Constructing Appropriate Models for Meta-Analyses

Nicola Greenlaw

A Dissertation Submitted to the

University of Glasgow

for the degree of

Master of Science

Department of Statistics

September 2009

Abstract

Meta-analysis combines individual studies or trials to achieve one overall treatment effect estimate and has come a long way since first appearing within medical literature 30 years ago.

Most articles examine how best to combine the individual trials and measure the combined estimate. A lot of articles also examine the different sources of variation, between study variation and within study variation, which occur when performing a meta-analysis, and how 'best' to account for the between study variation, the heterogeneity.

Very little information however, has been published on the relationship which occurs between the treatment effect estimate and the heterogeneity. Most publications examine these two measures individually, assuming they are independent, however further examination of this relationship brings this assumption of independence into question.

We have examined the relationship of the treatment effect estimates and their corresponding heterogeneities for 125 independent meta-analyses using the frequentist approach and note that the results indicate a relationship is present.

This relationship will have a resulting effect on how one measures the treatment effect estimates and their corresponding heterogeneity and is something that is considered here, using a Bayesian approach, along with a few other Bayesian modeling approaches.

Building on these Bayesian approaches, we consider whether a hierarchical model which would allow a meta-analysis of meta-analyses can be produced.

Acknowledgements

I would like to say a huge thank you to my supervisor, Professor Stephen Senn for his time, patience, comments and suggestions throughout the past two years which are very much appreciated.

I would also like to thank the pharmaceutical company GlaxoSmithKline for their funding which allowed me to do the research for this masters.

Thank you also to Christopher Schmid, one of the authors of the journal article "Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses" who kindly supplied me with the data for this thesis.

Many thanks to the girls in the office for their advice, helpful tips and the odd distraction at the end of a long day.

Last, but not least, I would also like to thank my family, especially my mum, dad and sister who were great at listening confusedly and nodding along anyway, as well as all of my friends, Elaine, Lesley, Laura, Adele and Cat, for their valuable support (and ears) they allowed me to take advantage of during my Masters. If ever you need to let off steam you know I'm here for you!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Many clinical trials, both past and present, aim to compare a new treatment with either a placebo, or an existing treatment which is already on the market, however they can often have results which are inconclusive. This may be because too small a sample size has been used, which is perhaps due to a poor intake of patients to the trial.

Even when a trial does appear to show a significant result which favours the new treatment, it is not necessarily considered for licensing, especially in the United States of America, unless there is at least one other independent clinical trial which also displays significant results in favour of the new treatment. For example, there is more convincing evidence if a trial in a different country obtains the same result.

This reasoning strengthens the argument in favour of this new treatment over

a placebo or any existing previous treatment by studying the same question on more than one occasion. It is less than ideal however, to continue repeating unnecessary clinical trials with the same new treatment until a significant difference is notable. This is partially for ethical reasons to prevent either the trials continuing indefinitely, or failure to notice the new treatment producing a significantly negative effect on the patients' health, but mainly to prevent accepting a false positive result in favour of the new treatment, which will occur in 5% of all trials.

1.2 Meta-Analysis

A meta-analysis is a statistical method which can be used to overcome these problems by systematically combining the results from several individual trials concerning the same treatment and similar constraints and formally summarising the results. The concept of meta-analysis is first known to have been used in the early 1900's (Pearson, 1904) for the combination of studies of typhoid vaccine effectiveness, before first appearing in medical literature during the 1980's. It is a technique which is becoming more frequently and widely used, as well as being recognised as an important step within the medical world.

Combinations of various independent trials allows an average treatment effect to be produced which will consequently have a greater precision than any of the individual estimates. Using the combined average treatment effect, a single overall null hypothesis which assumes no difference between the placebo/previously

existing treatment and the new treatment can then be tested against an alternative hypothesis which declares a difference.

This can prove useful for preventing any further unnecessary trials in view of a significant result which can determine early on (using the pre-existing clinical trials) whether or not there is evidence of a significant treatment difference, be this in favour of the new treatment or against it.

An area of great debate surrounds how best to model the overall treatment effect estimate. Some statisticians believe the use of the fixed effects model which assumes the size of the treatment effect is equal for each trial is best, whereas many believe the random effects model which allows the size of the treatment effect to differ between trials should be implemented. This random effects model will allow for two sources of variation, the within study variation which is also incorporated in a fixed effects model, and a different source of variation, that which occurs between the trials, also known as the heterogeneity. If a random effects model is produced and there is in fact no heterogeneity present, the results obtained would be similar to the results obtained if using a fixed effects model.

A common problem which one should be aware of when computing a meta-analysis is a possibility of bias occurring and subsequently care must be taken when computing a meta-analysis to avoid this bias as much as possible. Publication bias can occur if the researcher considers solely trial results which have been published, since many trial results (most commonly those which indicate a null result) fail to become published. The danger in publication bias is the majority of trials which are published are generally large trials, which can demonstrate

either a significant treatment effect or a non-significant treatment effect, since it is the large trials which are usually of most interest. Very few, if any, are small trials, and of the small trials which are published they are usually only published because they are significant. To overcome this form of bias, the researcher should ensure all studies are included in the meta-analysis as far as possible which may involve searching for studies which are unpublished. Registering a trial with a central register enables the trial to be tracked and allows the public to have access to it. At present it is recommended that every trial is registered, for both ethical reason and to attempt to prevent publication bias, however it is still not yet compulsory.

1.3 Bayesian Methods

There is a lot of debate about the 'best' way in which to model meta-analyses, which can roughly be categorised into either the frequentist approach or the Bayesian approach.

The frequentist approach uses solely the data that are provided, which can be in any form, e.g. individual patient level data, and calculates an overall single fixed estimate for the treatment effect. This can then be tested against a null hypothesis of there being no treatment difference for a trial which is being used to determine whether the new treatment is superior to the current treatment. The treatment effect estimate in this case, which is an unknown parameter, is treated as an unknown constant.

The Bayesian approach however, allows the parameters, in this instance the treatment effect and its variation, to be treated as random variables. Each of the random variables in the model are given prior distributions "a reasonable opinion concerning the plausibility of different values of the treatment effect *excluding* the evidence from the trial" (Spiegelhalter *et al.*, 2004) which can be informative, if one has a subjective opinion of the prior information for a random variable, or non-informative if one has no subjective prior belief concerning the parameter in question. Generally if one has no prior belief concerning the behaviour of a specified prior, one would model that specific variable as a non-informative prior. The priors are combined with the likelihood "the support for different values of the treatment effect based *solely* on data from the trial" (Spiegelhalter *et al.*, 2004), calculated from the data itself and this resulting combination produces the posterior distribution "a final opinion about the treatment effect" (Spiegelhalter *et al.*, 2004), the result. Use of non-informative priors therefore allows the data to 'speak for itself' and the resulting posterior will have no subjective opinion attached to it.

The frequentist approach allows calculations of 95% confidence intervals, whilst the Bayesian 'equivalent' produces a 95% credibility interval. Any percent can be used for these intervals, however the 95% is one of the most commonly used and corresponds to the 5% significance level. Whilst a 95% confidence interval will 95% of the time contain the true value of the parameter, the 95% credibility interval has a 95% probability of the true estimate being contained within it. The Bayesian credibility intervals are generally narrower due to the additional prior

information, although this will probably depend upon the model being used.

In this thesis the parameters for the treatment effect and its variation have both been given non-informative prior distributions to start with, before alternatives to these non-informative priors are then considered. The recent development of software for implementing the Bayesian approach allows these methods to be used relatively easily. The package used for the Bayesian analysis in this thesis is *WinBUGS version 1.4.3* (2007), which uses Markov Chain Monte Carlo methods (MCMC), and packages in *R version 2.6.1* (2007) have been used to allow WinBUGS to run from within R.

The results obtained via Bayesian methods can be similar in a numerical sense to those obtained using the frequentist methods, however there is the advantage that the Bayesian method can account for prior subjective opinion and the result is updated in light of the data.

1.4 Aims

The main aim of interest in this project is to determine appropriate models for meta-analyses. In doing this, the initial question of whether there is a relationship between the absolute treatment effect and the heterogeneity of a meta-analysis, all other things being equal will be examined. If there is no treatment effect, one would expect no variation between the studies; however no variation between the studies does not necessarily imply no treatment effect. A large treatment effect indicates there may be a lot of variation between studies, but there does

not have to be. If a relationship is seen to exist, the question of how this can be incorporated into the meta-analysis will then be considered

Another area of interest for this thesis is to determine which variables can feasibly be transferred between clinical trials. A meta-analysis is performed as mentioned previously on clinical trials which are similar in the sense that they look at the same new treatment, at the same concentration, and compare the results of this new treatment to the results from a single current treatment or placebo to determine whether or not this new treatment has a significant effect. This seems to be a reasonable comparison of results since each clinical trial which will be included in the meta-analysis hopes to come to the same conclusion as the remaining clinical trials. It does not appear reasonable to construct a meta-analysis which allows the inclusion of various clinical trials which have considered the difference between different treatments, since the main aim of a meta-analysis is to look at the treatment effect which can be transferred between the clinical trials. One might be interested however, in determining whether or not it would be reasonable to allow meta-analyses of clinical trials which consider different treatments if the treatment effect was not transferable between clinical trials. In this case, the meta-analyses could allow for the transfer of the variance.

1.5 Thesis Overview

To allow the reader to obtain some background knowledge of where the data for this thesis originated, Chapter 2 goes into some detail of what data are available and what it has already been used for, as well as a small literature review regarding publications concerning both meta-analyses and Bayesian approaches to meta-analyses.

Chapter 3 then looks more specifically at the relationship between the treatment effect and its variation, to determine whether these parameters should ideally be modelled as individual parameters, or whether this is not feasible.

More indepth analysis of the meta-analyses will be computed using a Bayesian analysis in Chapter 4, as well as the results from these analyses, before looking at ideas for possible future work which will be in Chapter 5.

Chapter 2

Meta-Analyses

2.1 Summary Estimates

As previously mentioned in Chapter 1, a meta-analysis produces one single overall treatment effect estimate for a group of clinical trials. This estimate can be measured using different metrics: the metric used depends on the type of data available and the format.

Normally distributed data makes use of the continuous data: the absolute difference between the means,

$$\hat{\theta} = \bar{y}_T - \bar{y}_C$$

where $\bar{y}_T$ represents the mean for the treatment group, likewise for the control group; the standardised difference between the means,

$$\hat{\theta} = \frac{\bar{y}_T - \bar{y}_C}{s}$$

where $s = \sqrt{s^2}$, s^2 representing the unbiased estimate of σ^2 based on the usual pooled sample variance, which will allow data from different scales to be combined.

Survival data uses the log-hazard ratio as an estimate for the treatment difference,

$$\theta = log\left(\frac{h_T(t)}{h_C(t)}\right),$$

where $h_T t$ and $h_C t$ represent the hazard functions for the treatment group and the control group respectively, whilst the ordinal data makes use of two types of the log-odds ratio as discussed in Whitehead (2002), one based on the proportional odds model,

$$\theta = log\left(\frac{Q_{kT}(1 - Q_{kC})}{Q_{kC}(1 - Q_{kT})}\right)$$

where θ is the log-odds of the treatment group having a greater success than the control group (Q_{kT} and Q_{kC} represent the cumulative probabilities of falling into category $k(k = 1, ..., m)$ or better for the treated and control groups respectively

(Whitehead, 2002)), and the other based on the continuation ratio model,

$$\theta = log\left(\frac{h_{kT}(1-h_{kC})}{h_{kC}(1-h_{kT})}\right)$$

with $h_{kT} = \frac{p_{kT}}{1-Q_{(k-1)T}}$ and $h_{kC} = \frac{p_{kC}}{1-Q_{(k-1)C}}$. θ produces a positive result when the treatment group is better than the control group.

Binary data, which is the form of data used in this thesis, uses three main metrics: the risk difference,

$$\theta = p_T - p_C$$

the logarithm of the odds ratio,

$$\theta = log\left(\frac{p_T(1-p_C)}{p_C(1-p_T)}\right)$$

and the logarithm of the relative risk,

$$\theta = log\left(\frac{p_T}{p_C}\right)$$

The notation used here is the proportion of patients who had the event of interest in the treatment group, p_T, and the proportion of patients who had the event of interest in the control group, p_C, and the estimates can be obtained using the Maximum Likelihood formulas.

The two metrics most commonly used for binary data are the risk difference,

and the natural logarithm of the odds ratio. The risk difference values are restricted to the interval [-1, 1], although "confidence intervals based on asymptotic theory can include points outside these limits" (Whitehead, 2002), whereas the log-odds-ratios can take values anywhere between $-\infty$ and $+\infty$, resulting in the log-odds-ratio generally being preferred. The log-odds-ratio is also known for its stability across a range of studies. The log-odds-ratio is usually used instead of the odds-ratio to allow the assumption of normality to be more reasonable (Spiegelhalter *et al.*, 2004).

The metric chosen depends greatly on the format of the data available. For example, if the data provided is Binary and is given in the summary form of the risk difference, then one is restricted to using the risk difference. If the data is provided in summary statistic form, with the number of events and non-events, there is more choice available, one could use any of the three mentioned metrics. The preference is to be able to produce a meta-analysis using the raw data, allowing analysis at the individual patient level so questions on other factors, for example, demographic issues may also be considered. This information is not generally available however, with most researchers using data provided in published papers of the trial results.

2.2 The Models

Once the data has been obtained and the decision upon which metric to use has been considered, one will then consider whether to use a fixed effects approach

or a random effects approach. Assuming a fixed effects model on a meta-analysis which consists of k independent studies with an overall treatment estimate θ, and individual study treatment estimates $\mathrm{Y}_i, i = 1, 2, \ldots, k$, $E(Y_i) = \theta$ and the variance of the individual treatment estimates for each study i, $\mathrm{Var}(\mathrm{Y}_i) = \sigma_i^2$, allowing the fixed effects model to be expressed as below.

$$Y_i = \theta + \epsilon_i \text{ where } \epsilon_i \sim N(0, \sigma_i^2)$$

$$Y_i \sim N(\theta, \sigma_i^2).$$

It is assumed here that σ_i^2 are known and equal to $\hat{\sigma}_i^2$. Then let the weights, ω_i, be the inverse variance of the summary statistic for each study i, $\omega_i = 1/(\sigma_i^2)$. The fixed effects model may then be expressed as

$$Y_i \sim N(\theta, 1/\omega_i).$$

The overall treatment effect, θ is estimated via an averaged treatment effect across the k trials,

$$\hat{\theta} = \frac{\sum_{i=i}^{k} \hat{\theta}_i \omega_i}{\sum_{i=1}^{k} \omega_i}. \tag{2.1}$$

Consider now the random effects model which allows the incorporation of heterogeneity into the previously stated fixed effects model. The random effects

model unlike the fixed effects model does not assume a common underlying treatment effect, denoted θ in the fixed effects model. Instead, it allows the treatment effect to differ between the studies.

Assuming once again a meta-analysis consists of k independent studies, with Y_i the individual study treatment estimates and $\hat{\sigma}_i^2$ the variance of $\mathrm{Y}_i, i = 1, 2, \ldots, k$. Letting θ_i represent the true treatment effect for study i,

$$Y_i = \theta_i + \epsilon_i \text{ where } \epsilon_i \sim N(0, \hat{\sigma}_i^2)$$

$$\theta_i = \mu + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, \tau^2)$$

for $i = 1, 2, \ldots, k$ and the ϵ_i and ε_i are assumed independent, allowing the random effects model to be rewritten as

$$Y_i = \theta_i + \epsilon_i + \varepsilon_i \text{ where } Y_i \sim N(\mu, \hat{\sigma}_i^2 + \tau^2)$$

where the extra variance component here, τ^2 is a measure of the heterogeneity, and μ is the overall treatment effect. The weights, ω_i are again the inverse variance of the summary statistic for each study i and incorporates the extra variance component for the random effects variance, $\omega_i = 1/(\sigma_i^2 + \tau^2)$. It is noticeable from this notation, that should there be no heterogeneity present, $\tau^2 = 0$, one will obtain the fixed effects model.

An important aspect of producing a meta-analysis is to ensure the check for

statistical heterogeneity between any trials is produced and accounted for, if necessary, before using the results obtained. Some statisticians believe that should heterogeneity be present in a meta-analysis a random effects model should be used rather than a fixed effects model to allow the combination of the individual trials whilst accounting for the heterogeneity within the model. Not everyone is of this opinion however, and some believe a fixed effects model (which corrects in some form for the heterogeneity) is still reasonable. Since heterogeneity can (and often does) occur, regardless of which model has been used to produce the meta-analysis, the heterogeneity has to be checked (and accounted for if present) to prevent invalid results. Assuming one has checked for heterogeneity and it is reasonable to discount this then a meta-analysis using a fixed effects approach seems feasible. If on the other hand, one performs a check for heterogeneity and discovers that there is significant heterogeneity in the meta-analysis, then this must be either included in the model, perhaps by means of a random effects model, or used to stratify the trials before producing overall treatment effect estimates for each strata.

The results throughout this thesis have used the random effects model which can take into account any heterogeneity that may be present, yet still return the results that would be obtained from a fixed effects model should there be no heterogeneity.

2.3 Bias

Publication bias, mentioned briefly in Chapter 1, is one of the main sources of bias which can occur when constructing a meta-analysis, but can be checked for without too much difficulty. Graphically, a funnel plot can be used to detect publication bias, however there are also statistical methods which can be used which are less subjective.

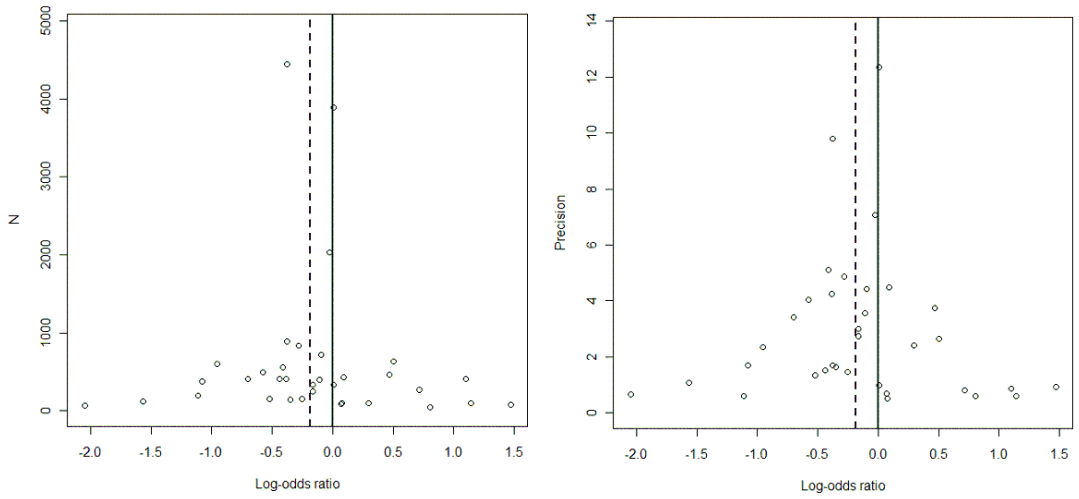

Figure 2.1. Funnel Plots for one of the Meta-Analyses with 34 trials

A funnel plot involves plotting the sample size of each of the individual trials against the trials' estimated treatment effect size and analysing the shape of the plot, however Whitehead (2002) makes note that instead of using the sample size one can use the precision of the treatment effect (the inverse of the standard error). Some examples of these funnel plots are displayed in figures 2.1 and 2.2. In these figures, the circles represent each of the individual trials within each meta-analysis, with the vertical black dashed line representing the overall fixed effects

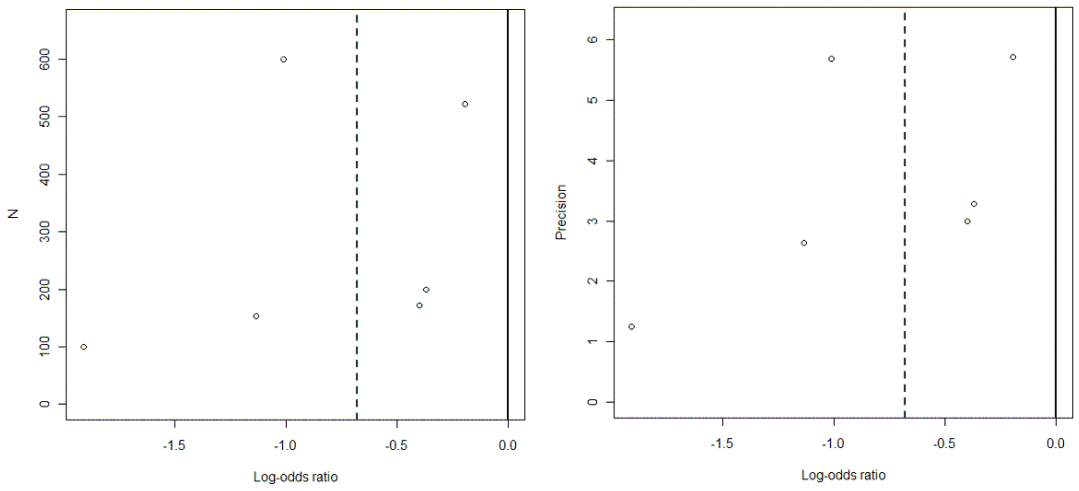

Figure 2.2. Funnel Plots for one of the Meta-Analyses with 6 trials

treatment estimate for the meta-analyses (using the Mantel-Haenszel method, Mantel and Haenszel (1959)) and the vertical black solid line representing no treatment effect. If there is no evidence of any publication bias, the plot should resemble an upside down symmetric funnel, with a large variation in the observed effect size being reasonable for trials with small sample sizes, and a gradually reducing variation about the true effect size as the sample sizes of the trials increases until the top of the plot displays very little variation about the true effect size for trials with a large sample size. Figure 2.1 is a good example of a funnel plot which indicates no evidence of any publication bias. In the event of publication bias occurring, one would tend to notice very few studies within the bottom right-hand corner, of the funnel plot, those which correspond to small numbers of trials and high values for the log-odds ratios (those trials which show a poor treatment effect). Figure 2.2 is the best example of this from the 125 meta-analyses considered in this thesis, however the assumption of no publication

bias for this specific meta-analysis based solely on the graph would be deemed reasonable since there are only six trials within this meta-analysis.

2.4 Measures of Heterogeneity

Using a random effects model allows the incorporation of heterogeneity, the between study variation, into the model. There are several ways in which to measure the heterogeneity and some of these are described here.

One of the most widely used measures of heterogeneity in practice today uses DerSimonian and Laird's method of moments (DerSimonian and Laird, 1986) which can be found by equating the sample statistic for the heterogeneity, Q_ω, with its expectation,

$$\tau^2_{DL} = max\left\{0, \left\{ \frac{Q_\omega - (k-1)}{\left[\sum_i \omega_i - \frac{(\sum_i \omega_i^2)}{\sum_i \omega_i}\right]} \right\}\right\} \tag{2.2}$$

The k represents the number of trials in the meta-analysis and the ω_i's are the weight assigned to trial$_i$. The test statistic for the heterogeneity, Q_ω, is calculated using the formula

$$Q_\omega = \sum_i \omega_i (y_i - \bar{y}_\omega)^2$$

Biggerstaff and Tweedie (1997) proposed a method of measuring the heterogeneity which is similar to the method of moments proposed by DerSimonian and Laird (1986), yet allows the incorporation of the uncertainty when estimating the

heterogeneity into the estimates for the treatment effect, it's standard error and the confidence intervals. Brockwell and Gordon (2007) compare these measures of heterogeneity, along with the method using maximum likelihood estimators outlined below, and note the similarities between DerSimonian and Laird (1986) and Biggerstaff and Tweedie (1997)

"there is very little difference between estimated coverages for the BT [Biggerstaff and Tweedie (1997)] method and those of the DL [DerSimonian and Laird (1986)] method... [and] the BT [Biggerstaff and Tweedie (1997)] method also requires numeric routines to obtain confidence intervals for μ".

Brockwell and Gordon (2007) note also that both the methods based on the method of moments (DerSimonian and Laird (1986) and Biggerstaff and Tweedie (1997)) "estimated coverage probabilities are frequently well below the nominal level of 0.95... particularly so when k is small", resulting in the confidence intervals being insufficiently small.

Hardy and Thompson use a method based on the use of Maximum Likelihood estimators (Hardy and Thompson, 1996), which uses an iterative procedure to obtain the overall treatment effect $\hat{\mu}$ and the heterogeneity τ^2_{HT},

$$\tau^2_{HT} = max\left\{0, \left\{\frac{\sum_{i=1}^{k} \omega_i^2\{(\hat{\theta}_i - \hat{\mu})^2 - \sigma_i^2\}}{\sum_{i=1}^{k} \omega_i^2}\right\}\right\} \tag{2.3}$$

where $\omega_i = \dfrac{1}{\sigma_i^2 + \tau^2_{HT}}$ and $\hat{\mu} = \dfrac{\sum_{i=1}^{k} \omega_i \hat{\theta}_i}{\sum_{i=1}^{k} \omega_i}$. Equation 2.3 is solved iteratively until it converges.

The maximum likelihood methods are noted to "underestimate variances"

(Thompson and Sharp, 1999) however, and so the method of restricted maximum likelihood (REML) estimates (Thompson and Sharp, 1999) are also used. The REML estimator, like the maximum likelihood estimators, uses an iterative procedure to obtain the same estimators $\hat{\mu}$ and τ^2_{REML} upon convergence using,

$$\tau^2_{REML} = max\left\{0, \frac{sum^k_{i=1}\omega_i^2\{(k/(k-1))(\theta_i - \hat{\mu})^2 - \sigma_i^2\}}{\sum^k_{i=1}\omega_i^2}\right\} \tag{2.4}$$

again with $\omega_i = \dfrac{1}{\sigma_i^2 + \tau^2_{REML}}$ and $\hat{\mu} = \dfrac{\sum^k_{i=1}\omega_i\hat{\theta}_i}{\sum^k_{i=1}\omega_i}$.

2.5 Background on the Data

2.5.1 Introduction

The data which has been used as the basis of this thesis is from a journal article in the Statistics in Medicine journal, Engels *et al.* (2000), titled "Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses" and has been kindly provided by Christopher Schmid, one of the authors.

This journal article studies the results of 125 individual meta-analyses, and examines specifically the heterogeneity in each of these meta-analyses and which metric, or metrics, should be used to measure the treatment effect, focusing on studies with binary outcomes. Engels *et al.* (2000) stated that it was "the largest systematic examination of meta-analyses... to our knowledge", although other similar studies on a smaller scale have been published (DerSimonian and Laird,

1986; Berlin *et al.*, 1989).

Engels *et al.* (2000) declare "there may be no metric that is 'best' for all circumstances" and that the "conclusions drawn from a meta-analysis may depend on which metric is used, especially when... heterogeneity is present".

The aim of the article was to answer three specific questions:

1. "How often are the collections of trials used in meta-analysis heterogeneous?"

2. "When do fixed effects and random effects methods give different estimates of treatment effect?"

3. "When does summarising risk differences give a different impression of treatment effect than summarising odds ratios?"

Journal	Number of Meta-Analyses
Ann Intern Med	7
Arch Intern Med	2
BMJ	28
Circulation	6
CCPC	45
Drugs	1
JAMA	8
Lancet	14
New Engl J Med	12
Journal Not Specified	2

Table 2.1. Journals from which the selected Meta-Analyses originated

The meta-analyses which have been included were selected from two sources, major medical journals which publish large numbers of meta-analyses and the

1994 Cochrane Pregnancy and Childbirth database (CCPC). The medical journals which were used are Annals of Internal Medicine, Archives of Internal Medicine, British Medical Journal, Circulation, Journal of the American Medical Association, Lancet, and New England Journal of Medicine, all of which were limited to the issues from years between 1990 and 1996. From the meta-analyses within this search criteria, the ones which were chosen were randomized controlled trial data with binary outcomes in the form of 2 x 2 tables.

Field	Number of Meta-Analyses which Focus on the specified Field
Cancer	17
Cardiovascular Disease	43
Diabetes	1
GI	6
Infectious Diseases	5
Neurology	1
Pediatrics	3
Perinatal	45
Psychiatry	4

Table 2.2. The Fields in which each of the Meta-Analyses primary focus occurred

The CCPC is the single source which accounts for most of the meta-analyses, 45 of the 125 available as displayed in table 2.1, all of which contain meta-analysis results for the same field, perinatal. However, not all of these meta-analyses focus on a single outcome, there are more than 20 different outcomes included across the 125 meta-analyses, from low birth weight to respiratory distress syndrome to mortality at various ages. The meta-analyses which come from the major medical

journals cover various different fields of medical interest, some of which include cancer, myocardial infarction, cardiovascular disease and infectious diseases, as well as varying outcomes and/or treatments within each of the fields of interest. Table 2.2 highlights the main field groups occurring in the 125 meta-analyses and a count of how many meta-analyses focused on each field.

2.5.2 Inclusion Criteria

The inclusion criteria used to obtain the meta-analyses required each of the meta-analyses to have at least six clinical trials to ensure each contained sufficient data which would result in a valid effect estimate. Table 2.3 displays some information about the summary statistics of the clinical trials which are used to obtain the 125 meta-analyses, which confirms that all of the meta-analyses included at least 6 clinical trials, although the numbers of clinical trials does differ greatly between meta-analyses, with the average meta-analysis containing 13 clinical trials.

It was also a requirement that each meta-analysis had at least one event in the control arm, and that the average number of events in the trial control arms be at least five. Table 2.5 summarises the number of events in both treatment arms, the treatment arm and the control arm. These summaries also confirm the inclusion criteria specified by Engels *et al.* (2000) concerning the average number of events in the trial control arms is indeed five. Assuming $n_T = n_C$, the average number of patients in one treatment group for the 125 meta-analyses is summarised in table 2.4. The mean number of patients in a treatment group is

462, with the smallest average number of patients in a single treatment group being 22.

If any trial had no events in either arm, a continuity correction, as used by various authors, Yates (1934), Anscombe (1956), and Cox and Snell (1989), was implemented: one half was added to each cell of the corresponding 2 x 2 table before calculating the summary effect statistics. The same continuity correction was also used when the number of events in any treatment arm in any given trial equaled the total number of patients corresponding to that treatment in that particular trial. Anscombe (1956) notes that by implementing this correction, the "bias can be very nearly removed", whilst improving the distributional assumption of Normality, Cox and Snell (1989). Very few trials contained zero, or n, events in either arm due to the strict criteria pre-determined for the inclusion of meta-analyses and so the authors were confident that this correction had very little effect on the overall results.

Summary Statistics	
Minimum	6
Median	10
Mean	13.06
Maximum	62

Table 2.3. Summary statistics for the numbers of clinical trials included in the 125 MA

Summary Statistics	
Minimum	22.29
Median	129.77
Mean	461.82
Maximum	13744.12

Table 2.4. Summary statistics for the average numbers of patients in one treatment group for the 125 MA, assuming the numbers of patients in the treatment and the control group are equal

There is one inclusion criterion which has been stated by Engels *et al.* (2000)

which does not appear to have been implemented: the exclusion of any trial with fewer than ten subjects in either arm. Engels *et al.* (2000) stated in their inclusion criteria "To avoid very small trials, we excluded from the meta-analysis any trial with fewer than ten subjects in either arm". Of the 125 meta-analyses, 21 of them included at least one trial with less than ten subjects in either treatment arm. These trials were removed and the summary statistics were then calculated. However, upon checking the data with the summary statistics which were provided by the authors, the results calculated differed slightly from the summary statistics which had been provided. It was then noticed that by not excluding the very small trials (those with less than ten subjects in either arm), the results appeared to match in almost all of the data (see Chapter 3 section 3.2).

	Control Group summary statistics	Treatment Group summary statistics
Minimum	5.29	2.8
Median	22.91	19.33
Mean	43.89	39.42
Maximum	578.33	536

Table 2.5. Summary statistics for the average numbers of events in both the control arm and the treatment arm for all 125 MA

2.5.3 Calculating the Summary Statistics

Since the data provided were in binary form, the metrics available are automatically limited to the three main metrics previously specified for binary data: the risk difference, the relative risk or the log-odds ratio. The two most common

methods used of these three are the risk difference and the log-odds-ratio, and as mentioned previously, the log-odds ratio is generally accepted as a fair measure of treatment effect due to its ease of transferability to the real line, therefore the log-odds-ratio has been the main focus of the analysis throughout this thesis.

Using the previously specified general formula for the calculation of the treatment effect estimates, 2.1, the summary estimates can be calculated for each metric, with the weights corresponding to the formula for the random effects weight, $\omega_i^* = 1/(\sigma_i^2 + \tau_i^2)$, which can be rewritten as $\omega_i^* = 1/((1/\omega_i) + \tau^2)$, where $\omega_i = 1/\sigma_i^2$, with the fixed effects estimates using the same formulae when $\tau^2 = 0$.

The fixed effects weights are calculated as follows,

$$\omega_{LORi} = \frac{1}{\frac{1}{n_{Ti}p_{Ti}(1-p_{Ti})} + \frac{1}{n_{Ci}p_{Ci}(1-p_{Ci})}} \quad \text{for the logarithm of the odds ratios}$$

$$\omega_{LRRi} = \frac{1}{\frac{(1-p_{Ti})}{n_{Ti}p_{Ti}} + \frac{(1-p_{Ci})}{n_{Ci}p_{Ci}}} \quad \text{for the logarithm of the relative risks}$$

$$\omega_{RDi} = \frac{1}{\frac{p_{Ti}(1-p_{Ti})}{n_{Ti}} + \frac{p_{Ci}(1-p_{Ci})}{n_{Ci}}} \quad \text{for the risk differences}$$

2.5.4 Results

Engels *et al.* (2000) examine both the odds ratio and the risk difference metrics, for both the fixed effects model and the random effects model to attempt to provide a possible preference, and to determine whether the two metrics produce the same results as one would hope. The value used for the heterogeneity is the

one-tailed p-value for the calculated Q-statistic. The risk difference would be classed as being more heterogeneous than the odds ratio if "if the p-value for the risk difference Q-statistic is less than the p-value for the odds ratio Q-statistic" and a meta-analysis would be termed "'heterogeneous' when the corresponding Q-statistic p-value is below a nominal cut-off, usually 0.05 or 0.10".

The heterogeneity among the two metrics were compared and it was noted that "the risk differences usually displayed more heterogeneity than the odds ratios", with the risk differences either "more heterogeneous than the odds ratios" or "the risk differences were judged heterogeneous when the odds ratios were not". There were a small number of cases (three) which noted "the odds ratios heterogeneous when the risk differences were not". Engels *et al.* (2000) noted however, that these results appeared to contradict results from two previous studies which reported "similar heterogeneity measures for risk differences and odds ratios", yet Engels *et al.* (2000) indicated the conflicting results could be due to a lack of "sufficient power to detect differences in heterogeneity between risk differences and odds ratios" since the number of meta-analyses used in each of the other studies were smaller, nine (DerSimonian and Laird, 1986) and 22 (Berlin *et al.*, 1989). Engels *et al.* (2000) also note another couple of reasons, including the sample may not be a good "representative of meta-analyses found in clinical research" and the small number of studies within the meta-analyses "(Berlin *et al.*, 1989) included nine meta-analyses with fewer than six trials, meta-analyses that (Engels *et al.*, 2000) would have excluded".

Engels *et al.* (2000) were interested in determining whether the results of a

meta-analysis were dependant on the metric chosen to produce the summary estimates. The summary estimates for each meta-analysis using the two different metrics were compared and there were no cases reported of the metrics being significant in opposite directions, allowing one to conclude the results "from meta-analyses are robust to changes of metric". Engels *et al.* (2000) then go on to examine different levels of significance for the different summary estimates (both fixed effects and random effects) for two-sided p-values, noting "that random effects estimates were often less significant than fixed effects estimates".

The meta-analyses were then split into four subgroups;

- "'Homogeneous'... in which neither the odds ratios nor the risk differences were heterogeneous"
- "Meta-analyses in which both risk differences and odds ratios were heterogeneous"
- "Meta-analyses in which only the risk differences were heterogeneous"
- "Meta-analyses in which only the odds ratios were heterogeneous"

with the definition of heterogeneous being "a Q-statistic p-value less than 0.10".

The majority of the meta-analyses fell into the 'homogeneous' subgroup (50.4 percent), with 32.8 percent of the meta-analyses having both metrics heterogeneous, 14.4 percent heterogeneous for just the risk differences and the remaining 2.4 percent heterogeneous for just the odds ratios.

The random effects summary estimates and standard errors allow for the extra heterogeneity parameter which the fixed effects estimates do not and so will lead

to different values when heterogeneity is present. Engels *et al.* (2000) noted the calculated estimates did not differ greatly between the fixed effect approach and the random effect approach, however the standard errors of the random effects estimates were greater than those of the fixed effects estimates and "the overall effect of heterogeneity was to make most random effects estimates less significant than the corresponding fixed effects estimates, for both the odds ratio and risk difference metrics".

The fixed effects estimates for the odds ratios and the risk differences "provided similar levels of significance" for the meta-analyses which were classed as 'homogeneous'. The same conclusion was declared for the random effects estimates for the meta-analyses in which both the risk differences and the odds ratios were heterogeneous, indicating "the choice of metric used to measure and summarize the treatment effect is not crucial", however Engels *et al.* (2000) state "the risk difference may not be the most appropriate metric to use in meta-analysis, because risk differences may be substantially heterogeneous among trials [and] the risk difference metric tends to give greatest weight to trials with low event rates... [although] trials with low event rates would seem to offer little information about treatment effects". Engels *et al.* (2000) also state their results "suggest that the odds ratio is more likely than the risk difference to remain constant across populations" and so this should be incorporated if an absolute measure of treatment effect is required.

Chapter 3

Examining the Relationship

3.1 Introduction

One of the main aims of this thesis is to determine whether the treatment effect estimates and their corresponding heterogeneity values are independent, or if some sort of relationship exists (Senn, 2007*b*). These two measures are often modelled as independent, however it is of interest to determine whether this is actually the case, or whether the heterogeneity varies with the absolute size of the treatment effect estimate. One might presume a meta-analysis with no treatment effect should indicate no heterogeneity, whilst a meta-analysis with a large absolute treatment effect could indicate anywhere between no heterogeneity and lots of heterogeneity.

3.2 Checking the Data

The first focus of interest is to produce a quality control study which ensures the statistics provided by Engels *et al.* (2000) coincide with those reproduced. As previously mentioned, the data provided were in binary form, and although the metric focused on is the log-odds-ratio, the checking of the values provided for the random effect mean estimate and its standard error were computed for each of the three metric choices available. This is to ensure the results obtained are consistent with those provided by the authors. After implementing the continuity correction, the random effect mean estimates and their standard deviations for each metric were calculated for each of the 125 meta-analyses in turn using scripts within R version 2.6.1. These computed values were then compared to the values in the data set which were provided.

As previously noted in Chapter 2 section 2 (Background on the Data), implementing one of the stated exclusions, removing any trials with fewer than ten subjects in either arm, caused deviations from the values for all the meta-analyses which this criteria affected. By ignoring this exclusion criteria, the values calculated matched almost exactly those provided, indicating that this specific criteria had not been implemented, although the differences are minimal and should not affect any results Engels *et al.* (2000) produced. Since the original values provided were calculated with this criteria not being implemented, the criteria has not been implemented for this thesis either.

The random effects summary estimates for the log-odds-ratio were calculated

using a modification of the method used by DerSimonian and Laird (1986), as used by Fleiss and Gross (1991). The results provided by Engels *et al.* (2000) are actually provided for the odds ratios and so the exponential of the calculated log-odds-ratios are compared with the odds ratio estimates provided.

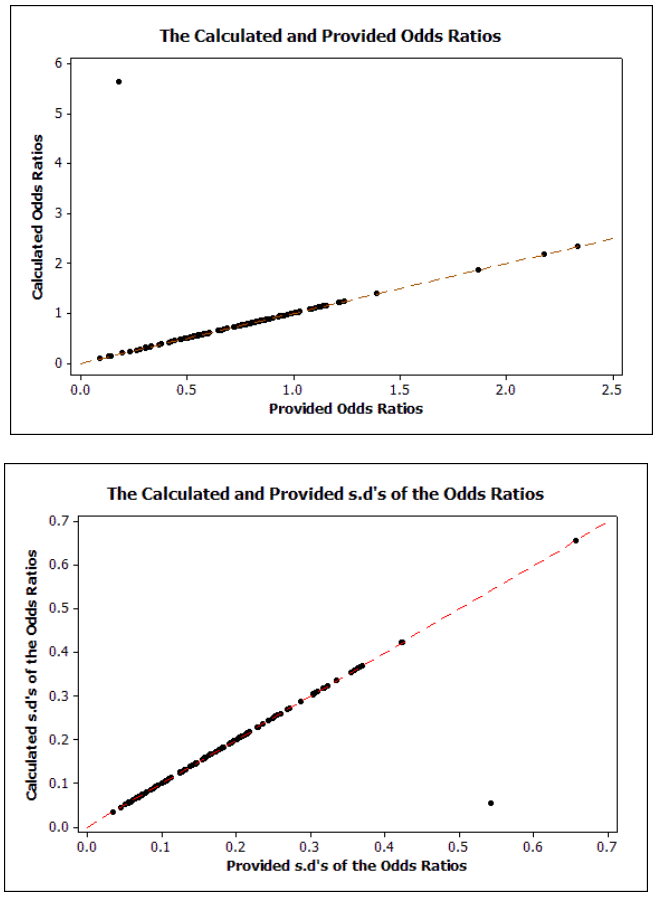

Figure 3.1. Plots comparing the calculated and given values for the odds-ratio summary estimates and the corresponding standard deviations of the odds-ratio summary estimates.

Of the 125 meta-analyses, there were only two discrepancies which occurred between the results from the quality control study and the data which was provided: one for the odds-ratio estimate, the data provided had an estimate for the odds-ratio of 0.177645, whilst the quality control estimate provided an odds-ratio

estimate of 5.6292; and the other discrepancy occurred for the standard deviation of the odds-ratio estimate, the data provided had an estimate for the standard deviation of 0.542, whilst the quality control estimate provided a standard deviation value of 0.0542. The exponential of the calculated log-odds ratios (and their corresponding standard deviations) are also compared graphically against those provided by Engels *et al.* (2000), Figure 3.1.

In the standard deviation instance, it appears the discrepancy is the result of an input error for that particular result, however at first glance, it remains unclear whether this is also the case for the different estimate values for the odds-ratio. Taking the natural logarithm of the value provided by Engels *et al.* (2000) (0.177645) produces a log-odds ratio value of -1.727968, whilst the computed log-odds ratio in the quality control study is 1.727975, indicating an input error in the results provided by Engels *et al.* (2000).

Examining the computed statistics for the risk difference metric, there are several more discrepancies (52 of the 125 meta-analyses contained discrepancies). The majority of these discrepancies occurred for both the risk difference value and it's standard error (40 of the 52 meta-analyses), however there are some cases in which only one of the calculated values, either the risk difference (10 of the 52 meta-analyses) or the standard error disagrees (2 of the 52 meta-analyses). Most of these appear on first inspection to be the result of either rounding error or input error since the values provided do not differ greatly from the values calculated. However, after a closer look at the results, it appears that some of the risk difference treatment estimates and their corresponding standard deviations,

which were provided, have been calculated without implementing the continuity correction to adjust for cases in which the number of events is equal to either zero or the number of patients in the corresponding treatment group. In fact, in all except two of the meta-analyses which included trials containing either zero or n events, the continuity correction had not been used.

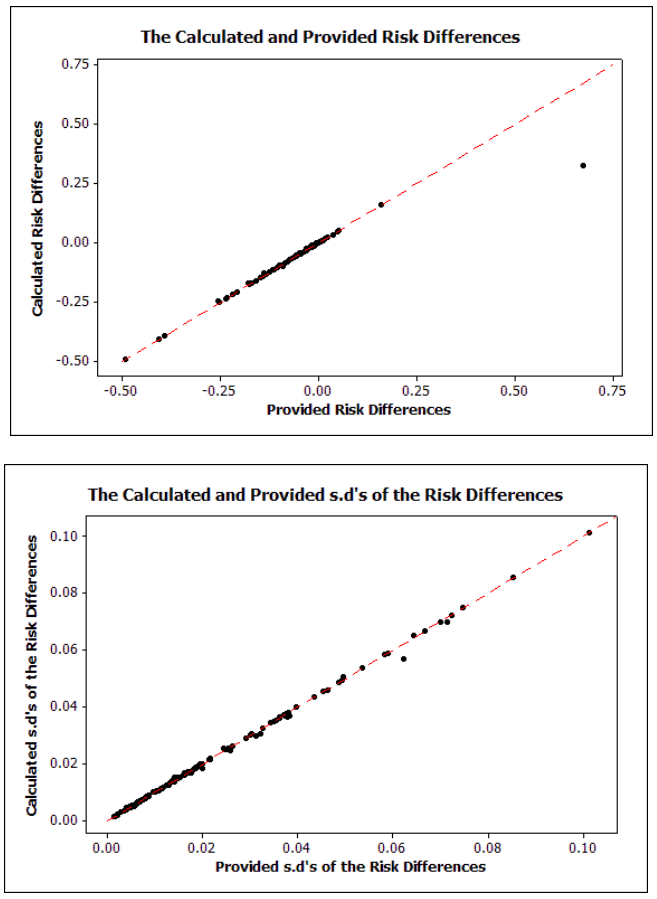

Figure 3.2. Plots comparing the calculated and given values for the risk difference summary estimates and the corresponding standard deviations of the risk difference summary estimates.

Ignoring the continuity correction, the majority of the results which did not previously match the values provided, now match exactly (60% of the 52 meta-analyses). The majority of the remaining meta-analyses which do not match the

results exactly after ignoring the continuity correction are similar in magnitude, implying the discrepancies may be due to a rounding error somewhere in the calculations.

The two meta-analyses for which the continuity correction had been applied produce results which disagree from those provided if the continuity correction is ignored. The results for the meta-analyses which had no trials with either zero or n events remain unchanged since the data for these meta-analyses has not been altered.

Plots comparing the calculated risk difference estimates (and their standard deviations) with those provided by Engels *et al.* (2000) are also produced here after the continuity correction has been implemented, Figure 3.2.

Lastly, checking the relative risk (rather than the log-relative risk) values provided by Engels *et al.* (2000), all 125 relative risk values and their corresponding standard error (s.e.$_i = \sqrt{1/\omega_i^*}$ for $i = 1, \ldots 125$) match to four decimal places after implementing the continuity correction.

The inclusion criteria stated appeared reasonable. To ensure the possibility of invalid effect estimates has been reduced, all meta-analyses with very few trials (less than six) were excluded, and only meta-analyses with an average of five events in the trial control arms were included.

3.3 Exploratory Analysis of a Relationship

An immediate reaction to examining the results of a meta-analysis is to automatically focus solely on the effect of the treatment. Perhaps one should not be interested in this alone, but interested in whether or not there is a relationship between the effect of the treatment and the heterogeneity. It might be of interest, for example, to attempt to determine whether meta-analyses which have a larger treatment effect coincide with larger between study variation.

Figure 3.3 shows the relationship between the treatment effect estimates and the heterogeneity for each of the three metrics considered. The heterogeneity in these graphs is estimated using the ratio Q/degrees of freedom, the same ratio as used by Hardy and Thompson (1998) in one of their practical examples, with Q corresponding to the Q-statistic calculated for each random effects metric respectively, and the treatment effect is taken as the absolute value of the treatment effect estimates, thus ignoring the direction of the treatment effect estimates.

These graphs clearly indicate there is some positive relationship between the treatment effect and the heterogeneity for all three metrics, with the corresponding correlation values provided in Table 3.1. Each of the correlations provided in Table 3.1 are highly significant with p-values produced for each correlation of <0.0001.

Significant positive correlations for each metric confirm that regardless of the metric chosen, and all other things equal, there does appear to be a significant

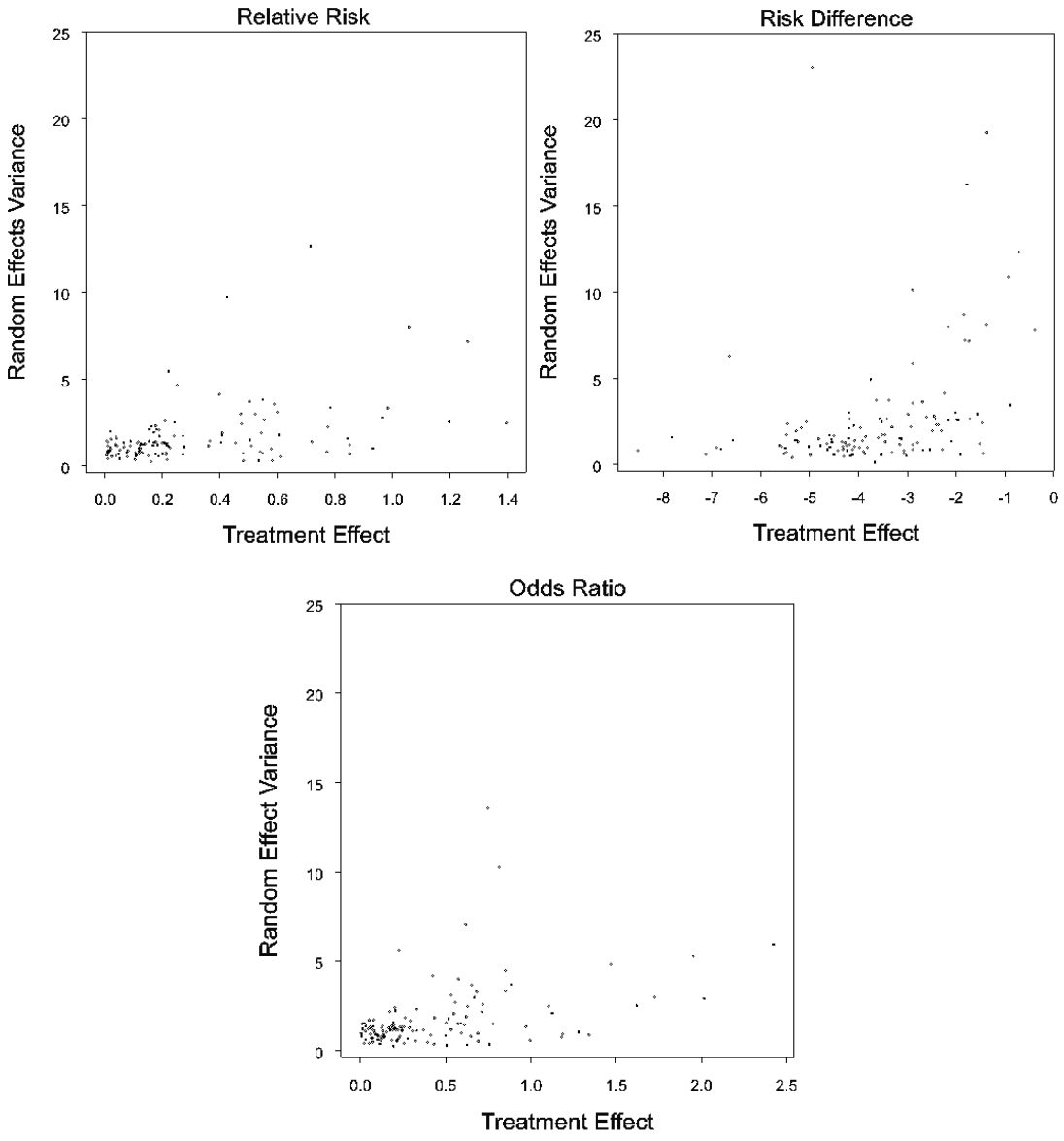

Figure 3.3. Plots showing the between study variation against the treatment effect for all three metrics: the relative risk; the risk difference; and the log-odds-ratio.

		Treatment Effect		
		Relative Risk	Risk Difference	Log-Odds-Ratio
Random Effects Between-study Variation	Relative Risk	0.4238	-	-
	Risk Difference	-	0.3524	-
	Odds Ratio	-	-	0.4194

Table 3.1. Correlations for the treatment effect estimates and their respective heterogeneity for each metric

relationship between the absolute treatment effect and the heterogeneity. As anticipated, the meta-analyses for which the estimated treatment effect was not significant also indicate very little, if any, variation between studies. The high between study variation (indicated by increasing values on the y-axis) arises in meta-analyses with a significant treatment effect (indicated by increasing values on the x-axis).

Existing articles on meta-analysis assume the treatment effect and the heterogeneity are independent, but the apparent relationship occurring between the two measures now causes some doubt about whether these previous analyses are correct, and indeed if they are not how best it would be to tackle this possible problem.

One author has already questioned the relationship between the treatment effect estimates and their heterogeneity (Senn (2007*b*), Senn (2008)) and offered a possible way to incorporate this relationship into the model. This will be considered later on in this thesis.

3.4 Summary

Upon examination of whether or not there is a relationship present between the absolute treatment effect estimates in a meta-analysis and their corresponding heterogeneity values, scatterplots indicated a possible positive relationship between these two measurements for each of the three metrics available for measuring binary data. Correlations produced for the relative risk, log-odds ratio and risk difference were 0.4238, 0.4194 and 0.3524 respectively and all were highly significant at the 5% significance level, with p-values of less than 0.001 produced for each of the correlations.

Chapter 4

Fitting a Bayesian Model

4.1 Introduction

Up to this point, a frequentist model which uses a Normal Normal model based on summary statistics has been used. It is of interest now to consider a Bayesian approach, which commonly uses one of two types of model, a Binomial Normal model, or less frequently a Normal Normal model. Both of these Bayesian models will be considered in this thesis.

As stated earlier, a Bayesian model treats the treatment estimate and the heterogeneity, the two main areas of interest here, as random variables rather than as fixed single estimates. All of the unknown random variables are modeled using priors.

For the moment, we will assume that there is no apparent prior information on any of the unknown random variables and so it is desirable to model these

priors as non-informative, allowing the data to speak for itself rather than attempting to influence the behaviour of the variables using informative priors.

Various non-informative priors are available, yet whether they are actually non-informative or not is still debatable. Lambert *et al.* (2005) wrote a paper on 13 non-informative priors and the results produced by them in a simulation study to determine how 'non-informative' these so called priors actually were. Lambert *et al.* (2005) noted the results produced by the priors which were believed to be non-informative were not necessarily so, with some priors performing "particularly poorly"and recommended a "sensitivity analysis", be performed to ensure feasible results are produced.

4.2 Creating a Model

Starting with the more commonly used Binomial Normal model and focusing on the log-odds-ratio, the following model, model 4.1, is constructed for the data available.

The notation used in model 4.1 demonstrates the number of events r_i for each $trial_i$ for the control group (c) and the treatment group (t) follow independent binomial distributions with parameters p_i, the proportion of events for $trial_i$ and n_i, the number of patients in $trial_i$, for each treatment group respectively.

Using the linear logistic model, the proportion of events in the control group for each $trial_i$ is equal to ψ_i, the overall effect for the control group for each $trial_i$, whilst the proportion of events in the treatment group for each $trial_i$ is set equal to

ψ_i plus δ_i, δ_i here representing the treatment effect for trial$_i$. Implementing the logit scale improves the normal approximation of the likelihood (Spiegelhalter *et al.*, 2004). Note that model 4.1 can be written with an intercept instead, however this would require also implementing a constraint on the ψ_i's, to ensure the coefficient matrix is of full rank, for example setting $\psi_1 = 0$.

$$
\begin{aligned}
r_{c_i} &\sim \text{Bin}(n_{c_i}, p_{c_i}) \\
r_{t_i} &\sim \text{Bin}(n_{t_i}, p_{t_i}) \\
\text{logit}(p_{c_i}) &= \psi_i \\
\text{logit}(p_{t_i}) &= \psi_i + \delta_i \\
\delta_i &\sim N(\mu, \tau^2) \\
\psi_i &\sim \text{N}(0, 1000000) \\
\mu &\sim \text{N}(0, 1000000) \\
\tau^2 &\sim \text{Inv} - \text{Gamma}(0.1, 0.1)
\end{aligned}
\tag{4.1}
$$

The treatment effect, δ_i for each trial$_i$ is the main focus of interest and is given a normal prior with parameters μ, the overall treatment effect, and τ^2, the variance of the treatment effect (the heterogeneity). The overall effect for each trial$_i$, ψ_i, and the hyperparameters, μ and τ^2, are all given non-informative flat priors over a range large enough to include all possible values, which allows the data to speak for itself since there is no prior knowledge for these parameters. Non-informative priors on the ψ_i's are implemented since the main interest here

lies on the treatment effect, δ_i, however hyperparameters can be used to define the ψ_i's.

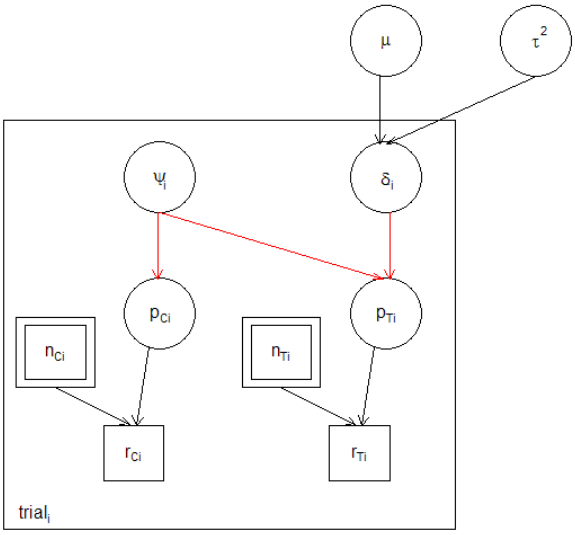

Figure 4.1. Graphical model for random effects meta-analysis using the Binomial model, model 4.1

One can construct a graphical model for model 4.1 as used by Whittaker (1990), Figure 4.1.

The nodes in Figure 4.1 represent the parameters of the model and the data. Single rectangles represent the observed variables (r_{Ci}, r_{Ti}), double rectangles represent constants fixed by the design of the study (n_{Ci}, n_{Ti}) and circles represent the unobserved parameters (p_{Ci}, p_{Ti}, ψ_i, δ_i, μ, τ^2). The arrows are drawn from the parental nodes to descendant nodes and indicate the models conditional independence assumptions, with the black line indicating stochastic dependence and the red line indicating a logical function (Smith *et al.*, 1995). In Figure 4.1, the treatment effect (δ_i) is conditionally independent of the trial effect (ψ_i), given μ and τ^2.

4.3 Published Data Check

4.3.1 Introduction

After having constructed the Bayesian model within WinBUGS, the first step was to check the model was working adequately, by using model 4.1 with data that has previously been published, allowing the comparison of the results from model 4.1 with the published results. Assuming the constructed model is working properly, the results produced should match the published results.

For this check, the published data used were obtained from an example on the selective decontamination of the digestive tract for patients in intensive care units (Abrams and Sansó, 1998), which have been used previously for a couple of meta-analyses (Digestive Tract Trialists' Collaborative Group (1993), Smith *et al.* (1995)).

Abrams and Sansó (1998) examine the data which consist of 22 randomised trials in binary form. Each trial has one treatment group and one control group, with the number of patients with a respiratory tract infection and the total number of patients in each group recorded. Initially, the data have been analysed using a fixed effects model (Digestive Tract Trialists' Collaborative Group, 1993), however previous authors (Smith *et al.*, 1995) noted a degree of heterogeneity in the data, implying the random effects model should perhaps appear more reasonably justified. Abrams and Sansó (1998) agreed with the random effects approach to modeling this data and produced a significant χ^2 statistic of 58.0 on 21 degrees of freedom ($p < 0.0001$) for testing the null hypothesis of no heterogeneity, and

so examined the data using a random effects model and focusing on the Bayesian analysis.

Study	Treated		Control		Odds Ratio
	Infections	Total	Infections	Total	
1	7	47	25	54	0.20
2	4	38	24	41	0.08
3	20	96	37	95	0.41
4	1	14	11	17	0.04
5	10	48	26	49	0.23
6	2	101	13	84	0.11
7	12	161	39	170	0.28
8	1	28	29	60	0.04
9	1	19	9	20	0.07
10	22	49	44	47	0.06
11	25	162	30	160	0.79
12	31	200	40	185	0.66
13	9	39	10	41	0.93
14	22	193	40	185	0.47
15	0	45	4	46	0.10
16	31	131	60	140	0.41
17	4	75	12	75	0.30
18	31	220	42	225	0.71
19	7	55	26	57	0.17
20	3	91	17	92	0.15
21	14	25	23	23	0.03
22	3	65	6	68	0.50
Pooled					0.36

Table 4.1. Respiratory tract infections of 22 studies in the control and treatment groups with individual and pooled estimates of odds ratios using the Mantel-Haenszel method (Mantel and Haenszel, 1959)

Abrams and Sansó (1998) include a Table displaying the original data in Binary format, also provided here in Table 4.1. Abrams and Sansó (1998) also included in their table the summarised individual odds ratio estimates, and so

the calculated odds ratio estimates are also contained in Table 4.1. Smith *et al.* (1995) produced a similar table also, however it is interesting to note that despite both Abrams and Sansó (1998) and Smith *et al.* (1995) using the same data in the same format, their calculations for the odds ratio estimates differ for 15 of the 22 studies. Smith *et al.* (1995) make a note of which method was used to calculate the odds ratios (the Mantel-Haenszel-Peto method), which one can reproduce (the calculations for this method are provided in Yusuf *et al.* (1985) and compare the observed number of deaths among the treated patients with the expected number of deaths), whilst Abrams and Sansó (1998) have "Odds Ratios T/C" as the title above the column containing the individual odds ratio values, which one would presume indicates their calculated odds ratios are obtained by dividing the treatment group by the control group. Further on in their paper, Abrams and Sansó (1998) note the use of the value n_i, where n_i is the number of patients in the treatment or the control group which are assumed to be equal, and in the case of $n_{Ti} \neq n_{Ci}$ an alternative is chosen, for example the average, the minimum, or the maximum of n_{Ti} and n_{Ci}.

$$y_i = log\frac{p_{Ti}(1-p_{Ci})}{p_{Ci}(1-p_{Ti})} \quad \text{with} \quad p_{Ti} = \frac{r_{Ti}}{n_{Ti}},\ p_{Ci} = \frac{r_{Ci}}{n_{Ci}} \tag{4.2}$$

$$y_i = log\frac{p_{Ti}(1-p_{Ci})}{p_{Ci}(1-p_{Ti})} \quad \text{with} \quad p_{Ti} = \frac{r_{Ti}}{n_i},\ p_{Ci} = \frac{r_{Ci}}{n_i},\ n_i = n_{Ti} + n_{Ci} \tag{4.3}$$

$$y_i = log\frac{p_{Ti}}{p_{Ci}} \quad \text{with} \quad p_{Ti} = \frac{r_{Ti}}{n_{Ti}},\ p_{Ci} = \frac{r_{Ci}}{n_{Ci}} \tag{4.4}$$

$$y_i = log\frac{p_{Ti}}{p_{Ci}} \quad \text{with} \quad p_{Ti} = \frac{r_{Ti}}{n_i},\ p_{Ci} = \frac{r_{Ci}}{n_i},\ n_i = n_{Ti} + n_{Ci} \tag{4.5}$$

Unfortunately, however, I could not reproduce the values obtained by Abrams

and Sansó (1998) for the individual odds ratios of the 22 studies. The values for the individual odds ratios do not correspond to those obtained by Smith *et al.* (1995), indicating the Mantel-Haenszel-Peto method has not been used. Different formulas to calculate the individual log-odds-ratios were implemented (equations 4.2, 4.3, 4.4 and 4.5) with none of them reproducing the results obtained by Abrams and Sansó (1998) exactly.

4.3.2 The Model Used

The model Abrams and Sansó (1998) used is slightly different to model 4.1. Abrams and Sansó (1998) use a random effects model, as displayed in model 4.6 with y_i corresponding to the summarised logarithm of the odds ratios, which are assumed to have normal distributions, rather than allowing the raw data to come from Binomial distributions. This is the Normal Normal model mentioned previously rather than the Binomial Normal model already stated.

$$y_i \sim N(\theta_i, \sigma_i^2/n_i) \tag{4.6}$$

$$\theta_i \sim N(\mu, \tau^2) \quad i = 1, ..., k$$

In model 4.6, n_i is the number of patients in either group, assuming the number in each group is equal, and

$$\sigma_i^2 = \frac{1}{p_{Ci}(1 - p_{Ci})} + \frac{1}{p_{Ti}(1 - p_{Ti})} \tag{4.7}$$

where p_{Ci} and p_{Ti} correspond to the proportion of events in the control group and the treatment group in trial$_i$ (p_{Ci} and p_{Ti} are calculated using a common n_i). If the number of patients in each group is not equal, several choices for n_i can be made instead: an average, the minimum, or the maximum number of patients in the treatment group and the control group, from $i = 1, 2, \ldots, k$. For this specific meta-analysis, the numbers of patients were not equal across the groups and therefore the average number of patients were used.

Using this Normal-Normal model within WinBUGS to obtain the Bayesian results should produce very similar results to the frequentist calculations which also assume a Normal-Normal model.

Model 4.6 can be rewritten as

$$\begin{gathered} y_i \sim N(\theta_i, \sigma_i^2) \\ \theta_i \sim N(\mu, \tau^2) \quad i = 1, ..., k \\ \text{with } \sigma_i^2 = \frac{1}{rt_i} + \frac{1}{n_i - rt_i} + \frac{1}{rc_i} + \frac{1}{n_i - rc_i} \end{gathered} \tag{4.8}$$

after some simple rearrangement, where n_i is the average number of patients in trial i, $i = 1, 2, \ldots, k$.

The priors used by Abrams and Sansó (1998) for μ and τ^2 were non-informative Uniform(-10, 10) and Inverse-Gamma(3, 1) priors respectively. An Inverse-Gamma(0, 2) prior for τ^2 was also implemented, presumably to demonstrate some of the results of the sensitivity analysis produced.

This Normal Normal model can be written in the Bayesian format with the

priors, as has previously been done for the Binomial Normal model (model 4.1) and is displayed in model 4.9.

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\theta_i, \sigma_i^2) \\
\theta_i &\sim N(\mu, \tau^2) \quad i = 1, ..., k \\
\mu &\sim \mathrm{Unif}(-10, 10) \\
\tau^2 &\sim \mathrm{Inv} - \mathrm{Gamma}(3, 1) \\
\text{where } \sigma_i^2 &= \frac{1}{rt_i} + \frac{1}{n_i - rt_i} + \frac{1}{rc_i} + \frac{1}{n_i - rc_i}
\end{aligned}
\tag{4.9}
$$

4.3.3 Looking at the Results

The results displayed in Table 4.2 include: the results obtained by Abrams and Sansó (1998) for the priors previously stated; the results one obtains by reproducing the published results using the same model, data and priors (with a burn in of 5000 and a further 25,000 iterations); the results one obtains using just a different vague prior for μ ($\mu \sim$ N(0.0, 1,000,000)); the results one obtains using just a different prior for τ^2 ($\tau \sim$ Unif(0, 10)); and the results one obtains using the different priors for μ and τ^2 together. Further vague priors were also used to ensure the sensitivity analysis results are suitable, however these are not displayed here.

The reproduced results in Table 4.2 for the expected values of the treatment effect estimate (μ) and the heterogeneity (τ^2) are notably different to the results

Parameter	Result in published data	Reproduced result	Result using a different vague prior for μ	Result using a different vague prior for τ^2	Result using different vague priors for μ and τ^2 together
E(μ)	-1.488	-1.254	-1.256	-1.281	-1.281
SD(μ)	0.230	0.1908	0.1915	0.2148	0.2143
E(τ^2)	1.090	0.4346	0.4381	0.5801	0.5738
SD(τ^2)	0.375	0.1937	0.1966	0.3347	0.3418

Table 4.2. Results for the published Bayesian analysis for model 4.6

published in Abrams and Sansó (1998). The expected value for the treatment effect estimate in Abrams and Sansó (1998) is -1.488, with an expected value of 1.090 for the heterogeneity, whereas the reproduced results using the same priors and data are -1.254 for the treatment effect estimate and 0.4346 for the heterogeneity. The sensitivity results for the model to reproduce the treatment effect estimate and the heterogeneity are quite consistent with the reproduced values, indicating the priors are non-informative and the results come from the data itself, however the underlying reason for the results not matching those published still has to be determined.

Upon further examination into the discrepancies, it appears similar results can be obtained (again using the same priors and the same data) if a slightly different formula for σ_i^2 (equation 4.10) is used.

$$\sigma_i^2 = 1/p_{Ci}(1 - p_{Ci}) + 1/p_{Ti}(1 - p_{Ti}), \tag{4.10}$$

Equation 4.10 looks similar to the equation used previously (equation 4.7),

however typing equation 4.10 into R exactly as it is written fails to give the expected result, that obtained via equation 4.7. This is due to the lack of brackets in equation 4.10 which determine the correct ordering of the elements of the equation.

Parameter	Result in published data	Reproduced result	Result using a different vague prior for μ	Result using a different vague prior for τ^2	Result using different vague priors for μ and τ^2 together
E(μ)	-1.488	-1.428	-1.453	-1.446	-1.439
SD(μ)	0.230	0.2118	0.215	0.245	0.2364
E(τ^2)	1.090	07612	0.7528	1.018	1.021
SD(τ^2)	0.375	0.2697	0.248	0.4105	0.415

Table 4.3. Results for the published Bayesian analysis for model 4.6 using equation 4.10

The expected treatment effect estimate and heterogeneity are again calculated as before with the only difference being the change of equation, from equation 4.7 to equation 4.10. The results are displayed in Table 4.3. One can clearly see from these results that the estimates produced for the treatment effect and the heterogeneity match the results published by Abrams and Sansó (1998) more than the previous results obtained using equation 4.7 for the calculation of the σ_i^2's.

4.4 Comparing Variances

As previously mentioned, upon attempting to reproduce the results obtained by Abrams and Sansó (1998) for their Bayesian Normal-Normal model, there were

some conflicting results. The reproduced overall treatment effect estimate was quite a bit smaller than the one obtained by Abrams and Sansó (1998) (Table 4.2), as was the heterogeneity estimate. It was then discovered that altering the formula for the σ_i^2's slightly by using equation 4.10 instead of equation 4.7, resulted in values for the treatment effect estimate and the heterogeneity which were closer to those produced by Abrams and Sansó (1998). This formula for the σ_i^2's however, cannot be rearranged into the usual variance formula,

$$\sigma_i^2 = \frac{1}{rt_i} + \frac{1}{nt_i - rt_i} + \frac{1}{rc_i} + \frac{1}{nc_i - rc_i}$$

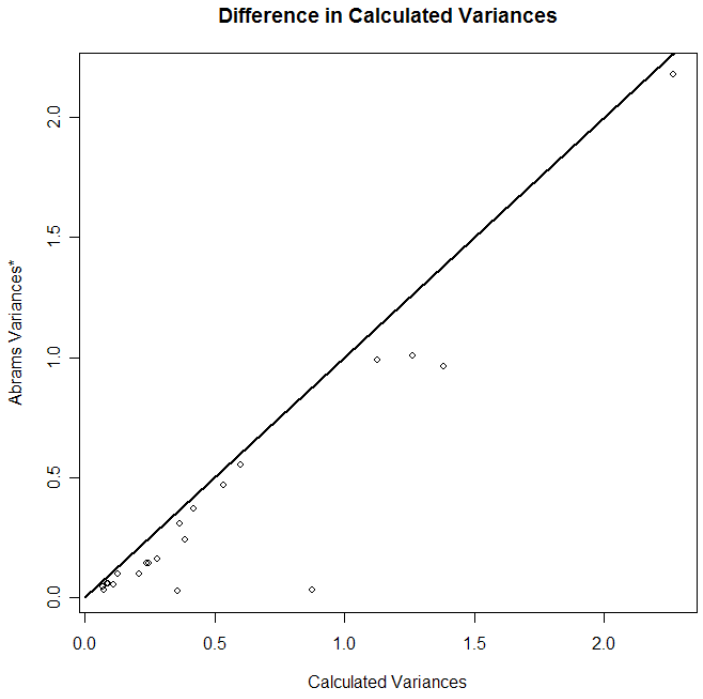

Figure 4.2. Comparing the variances calculated using Abrams and Sansó (1998) formula for the variances and the common formula for the variances.
*The variances calculated using the formula in Abrams and Sansó (1998)

Plotting the variances which have been calculated using the equation in Abrams and Sansó (1998) against the variances calculated using the usual formula for the variances (equation 4.7) and inserting a line of equality, one can see from the graph (Figure 4.2) that there is a notable difference between the two equations, with the variance estimates calculated using the usual equation being larger in magnitude.

Some of the estimates calculated using the equation in Abrams and Sansó (1998) correspond fairly well with those obtained using equation 4.7, however some estimates differ quite noticeably.

4.5 Bayesian and Frequentist Heterogeneity

4.5.1 Introduction

Having fitted two Bayesian models to the data, a Binomial-Normal model, model 4.1 and a Normal-Normal model, model 4.9, it will be of interest to examine the results produced via the frequentist methods and the results produced via the Bayesian methods to determine which differences occur, if any. It seems reasonable to compare the Bayesian and the frequentist methods individually using both the Normal-Normal model and the Binomial-Normal model.

4.5.2 The Binomial-Normal Model

All of the Bayesian results, a sample of which are displayed in Table 4.4, used a burn in of 5,000 and then a further 25,000 iterations for each of the two

Markov chains. Two Markov chains were used to check visually that convergence of each parameter did not produce any unexpected problems. The frequentist values for the random effects between study variation, also displayed in Table 4.4 are calculated using a nlmixed macro in SAS which incorporates the Binomial Normal design used for the Bayesian model. The Bayesian results are obtained using 4.1.

Data No.	No. of Clinical Trials	Average no. of patients in each trial	Bayesian Mean of Variation	Bayesian S.D. of Variation	Frequentist value for the between-study variation	95% C. I. for Bayesian Estimate of variation
23	51	311	0.01192	0.01782	0	(0.00052, 0.06346)
116	34	610	0.02528	0.03107	0	(0.00082, 0.1109)
123	42	356	0.00746	0.007848	0	(0.00057, 0.02859)
79	23	87	0.2243	0.1463	0.0843	(0.02736, 0.588)
86	41	22	2.452	0.8812	1.4639	(1.178, 4.541)
115	62	49	0.121	0.06577	0	(0.03534, 0.2876)
26	6	1093	0.01317	0.02799	0	(0.00053, 0.07176)
70	6	616	0.6199	1.228	0.1750	(0.04376, 3.059)
74	6	6082	0.0374	0.0704	0.0072	(0.00133, 0.184)
41	6	68	1.157	1.566	0.4896	(0.159, 4.569)
107	6	61	0.04834	0.1635	0	(0.00066, 0.3164)
113	6	64	3.996	5.363	1.4226	(0.4144, 16.59)

Table 4.4. Comparing the frequentist and Bayesian methods for the heterogeneity (using model 4.1), with groupings to emphasise both the number of trials in each Meta-analysis and the average number of patients within each trial

The results produced in table 4.4 for meta-analyses which have a large number of trials as well as a large number of patients within each trial on average (the

first three rows of results) indicate the method used to calculate the frequentist heterogeneity values seems to provide a fairly accurate estimate for the heterogeneity, with the results corresponding fairly well to the Bayesian heterogeneity results produced for the meta-analyses using WinBUGS. Both the Bayesian method and the frequentist method produce a value close to zero for these three meta-analyses.

There are however some differences in the estimates for the frequentist approach and the Bayesian approach when the number of trials is fairly large and the number of patients in each trial is quite small on average which one would not expect. For example, data number 86 which has 41 clinical trials with an average of 22 patients in each trial has a produced frequentist estimate for the heterogeneity of 1.4639 whilst the Bayesian estimate is quite a bit higher (2.452), although the frequentist estimate is still contained within the 95% Confidence Interval for the Bayesian heterogeneity (1.178, 4.541). For meta-analyses with large numbers of trials, one would generally expect the frequentist value for the heterogeneity to produce similar results to those obtained computing the heterogeneity via Bayesian methods. There are also differences between the Bayesian and the frequentist heterogeneity estimates when the number of trials is small, although all of the frequentist estimates, apart from those which are calculated as zero exactly, are included in the corresponding Bayesian heterogeneity confidence intervals.

One can note however, that all of the Bayesian results are slightly higher than the frequentist results and the greater the frequentist heterogeneity result, the

greater the difference between the heterogeneity estimates.

The Bayesian results for the treatment effect estimates and the heterogeneity using the Binomial-Normal model (model 4.1) can now be examined graphically against the corresponding values obtained via the frequentist method, Figure 4.3, providing some indication as to whether the results do coincide as one believes they should. Assuming the results agree exactly, one would anticipate the points to lie on the line of equality, the dashed line on both of the plots. Both plots use the number of studies as the grouping variable, with the legend for the labelling of the groups beside Figure 4.3.

Examining Figure 4.3, it appears the Bayesian method and the frequentist method correspond well using the Binomial-Normal model to calculate the treatment effect estimates with all of the estimates lying nearly perfectly along the line of equality. Producing a linear regression results in a slope parameter of 1.01 with an intercept close to zero, -0.0112.

The Bayesian heterogeneity estimates for the Binomial Normal model however do not correspond as well to the frequentist heterogeneity estimates using the nlmixed macro in SAS for the Bayesian Normal model, with hardly any of the estimates lying on the line of equality. In all of the cases where the Bayesian and frequentist heterogeneity estimates differ, the Bayesian estimates produced are always greater than those obtained via the frequentist method. In fact, fitting a linear regression to the estimates, the Bayesian heterogeneity estimates are equal to 0.107 plus 2.04 times the frequentist heterogeneity estimates.

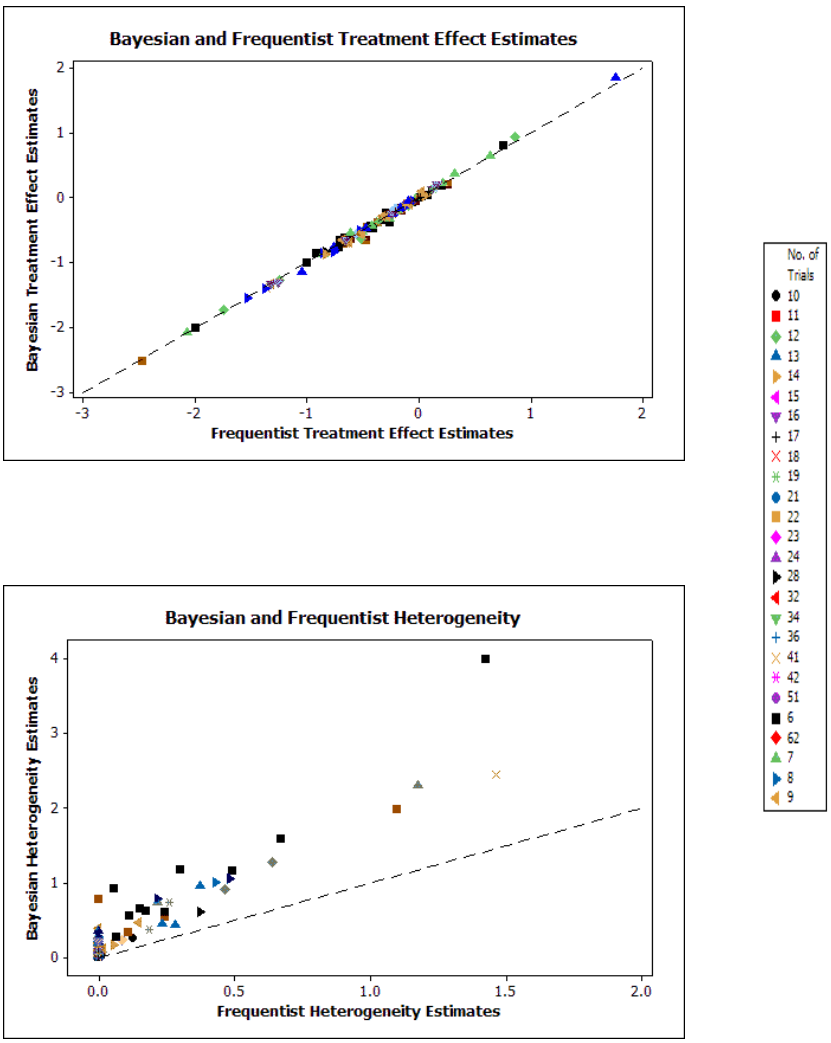

Figure 4.3. Comparing the Bayesian estimates with the frequentist estimates (calculated using a macro in SAS) for the Binomial-Normal model

4.5.3 The Normal-Normal Model

Previously, the results from the Bayesian Binomial-Normal model (model 4.1) were compared to the frequentist values. Here, we compare the values produced for the heterogeneity from the Bayesian Normal-Normal model (model 4.9) and the frequentist results. Non-informative priors were used for the overall treatment effect ($\mu \sim$ N(0.0, 1,000,000)) and the heterogeneity ($\tau^2 \sim$ IG(0.001, 0.001)).

Note model 4.9 uses the common n_i's to calculate the σ_i^2's, however the results produced in Table 4.5 use the given n_{Ci}'s and n_{Ti}'s, so the σ_i^2 here are calculated using

$$\sigma_i^2 = \frac{1}{rt_i} + \frac{1}{nt_i - rt_i} + \frac{1}{rc_i} + \frac{1}{nc_i - rc_i}$$

All of the Bayesian results, displayed in Table 4.5 used a burn in of 5, 000 and then a further 25, 000 iterations for each of the two Markov chains. Once again, two Markov chains were used to check visually that convergence of each parameter did not produce any unexpected problems. The frequentist values for the random effects between study variation, also displayed in Table 4.5 are calculated using the moment estimate as used by DerSimonian and Laird (1986), equation 2.2.

The results produced in table 4.5 indicate the Bayesian results using a Normal-Normal model correspond quite well across the selected meta-analyses with the two different frequentist results obtained via the methods used by DerSimonian and Laird (1986) (equation 2.2) and Hardy and Thompson (1996) (equation 2.3). All of the frequentist heterogeneity estimates (excluding those estimated as zero) are contained within the 95% C.I. for the Bayesian heterogeneity estimate. The confidence intervals for the Bayesian heterogeneity estimates for the meta-analyses for which the frequentist heterogeneity estimates are zero all have a lower estimate very close to zero (zero to two decimal places).

For meta-analyses with large numbers of trials and a large number of patients

Data No.	No. of Clinical Trials	Average no. of patients in each trial	Bayesian Mean of Variation	Bayesian S.D. of Variation	Frequentist value for the between-study variation (DL)	Frequentist value for the between-study variation (HT)	95% C. I. for Bayesian Estimate of variation
23	51	311	0.00994	0.01474	0	0	(0.00054, 0.05093)
116	34	610	0.01916	0.0252	0.015	0.0056	(0.0007, 0.0887)
123	42	356	0.00674	0.00723	0.0085	0.0022	(0.00051, 0.02664)
79	23	87	0.1744	0.1368	0.17	0.1672	(0.00418, 0.5193)
86	41	22	0.9552	0.4972	0.97	0.8693	(0.2529, 2.146)
115	62	49	0.03323	0.04142	0.064	0.0363	(0.00071, 0.1487)
26	6	1093	0.01226	0.02464	0	0	(0.00052, 0.06718)
70	6	616	0.5791	1.461	0.29	0.1992	(0.03452, 2.7)
74	6	6082	0.03701	0.07253	0.021	0.0117	(0.00124, 0.1796)
41	6	68	1.116	1.793	0.85	0.5547	(0.1355, 4.286)
107	6	61	0.04491	0.1237	0	0	(0.00063, 0.296)
113	6	64	2.938	4.475	1.91	1.4896	(0.03306, 13.27)

Table 4.5. Comparing the DerSimonian and Laird frequentist method, the Hardy and Thompson method, and the Bayesian method for the heterogeneity (using model 4.8), with emphasis on the number of trials in each Meta-analysis and the average number of patients within each trial

in each trial on average, one would generally expect the frequentist values for the heterogeneity to produce similar results to those obtained computing the heterogeneity via Bayesian methods, as seen here. It is interesting to note however that when the heterogeneity is zero, or is close to zero, the Bayesian and frequentist values appear to agree fairly well, although more so if the number of clinical trials is large. If the heterogeneity is not close to zero however, it appears the Bayesian method produces a much larger estimate for the heterogeneity than the frequentist method does.

The method used to calculate the frequentist heterogeneity values, equation 2.2, as used by DerSimonian and Laird (1986) seems to provide an accurate estimate for the heterogeneity from a frequentist view with the results corresponding well to the frequentist heterogeneity results produced for the meta-analyses using the maximum likelihood approach as used by Hardy and Thompson (1996), equation 2.3.

Plotting the Bayesian results for the treatment effect estimates and the heterogeneity using the Normal-Normal model (model 4.9, with $\sigma_i^2 = \frac{1}{rt_i} + \frac{1}{nt_i - rt_i} + \frac{1}{rc_i} + \frac{1}{nc_i - rc_i}$) against the corresponding values obtained via the two frequentist methods, Figures 4.4, 4.5, and 4.6, will indicate whether the results do coincide as one believes they should. Assuming the results agree exactly, one would anticipate the points to lie on the line of equality, the dashed line in each of the six plots. All of the plots use the number of studies as the grouping variable, with the legend for the labelling of the groups beside each figure.

The first figure, Figure 4.4 beolw, indicates the estimates for the treatment

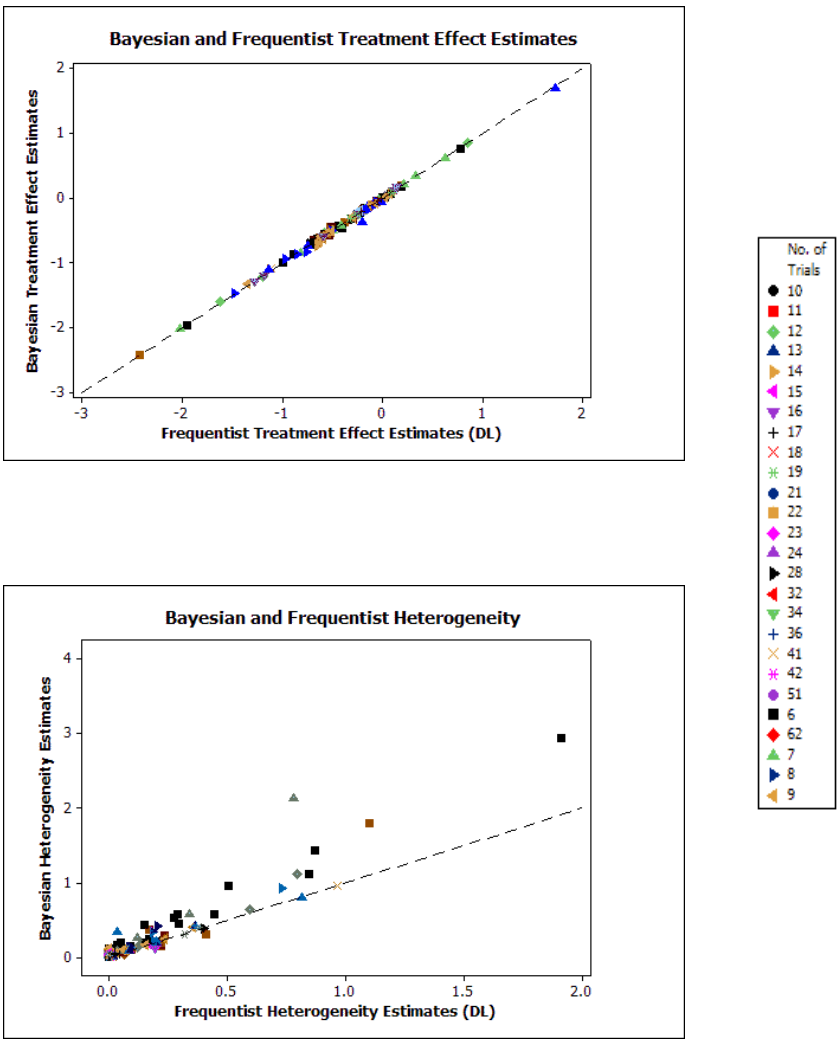

Figure 4.4. Comparing the Bayesian estimates with the frequentist estimates calculated using DerSimonian and Laird's method (equation 2.2)

effect using the Bayesian model (model 4.9 without using an averaged n_i) coincide almost exactly with the treatment effect estimates obtained using the method used by DerSimonian and Laird (1986), with the estimates barely deviating from the line of equality. Examining now the scatterplot for the heterogeneity values calculated, it appears the Bayesian estimates are in general slightly greater than those obtained using the frequentist method. Fitting a regression line to the

heterogeneity estimates, the Bayesian heterogeneity values are equal to 0.0193 plus 1.43 times the frequentist heterogeneity values. The slope parameter, 1.43, confirms the initial impression that the Bayesian estimates for the heterogeneity are greater than those obtained via the frequentist method.

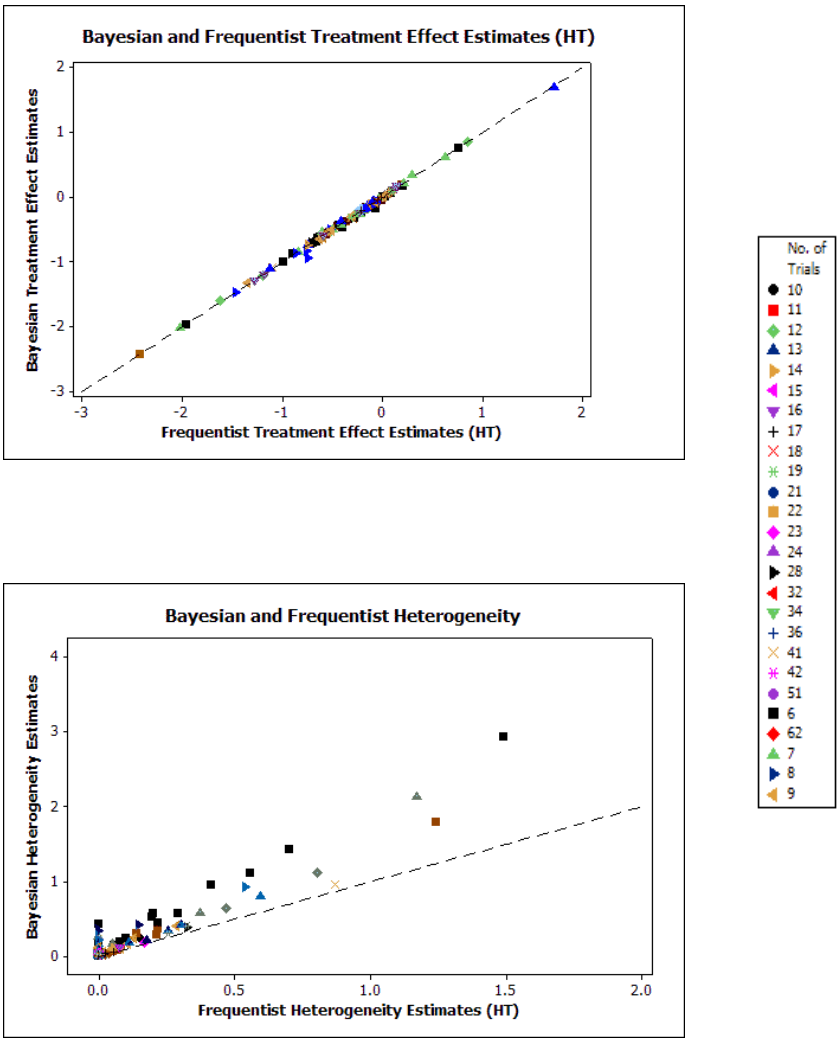

Figure 4.5. Comparing the Bayesian estimates with the frequentist estimates calculated using Hardy and Thompson's method (equation 2.3)

Figure 4.5, which compares the Bayesian estimates with the frequentist estimates calculated using the method used by Hardy and Thompson (1996) produces

similar results to those noted from Figure 4.4. Like Figure 4.4, the treatment effect estimates computed using the Bayesian method, model 4.9, match almost identically those computed using the method as used by Hardy and Thompson (1996). The heterogeneity values however appear to differ between the two methods with the Bayesian values again being slightly greater than the corresponding heterogeneity values. The fitted regression for these two methods, Bayesian = 0.0431 + 1.61*frequentist, again confirms this, with a slightly larger slope parameter, 1.61.

Comparing now the two frequentist methods (DerSimonian and Laird (1986) and Hardy and Thompson (1996)) displayed in Figure 4.6, the treatment effect estimates appear to agree in almost every meta-analysis with just a small difference occurring in the meta-analyses which do not agree. The estimates for the heterogeneity do not agree as well as noted in the treatment effect estimates, however fitting a regression line for the estimates computed using the method used by DerSimonian and Laird (1986) against the method used by Hardy and Thompson (1996), the intercept produced is 0.0249, with a slope parameter of 1.06, which is fairly close to one.

Figures 4.4, 4.5 and 4.6 indicate that no two methods of the three considered agree exactly with any other 100% of the time, however of the three methods, the two which match the closest appear to be the two frequentist methods, especially for the heterogeneity estimates, with the Bayesian heterogeneity estimates appearing to have slightly larger estimates than either of the two frequentist methods.

It is apparent that the results via both the frequentist methods and the Bayesian method for the treatment effect estimates correspond fairly well in the majority of the meta-analyses. In the event of a difference occurring, the magnitude of the difference is minimal.

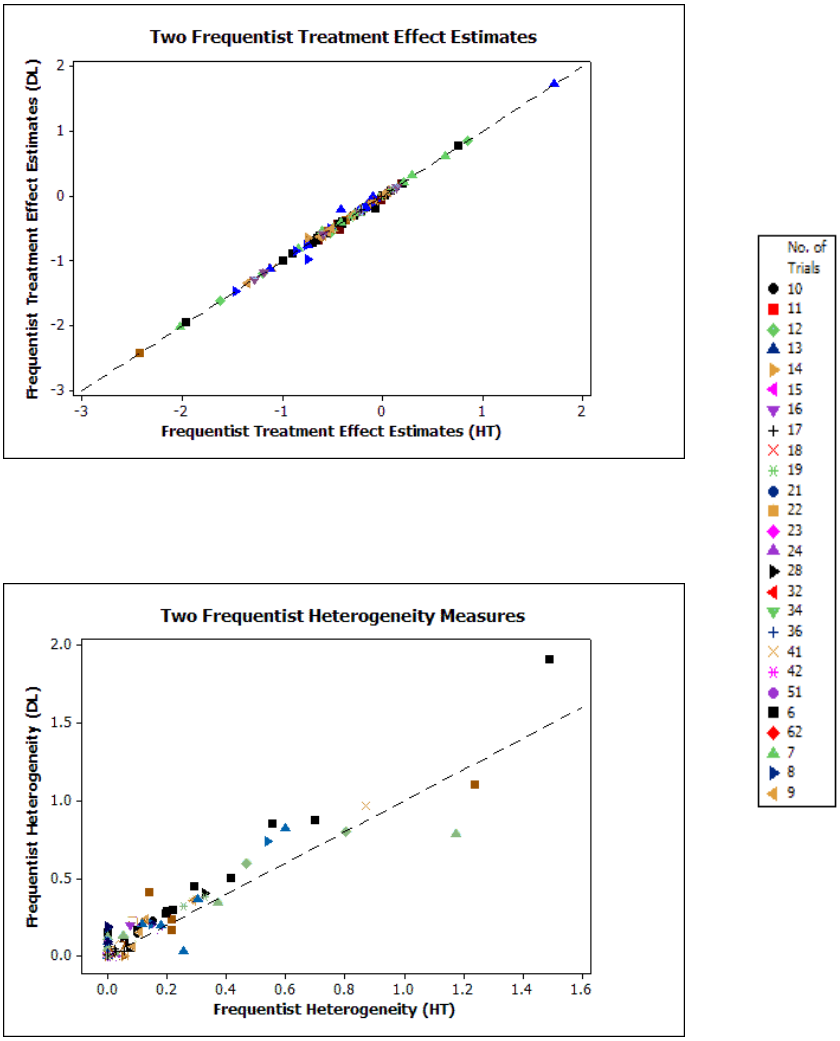

Figure 4.6. Comparing the two frequentist approaches using estimates calculated using DerSimonian and Laird's method (equation 2.2) and Hardy and Thompson's method (equation 2.3)

On examination of the results for the heterogeneity, there is more evidence of disagreements occurring between the different methods. The Bayesian method

appears to produce slightly larger measures for the heterogeneity than either the DerSimonian and Laird method for calculating the heterogeneity or the method of maximised likelihood as used by Hardy and Thompson.

4.5.4 Summary

Two different Bayesian models, a Binomial-Normal model and a Normal-Normal model were fitted to the data for all 125 meta-analyses, along with the corresponding frequentist models using a macro within SAS (nlmixed) for the Binomial-Normal model and two different methods for calculating the heterogeneity in a Normal-Normal model (DerSimonian and Laird (1986) and Hardy and Thompson (1996)). The results produced were then examined using different methods and models. Since each model used the same data, one would expect the results to be very similar, irrespective of which method was used to obtain them. The results produced for both the overall treatment effect and the heterogeneity for each meta-analysis were plotted in separate scatter plots allowing the results from the different methods used to be compared.

The scatter plots indicated the overall treatment effect estimates for each meta-analysis were very similar regardless of the method used to calculate them. The estimates of the heterogeneity however differed slightly depending on which method was used to estimate them, with the two frequentist methods used for the Normal-Normal model (DerSimonian and Laird (1986) and Hardy and Thompson (1996)) producing the most similar results. These results indicate when estimating the heterogeneity for a meta-analysis the results can differ slightly depending

on which method one has used. The estimates produced using a Bayesian method tend to be slightly larger than the corresponding estimates produced using a frequentist method.

4.6 Comparing the Models

It is of interest now to compare all of the results produced for the data in Table 4.1 for the different models, the Binomial-Normal and the Normal-Normal model, as well as comparing the results that have been produced from the two different methods, the Bayesian method and the frequentist methods. For simplicity here, and since the previous results (Figure 4.6) indicated the two frequentist methods produced similar results, the results from only one frequentist method are produced here, that of DerSimonian and Laird (1986).

	Model	
	Binomial-Normal Model	Normal-Normal Model
Frequentist	$E(\delta)$ = -1.3944 $SD(\delta)$ = 0.1929 $E(\tau^2)$ = 0.4275	$E(\delta)$ = -1.3387 $SD(\delta)$ = 0.1863 $E(\tau^2)$ = 0.4428
Bayesian	$E(\delta)$ = -1.419 $SD(\delta)$ = 0.2065 $E(\tau^2)$ = 0.555 $SD(\tau^2)$ = 0.24	$E(\delta)$ = -1.346 $SD(\delta)$ = 0.202 $E(\tau^2)$ = 0.5037 $SD(\tau^2)$ = 0.2234

Table 4.6. Results for the data used by Abrams and Sansó (1998) using the two models 4.1 and 4.9 for the two different methods, the frequentist and the Bayesian.

A summary of the overall treatment effect estimate and the heterogeneity for

the two different models and the two different methods are produced in Table 4.6. The Bayesian model used for the Normal-Normal model is model 4.9, with the σ_i^2's calculated as in equation 4.8, whilst the Bayesian Binomial-Normal model results are produced using model 4.1, with the non-informative priors which Abrams and Sansó (1998) used for μ (Uniform(-10, 10))and σ_i^2 (Inverse-Gamma(3, 1)).

Examining the results produced in Table 4.6, the estimates for the overall treatment effect are most similar within the columns, i.e. there is a larger difference in the size of the overall treatment effect estimate when comparing the different types of models, the Binomial-Normal with the Normal-Normal, than if one compares the different methods (the frequentist and the Bayesian) for each model.

The estimate for the heterogeneity however, indicates the estimates are most similar within the rows, i.e. there is a larger difference in the size of the heterogeneity when comparing the different methods (the frequentist and the Bayesian) than if one compares the different models, the opposite of what was indicated from the estimate of the overall treatment effect.

The results for the different estimates of the heterogeneity are not what one would expect. One would believe any differences which occur in the calculation of these estimates would be due to the decision of which type of model used to calculate the estimates and expect the results to be very similar if not the same regardless of which method was used. In actual fact, it appears that the results for the treatment effect estimate agree with this reasoning, yet the results for the heterogeneity indicate that regardless of which model is used, the Bayesian

results for the heterogeneity will always be greater than those corresponding to the frequentist methods.

4.7 Fitting an Informative Prior

4.7.1 Introduction

So far in this thesis, only non-informative priors which assume no prior knowledge of how a variable will behave have been implemented into the Bayesian models, however informative priors are also commonly used.

In a meta-analysis context, if one is creating a model to produce estimates for the overall treatment effect and the heterogeneity, as is the case here, some believe an informative prior should be used for the heterogeneity since there is in fact some prior knowledge about it.

For example, since the heterogeneity is a measure of variance, it would seem reasonable to expect it to take values in the interval $[0, \infty)$ and so a prior which restricts the values produced to within this interval could be appropriate. (Spiegelhalter *et al.*, 2004)

Senn (2007*b*) wrote a response to Lambert *et al.* (2005) which claimed "no applied statistician believes that [the treatment effect and the heterogeneity] are independent", in which case non-informative priors on both the treatment effect and the heterogeneity would not be of use either. Upon examination of the relationship between the treatment effect and the heterogeneity earlier in this thesis (Chapter 3), there did appear to be a relationship between the two, with a larger

range of heterogeneity values occurring for larger treatment effects.

4.7.2 Creating a New Model

$$f(\tau|\mu;\beta,\alpha) = \frac{1}{\alpha+\beta|\mu|}\, exp\left\{-\frac{\tau}{\alpha+\beta|\mu|}\right\} \tag{4.11}$$

If now we consider creating a new informative prior which takes into account the prior information that the heterogeneity should in fact be restricted to the interval $[0,\infty)$ as well as the treatment effect and the heterogeneity have some sort of relationship, we might consider a prior which Senn (2007*b*) proposed, as mentioned earlier.

$$\begin{gathered}
r_{c_i} \sim \text{Bin}(n_{c_i}, p_{c_i}) \\
r_{t_i} \sim \text{Bin}(n_{t_i}, p_{t_i}) \\
\text{logit}(p_{c_i}) = \psi_i \\
\text{logit}(p_{t_i}) = \psi_i + \delta_i \\
\delta_i \sim N(\mu, \tau^2) \\
\psi_i \sim \text{N}(0, 1000000) \\
\mu \sim \text{N}(0, 1000000) \\
\tau \sim \text{Exp}(\lambda) \\
\lambda = \frac{1}{\alpha+\beta|\mu|}
\end{gathered} \tag{4.12}$$

In his paper, Senn (2007*b*) proposed an exponential prior for τ using the "conditional prior distribution", noted here in equation 4.11.

This is an exponential prior which "allows dependence of τ on μ [the pooled log odds ratio] (Senn, 2007*b*)" and the resulting new model is displayed in model 4.12 for using this prior with the Binomial Normal model and displayed in model 4.13 for using this prior with the Normal Normal model.

In models 4.12 and 4.13, α and β are both given constants, with Senn (2007*b*) suggesting values of $\beta < 1$ and small values of α. Here, a value for α of 0.05 and a value for β of 0.1 is used.

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\theta_i, \sigma_i^2) \\
\theta_i &\sim N(\mu, \tau^2) \quad i = 1, ..., k \\
\mu &\sim \mathrm{Unif}(-10, 10) \\
\tau &\sim \mathrm{Exp}(\lambda) \\
\lambda &= \frac{1}{\alpha + \beta|\mu|}
\end{aligned}
\tag{4.13}
$$

Using the data from Abrams and Sansó (1998) which has previously been used for the Binomial Normal model (model 4.1) and the Normal Normal model (model 4.9) with non-informative priors, one can implement these models again, this time with the informative exponential prior (models 4.12 and 4.13).

4.7.3 Examining the Results

The results produced using this new informative exponential prior, for both the Binomial Normal Bayesian model (model 4.12) and the Normal Normal Bayesian model (model 4.13) are displayed here in Table 4.7.

	Model	
	Binomial-Normal Model	Normal-Normal Model
Bayesian	E(μ) = -1.461 SD(μ) = 0.2099 E(τ^2) = 0.5278 SD(τ^2) = 0.2655	E(μ) = -1.37 SD(μ) = 0.2075 E(τ^2) = 0.4508 SD(τ^2) = 0.2419

Table 4.7. Results for the data used by Abrams and Sansó (1998) using the two models with the informative exponential prior (models 4.12 and 4.13) for the Bayesian method.

One can see from Table 4.7 that although the Binomial Normal results are not hugely different to the results obtained using the Normal Normal model, the Binomial Normal results are larger for both the overall treatment effect estimate and the heterogeneity. This is consistent with the results found for the two Bayesian models using the non-informative prior for the heterogeneity.

It is interesting to note that on using the informative exponential prior, the mean value for μ is slightly larger than the equivalent results using the non-informative priors, however the mean value for τ^2 is slightly smaller than the equivalent results using the non-informative priors.

A sensitivity analysis for different values of α and β (α small and β <1) in models 4.12 and 4.13 were also computed to ensure the prior used was not

sensitive to the choices of α and β.

Ideally, one would like to be able to produce results using models 4.12 and 4.13 with the values for α and β coming from a distribution rather than using set constants.

Altering model 4.12 slightly to illustrate how this would be achieved, the α and the β have been replaced with $\beta[1]$ and $\beta[2]$ respectively to allow them to come from a bivariate log-Normal distribution, as displayed in model 4.14.

$$
\begin{aligned}
r_{c_i} &\sim \text{Bin}(n_{c_i}, p_{c_i}) \\
r_{t_i} &\sim \text{Bin}(n_{t_i}, p_{t_i}) \\
\text{logit}(p_{c_i}) &= \psi_i \\
\text{logit}(p_{t_i}) &= \psi_i + \delta_i \\
\delta_i &\sim N(\mu, \tau^2) \\
\psi_i &\sim \text{N}(0, 1000000) \\
\mu &\sim \text{N}(0, 1000000) \\
\tau &\sim \text{Exp}(\lambda) \\
\lambda &= \frac{1}{\text{expbeta}[1] + \text{expbeta}[2]|\mu|} \\
\beta[1:2] &\sim \text{MN}(\mu_2, \Sigma) \\
\mu_2 &\sim \text{MN}(\mu_3, \Sigma_2) \\
\text{expbeta}[i] &= \exp(\beta[i]) \\
\Sigma &\sim \text{Wishart}_2(\Omega)
\end{aligned}
\tag{4.14}
$$

To allow $\beta[1]$ and $\beta[2]$ to come from a bivariate log-normal distribution, the exponentials of the β_i's need to be used in the calculation of λ, the parameter for the exponential prior on τ. The Bivariate Normal distribution for the β's has parameters μ_2, a vector of length 2 which comes from a Bivariate Normal, and Σ, a 2 x 2 matrix which has a Wishart distribution here.

The priors placed on μ_3, Σ_2 and Ω for model 4.17 above are restricted so that the values obtained for expbeta[1] and expbeta[2] are restricted to follow the suggested values of expbeta[1] small and expbeta[2] < 1 as discussed in Senn (2007*b*).

Implementing model 4.14 using again the data from Abrams and Sansó (1998), the matrix parameter for the Wishart distribution Ω is diag(1), μ_3 is a zero vector of length 2 and Σ_2 is a diag(100) matrix. The results are displayed in Table 4.8.

One can see on comparison with the previous results which used given values for the hyperparameters, α and β (0.05 and 0.1 respectively), the computed values for the treatment effect, μ, and the heterogeneity, τ^2 are again quite similar.

	Model
	Binomial-Normal Model
Bayesian	E(μ) = -1.446
	SD(μ) = 0.2332
	E(τ^2) = 0.7626
	SD(τ^2) = 0.4099

Table 4.8. Results for the data used by Abrams and Sansó (1998) using the informative exponential prior for τ with non-informative hyperparameters (model 4.14).

The value for the treatment effect, -1.446, remains close to the estimates reported in Table 4.7, whilst the value for the heterogeneity, 0.7626, is just slightly greater than both of the previous results for the heterogeneity in Table 4.7. Although the size of the heterogeneity estimate is fairly similar to the previous estimates, it could be slightly bigger due to the model being more complex and only the same amount of information being available.

4.7.4 Summary

As previously discussed in Chapter 3, there appears to be a relationship between the overall treatment effect of a meta-analysis and its corresponding heterogeneity. Senn (2007*b*) had also previously considered this and suggested the use of an exponential prior on τ (the square root of the heterogeneity) which was also conditional on the overall treatment effect for that meta-analysis (equation 4.11). Senn (2007*b*) suggested the values for α to be small and $\beta < 1$. This new prior for τ was implemented using both a Binomial-Normal model (model 4.12) and a Normal-Normal model (model 4.13). The results were fairly consistent with the results previously calculated in Section 4.3, with the Binomial-Normal model having a slightly larger value for both the overall treatment effect estimate and the heterogeneity.

Hyperparameters were then added to allow the estimates for α and β to be obtained from their own parameters. A Bivariate log-Normal distribution for α and β was implemented with similar results to the previous results from the model with the exponential prior on τ but with given values for α and β. The

obtained value for the heterogeneity was slightly larger, although that could be due to a more complex model being fitted with no more additional data.

4.8 A Meta-Analysis of Meta-Analyses

4.8.1 Introduction

So far, meta-analyses of various clinical trials concerning the same treatment have been examined using two approaches, the frequentist and the Bayesian. It is of interest now however to consider whether or not a meta-analysis of meta-analyses can be computed using a Bayesian hierarchical model.

4.8.2 Creating the Model

To produce a meta-analysis of meta-analyses, the model will follow a hierarchical format with k clinical trials ($j = 1, \ldots, k$) and m meta-analyses ($i = 1, \ldots, m$), which is displayed in model 4.15.

This model allows the treatment effect, $\delta_{i,j}$ in model 4.15, to differ by both meta-analysis and clinical trial. The μ_i's correspond to the overall treatment effect for meta-analysis i, whilst the τ_i^2's correspond to the heterogeneity for meta-analysis i. The non-informative priors on each μ_i and each τ_i^2 allow these estimates to vary independently between each meta-analysis assuming there is no prior knowledge on how these parameters should behave.

$$
\begin{aligned}
r_{c_{i,j}} &\sim \text{Bin}(n_{c_{i,j}}, p_{c_{i,j}}) \\
r_{t_{i,j}} &\sim \text{Bin}(n_{t_{i,j}}, p_{t_{i,j}}) \\
\text{logit}(p_{c_{i,j}}) &= \psi_{i,j} \\
\text{logit}(p_{t_{i,j}}) &= \psi_{i,j} + \delta_{i,j} \\
\psi_{i,j} &\sim \text{N}(0, \eta_{\text{i}}) \\
\eta_i &\sim \text{IG}(0.001, 0.001) \\
\delta_{i,j} &\sim N(\mu_i, \tau_i^2) \\
\mu_{\text{i}} &\sim \text{N}(0, 1000000) \\
\tau_i^2 &\sim \text{IG}(0.001, 0.001)
\end{aligned}
\tag{4.15}
$$

4.8.3 Examining the Results

Using the data which has been provided for 125 independent meta-analyses, some, or even all of the data available, can be used with this new model, model 4.15 displayed above.

Selecting a subset of the meta-analyses available, for example all 11 of the meta-analyses which have the field myocardial infarction, these can be used with the latest model, model 4.15, the results of which are displayed in Table 4.9.

Examining the results from the meta-analysis of the 11 myocardial infarction meta-analyses, the meta-analyses with a large absolute treatment effect (μ_i) also have a corresponding large variance (τ_i^2) as one would expect. In table 4.9,

Number of trials in MA i	μ_i		τ_i^2	
51	$E(\mu_1)$ = -0.1411	$SD(\mu_1)$ = 0.0635	$E(\tau_1^2)$ = 0.0105	$SD(\tau_1^2)$ = 0.0163
17	$E(\mu_2)$ = -0.2201	$SD(\mu_2)$ = 0.0622	$E(\tau_2^2)$ = 0.0179	$SD(\tau_2^2)$ = 0.0213
16	$E(\mu_3)$ = 0.0877	$SD(\mu_3)$ = 0.1358	$E(\tau_3^2)$ = 0.0290	$SD(\tau_3^2)$ = 0.0636
6	$E(\mu_4)$ = 0.00096	$SD(\mu_4)$ = 0.0716	$E(\tau_4^2)$ = 0.0126	$SD(\tau_4^2)$ = 0.0326
7	$E(\mu_5)$ = -0.2798	$SD(\mu_5)$ = 0.1147	$E(\tau_5^2)$ = 0.0250	$SD(\tau_5^2)$ = 0.0676
13	$E(\mu_6)$ = -0.178	$SD(\mu_6)$ = 0.0801	$E(\tau_6^2)$ = 0.0192	$SD(\tau_6^2)$ = 0.0377
11	$E(\mu_7)$ = -0.7692	$SD(\mu_7)$ = 0.319	$E(\tau_7^2)$ = 0.5516	$SD(\tau_7^2)$ = 0.5249
13	$E(\mu_8)$ = -0.5274	$SD(\mu_8)$ = 0.2567	$E(\tau_8^2)$ = 0.4658	$SD(\tau_8^2)$ = 0.3883
7	$E(\mu_9)$ = -0.6276	$SD(\mu_9)$ = 0.3551	$E(\tau_9^2)$ = 0.2446	$SD(\tau_9^2)$ = 0.6125
12	$E(\mu_{10})$ = -0.2517	$SD(\mu_{10})$ = 0.1052	$E(\tau_{10}^2)$ = 0.0203	$SD(\tau_{10}^2)$ = 0.0396
34	$E(\mu_{11})$ = -0.2856	$SD(\mu_{11})$ = 0.0683	$E(\tau_{11}^2)$ = 0.0297	$SD(\tau_{11}^2)$ = 0.0371

Table 4.9. Results for the Meta-Analysis of the 11 Meta-Analyses which all have the same field, myocardial infarction, using non-informative priors.

meta-analyses 7, 8 and 9 all have quite a large absolute treatment effect and corresponding variance which is also quite large compared to the variances of the meta-analyses with smaller absolute treatment effects.

4.8.4 A More Complex Model

The previous model (model 4.15) has been created using non-informative priors, however these can be changed to informative priors if one believes there is some prior information, as has been demonstrated previously in Chapter 4 section 7 (model 4.12).

The informative prior used previously for τ (the square root of the heterogeneity) in model 4.12 was an exponential prior and this may be used here instead of

the non-informative Inverse-Gamma prior for the heterogeneity.

Computing this more complex model to now include an informative exponential prior, as noted in model 4.16 below, the same subset of meta-analyses which was used for the previous model, model 4.15, can be used to compute results for implementing a more informative prior.

$$
\begin{aligned}
r_{c_{i,j}} &\sim \text{Bin}(n_{c_{i,j}}, p_{c_{i,j}}) \\
r_{t_{i,j}} &\sim \text{Bin}(n_{t_{i,j}}, p_{t_{i,j}}) \\
\text{logit}(p_{c_{i,j}}) &= \psi_{i,j} \\
\text{logit}(p_{t_{i,j}}) &= \psi_{i,j} + \delta_{i,j} \\
\psi_{i,j} &\sim \text{N}(0, \eta_{\text{i}}) \\
\eta_i &\sim \text{IG}(0.001, 0.001) \\
\delta_{i,j} &\sim N(\mu_i, \tau_i^2) \\
\mu_{\text{i}} &\sim \text{N}(0, 1000000) \\
\tau_i &\sim \text{Exp}(\lambda_{\text{i}}) \\
\lambda_i &= \frac{1}{\alpha + \beta|\mu_i|}
\end{aligned}
\tag{4.16}
$$

Model 4.16 uses the same restrictions for the values of α and β that model 4.12 used, with α and β taking the same values here, 0.05 and 0.1 respectively.

Looking at the results produced using model 4.16, displayed in Table 4.10

Number of trials in MA i	μ_i		τ_i^2	
51	$E(\mu_1)$ = -0.1384	$SD(\mu_1)$ = 0.0609	$E(\tau_1^2)$ = 0.0035	$SD(\tau_1^2)$ = 0.0070
17	$E(\mu_2)$ = -0.2212	$SD(\mu_2)$ = 0.0557	$E(\tau_2^2)$ = 0.0081	$SD(\tau_2^2)$ = 0.0116
16	$E(\mu_3)$ = 0.0925	$SD(\mu_3)$ = 0.1105	$E(\tau_3^2)$ = 0.0054	$SD(\tau_3^2)$ = 0.0123
6	$E(\mu_4)$ = 0.0026	$SD(\mu_4)$ = 0.0595	$E(\tau_4^2)$ = 0.0034	$SD(\tau_4^2)$ = 0.0071
7	$E(\mu_5)$ = -0.2714	$SD(\mu_5)$ = 0.0991	$E(\tau_5^2)$ = 0.0075	$SD(\tau_5^2)$ = 0.0149
13	$E(\mu_6)$ = -0.1767	$SD(\mu_6)$ = 0.0694	$E(\tau_6^2)$ = 0.0054	$SD(\tau_6^2)$ = 0.0106
11	$E(\mu_7)$ = -0.7742	$SD(\mu_7)$ = 0.2877	$E(\tau_7^2)$ = 0.2594	$SD(\tau_7^2)$ = 0.2153
13	$E(\mu_8)$ = -0.5458	$SD(\mu_8)$ = 0.2562	$E(\tau_8^2)$ = 0.2098	$SD(\tau_8^2)$ = 0.1772
7	$E(\mu_9)$ = -0.6516	$SD(\mu_9)$ = 0.3023	$E(\tau_9^2)$ = 0.0288	$SD(\tau_9^2)$ = 0.0764
12	$E(\mu_{10})$ = -0.2427	$SD(\mu_{10})$ = 0.0929	$E(\tau_{10}^2)$ = 0.0061	$SD(\tau_{10}^2)$ = 0.0120
34	$E(\mu_{11})$ = -0.2668	$SD(\mu_{11})$ = 0.0588	$E(\tau_{11}^2)$ = 0.0133	$SD(\tau_{11}^2)$ = 0.0195

Table 4.10. Results for the Meta-Analysis of the 11 Meta-Analyses which all have the same field, myocardial infarction, using an informative exponential prior for τ.

below, one can see the results for the absolute treatment effects for the same meta-analyses are quite similar to the previous results for the same meta-analyses using a non-informative prior on the heterogeneity (Table 4.9), as one would expect. One can also note the apparent reduction (by at least a half) for the expected heterogeneity values using the informative exponential prior compared to using the non-informative Inverse-Gamma prior.

Ideally, as mentioned before, a model which allows the parameters of the exponential prior, α and β, to come from a bivariate log-normal distribution rather than using given values would be preferable, as displayed in model 4.17.

Using the same 11 meta-analyses that have previously been used for model

4.16, but this time instead of specifying constant values for α and β in model 4.16, which correspond to expbeta[1] and expbeta[2] here respectively, model 4.17 below allows these parameters to come from a non-informative Bivariate Log-Normal model.

$$
\begin{aligned}
r_{c_{i,j}} &\sim \text{Bin}(n_{c_{i,j}}, p_{c_{i,j}}) \\
r_{t_{i,j}} &\sim \text{Bin}(n_{t_{i,j}}, p_{t_{i,j}}) \\
\text{logit}(p_{c_{i,j}}) &= \psi_{i,j} \\
\text{logit}(p_{t_{i,j}}) &= \psi_{i,j} + \delta_{i,j} \\
\psi_{i,j} &\sim \text{N}(0, \eta_{\text{i}}) \\
\eta_i &\sim \text{IG}(0.001, 0.001) \\
\delta_{i,j} &\sim N(\mu_i, \tau_i^2) \\
\mu_{\text{i}} &\sim \text{N}(0, 1000000) \\
\tau_i &\sim \text{Exp}(\lambda_{\text{i}}) \\
\lambda_i &= \left\{ \frac{1}{\text{expbeta}[1] + (\text{expbeta}[2] | \mu_i |)} \right\} \\
\beta[1:2] &\sim \text{MN}(\mu_2, \Sigma) \\
\mu_2 &\sim \text{MN}(\mu_3, \Sigma_2) \\
\text{expbeta}[i] &= \exp(\beta[i]) \\
\Sigma &\sim \text{Wishart}_2(\Omega)
\end{aligned}
\tag{4.17}
$$

Implementing the same priors on μ_3, Σ_2 and Ω as previously done for model

4.14 to ensure the obtained values for expbeta[1] and expbeta[2] follow the values as previously suggested (Senn, 2007*b*), the results using model 4.17 below are displayed in Table 4.11.

One can see from these results that the values produced for both μ and τ are very similar to the earlier computed results when using given values for α and β (displayed in Table 4.10). The computed values for the parameters expbeta[1] and expbeta[2], which correspond to α and β respectively from model 4.16 are 0.0497 and 0.4538, which are quite similar to the values for α and β which were used to produce the results displayed in Table 4.10.

Number of trials in MA i	μ_i		τ_i^2	
51	$E(\mu_1) = -0.1427$	$SD(\mu_1) = 0.0575$	$E(\tau_1^2) = 0.0054$	$SD(\tau_1^2) = 0.0105$
17	$E(\mu_2) = -0.2232$	$SD(\mu_2) = 0.0584$	$E(\tau_2^2) = 0.0132$	$SD(\tau_2^2) = 0.0180$
16	$E(\mu_3) = 0.0834$	$SD(\mu_3) = 0.1234$	$E(\tau_3^2) = 0.0111$	$SD(\tau_3^2) = 0.0275$
6	$E(\mu_4) = -0.0002$	$SD(\mu_4) = 0.0632$	$E(\tau_4^2) = 0.0042$	$SD(\tau_4^2) = 0.0118$
7	$E(\mu_5) = -0.2700$	$SD(\mu_5) = 0.1086$	$E(\tau_5^2) = 0.0148$	$SD(\tau_5^2) = 0.0334$
13	$E(\mu_6) = -0.1745$	$SD(\mu_6) = 0.0749$	$E(\tau_6^2) = 0.0099$	$SD(\tau_6^2) = 0.0232$
11	$E(\mu_7) = -0.8135$	$SD(\mu_7) = 0.3033$	$E(\tau_7^2) = 0.4503$	$SD(\tau_7^2) = 0.4061$
13	$E(\mu_8) = -0.5956$	$SD(\mu_8) = 0.2458$	$E(\tau_8^2) = 0.3617$	$SD(\tau_8^2) = 0.2977$
7	$E(\mu_9) = -0.666$	$SD(\mu_9) = 0.3073$	$E(\tau_9^2) = 0.1135$	$SD(\tau_9^2) = 0.2900$
12	$E(\mu_{10}) = -0.2335$	$SD(\mu_{10}) = 0.0955$	$E(\tau_{10}^2) = 0.0123$	$SD(\tau_{10}^2) = 0.0243$
34	$E(\mu_{11}) = -0.2771$	$SD(\mu_{11}) = 0.0650$	$E(\tau_{11}^2) = 0.0230$	$SD(\tau_{11}^2) = 0.0306$
	expbeta[1] = 0.0497		expbeta[2] = 0.4538	

Table 4.11. Results for the Meta-Analysis of the 11 Meta-Analyses which all have the same field, myocardial infarction, using an informative exponential prior for τ with Bivariate Log-Normal hyperparameters.

4.8.5 Summary

Having previously used several different models to obtain meta-analysis results, it was of interest to determine whether a meta-analysis of several meta-analyses could be obtained. A simple Bayesian model was produced to model this which first used non-informative priors for the estimates of interest, the overall treatment effects and the heterogeneities for each meta-analysis. Using 11 of the available 125 meta-analyses which all had the same field, myocardial infarction, the model was implemented and results were obtained. The results for each of the individual meta-analyses indicated the parameters of interest using the meta-analysis of meta-analyses resulted in very similar estimates as those produced via individual meta-analyses, the results of which have been attached in Appendix A. The estimates which did differ the most between a simple meta-analysis model and a meta-analysis of meta-analyses model occurred for the meta-analyses which had larger corresponding standard deviations unsurprisingly. This did not seem to occur for only the meta-analyses which had a small number of clinical trials included.

A more complex model was then fitted which included an exponential prior on τ as previously examined in Section 4.7, firstly with given values for α and β and secondly allowing α and β to come from a Bivariate log-Normal model with hyperparameters.

Implementing the exponential prior on τ resulted in a reduction in the size of the heterogeneity for each meta-analysis, with the model using given values

for the hyperparameters α and β reducing the heterogeneities slightly more than the Bivariate-Log-Normal distribution on the hyperparameters. The estimates for the treatment effect estimates were all similar regardless of the model used. This was expected since the parameter for the treatment effect estimates did not change from a non-informative prior.

Chapter 5

Conclusions and Discussions

Meta-analysis is a technique which can be used to combine the individual treatment effects for several individual studies to obtain a single overall treatment effect estimate. They are used for various different fields of interest - medical, agricultural and educational research to name a few - and suit data of any form, normally distributed data, ordinal data, survival data and binary data.

The methodology for combining the individual trial estimates, regardless of the format in which the data arises, includes using a fixed effects model, or a random effects model, with the random effects model incorporating the heterogeneity into the fixed effects model. The choice of which model should be used is still debateable, with some statisticians believing the fixed effects approach suits their work better, whereas another might prefer to use the random effects approach. Regardless of the approach used, the heterogeneity should be checked for, and if need be, incorporated into the model to prevent an invalid result.

The treatment effect estimates may be calculated numerically using a frequentist approach such as the Mantel-Haenszel odds ratio for binary data, with an additional random effects weight should the random effect model be used, or using a more complex Bayesian approach which calculates estimates using the data provided and specified priors.

The Mantel-Haenszel odds ratio calculates the odds ratio using a Normal approximation of the binary data, effectively losing the Binary format of the data, however the Bayesian model would account for the Binary format from which the data initially arose.

There are several measures of heterogeneity. Perhaps two of the better known frequentist methods are DerSimonian and Laird's method of moments (DerSimonian and Laird, 1986) and Hardy and Thompson's method based on the Maximum Likelihood estimators (Hardy and Thompson, 1996). The data used for this thesis used both of these methods when obtaining frequentist heterogeneity estimates and found the methods coincided extremely well.

The priors for a Bayesian model can be non-informative if no prior knowledge is held or informative if there is some prior knowledge which the priors should account for. There are several choices of non-informative priors available, however just how non-informative these priors actually are has been questioned (Lambert *et al.*, 2005). The general agreement is regardless of the prior chosen, a sensitivity analysis should be performed to check the robustness of the results across various non-informative priors.

Previously when meta-analyses have been performed, the two main areas of

interest, the overall treatment effect estimate and the heterogeneity, if a random effects model has been used, are considered to be independent estimates and have generally been modelled as such (Smith *et al.* (1995), Abrams and Sansó (1998)), although this assumption has in fact been questioned (Senn, 2007*b*).

These estimates were calculated for 125 independent meta-analyses and upon examination of these estimates, it appears a relationship between them does occur.

The data used for the purpose of this thesis was binary and estimates for all three main metrics for binary data, the log-odds-ratio, the log-relative risk and the risk difference, were calculated using the frequentist approach. The results when graphed indicated a positive correlation could occur between the treatment effect estimate and the heterogeneity, regardless of which metric was chosen, with increasing heterogeneity occurring for an increasing treatment effect estimate. Correlations between the treatment effect estimates and the heterogeneity for each metric separately were also computed and the results indicated a significant positive correlation of approximately 0.4 occurred irrespective of the metric chosen.

A Bayesian model was then considered with separate non-informative priors initially for the treatment effect estimate and the heterogeneity as have previously been used (Lambert *et al.* (2005), (Gelman, 2006), Abrams and Sansó (1998), Smith *et al.* (1995)) before going on to fit an informative exponential prior on the square root of the heterogeneity which would allow the variance to depend upon the overall treatment effect estimate.

The first Bayesian model which made use of non-informative priors for the parameters of interest and allowed the data to retain its binary format indicated the treatment effect estimates using the Bayesian model with non-informative priors corresponded very well to the estimates obtained using a frequentist Binomial-Normal model. The estimates for the heterogeneity however indicated a slight discrepancy between the Bayesian method and the frequentist method, with the Bayesian method resulting in slightly larger estimates which increased with increasing heterogeneity.

A Bayesian model using a Normal approximation for the log-odds ratios was then implemented and compared to the results obtained using two frequentist methods for the heterogeneity. Again the results for the treatment effect estimates corresponded very well whether the Bayesian method or the frequentist method was used. The estimates of the heterogeneity for the Bayesian method were again slightly larger than those obtained using a frequentist method, although these estimates did not increase quite as much with the increasing heterogeneity values compared to the models which kept the data in its Binary format.

Data for a meta-analysis using 22 studies which had been previously published (Abrams and Sansó, 1998) was then used with all four combinations of model: a Bayesian model using the Binary format, a Binomial-Normal model; a frequentist model using the Binary format, a Binomial-Normal model; a Bayesian model using a Normal approximation, a Normal-Normal model; a frequentist model using a Normal approximation, a Normal-Normal model. The results for the overall treatment effect estimate appeared to differ the most between the models

used - the Binomial-Normal and the Normal-Normal, regardless of whether the Bayesian approach or the frequentist approach was used, whereas the heterogeneity estimate differed the most between the approaches used - the Bayesian or the frequentist.

Having examined the results for the different methods and models, it would appear reasonable for one to conclude that if the person constructing the meta-analysis is doing so under the assumption that the treatment effect estimate and the heterogeneity are independent, then the advantage of the Bayesian model is minimal and is perhaps not worth the extra effort since the treatment effect estimates reported for the frequentist method and the Bayesian method are very similar.

However, the estimates of interest appear to be related, and so the exponential prior for the square root of the heterogeneity which is dependant upon the overall treatment effect estimate was implemented. Using given suggested values for the two hyperparameters α and β (Senn, 2007*b*), the results obtained were similar to those which were produced using the non-informative priors for both the treatment effect estimate and the heterogeneity, with the Binomial-Normal model having slightly larger estimates than the Normal-Normal model.

Despite the results after implementing the informative exponential prior remaining similar to those obtained using non-informative independent priors for the treatment effect estimate and the heterogeneity, I would recommend if performing a meta-analysis that one should use the Bayesian method, which allows the user to include a subjective opinion, with the informative exponential prior

rather than the non-informative independent priors since a relationship between the treatment effect estimate and the heterogeneity has been indicated and therefore should not be ignored.

After then introducing a Bivariate-Log-Normal prior for the hyperparameters α and β, the results were again similar, with the heterogeneity value remaining slightly larger. This may however be the result of a more complex model being fitted with no more additional data becoming available. The priors for the hyperparameters for the Bivariate-Log-Normal distribution were carefully selected so that the parameters corresponding to the previous α and β parameters were restricted so as not to deviate from the suggested given values (Senn, 2007*b*), however more work focusing on the values which α and β take, whether they are specified as given in the model or if they are the result of fitting priors to these values, should perhaps be done in an attempt to determine how sensitive a model as complex as this one is.

An appropriate model for meta-analyses, I believe, should take into consideration that the treatment effect estimate and the heterogeneity appear to be related. Using Bayesian methods will allow this prior knowledge to be incorporated into the model as has been done in this thesis whilst using an informative exponential prior.

As previously mentioned in the aims of this thesis, a meta-analysis involves combining several independent clinical trials which are all alike allowing the treatment effect to be transferred between the clinical trials. It may also be reasonable to consider trials which have the same field but different treatments, so long as

care is taken when stating the hypotheses, looking at the modelling and the interpretation of the results as discussed in Senn (2007*a*). By carefully considering the modelling of the meta-analysis, one allows the treatment effect to be transferred only between the trials which consider the same treatment effect.

11 Meta-analyses which all had the same field, myocardial infarction, were then used in a meta-analysis of meta-analyses. Although the clinical trials which make up the 11 meta-analyses here all have the same field, they do not necessarily all examine the same treatments or the same concentrations. The results for the same models as were previously used (one which uses non-informative priors for the overall treatment effect estimates and the heterogeneity's, another which uses the exponential prior for the square root of the heterogeneity and given values for the hyperparameters α and β, and another which allows α and β to come from a Bivariate-Log-Normal model) were obtained.

All three models produced results for the treatment effect estimates and the heterogeneity's which were similar to the results obtained via individual meta-analyses. Some estimates differed a bit more between the models than others and these appeared to be the meta-analyses which had larger corresponding standard deviations for their estimates and not necessarily due to the number of clinical trials which were included in the meta-analyses. As mentioned previously, the final most complex model which had a Bivariate-Log-Normal for the α and β parameters were restricted so that the α and β did not differ too much from suggested values (Senn, 2007*b*), although more work on these suggested values could be done to check this restriction does not limit the results. It might also be

worthwhile considering whether a different distribution would better suit these priors and what effect, if any, the change in prior alters the parameters and if this alters the results found here.

An idea for future work would be to further examine constructing meta-analyses which consider clinical trials with different treatments. This could be done by examining the variance of a clinical trial when it is included in a meta-analysis which looks at trials with the same treatment and comparing the results to a meta-analysis which looks at trials with different treatments.

Appendix A

Individual Meta-Analysis Results

The results from the 11 individual meta-analyses, which all have the same field myocardial infarction, which were subsequently used in the meta-analysis of meta-analyses are displayed here in table A.1. Model 4.1* was used to produce the results and a burn-in of 5000 was used for each meta-analysis followed by another 25,000 iterations using two chains.

*The prior for τ^2 is slightly different to that in model 4.1, using an IG$(0.001, 0.001)$

Number of trials in MA i	μ_i		τ_i^2	
51	$E(\mu_1) = -0.1125$	$SD(\mu_1) = 0.0658$	$E(\tau_1^2) = 0.0119$	$SD(\tau_1^2) = 0.0178$
17	$E(\mu_2) = -0.2100$	$SD(\mu_2) = 0.0634$	$E(\tau_2^2) = 0.0179$	$SD(\tau_2^2) = 0.0212$
16	$E(\mu_3) = 0.1344$	$SD(\mu_3) = 0.1388$	$E(\tau_3^2) = 0.0314$	$SD(\tau_3^2) = 0.0722$
6	$E(\mu_4) = 0.0062$	$SD(\mu_4) = 0.0726$	$E(\tau_4^2) = 0.0132$	$SD(\tau_4^2) = 0.0280$
7	$E(\mu_5) = -0.2589$	$SD(\mu_5) = 0.1165$	$E(\tau_5^2) = 0.0247$	$SD(\tau_5^2) = 0.0679$
13	$E(\mu_6) = -0.1591$	$SD(\mu_6) = 0.0805$	$E(\tau_6^2) = 0.0211$	$SD(\tau_6^2) = 0.0444$
11	$E(\mu_7) = -0.7069$	$SD(\mu_7) = 0.3066$	$E(\tau_7^2) = 0.5387$	$SD(\tau_7^2) = 0.5596$
13	$E(\mu_8) = -0.4691$	$SD(\mu_8) = 0.2593$	$E(\tau_8^2) = 0.4568$	$SD(\tau_8^2) = 0.3970$
7	$E(\mu_9) = -0.5465$	$SD(\mu_9) = 0.3752$	$E(\tau_9^2) = 0.2867$	$SD(\tau_9^2) = 0.7324$
12	$E(\mu_{10}) = -0.2246$	$SD(\mu_{10}) = 0.1016$	$E(\tau_{10}^2) = 0.0196$	$SD(\tau_{10}^2) = 0.0358$
34	$E(\mu_{11}) = -0.2610$	$SD(\mu_{11}) = 0.0614$	$E(\tau_{11}^2) = 0.0253$	$SD(\tau_{11}^2) = 0.0311$

Table A.1. Results for the 11 individual Meta-Analyses which all have the same field, myocardial infarction, using a slightly altered model 4.1 (see footnote)

References

Abrams, K. and Sansó, B. (1998), 'Approximate Bayesian inference for random effects meta-analysis', *Statistics in Medicine* **17**(2), 201–218.

Anscombe, F. J. (1956), 'On estimating binomial response relations', *Biometrika* **43**, 461–464.

Berlin, J. A., Laird, N. M., Sacks, H. S. and Chalmers, T. C. (1989), 'A comparison of statistical methods for combining event rates from clinical trials', *Statistics in Medicine* **8**(2), 141–151.

Biggerstaff, B. J. and Tweedie, R. L. (1997), 'Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis', *Statistics in Medicine* **16**(7), 753–768.

Brockwell, S. E. and Gordon, I. R. (2007), 'A simple method for inference on an overall effect in meta-analysis', *Statistics in Medicine* **26**(25), 4531–4543.

Cox, D. R. and Snell, E. J. (1989), *Analysis of Binary Data*, second edn, Chapman and Hall.

DerSimonian, R. and Laird, N. (1986), 'Meta-analysis in clinical trials', *Controlled Clinical Trials* **7**, 177–188.

Digestive Tract Trialists' Collaborative Group (1993), 'Meta-analysis of randomised controlled trials of selective decontamination of the digestive tract', *British Medical Journal* **307**, 525–532.

Engels, E. A., Schmid, C. H., Terrin, N., Olkin, I. and Lau, J. (2000), 'Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses', *Statistics in Medicine* **19**(13), 1707–1728.

Fleiss, J. L. and Gross, A. J. (1991), 'Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique', *Journal of Clinical Epidemiology* **44**(2), 127–139.

Gelman, A. (2006), 'Prior distributions for variance parameters in hierarchical models', *Bayesian Analysis* **1**(3), 515–533.

Hardy, R. J. and Thompson, S. G. (1996), 'A likelihood approach to meta-analysis with random effects', *Statistics in Medicine* **15**(6), 619–629.

Hardy, R. J. and Thompson, S. G. (1998), 'Detecting and describing heterogeneity in meta-analysis', *Statistics in Medicine* **17**, 841–856.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. and Jones, D. R. (2005), 'How vague is vague? A simulation study of the impact of the use of

vague prior distributions in MCMC using WinBUGS', *Statistics in Medicine* **24**(15), 2401–2428.

Mantel, N. and Haenszel, W. (1959), 'Statistical aspects of the analysis of data from retrospective studies of disease', *Journal of the National Cancer Institute* **22**(9), 719–748.

Pearson, K. (1904), 'Report on certain enteric fever inoculation statistics', *British Medical Journal* **2**(2288), 1243–1246.

R version 2.6.1 (2007). The R Foundation for Statistical Computing.

Senn, S. (2007*a*), *Statistical Issues in Drug Development*, second edn, John Wiley and Sons Ltd.

Senn, S. (2007*b*), 'Trying to be precise about vagueness', *Statistics in Medicine* **26**(7), 1417–1430.

Senn, S. (2008), 'Authors' reply', *Statistics in Medicine* **27**(4), 622–624.

Smith, T. C., Spiegelhalter, D. J. and Thomas, A. (1995), 'Bayesian approaches to random-effects meta-analysis: A comparative study', *Statistics in Medicine* **14**(24), 2685–2699.

Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, first edn, John Wiley and Sons Ltd.

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

Whitehead, A. (2002), *Meta-Analysis of Controlled Clinical Trials*, first edn, John Wiley and Sons Ltd.

Whittaker, J. (1990), *Graphical Models in Applied Multivariate Statistics*, first edn, John Wiley and Sons Ltd, Chichester.

WinBUGS version 1.4.3 (2007). WinBUGS with DoodleBUGS.

Yates, F. (1934), 'Contingency tables involving small numbers and the χ^2 test', *Journal of the Royal Statistical Society (Supplement)* **1**(2), 217–235.

Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. (1985), 'Beta blockade during and after myocardial infarction: an overview of the randomized trials', *Progress in Cardiovascular Diseases* **XXVII**(5), 335–371.