parameters to get the functional parameter. From elementary statistics we know that if two events are independent then their variance is $Var(A + B) = Var(A) + Var(B)$. We know that

$$lnOR_{BC} = lnOR_{AC} - lnOR_{AB},$$

by the consistency assumption. Hence standard error for $lnOR_{BC}$ is

$$SE(lnOR_{BC}) = \sqrt{SE(lnOR_{AC})^2 + SE(lnOR_{AB})^2}.$$

Therefore we can calculate its credible interval as

$$e^{lnOR_{BC} \pm (1.96 * SE(lnOR_{BC}))}$$

$$e^{lnOR_{BC} \pm (1.96 * \sqrt{SE(lnOR_{AC})^2 + SE(lnOR_{AB})^2})}$$

$$e^{(-0.95 + 0.7743) \pm (1.96 * \sqrt{0.5963^2 + 0.6802^2})}$$

$$e^{-0.1757 \pm 1.7730}.$$

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
d[2]	-0.9783	0.8977	0.0224	-2.8970	-0.9052	0.5200	30001	90000
d[3]	-1.0880	0.7712	0.0181	-2.7690	-1.0150	0.1839	30001	90000
d[4]	-0.9040	0.6958	0.0170	-2.2390	-0.8961	0.4239	30001	90000
τ^2	118.30	321.30	7.9820	0.1646	9.2460	997.70	30001	90000

Table 4.6: Estimates of the basic parameters from the Anaesthetic drug dataset using the random effects model

The other functional parameter credible intervals are calculated as above. Also, we perform the analysis using the random effects model, the annotated code for the analysis of the fixed effects and random effects model is made available in Appendix *A*. After using Winbugs to run the MCMC, we had estimates of the basic parameters after a burn-in of 30001 and iteration of 90000 as given in Table 4.6. We then calculate the odds ratio's using the same approach as above and calculating the functional parameters as we did in the fixed effects model. However, we use different notations in random effects model. This is given by $\delta_{AB} = d[2]$, $\delta_{AC} = d[3]$ and $\delta_{AD} = d[4]$ for the basic parameters. Table 4.7 shows the odds ratios and 95% credible intervals of the the treatment comparison under the fixed effects model and random effects model. From Table 4.5 of the fixed effects model we realize the credible intervals of the basic parameters are relatively smaller as compared to Table 4.6 of the random effects model. This is because the random effects model quantifies the heterogeneity that might exist across studies. This consequently makes the credible intervals of the odd ratios of the random effects model to be relatively larger as compared to the fixed effects model in Table 4.7.

Also, we look at another dataset called the Thrombolytic drug dataset (Lu and Ades, 2006), which is made up of 8 drugs, reteplase (Ret-6), streptokinase

Fixed Effects Model		
	Odds ratio	95% credible interval
d_{AB}	0.46	0.12-1.75
d_{AC}	0.39	0.12-1.24
d_{AD}	0.42	0.13-1.34
d_{BC}	0.84	0.14-4.94
d_{BD}	0.91	0.16-5.34
d_{CD}	1.09	0.21-5.56
Random Effects Model		
	Odds ratio	95% credible interval
d_{AB}	0.38	0.06-2.18
d_{AC}	0.34	0.07-1.53
d_{AD}	0.40	0.10-1.58
d_{BC}	0.90	0.09-2.21
d_{BD}	1.08	0.12-9.98
d_{CD}	1.20	0.16-9.21

Table 4.7: The odds ratios and 95% credible interval of the treatment comparison under the fixed effects and random effects consistency models.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
d[2]	-0.1651	0.0440	6.66E-4	-0.2515	-0.1648	-0.0804	20001	30000
d[3]	0.0017	0.0304	0.0181	3.365E-4	0.0016	0.0612	20001	30000
d[4]	-0.0459	0.0463	4.61E-4	-0.1374	-0.0458	0.0445	20001	30000
d[5]	-0.1606	0.0767	0.0011	-0.3095	-0.1610	-0.0108	20001	30000
d[6]	-0.1152	0.0599	8.201E-4	-0.2314	-0.1151	0.0027	20001	30000
d[7]	-0.1952	0.2181	0.0021	-0.6205	-0.1955	0.2304	20001	30000
d[8]	0.0153	0.0367	3.779E-4	-0.057	0.0153	0.0875	20001	30000

Table 4.8: Estimates of the basic parameters from the Thrombolytic drug dataset using the fixed effects consistency model

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
d[2]	-0.2094	0.1130	0.003001	-0.4963	-0.1915	-0.0408	20001	30000
d[3]	-0.0099	0.0811	0.001739	-0.2017	-0.0033	0.1299	20001	30000
d[4]	-0.0543	0.1204	0.001539	-0.3066	-0.0524	0.1847	20001	30000
d[5]	-0.2036	0.1770	0.003175	-0.6238	-0.1878	0.0955	20001	30000
d[6]	-0.1585	0.1260	0.002795	-0.4690	-0.1438	0.0407	20001	30000
d[7]	-0.2344	0.2391	0.011140	-0.7105	-0.2303	0.2391	20001	30000
d[8]	0.0408	0.0954	0.001946	-0.1381	0.0347	0.2564	20001	30000
τ^2	376.9	515.2	16.27	10.61	192.0	1804.0	20001	30000

Table 4.9: Estimates of the basic parameters from the Thrombolytic drug dataset using the random effects consistency model

(SK-1), urokinase (UK-7), alteplase (tPA-3), anistreptilase (ASPAC-8), accelerated alteplase (AtPA-2), tenecteplase (Ten-5), and streptokinase plus alteplase (SK + tPA-4) and carried out in twenty-eight trials. We calculate its basic parameters and the corresponding odds ratios as we did in the previous example. In Table 4.8, we show the basic parameters estimates from the fixed effects consistency model. Table 4.9 shows the basic parameter estimates from the random effects consistency model. We obtain the odds ratio's from both models in Table 4.10

Fixed Effects Model					
	Odds ratio	95% credible interval		Odds ratio	95% credible interval
d_{12}	0.85	0.78–0.92	d_{35}	0.85	0.72–1.00
d_{13}	1.00	0.94–1.06	d_{36}	0.89	0.78–1.01
d_{14}	0.96	0.87–1.05	d_{37}	0.82	0.53–1.26
d_{15}	0.85	0.73–0.99	d_{38}	1.01	0.92–1.11
d_{16}	0.89	0.79–1.00	d_{45}	1.03	0.75–1.06
d_{17}	0.82	0.54–1.26	d_{46}	0.93	0.80–1.08
d_{18}	1.02	0.94–1.09	d_{47}	0.86	0.56–1.33
d_{23}	1.18	1.06–1.31	d_{48}	1.06	0.95–1.19
d_{24}	1.12	0.99–1.28	d_{56}	1.04	0.86–1.27
d_{25}	1.00	0.84–1.19	d_{57}	0.97	0.61–1.52
d_{26}	1.05	0.91–1.22	d_{58}	1.19	1.01–1.41
d_{27}	0.97	0.63–1.50	d_{67}	0.92	0.59–1.45
d_{28}	1.20	1.07–1.34	d_{68}	1.14	0.96–1.35
d_{34}	0.95	0.86–1.06	d_{78}	1.23	0.80–1.90

Table 4.10: The odds ratios and 95% credible interval of the treatment comparison under the fixed effects consistency models.

Random Effects Model					
	Odds ratio	95% credible interval		Odds ratio	95% credible interval
d_{12}	0.81	0.65–1.01	d_{35}	0.82	0.56–1.21
d_{13}	0.99	0.84–1.16	d_{36}	0.86	0.64–1.16
d_{14}	0.95	0.75–1.20	d_{37}	0.80	0.49–1.31
d_{15}	0.82	0.58–1.15	d_{38}	1.05	0.82–1.34
d_{16}	0.85	0.67–1.09	d_{45}	0.86	0.57–1.31
d_{17}	0.79	0.50–1.26	d_{46}	0.90	0.64–1.27
d_{18}	1.04	0.86–1.26	d_{47}	0.84	0.49–1.41
d_{23}	1.22	0.93–1.60	d_{48}	1.10	0.81–1.49
d_{24}	1.17	0.84–1.61	d_{56}	1.05	0.68–1.60
d_{25}	1.01	0.67–1.52	d_{57}	0.97	0.54–1.74
d_{26}	1.05	0.76–1.47	d_{58}	1.28	0.86–1.89
d_{27}	0.98	0.58–1.64	d_{67}	0.93	0.55–1.57
d_{28}	1.28	0.96–1.72	d_{68}	1.22	0.90–1.66
d_{34}	0.96	0.72–1.27	d_{78}	1.31	0.80–2.18

Table 4.11: The odds ratios and 95% credible interval of the treatment comparison under the random effects consistency models.

4.3 Example 3: Model Selection and Ranking of Treatments.

Now from the Anaesthetic drug agent example, we see the DIC output from this dataset, as given in Table 4.12 for the fixed effects and random effects models. We see that the DIC for the fixed effects and the random effects is close, with an absolute difference of 0.086. Hence, by the principle of parsimony, we settle for the fixed effects model since it's much more conservative. Also with the Thrombolytic drug dataset, we compute the DIC for both the fixed effects and random effects models and decide which model to choose.

Fixed effects			
Dhat	Dbar	pD	DIC
71.690	88.516	16.826	105.342
Random effects			
Dhat	Dbar	pD	DIC
68.024	86.726	18.702	105.428

Table 4.12: DIC for the Fixed effects and Random effects model for the Anaesthetic dataset.

Fixed effects			
Dhat	Dbar	pD	DIC
300.53	335.65	35.12	370.76
Random effects			
Dhat	Dbar	pD	DIC
295.95	333.61	37.66	371.28

Table 4.13: DIC for the Fixed effects and Random effects model for the Thrombolytic drug dataset.

We see from Table 4.13 that the DIC results for the Thrombolytic drug dataset are also close, with an absolute difference of 0.520. Hence, by the principle of parsimony, we settle for the fixed effects model since it is much more conservative in the sense that it makes fewer assumptions and few underlying parameters.

Example of Using the Posterior Distribution of the Sum of Residual Deviance.

From the Anaesthetic agent example, where we have 4 treatments, we performed network meta-analysis using the fixed effects and random effects model and settled

Node	Probability	Start	Sample
p	0.09972	20001	60000
probability[1]	0.5892	20001	60000
probability[2]	0.5021	20001	60000
probability[3]	0.414	20001	60000
probability[4]	0.3282	20001	60000
probability[5]	0.2509	20001	60000
probability[6]	0.1853	20001	60000
probability[7]	0.1316	20001	60000

Table 4.14: Output for Residual Deviance Difference and Some Probabilities of β.

to use the fixed effects model. Here we try to find out if modeling the fixed effects model incorporating inconsistency is supported by the data. Let model 1 denote the fixed effects model assuming the network is already consistent. Let model 2 denote the fixed effects model assuming inconsistency. From the WinBugs program we performed MCMC simulations. With the fixed effect model assuming the network is already consistent, it estimated 3 parameters and with the fixed effects model assuming inconsistency, it estimated 6 parameters. Since we know the β is approximately Chisquare with degrees of freedom being the difference between the parameters that will be estimated from the models, we have $d.f. = 6 - 3 = 3$. Thus we have $\beta = \chi^2_3 = 7.81$. The residual deviance difference was obtained from the MCMC simulations, $\bar{D}_{1,2} = \text{diff} = 1.81$. Hence from Table 4.14 we had $P[\bar{D}_{1,2} > \beta(\bar{D})] = p = 0.09972$ which is quite small. Hence having such small probability we have little evidence in favour of model 2. As mentioned earlier, the other way to help in the model selection is to use the Kass and Raftery table by finding which value of β that will enable $P[\bar{D}_{1,2} > \beta(\bar{D})] = 0.5$. From the above

table we realize the value of β that will allow the above probability to be equal to 0.5 is 2. Hence the value 2 in the Kass and Raftery table corresponds to "Not worth more than a bare mention" evidence in favour of model 2. Thus it does not support model 2, which is the fixed effects model assuming one is accounting for inconsistency.

Also we look at the Thrombolytic drug dataset, where we have 8 treatments. We performed network meta-analysis using the fixed effects and random effects model and also settled on the fixed effects model. Here we try to find out if modeling the fixed effects model incorporating inconsistency and comparing it to the fixed effects consistency model is supported by the data. Again, let model 1 denote the fixed effects model assuming the network is already consistent. Fixed effects model assuming one is accounting for inconsistency, is denoted by model 2. From the WinBugs program we performed MCMC simulations based on the data. The fixed effects model assuming the network is already consistent, estimated 7 parameters and the fixed effects model accounting for inconsistency, estimated 28 parameters. We set the $\beta = \chi^2_{21} = 32.67$ since its $d.f = 28 - 7 = 21$. In Table 4.15, the residual deviance difference was obtained from the MCMC simulations, $\bar{D}_{1,2} = \text{diff} = 6.32$, Hence we had $P[\bar{D}_{1,2} > \beta(\bar{D})] = p = 0.6664$ which is quite large. Hence having such large probability we have sufficient evidence in favour of the fixed efffects inconsistency model. Another method we can use to support our conclusion is to use use the Kass and Raftery table by finding which value of β will ensure that $P[\bar{D}_{1,2} > \beta(\bar{D})] = 0.5$. From Table 4.15 we realize the value of β that will allow the above probability to be equal to 0.5 is 6. The value 6 in the Kass and Raftery table corresponds to "a strong"

Node	Probability	Start	Sample
p	6.667E-1	20001	60000
probability[1]	0.7684	20001	60000
probability[2]	0.7242	20001	60000
probability[3]	0.6759	20001	60000
probability[4]	0.6245	20001	60000
probability[5]	0.5712	20001	60000
probability[6]	0.5160	20001	60000
probability[7]	0.4597	20001	60000
probability[8]	0.4047	20001	60000
probability[9]	0.3527	20001	60000
probability[10]	0.3026	20001	60000

Table 4.15: Output for Residual Deviance Difference and Some Probabilities.

evidence in favour of model 2. Thus, here also, it does show support for model 2, suggesting inconsistency is present. With inconsistency being present, it will be good to determine the source of the inconsistency. Hence some diagnostic techniques must be performed on the network to know if this presence of inconsistency will affect the conclusion one will come to with this analysis and determine the source and whether it is just an influential or an outlier observation.

Ranking of Treatments Using the SUCRA Index.

Here we look at a table of the SUCRA generated as output from WINBUGS in Table 4.16 for the Anaesthetic drug dataset, where it shows the cummulated probabilities of the ranks of each treatment. Here we see that treatment C has the highest SUCRA index of 69.7% suggesting it was the best treatment, followed by treatments D, B and A being the least effective. To get the SUCRA, we first have

	Treatment j			
Rank z	A	B	C	D
1 (best)	0.005 (0.005)	0.266 (0.266)	0.418 (0.418)	0.311 (0.311)
2	0.035 (0.040)	0.304 (0.570)	0.295 (0.713)	0.366 (0.677)
3	0.158 (0.198)	0.318 (0.888)	0.248 (0.961)	0.276 (0.953)
4 (worst)	0.802 (1.000)	0.111 (1.000)	0.039 (1.000)	0.048 (1.000)
SUCRA(%)	8.100	57.500	69.700	64.700

Table 4.16: Ranking of Treatment A, B, C and D using the SUCRA Index.

to calculate the cummlative probabilities to a certain rank which is shown in the brackets above. They are then summed and divided by the number of treatments less 1. Hence, for treatment A, we have $(0.005 + 0.04 + 0.198)/3 = 0.081$. The rest follows similarly. Also, here we see the rankograms of the treatments after applying the SUCRA formular to the ranking probabilities in Table 4.16. The pictorial view of the SUCRA index is shown in Figure 4.2 for the anaesthetic drug dataset.

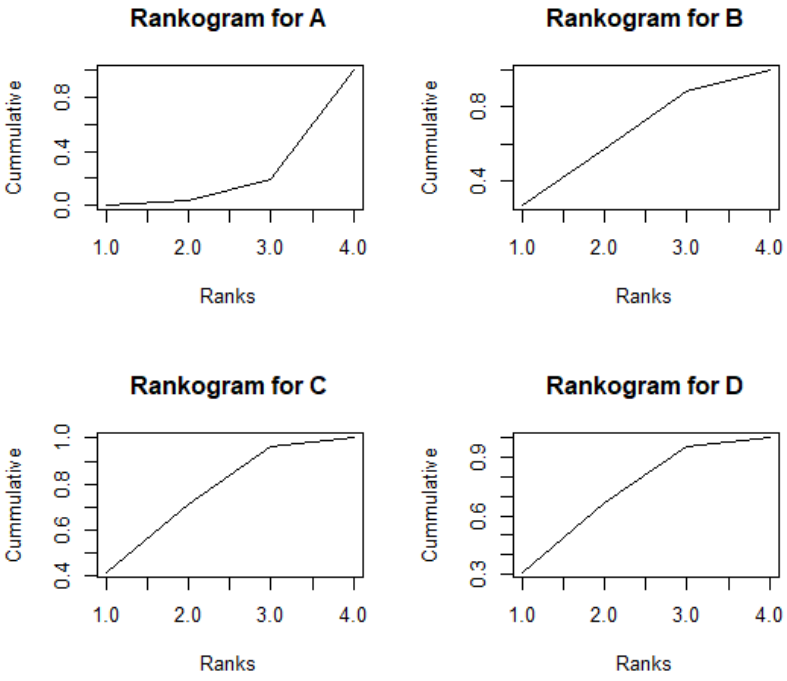

Figure 4.2: The Rankogram diagrams of the treatments

4.4 Example 4: Inconsistency Detection in Thrombolytic drug dataset and Simulated data

This dataset seeks to compare eight thrombolytic drugs for use after acute myocardial infarction with the primary outcome being 30 to 35 day mortality. There were twenty-eight trials which were conducted to study eight drugs: reteplase (Ret), streptokinase (SK), urokinase (UK), alteplase (tPA), anistreptilase (ASPAC), accelerated alteplase (AtPA), tenecteplase (Ten), and streptokinase plus alteplase (SK + tPA). There is only one three-arm comparison in trial 2 and the rest of the 27 trials are all two-arm comparisons. This dataset is used as a motivating example because from the existing literature (Lu and Ades, 2006) the authors detected the presence of inconsistency in this dataset using their novel contrast based approach. They found out that trials 22 and 23 of treatments 2 are the potential sources of inconsistencies. Here we will look at how the arm based approach will be used to identify these sources of inconsistencies. The first step to investigate the presence of inconsistency using the fixed effects in the arm based model. Here there is the need to identify the two treatments in which one suspects the inconsistency in the analysis. By doing so, one must group the treatments into the four groups discussed above. In Lu and Ades (2006), the trials introducing inconsistency are made up of a comparison of treatment AtPA(2) and ASPAC(8) using the contrast based model. Hence, with the arm based model, we have the discrepancy factor given by

$$\Delta_{28} = (\mu_2^{(i)} - \mu_8^{(i)}) - (\mu_2^{(ii)} - \mu_8^{(iii)}),$$

where $\mu_2^{(i)}$ and $\mu_8^{(i)}$ are the trials that compare both treatment 2 and 8 whilst $\mu_2^{(ii)}$

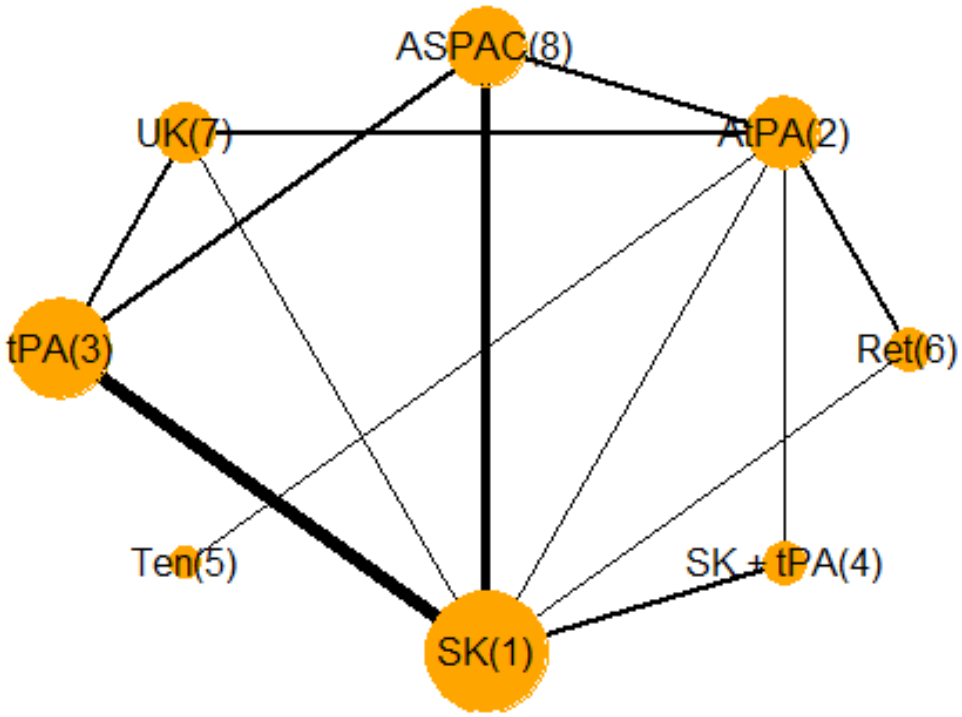

Figure 4.3: The Network graph of the Thrombolytic drug dataset where the drugs are reteplase (Ret), streptokinase (SK), urokinase (UK), alteplase (tPA), anistreptilase (ASPAC), accelerated alteplase (AtPA), tenecteplase (Ten) and streptokinase plus alteplase (SK + tPA).

are trials that compare treatment 2 with other treatments except treatment 8. Finally $\mu_8^{(iii)}$ are trials that compare treatment 8 with other treatment except treatment 2. Hence from the statistical software R, we call OpenBugs using the BRugs package which is used to run Markov Chain Monte Carlo (MCMC) in Bayesian analysis. We obtain estimate for the discrepancy factor comprising of comparing the direct and indirect evidences of treatment 2 and 8 with their corresponding 95% Bayesian credible interval (BCI). From the BCI we realize a deviation from zero (the value zero does not fall in the BCI) suggesting a discrepancy between treatment 2 and 8, hence indicating the presence of inconsistency. This can be seen in Table 4.17. Now that we have identified the presence of inconsistency, we try to investigate the source of the inconsistency to know which trial is causing it. We use the random effects approach of the arm based to investigate this. Here we shall use the yardstick proposed in (Zhao et al., 2016), where the authors suggested that we assess the top ranked 5% of the absolute values of the random effects. From this dataset with 28 trials and 8 treatments, we will have 224 (28×8) random effects. Hence the top ranked 5% of the absolute values of the random effects will be 11 random top ranked random effects. This shown in Table 4.18. From Table 4.18, we realize trial 2 with treatment 3, 8 and 1 have large random effects corresponding to 0.5271, 0.4562 and 0.4035 respectively. Also, we have trials 6 and 22 with their corresponding random effects of treatment 3 and 2 to be 0.3676 and -0.2846 respectively and so on. Now identifying these sources of inconsistencies, it will be prudent to use the fixed effects approach of the arm based model to investigate which treatment combinations give discrepancy factors that have zero further away at the tail of the posterior distribution. From

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
dis_28	-1.48	0.61	0.02	-2.74	-1.46	-0.34	30001	20000

Table 4.17: The discrepancy factor for treatment 2 and 8 from the Thrombolytic drug dataset.

$\eta[i,k]$	$\eta[2,3]$	$\eta[2,8]$	$\eta[2,1]$	$\eta[6,3]$	$\eta[22,2]$	$\eta[11,1]$	$\eta[20,7]$	$\eta[23,2]$
Posterior mean	0.5271	0.4562	0.4035	0.3676	-0.2846	0.2713	0.2586	-0.2427
	$\eta[18,2]$	$\eta[11,6]$	$\eta[19,6]$					
	0.2231	0.2206	-0.2089					

Table 4.18: The top ranked 5% of the absolute values of the random effects of study i with treatment k.

the above discrepancy factor, we realized that treatment 2 and 8 were introducing inconsistencies in the studies. Hence potential sources are from trials 2, 22, 23 and 18. The paper (Lu and Ades, 2006), shows that trials 22 and 23 are potential sources of inconsistencies which are comprised of treatments 2 and 8 using the contrast based model. Therefore the arm-based method for inconsistency detection did a good job in identifying these sources of inconsistencies.

4.4.1 Simulation Setting of Inconsistency Detection.

Here we develop a simulation setting to show how the arm-based model can detect inconsistency as well. We use the contrast based model to generate a dataset where we model it by using an inconsistency factor labelled as w. We will define a set of equations considering four treatments with direct estimates to be d_{12}, d_{13} and d_{14}. Using the consistency assumption with inconsistency factors, we define the indirect

estimate as below:

$$d_{23} = d_{13} - d_{12} + w_{123}$$

$$d_{24} = d_{14} - d_{12} + w_{124}$$

$$d_{34} = d_{14} - d_{13} + w_{134}.$$

With this kind of set up, let us introduce some inconsistency by assigning $w_{123} = w_{124} = 0.01$ and $w_{134} = 2.5$ with $d_{12} = 0.5$, $d_{13} = 0.9$ and $d_{14} = 1.2$. Hence indirect estimates using the above set of equations will be $d_{23} = 0.41$, $d_{24} = 0.71$, and $d_{34} = 2.8$. Also all the number of observations in study i using treatment k were all assigned 100, thus $n_{ik} = 100$. Now we set up 30 trials where α_{iB} with $i = 1, ..., 30$ will be assigned values from -2 to -3 of equal intervals. With four treatments, we will have 6 treatment combinations where each combination will be assigned 5 studies. Thus we will have treatment combination $1 - 2$ to have trials $1 - 5$, treatment combination $1 - 3$ to have trials $6 - 10$, treatment combination $1 - 4$ to have trials $11 - 15$, treatment combination $2 - 3$ to have trials $16 - 20$, treatment combination $2 - 4$ to have trials $21 - 25$ and treatment combination $3 - 4$ to have trials $26 - 30$. With the inconsistency factor of $w_{134} = 2.5$, treatment loop of 134 will have some level of inconsistency and in particular between treatment 3 and 4. Simulated data is then generated with this set up using the contrast based model and the binomial likelihood where inconsistency checks are done using the arm-based model. First we tend to check the discrepancy factor between treatment 3 and 4 using the fixed effects and realise it is -2.27 with a 95% BCI of $(-2.723, -1.817)$. This provides evidence that indeed inconsistency does exist in this study. This is shown in Table 4.19. Hence, in the second step, we try to find out where the sources

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
dis_34	-2.27	0.2312	0.002137	-2.723	-2.2690	-1.817	50001	20000

Table 4.19: The discrepancy factor for treatment 3 and 4 from the Simulated dataset.

$\eta[i,k]$	$\eta[27,4]$	$\eta[30,4]$	$\eta[26,4]$	$\eta[28,4]$	$\eta[29,4]$	$\eta[26,3]$
Posterior mean	0.92080	0.9074	0.8985	0.8514	0.8396	-0.5574

Table 4.20: The top ranked 5% of the absolute values of the random effects of study i with treatment k in the Simulated dataset.

of inconsistencies might be occuring most. We then use the random effects arm-based model to investigate this as shown above. Here we will see the top ranked 5% of the absolute values of the random effects where we will have 120 (30×4) random effects. Thus, we will have 6 top random effects. These 6 consisted of trials $27, 30, 26, 28$ and 29 comprising of treatment 4 and the sixth one being trial 26 of treatment 3. From the setup we realised the loop of treatment combination 134 of the inconsistency was assigned trials $26 - 30$, and the random effects arm-based model was able to detect these, suggesting this technique of inconsistency detection is effective. This is shown in Table 4.20. Figure 4.4 shows the evidence network of the simulation study for the analysis showing the four treatments.

4.5 Example 5: Diagnostics of Network Meta-Analysis after Inconsistency Detection.

In the previous example, we used the Thrombolytic drug dataset to perform some network meta-analysis diagnostics. We were able to determine or detect some level of inconsistencies present in the network. Here, we try to perform some diagnostics

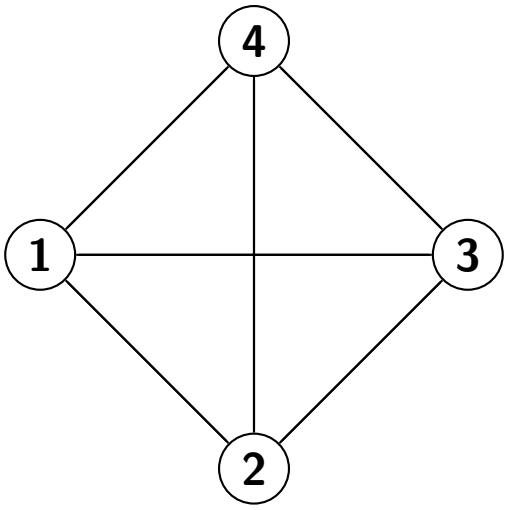

Figure 4.4: The simulation evidence network of the four treatments.

to this network to assess the impact of this inconsistencies identified. We know the network meta-analysis used is a classic example of Bayesian hierarchical model. Hence we shall try two methods for diagnostics of the network. These methods are deleting the suspected sources of inconsistencies and re-running the MCMC. The first one is called the exact-approach for the diagnostics and the other method is the linear approximation method by converting the Bayesian hierarchical model to a linear model. Hence we first look at the exact method approach using the relative change (RC) as the yardstick to determine if a source of inconsistency is influential or not by checking if the absolute value of the RC is greater than or equal to 2, that is $|RC| \geq 2$. Table 4.21 shows the exact effects from the full dataset of the Thrombolytic dataset after running the MCMC. Now from the previous example, we realized trials 22 and 23 introduced some level of inconsistency in the network, hence we delete each trial and assess if it is influential or not using the RC formula,

in which we have the relative change as given by

$$RC\{\theta; r\} = \frac{\hat{\theta}_{(r)} - \hat{\theta}}{psd\{\theta|Y\}},$$

with $\hat{\theta}_{(r)}$ being the deleted rth case from the dataset, $\hat{\theta}$ the estimate of the full dataset and finally the denominator being the posterior standard deviation from the full dataset. Also after deleting trial 22 (since it was suspected to be a source of inconsistency), we re-run the MCMC cycle and get estimates from the reduced dataset. This is shown in Table 4.22. Hence we can calculate the RC for treatments 2 and 8 since they make up trial 22, which is shown below. Trial 23 (which was a source of inconsistency) was also tested to see if its influential or not by calculating the RC. It was calculated like trial 22 below. Hence the RC for treatment 2 was, $RC\{\mu_2\} = 0.3984$ and, for treatment 8, it was $RC\{\mu_8\} = -0.3149$ for trial 23. The absolute value for the RC's for trials 22 and 23 of treatment 2 and 8 suggest they are not influential (since they are all less than 2), although they are sources of inconsistencies.

$$RC\{\mu_2\} = \frac{-2.722 - (-2.788)}{0.1230} = 0.5366,$$

$$RC\{\mu_8\} = \frac{-2.642 - (-2.598)}{0.1588} = -0.2771.$$

Also, looking at the Thrombolytic drug dataset, we realized treatment 5 was only introduced in trial 17, so we decided to check if this treatment was influential by deleting trial 17 and re-running the MCMC and calculating the RC. The RC was calculated as, $RC\{\mu_5\} = 9.7788$, which is much greater than 2. Hence this trial,

i	μ_i	Sd
1	-2.535	0.0906
2	-2.788	0.1230
3	-2.674	0.1335
4	-2.625	0.2151
5	-2.735	0.2536
6	-2.535	0.1838
7	-2.988	0.1973
8	-2.598	0.1588

Table 4.21: The exact effects from the full dataset of the Thrombolytic dataset.

i	μ_i	Sd
1	-2.547	0.0925
2	-2.722	0.1143
3	-2.697	0.1324
4	-2.641	0.2304
5	-2.745	0.2542
6	-2.563	0.1947
7	-2.994	0.2067
8	-2.642	0.1739

Table 4.22: The exact effects after deleting trial 22 of the Thrombolytic dataset.

although not a source of inconsistency, but is influential. This might be the case because the treatment 5 is primarily driven by the prior of itself. We have realized that by deleting suspected sources of inconsistencies and re-running the MCMC, we are able to calculate the RC's and therefore come to a conclusion which trials are influential or not. This process is most efficient but the issue is that, if we have a very large network consisting of many trials, then deleting and re-running the MCMC will be a herculean task. Hence a linear approximation method can be useful to estimate the numerator component of the RC formular.

Here, we will look at how to use the linear approximation method where we will consider three scenarios with matrix A and B where below is the set up for the η covariance matrix. From Equation 3.8, the H matrix is formed as per how the treatments are compared in a particular trial. For the rows in H corresponding to the error term, η, the formulation of the variances of η will result in the three scenarios of the linear approximation method we will be looking at. From Equation 3.9, we multiply it with inverse square root of Γ. Hence the constituent of Γ is made up of the variances of ϵ, η and υ. The variance of ϵ is made up from the transformed data and the variances of υ will all be assigned 1000. Here we are using the arm based model and as such we can consider deletion of individual data cases. Thus, for example, if we have a two arm trial where we delete one arm, only one treatment arm is left. With this single arm left we can keep it in the arm based model since it still contributes to the likelihood function under the missing data framework (Lin et al., 2016). For scenario one, we generate once, variances from the Wishart distribution with parameters 58 as the degrees of freedom and a 58×58 matrix consisting of 0.1 in the diagonals and the off diagonals consisting of 0.005.

After a random 58×58 matrix is generated from the Wishart distribution in software R, we select the diagonal elements of this matrix to form the variances of the η with its off diagonals setting them to 0, hence using matrix A below. And the variances of the ϵ and υ are generated as above and set both their off diagonals to 0. Combining all these constituents of ϵ, η and υ forms the $\Gamma_{(1)}$ matrix of 124×124, since the ϵ is also a matrix of 58×58 and the matrix of υ is 8×8.

$$\boldsymbol{A} = \begin{bmatrix} \sigma_{11} & 0 & 0 & 0 & & & \dots & 0 \\ 0 & \sigma_{22} & 0 & 0 & & & \dots & 0 \\ 0 & 0 & \sigma_{44} & 0 & & & \dots & 0 \\ 0 & 0 & 0 & \sigma_{11} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \sigma_{33} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{88} & \dots & 0 \\ & & & \vdots & & & & \\ 0 & 0 & 0 & \dots & & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \dots & & 0 & 0 & \sigma_{88} \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{14} & 0 & & & \dots & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{24} & 0 & & & \dots & 0 \\ \sigma_{41} & \sigma_{42} & \sigma_{44} & 0 & & & \dots & 0 \\ 0 & 0 & 0 & \sigma_{11} & \sigma_{13} & \sigma_{18} & \dots & 0 \\ 0 & 0 & 0 & \sigma_{31} & \sigma_{33} & \sigma_{38} & \dots & 0 \\ 0 & 0 & 0 & \sigma_{81} & \sigma_{83} & \sigma_{88} & \dots & 0 \\ & & & \vdots & & & & \\ 0 & 0 & 0 & \dots & & 0 & \sigma_{33} & \sigma_{38} \\ 0 & 0 & 0 & \dots & & 0 & \sigma_{83} & \sigma_{88} \end{bmatrix}$$

For scenario 2, we generate 10000 random covariance matrices from a Wishart distribution with parameters 58 as the degrees of freedom and a 58×58 matrix consisting of 0.1 in the diagonals and the off diagonals consisting of 0.005. With these 10000 covariance matrices we find the mean of each entries of the matrices, then the mean covariance matrix we take the diagonal entries of these elements which

will be the variances of the η matrix. The variances of the ϵ and υ are generated as the first scenario, hence for another $\Gamma_{(2)}$ matrix of dimension 124×124.

Finally for scenario 3, we formulate the covariance matrix of η by using matrix B above, where we take into consideration the dependencies that exist between the random components in each trial. Trials 1 and 2 are 3-arm trials, with trial 1 consisting of treatments 1, 2 and 4 and trial 2 consisting of treatments 1, 3 and 8, with the rest being two arm trials. For trial 1, we generated 10000 random covariance matrices from a Wishart distribution with parameters 3 as the degrees of freedom and a 3×3 matrix consisting of 0.1 in the diagonals and the off diagonals consisting of 0.005. From these 10000 3×3 covariance matrices generated from the Wishart distribution, we find the mean of each entry and the new covariance matrix will be for trial 1. This procedure is also done for trial 2 as it is also a 3 arm trial. For the rest which are made of 2 arm trials, we generated 10000 random covariance matrices from a Wishart distribution with parameters 2 as the degrees of freedom and a 2×2 matrix consisting of 0.1 in the diagonals and the off diagonals consisting of 0.005. From these 10000 2×2 covariance matrices generated from the Wishart distribution for each of the 26 remaining trials, we find the mean of each entry and the new block of 26 covariance matrix will be for the rest of the trials. We get a new 58×58, η matrix and combine it with ϵ and υ to form another $\Gamma_{(3)}$ matrix of dimension 124×124.

Now we pre multiply Equation 3.9 by the three new gamma's developed from the above scenarios. This gives us new sets of equations for which we will use in the linear approximation method for the numerator of the RC formular. The constituents of this new set of equations will be used in the approximation of Equation 3.15

Omitted case	Trial 22,trt 2	Trial 22, trt 8	Trial 23,trt 2	Trial 23, trt 8	Trial 17, trt 5
Parameter	μ_2	μ_8	μ_2	μ_8	μ_5
Rerun MCMC	0.5366	-0.2771	0.3984	-0.3149	9.7788
Linear approximation					
Scenario 1	0.9482	-0.2011	0.7629	-0.3531	10.6800
Scenario 2	0.8773	-0.2150	0.9479	-0.2952	10.6600
Scenario 3	0.7087	-0.2116	0.5634	-0.2965	-4.0367

Table 4.23: RC's using the arm-based model for the respective treatments and corresponding trials .

as the numerator of the RC. From Table 4.23, we see the various approximations of the three scenarios give the same conclusion as the exact method of re-running the MCMC. Hence the linear approximation can be used in place of re-running the MCMC when the network is very large to reduce the herculean task involved in re-running the MCMC when a deleted trial is made. But scenario 3 should be used often since it incorporates the dependencies among the random components in each trial which is usually the case in real world application. Thus, in the linear approximation method, we shall use

$$\hat{\boldsymbol{\Theta}}_{\boldsymbol{r}} - \hat{\boldsymbol{\Theta}} \approx -\frac{(\boldsymbol{X}^{*\top}\boldsymbol{X}^{*})^{-1}\boldsymbol{x}_{\boldsymbol{r}}^{*\top}}{1 - v_r}\hat{E}_r^{*}, \tag{4.2}$$

where $\boldsymbol{X}^{*}$ is transformed using the posterior mean of $\boldsymbol{\Gamma}$ as in equation 3.9, $\hat{\boldsymbol{\Theta}}$ is the posterior mean of $\boldsymbol{\Theta}$ using the full dataset, $\hat{\boldsymbol{\Theta}}_{\boldsymbol{r}}$ is the posterior mean of Θ after deleting the r^{th} case, $\hat{E}^{*}{}_r$ is the r^{th} row of 3.11 which is a scalar. Also v_r is the r^{th} diagonal element of $\boldsymbol{V}$ and also a scalar, for the numerator of the RC formula, where

$$RC\{\theta; r\} = \frac{\hat{\theta}_{(r)} - \hat{\theta}}{psd\{\theta|Y\}}. \tag{4.3}$$

4.5.1 Outlier Detection from the Linear Approximation Model

Here we look at the treatments and their trials that may be potential outliers. We see that trials 22 and 23 of treatment 2 which were sources of inconsistency are actually outliers. These values were actually among the top ranked standardized residuals. This is given in Table 4.24

Trial	Treatment	Standardized Residuals
20	7	1.83
22	2	-1.55
23	2	-1.44
2	2	1.15
6	3	1.10
19	6	-1.06
13	1	-1.01
10	1	-1.00

Table 4.24: Outlier detection for the Thrombolytic drug dataset.

Chapter 5

Discussion and Future Work

In the clinical world, clinicians have the need to compare new interventions/treatments to existing ones and also new interventions/treatments to new intervetions/treatments to see which ones are of outmost benefit to a medical condition. Hence, with the necessary comparison being made among interventions, clinicians need to make decisive decisions keeping in mind the most cost effective approach. Therefore a statistical procedure called network meta-analysis was used with the available data of the comparison of treaments/interventions. This method has been used over the years to answer these questions asked by clinicians by performing just a single analysis.

The frequentist method had been used extensively for this method but as things got complicated, the Bayesian paradigm proves to provide a more robust approach to model such complexity and provides a suitable approach to rank interventions/treatments and even perform better in a missing data framework.

We looked at how to perform network meta-analysis using both the frequentist

approach and the bayesian approach. This incorporated a continuous and binary data framework where we used the frequentist approach to analyse the continuous data and the bayesian approach to analyse the binary data. We realized that, with the frequentist approach, ranking of the treatments to know which one has the optimal benefit was difficult to incorporate. Also, in a missing data framework, it will under-perform. Hence we needed to analyse a network meta-analysis data using the Bayesian approach. We saw how the treatments were able to be ranked and how historical data could be used to update our new results.

We further used some model selection techniques to discriminate between two models and suggested the optimum model to use in our network meta-analysis. For this purpose, we used deviance information criteria (DIC). Further, we made use of the paper (Kass and A.E., 1995) for model selection as well.

There were certain assumptions that needed to be met, such as the consistency assumption. If this assumption is not met, conclusions drawn out of such network meta-analysis will not be appropriate. We investigated to know if there is a presence of inconsistency in a network. If we found inconsistency, we then find the sources of these inconsistencies. Hence, we tried to detect the presence of inconsistencies as well as to identify the sources.

We further considered the fact that, a network meta-analysis is a classic example of linear hierarchical model. Hodges (1998) suggested that every linear hierarchical model could be approximated by a linear model. Hence, if the linear model is formulated, some linear model approaches can be implemented on this approximated

hierarchical model. This helped us in identifying influential observations and outlier observations. We investigated whether the inconsistencies were in our network, and if so we checked whether there were influential observations outliers or both. For identifying influential observations, we used a linear model theory approach called the relative change (RC) and evaluated its absolute magnitude. We look at using the standardized residuals from our model for the assessment of outlying observations.

We encountered some challenges in our analysis, we tried to establish the optimum covariance matrix structure for the error term components of the approximate linear model of the linear hierarchical model. In particular, we considered three scenarios to see which will be the optimum in our analysis. Also, acquiring datasets for our analysis was difficult to see how robust this linear approximation method is.

For future work, we would like to study if we can extend this approach to a continuous dataset and see how the implementation of a varying our prior distributions affect the robustness of the model. We hope to explore our model by using the Dirichlet prior and see how best it works with this prior distribution. This is a meta-analysis in which multiple treatments are compared in a multivariate setup, where we assume that there is a probability distribution over a measurable space. Then a Dirichlet process is a probability distribution over all the distributions of the subsets of the measurable space (Sethuraman, 1994).

We also hope to extend our methods to other types of outcomes such as time-to-event outcome. However, this will require individual-level patient data which could be incorporated with the aggregated data summaries. We would like to pursue

with details in future.

Appendices

Appendix A

Code for Anaesthetic Drug Dataset

```
#################################################################
#Author: Kamso Mohammed Mujaab
#The fixed effects model for the Anaesthetic drug
# dataset In WINBUGS
#################################################################
model{
#-----------------------------------------------------------------
for(i in 1:ns){
for(k in 1:na[i]){
#Binomial likelihood
r[i,k]~dbin(p[i,k], n[i,k])
#Model specification
#Each treatment effect estimate is defined by difference
#between arm k and arm 1 (reference group)
#The difference d[t[i,k]] - d[t[i,1]] is a fixed effect
logit(p[i,k])<-mu[i] + d[t[i,k]] - d[t[i,1]]
}
}
#-----------------------------------------------------------------
#Prior definition
#Vague prior for the reference treatment in each ns trials
#Normal distribution is specified in term
```

```
#of mean and precision
for(i in 1:ns){mu[i]~dnorm(0, 0.0001)}
#The treatment effect difference between a treatment
#and itself is setted to zero
d[1]<-0
#Vague prior for nt-1 basic parameters:
#treatment effect difference
# between treatment k and reference group
for(k in 2:nt){d[k]~dnorm(0,0.0001)}
#-----------------------------------------------------------
}

#DATA
#LOAD DATA
list(nt=4, ns=30)
t[,1] t[,2] t[,3] n[,1] n[,2] n[,3] r[,1] r[,2] r[,3] na[] #ID
1 4 NA 19 21 NA 0 0 NA 2 # study 1
1 4 NA 50 50 NA 2 1 NA 2 # study 2
1 4 NA 22 22 NA 0 0 NA 2 # study 3
1 4 NA 11 12 NA 0 0 NA 2 # study 4
1 4 NA 15 15 NA 0 0 NA 2 # study 5
1 3 4 15 15 15 1 0 0 3 # study 6
1 3 4 80 80 80 1 0 0 3 # study 7
1 4 NA 50 150 NA 0 0 NA 2 # study 8
3 4 NA 137 132 NA 2 1 NA 2 # study 9
1 2 NA 35 35 NA 0 0 NA 2 # study 10
1 2 NA 49 51 NA 0 0 NA 2 # study 11
1 3 NA 55 57 NA 1 0 NA 2 # study 12
1 4 NA 50 50 NA 0 1 NA 2 # study 13
2 4 NA 20 20 NA 0 0 NA 2 # study 14
1 2 NA 30 30 NA 0 0 NA 2 # study 15
1 2 NA 10 20 NA 0 0 NA 2 # study 16
1 2 NA 33 39 NA 0 0 NA 2 # study 17
1 3 NA 61 59 NA 2 0 NA 2 # study 18
1 2 NA 20 20 NA 1 1 NA 2 # study 19
1 3 NA 14 14 NA 0 0 NA 2 # study 20
1 2 NA 12 12 NA 0 0 NA 2 # study 21
1 3 NA 91 91 NA 0 0 NA 2 # study 22
1 4 NA 62 62 NA 0 2 NA 2 # study 23
```

```
2 4 NA 142 142 NA 4 1 NA 2 # study 24
1 2 4 120 120 120 2 0 1 3 # study 25
1 2 NA 22 23 NA 1 0 NA 2 # study 26
2 3 NA 20 21 NA 1 2 NA 2 # study 27
1 3 NA 55 52 NA 3 1 NA 2 # study 28
1 3 NA 75 75 NA 1 1 NA 2 # study 29
1 2 4 20 20 20 0 0 0 3 # study 30
END


#INITS VALUES
#chain 1
list( d=c(NA, 0, 0, 0), mu=c(0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0) )
#chain 2
list( d=c(NA, -1, 4, 2), mu=c(-2, 2, -2, 2, -2, 2, -2,
2, -2, 2,-2, 2, -2, 2, -2, 2, -2, 2, -2, 2,-2, 2, -2, 2,
 -2, 2, -2, 2,-2,2))
#chain 3
list( d=c(NA, -2, 2, 2), mu=c(-2, 3, -1, -3, 5, -2, 3,
-1, -3, 5,-2, 3, -1, -3, 5, -2, 3, -1, -3, 5, -2, 3, -1,
-3, 5, -2, 3, -1, -3, 5))

#################################################################
#Author: Kamso Mohammed Mujaab
#The random effects model for the Anaesthetic drug
# dataset In WINBUGS
#################################################################
model{
#-----------------------------------------------------------
for(i in 1:ns){
for(k in 1:na[i]){
#Binomial likelihood
r[i,k]~dbin(p[i,k], n[i,k])
#Model specification
#Each treatment effect estimate is defined
#by difference between
#arm k and arm 1 (reference group)
#The node delta[i,k] is a random effect
```

```
logit(p[i,k])<-mu[i]+delta[i,k]
}
#-------------------------------------------------------
# Adjustment for multi-arm trial
for(k in 2:na[i]){
md[i,k]<-d[t[i,k]] - d[t[i,1]] + sw[i,k]
taud[i,k]<-tau *2*(k-1)/k
w[i,k]<-(delta[i,k] - d[t[i,k]] + d[t[i,1]])
sw[i,k]<-sum(w[i,1:k-1])/(k-1)
}
}
#-------------------------------------------------------
#Prior definition
#Vague prior for the reference treatment in each ns trials
for(i in 1:ns){
w[i,1]<-0
delta[i,1]<-0
mu[i]~dnorm(0, 0.0001)
#Vague prior for the random effect node within-trial
for(k in 2:na[i]){delta[i,k]~dnorm(md[i,k],taud[i,k])}
}
#Vague prior for nt-1 basic parameters:
#treatment effect difference
#between treatment k and reference group
#The treatment effect difference between
#a treatment and itself
#is setted to zero
d[1]<-0
for(k in 2:nt) {d[k]~dnorm(0,0.0001)}
#Vague Gamma prior for random-effect precision
#(useful in case of sparse data)
tau~dgamma(0.001,0.001)
sd<-pow(tau,-0.5)
#Alternative: vague prior for random-effect
#standard deviation
#sd~dunif(0,2)
#tau<-pow(sd, -2)
}
#-------------------------------------------------------
```

```
#DATA
#copy the data from the fixed effects model

#INITS VALUES
#chain 1
list( d=c(NA, 0, 0, 0),tau=0.1, mu=c(0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0) )
#chain 2
list( d=c(NA, -1, 4, 2),tau=0.1,  mu=c(-2, 2, -2, 2, -2, 2,
-2, 2,-2, 2,-2, 2, -2, 2, -2, 2, -2, 2, -2, 2,-2, 2, -2, 2,
 -2, 2, -2, 2,-2,2))
#chain 3
list( d=c(NA, -2, 2, 2),tau=0.1, mu=c(-2, 3, -1, -3, 5, -2,
3, -1, -3, 5,-2, 3, -1, -3, 5, -2, 3, -1, -3, 5, -2, 3,
-1, -3, 5, -2, 3, -1, -3, 5))
```

Appendix B

Code for Thrombolytic Drug Dataset

```
##############################################
#Author: Kamso Mohammed Mujaab
#Finding the discrepancy between treatment
#2 and 8 in the Thrombolytic drug #dataset
##############################################
library("BRugs") # loading BRugs
## Now setting the working directory to the temporary one:
oldwd <- getwd()
setwd("C:/Users/Mujaab/Dropbox/Thesis2/final work/
finalnew/Brugs/Thrombo")
## some usual steps (like clicking in WinBUGS):
modelCheck("Thrombomodel.txt") # check model file
modelData("Thrombodata.txt") # read data file
modelData("Thrombodata2.txt") # read data file
modelCompile(numChains=1) # compile model with 1 chains
modelGenInits()
#modelInits("Thromboinits.txt",2) # read init data file
modelUpdate(30000) # burn in
samplesSet("dis_28") # dis_28 should
#be monitored for the descrepancy of treatment 2 and 8
modelUpdate(20000) # 20000 more iterations ....
samplesStats("*")# the summarized results
```

```
xtable(samplesStats("*") )#tabulated results
## some plots
samplesHistory("*", mfrow = c(4, 2)) # plot the chain,
samplesDensity("dis_28") # plot the densities,
samplesBgr("dis_28") # plot the bgr statistics, and
samplesAutoC("dis_28", 1) # plot autocorrelations of
#1st chain
## switch back to the previous working directory:
setwd(oldwd)


########################################
#Author: Kamso Mohammed Mujaab
#Finding the highest ranked random
#effects in the Thrombolytic drug #dataset
########################################
library("BRugs") # loading BRugs
## Now setting the working directory to the temporary one:
oldwd <- getwd()
setwd("C:/Users/Mujaab/Dropbox/Thesis2/
finalwork/finalnew/Brugs/ThromboRandomModel")
## some usual steps (like clicking in WinBUGS):
modelCheck("ThromboRandomModel.txt") # check model file
modelData("ThromboRandomData1.txt") # read data file
modelData("ThromboRandomData2.txt") # read data file
modelCompile(numChains=1) # compile model with 1 chains
modelGenInits()
#modelInits("ThromboRandomInits.txt") # read init data file
modelUpdate(60000) # burn in
samplesSet(c("Dbar","v")) # r should be monitored
#samplesSet("v")
modelUpdate(20000) # 1000 more iterations ....
samplesStats("*") # the summarized results
v<-as.vector(samplesStats("v")[,1])#r values
v[137:144]#v for study 18
v[145:152]#v for study 19
v[153:160]#v for study 20
v[169:176]#v for study 22
v[177:184]#v for study 23
```

```
v_excel<-matrix(v,224,1)
v_excel
v_excel[order(abs(v_excel))] #ordering the random effects
## some plots
samplesHistory("*", mfrow = c(4, 2)) # plot the chain,
samplesDensity("r") # plot the densities,
samplesBgr("r[1:6,]") # plot the bgr statistics, and
samplesAutoC("r[1:6,]", 1) # plot autocorrelations
#of 1st chain
## switch back to the previous working directory:
setwd(oldwd)
```

Appendix C

Code for Linear Approximation Method

```
#########################################################
#Author: Kamso Mohammed Mujaab
#Linear approximation method for scenario 3
#
#########################################################

#incorporating dependencies among random components

#setting seed
set.seed(5555)

#formulating the design matrix
x1<-diag(58)
x2<-matrix(0,58,8)
x3<—diag(58)
x4<- matrix(c(1,rep(0,8),
1,rep(0,9),1,rep(0,4),
1,rep(0,8),
1,rep(0,13),1,
1,rep(0,9),1,rep(0,5),
1,rep(0,9),1,rep(0,5),
1,rep(0,9),1,rep(0,5),
```

```
1,rep(0,9),1,rep(0,5),
1,rep(0,9),1,rep(0,5),
1,rep(0,9),1,rep(0,5),
1,rep(0,9),1,rep(0,5),
1,rep(0,10),1,rep(0,4),
1,rep(0,12),1,rep(0,2),
1,rep(0,14),1,
1,rep(0,14),1,
1,rep(0,14),1,
1,rep(0,14),1,
1,rep(0,13),1,
rep(0,2),1,rep(0,10),1,rep(0,4),
1,rep(0,11),1,rep(0,3),
1,rep(0,11),1,rep(0,3),
1,rep(0,12),1,rep(0,2),
1,rep(0,12),1,rep(0,2),
1,rep(0,13),1,
0,1,rep(0,13),1,
0,0,1,rep(0,11),1,rep(0,3),1,
rep(0,11),1,rep(0,3),1,
rep(0,11),1,rep(0,3),1,rep(0,12),1,
rep(0,2),1,rep(0,12),1),58,8, byrow = TRUE)
x5<-matrix(0,8,58)
x6<—diag(8)

e<-cbind(x1,x2)
d<-cbind(x3,x4)
f<-cbind(x5,x6)
dd1<-rbind(e,d)
X<-rbind(dd1,f)
dim(X)
inv<-solve(t(X)%*%X)
dim(inv)
V<-X%*%inv%*%t(X)
dim(V)

#formulating the Y data-Loading the
#thrombolytic drug dataset
r<-c(1472,652,723,1455,1418,1448,9,6,5,2,3,3,887,929,
```

```
7,4,12,7,10,5,4,6,285,270,3,2,3,6,3,2,13,11,10,7,522,
523,356,757,13,7,2,7,12,16,5,17,3,13,8,4,10,6,6,5,13,10,7,5)

n<-c(20173,10344,10328,13780,13746,13773,130,123,63,59,
65,64,10396,10372,85,86,147,143,135,135,107,109,2992,2994,
58,52,86,89,58,58,182,188,203,198,8488,8461,4921,
10138,155,169,26,54,268,350,210,211,138,147,132,66,
164,166,124,121,164,161,93,90)

#creating a dataframe for the dataset
df<-matrix(c(r,n),58,2)
df
dim(df)#dimension of the dataset

#Testing for the necessary condition before converting
#dataset of binary to a normal
y<-numeric(nrow(df))
for(i in 1:nrow(df)){
if((df[i,1]== df[i,2]) || df[i,1]==0){
y[i]<-log((df[i,1]+ 0.5)/((df[i,2]-df[i,1]) + 0.5))
}else{
y[i]<-log(df[i,1]/(df[i,2]-df[i,1]))
}
}

#creating the 124*1 Y vector
y
y<-matrix(y,58,1, byrow = T)
z<-matrix(rep(0,66),66,1)
Y<-rbind(y,z)
Y


#formulating the Gamma matrix

#transformed variance of e_ik
sigma<-numeric(nrow(df))
for(i in 1:nrow(df)){
sigma[i]<-(1/(df[i,2]*(df[i,1]/df[i,2])*
```

```
(1-df[i,1]/df[i,2])))
}
sigma
e_new<-diag(sigma)


#creating 3*3 wishart prior
rr<-rep(1.00E-01,3)
offdiag<-numeric(0)
for(i in 1:2){
offdiag<-c(offdiag,rep(5.00E-03,(3-i)))
}
offdiag
m124 <- matrix(NA, ncol = 3, nrow = 3)
m124[lower.tri(m124)] <- offdiag
m124[upper.tri(m124)] <- t(m124)[upper.tri(t(m124))]
diag(m124) <- rr
m124
dim(m124)

#covariance matrix for treatment 124
cov_trt124<-list(10000)
for(k in 1:10000){
cov_trt124Mat<-matrix(as.vector(rWishart(1,3,m124)),
3,3,byrow = TRUE)
cov_trt124[[k]]<-cov_trt124Mat
}
eta_ik124<-apply(simplify2array(cov_trt124), 1:2, mean)

#covariance matrix for treatment 128
cov_trt128<-list(10000)
for(k in 1:10000){
cov_trt128Mat<-matrix(as.vector(rWishart(1,3,m124)),
3,3, byrow = TRUE)
cov_trt128[[k]]<-cov_trt128Mat
}
eta_ik128<-apply(simplify2array(cov_trt128), 1:2, mean)

#creating 2*2 wishart prior
```

```
m52rest<-matrix(c(1.00E-01,5.00E-03,5.00E-03,1.00E-01),2,2)

#covariance matrix for treatment for remaing 52 matrices
eta_ikrest<-list(26)
for(i in 1:26){
cov_trtrest<-list(10000)
for(k in 1:10000){
cov_trtrestMat<-matrix(as.vector(rWishart(1,2,m52rest))
,2,2, byrow = TRUE)
cov_trtrest[[k]]<-cov_trtrestMat
}
eta_ikrest[[i]]<-apply(simplify2array(cov_trtrest),
1:2, mean)
}
install.packages("magic")
library("magic")
eta_restNew<-adiag(eta_ikrest[[1]],eta_ikrest[[2]],
eta_ikrest[[3]],eta_ikrest[[4]],eta_ikrest[[5]],
eta_ikrest[[6]],eta_ikrest[[7]],eta_ikrest[[8]],
eta_ikrest[[9]],eta_ikrest[[10]],eta_ikrest[[11]],
eta_ikrest[[12]],eta_ikrest[[13]],eta_ikrest[[14]],
eta_ikrest[[15]],eta_ikrest[[16]],eta_ikrest[[17]],
eta_ikrest[[18]],eta_ikrest[[19]],eta_ikrest[[20]],
eta_ikrest[[21]],eta_ikrest[[22]],eta_ikrest[[23]],
eta_ikrest[[24]],eta_ikrest[[25]],eta_ikrest[[26]])
eta_restNew
dim(eta_restNew)

#variance for last 8*8
v_new1<-diag(rep(1000,8))

etaadvance<-adiag(eta_ik124,eta_ik128,eta_restNew)

#Creating Gamma_4 matrix for scenario 4
gammaadvance<-adiag(e_new,etaadvance,v_new1)
dim(gammaadvance)

#loading library for square-root function for matrices
library("expm")
```

```
#Inverse square-root Gamma matrix
gamma_rtInvadvance<-solve(sqrtm(gammaadvance))

#Transforming the X and Y matrix with
#the Inverse square-root Gamma
Y_new<-gamma_rtInvadvance%*%Y
X_new<- gamma_rtInvadvance%*%X

#creating the Hat-matrix using the transformed X matrix
inv<-solve(t(X_new)%*%X_new)
V<-X_new%*%inv%*%t(X_new)
dim(V)

#Creating a 124*124 identity matrix
I<-diag(124)

#creating the residual vector
E<-(I - V)%*%Y_new
E
dim(E)

#for deleted trial 22 of treatment 2
xr<-matrix(X_new[103,],1,66)
xr
w<--inv%*%t(xr)
k1<-w*(E[103,1]/(1-V[103,103])) #RC=0.7087
k1[60,1]/0.123

#for deleted trial 22 of treatment 8
xr<-matrix(X_new[104,],1,66)
xr
w<--inv%*%t(xr)
k2<-w*(E[104,1]/(1-V[104,104])) #RC= -0.2116
k2[66,1]/0.1588

#for deleted trial 17 of treatment 5
xr<-matrix(X_new[94,],1,66)
xr
```

```
w<-inv%*%t(xr)
k3<-w*(E[94,1]/(1-V[94,94])) #RC= -4.0367
k3[63,1]/0.2536

#for deleted trial 23 of treatment 2
xr<-matrix(X_new[105,],1,66)
xr
w<-inv%*%t(xr)
k4<-w*(E[105,1]/(1-V[105,105])) #RC= 0.5634
k4[60,1]/0.123

#for deleted trial 23 of treatment 8
xr<-matrix(X_new[106,],1,66)
xr
w<-inv%*%t(xr)
k5<-w*(E[106,1]/(1-V[106,106])) #RC= -0.2965
k5[66,1]/0.1588
################################################
#creating a standardized residual matrix
#for outlier detection
g<-(I - V)%*%t((I - V))
dim(g)
E[98,1]/sqrt(g[98,98])#trt6 19
E[79,1]/sqrt(g[79,79])#trt1 10
E[85,1]/sqrt(g[85,85])#trt1 13
E[100,1]/sqrt(g[100,100])#trt7 20
E[105,1]/sqrt(g[105,105])#trt2 23
E[103,1]/sqrt(g[103,103])#trt2 22
E[103,1]/sqrt(g[103,103])#trt2 22
E[72,1]/sqrt(g[72,72])#trt3 6
E[63,1]/sqrt(g[63,63])#trt2 2
```

Bibliography

Ades, A. (2003). A chain of evidence with mixed comparison models for multi-parameter evidence synthesis and consistency of evidence. *Statistics in medicine 22*, 2995–3016. (Cited on page 14.)

Birge, R. T. (1932). The calculation of errors by the method of least squares. *Physical Review 40*, 207–227. (Cited on page 10.)

Borenstein, M., H. J. Hedges, L. V., and H. R. Rothstein (2010). A basic introduction to fixed-effects and random-effects models for meta-analysis. *Research Synthesis Methods 1(2)*, 97–111. (Cited on pages 3, 5 and 9.)

Borenstein, M., L. Hedges, and H. Rothstein (2015). www.meta-analysis.com. *Springer International Publishing, Switzerland*. (Cited on pages 1, 3 and 4.)

Bryk, S., W. Raudenbush, and S. Anthony (2002). Hierarchical linear models: applications and data analysis methods (2. ed., [3. dr.] ed.). *Sage Publications. ISBN 0-7619-1904-X 1(2)*, 97–111. (Cited on page 46.)

Bucher, H. C., G. H. Guyatt, L. E. Griffith, and S. D. Walter (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized

controlled trials. *Journal of clinical epidemiology 50(6)*, 683–691. (Cited on pages 11 and 12.)

DerSimonian, R. and N. Laird (1986). Meta-analysis in clinical trials. *Control Clininical Trials 7*, 177–88. (Cited on page 6.)

Dias, S., N. J. Welton, A. J. Sutton, D. M. Caldwell, G. Lu, and A. Ades (2011). Nice dsu technical support document 4: inconsistency in networks of evidence based on randomised controlled trials. *NICE Decision Support Unit: Sheffield, UK*. (Cited on page 41.)

Greco, T., G. Landon, G. Biondi-Zoccai, F. D'Ascenzo, and A. Zangrillo (2013). A bayesian network meta-analysis for binary outcome: how to do it. *Statistical Methods in Medical Research*. (Cited on pages 15 and 63.)

Higgins, J. and S. G. Thompson (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine 21)*, 1539–1558. (Cited on page 10.)

Higgins, J. and A. Whitehead (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in medicine 15-24*, 2733–2749. (Cited on pages 30 and 31.)

Hodges, J. S. (1998). Some algebra and geometry for hierarchical models applied to diagnostics. *Statistics in medicine*. (Cited on pages 49 and 92.)

Jeroen, J. P., R. Fleurence, B. Devine, R. Itzler, B. Annabel, N. Hawkins, L. Karen, C. Boersma, A. Lieven, and J. C. Cappelleri (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*. (Cited on pages 11, 14 and 15.)

Kass, R. and R. A.E. (1995). Bayes factors. *J Am Stat Assoc 90*, 773–795. (Cited on pages 39 and 92.)

Kiefer, C., S. Sibylle, and B. Ralf (2015). Indirect comparisons and network meta-analysis: Estimation of effects in the absence of head-to-head trials. *Part 22 of a Series on Evaluation of Scientific Publication*. (Cited on page 14.)

Lin, S., S. Chu, and S. Hodges (2016). Sensitivity to excluding treatments in network meta-analysis. *Epidemiology 27(4)*, 562–569. (Cited on page 86.)

Lu, G. and A. Ades (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association 101*, 447459. (Cited on pages 41, 66, 77 and 80.)

Lumley, T. (2002). Network meta-analysis or indirect treatment comparisons. *Statistics in medicine 21*, 2313–2324. (Cited on pages 4 and 11.)

Molitierno, J. J. (2012). Applications of combinatorial matrix theory to laplacian matrices of graphs. *A Chaman and Hall book*. (Cited on pages 17, 19 and 20.)

Rcker, G., G. Schwarzer, U. Krahn, and J. Knig (2018). Network meta-analysis using frequentist methods. `https://github.com/guido-s/netmeta`. (Cited on page 58.)

Salanti, G., J. P. Higgins, A. E. Ades, and J. P. Ioannidis (2008). Evaluation of networks of randomized trials. *Statistical Methods in Medical Research 17 17*, 279–301. (Cited on page 32.)

Salanti, G., J. P. Higgins, A. E. Ades, and J. P. Ioannidis (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal Clinical Epidemiology 64*, 163–171. (Cited on page 40.)

Schwarzer, G. (2015). Meta analysis with r,use r. *Journal of clinical epidemiology*. (Cited on pages 8, 9, 24 and 25.)

Senn, S., F. Gavini, D. Magrez, and A. Scheen (2013). Issues in performing a network meta-analysis. *Statistical Methods in Medical Research 22*, 169189. (Cited on pages 57, 58 and 61.)

Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica Sinica 4*, 639–650. (Cited on pages 27 and 93.)

Shuttleworth, M. (2009). Meta analysis. *Internet:Explorable.com*, 35. (Cited on page 2.)

Spielgelhalter, D., N. Best, and B. Carlina (1998). Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. *Technical Report, MRc Biostatistics Unit, Cambridge, UK*. (Cited on page 36.)

Spielgelhalter, D.J.and Abrams, K. and J. Myles (2004). Bayesian approaches to

clinical trials and health-care evaluation. *New York: Wiley*. (Cited on pages 28 and 36.)

Stegenga, H., G. Rogers, P. Kandaswamy, and J. Dylan (2008). National institute for health and clinical excellence (nice).-guide to the methods of technology appraisal. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf. (Cited on pages 11 and 12.)

Sutton, A., K. Abram, D. Jones, T. Sheldon, and F. Song (2000). Methods for meta-analysis in medical research. *London: Wiley*. (Cited on pages 1, 7 and 8.)

Whitehead, A. (2002). Meta-analysis of controlled clinical trials. *Chichester, Uk*. (Cited on pages 1 and 2.)

Zhao, H., J. Hodges, and B. Carlin (2017). Diagnostic for generalized liear hierarchical models in network meta-analysis. *Res Syn Meth. 8*, 333–342. (Cited on pages 48, 49, 52 and 55.)

Zhao, H., J. Hodges, H. Ma, Q. Jiang, and B. Carlina (2016). Hierarchical bayesian approaches for detecting inconsistency in network meta-analysis. *Statistics in medicine*. (Cited on pages 40, 42, 43 and 79.)