Development and application of multivariate meta-analysis in medical research: borrowing strength across multiple correlated outcomes

Miriam Susanna MacDonald Hattle

A thesis submitted for the degree of Doctor of Philosophy

December 2019

Keele University

Abstract

Multivariate meta-analysis methods combine effect estimates for multiple correlated outcomes (such as systolic and diastolic blood pressure) from independent studies, utilising their between-study and within-study correlations. In contrast, a univariate meta-analysis pools effect estimates for each outcome independently. By using the multivariate over the univariate approach, there is a potential gain in information toward summary meta-analysis results, quantified by the Borrowing of Strength (BoS) statistic, a percentage reduction in the variance for the summary effect between the two approaches. This thesis examines BoS and multivariate meta-analysis applications in detail.

Firstly, multivariate meta-analysis is applied to an individual participant data meta-analysis examining the effect of diet and exercise interventions during pregnancy. This shows results from the univariate and multivariate meta-analyses are similar. However, a review of 43 Cochrane reviews concludes that although results between the univariate and multivariate are often similar, a few multivariate meta-analyses do give important differences to results from univariate meta-analyses, and these are shown to have a larger magnitude of BoS.

This motivates research to identify predictors of BoS and to develop a model to predict BoS (in advance of analysis) to flag when researchers should consider multivariate meta-analysis. Additionally, an interactive tool is developed to investigate the relationship between the various characteristics and the magnitude of BoS. The magnitude of BoS is shown mathematically to be approximately bounded by the percentage of missing data for the outcome of interest.

A novel application of bivariate meta-analysis is then proposed for trials with continuous outcomes analysed with final score or ANCOVA models, with examination in

real examples and a simulation study.

In conclusion, multivariate meta-analysis may provide important differences to univariate meta-analysis when BoS is large, and so researchers should consider the approach when BoS is expected to be large.

Acknowledgements

I would first like to express my gratitude to my PhD supervisors Prof. Richard Riley and Dr. Danielle Burke for their support and encouragements throughout my PhD. I feel fortunate to have had a supervisory team that is so knowledgeable, experienced and approachable. I am grateful to Richard for always listening to my ideas and encouraging me to explore these ideas. I look forward to working together in the future. I would like to express my gratitude to Danielle, for listening to all my PhD stresses and offering her invaluable advice. I would also like to thank Dr. Sam Hider for her advice and support throughout the PhD process.

I am grateful to Keele University for funding this work through an ACORN studentship and providing the best environment for both my undergraduate and PhD studies.

Thank you to the Research Institute for Primary Care and Health Sciences and colleagues for providing me with the perfect environment to enable me to develop as a researcher through conversations, opportunities to present my work, seminars and book/journal clubs. I am also thankful to my fellow PhD students who I have discussed work with, laughed with and who have provided a sympathetic ear.

To Mum and Dad, who have always cheered me on the loudest. Thank you for always encouraging me to do my best and having an interest in everything I do and achieve. Thank you to Naomi for our 'quick' three-hour phone calls and encouraging me to keep going, the card had better be worth it! Thank you to all my friends for their support and understanding, particularly when I took too long to reply to messages.

Stuart, a couple of lines will never truly express how grateful I am to you. This work would not have been possible without your love and support. Thank you for never doubting me and enduring me talking about statistics.

Contents

List of Figures

List of Tables

List of Abbreviations

ANCOVA	Analysis of covariance
BoS	Borrowing of strength
CI	Confidence Interval
CVD	Cardiovascular Disease
DBP	Diastolic Blood Pressure
DL	Dersimonian-Laird
H-K	Hartung-Knapp
HR	Hazard ratio
HTA	Health Technology Assessment
IPD	Individual Participant Data
IQR	Interquartile range
iWiP	International Weight Management in Pregnancy
LGA	Large for gestational age
MAR	Missing at random
MM	Method of moments
MNAR	Missing not at random

NICU	Neonatal Intensive Care Unit
OR	Odds ratio
PE	Pre-eclampsia
PIH	Pregnancy induced hypertension
RCT	Randomised Control Trial
REML	Restricted Maximum Likelihood
SBP	Systolic Blood Pressure
SGA	Small for gestational age

Chapter 1

Introduction

1.1 Overview

Evidence-based medicine is an approach to practising medicine that combines two components; individual clinical expertise and external evidence (Sackett 1997). With the quantity of published evidence ever growing, it is impossible for doctors to evaluate all of the available evidence. The aim of evidence-based medicine is to make decisions based on the best currently available evidence (Masic et al. 2008). Evidence-based medicine requires a clear clinical question, proposed using individual clinical expertise, followed by a comprehensive systematic search of the literature and evaluation of the external evidence (Rosenberg & Donald 1995). The findings from the evaluation of the external evidence need to be implemented using clinical expertise. This requires results from medical research studies to be identified and summarised, to quantify important measures such as treatment effects, that conclude whether a treatment is effective or not. This process usually requires the results from the multiple studies to be combined, in a so-called meta-analysis.

This thesis evaluates novel 'multivariate meta-analysis' methods, which allow multiple treatment effects to be obtained for multiple outcomes, so that evidence-based medical decisions can be made on a more complete set of evidence. In particular, this thesis evaluates if and when multivariate meta-analyses are actually needed, and the statistical benefits of the approach. In this chapter, the foundations and principles of

meta-analysis are introduced, and the rationale for subsequent chapters then described.

1.2 Systematic reviews

A systematic review is the process by which the available external evidence is collated and evaluated (Cook et al. 1997, Akobeng 2005). It can be used to review medical research papers on a specific topic to answer a pre-specified clinical question. It is a formal process that involves searches of all available literature, using a predefined search strategy and eligibility criteria, described in the systematic review's protocol, which should also be published (Moher et al. 2015). The aim of the systematic review is to capture all available literature to answer the clinical question. Using systematic search strategies, bias (from the decision of which papers to include) is minimised (Mulrow 1994). Systematic reviews can identify areas with a lack of evidence which can inform future research (Egger et al. 2008).

1.2.1 Cochrane

The Cochrane library contains the Cochrane Database of Systematic Reviews made up of Cochrane reviews and protocols (Higgins & Green 2008, Cochrane Collaboration UK n.d., Egger et al. 2008). A Cochrane review is a systematic review that has been undertaken by a Cochrane Review Group. Cochrane is an international organisation that aims to provide accurate credible evidence based information to help support decision making in the medical field.

1.2.2 Meta-analysis

The term 'meta' is Greek for 'after', 'above', or 'transcending'. Meta-analysis is a statistical method that combines results from multiple independent studies in order to address the research question (Glass 1976, Borenstein et al. 2009). The multiple independent studies must be relevant to the research question proposed in the meta-analysis. For example, if the aim is to summarise the effect of a treatment, then the

studies must have evaluated the treatment's effect.

Meta-analysis is important in order to increase the power and improve the precision of an answer to a research question (Higgins & Green 2008). Meta-analysis can be used to draw conclusions from studies that provide contradictory results. While undertaking a meta-analysis, there needs to be some caution around combining studies that are at high risk of bias or studies that are clinically diverse, such that summary results may not be meaningful or reliable (Higgins & Green 2008).

1.3 Univariate meta-analysis

Univariate meta-analysis synthesises a singular summary effect estimate from multiple independent studies with one outcome, such as a treatment effect, for one particular outcome. For a univariate meta-analysis, a decision regarding the assumption for the effect estimates across all studies is required that determines which approach is used, a fixed-effect meta-analysis or a random-effects meta-analysis (Riley et al. 2011, Borenstein et al. 2009, Higgins & Green 2008). The details for both approaches are provided below with examples.

1.3.1 Univariate fixed-effect meta-analysis

A fixed-effect meta-analysis assumes that all the studies in the meta-analysis share a common true effect size (Riley et al. 2011, Borenstein et al. 2009, Higgins & Green 2008). Therefore, any variation observed in the effect sizes is only due to random sampling errors. Mathematically this is expressed as (Borenstein et al. 2009):

$$Y_i = \theta + e_i \tag{1.1}$$

where θ is the true treatment effect for all studies, e_i is the sampling error in study i and Y_i is the observed value in study i.

The univariate fixed-effect meta-analysis can be modelled using a normal distribution with mean θ and variance s_i^2, the within study variance (Kulinskaya et al. 2008, van Houwelingen et al. 2002).

$$Y_i \sim N(\theta, s_i^2) \tag{1.2}$$

The studies included in a fixed-effect meta-analysis are assigned a weight which is utilised during the modelling process (Borenstein et al. 2009). The magnitude of the weight for a study is dependent upon the magnitude of the precision of the treatment effect. The magnitude of the precision of the treatment effect for a study is determined by the sample size and the design of the study. The weight for the inverse-variance method is calculated using (Higgins & Green 2008, Borenstein et al. 2009):

$$W_i = \frac{1}{s_i^2} \tag{1.3}$$

where s_i^2 is the within study variance for study i, that is the variance of the effect estimate in study i. The relationship between the weight and the within study variance can be described as the greater the within study variance, the smaller the weight assigned to the study. So generally larger studies have greater weight toward the pooled (summary) effect estimate from the meta-analysis.

Using the study weights, the summary effect estimate from all the studies can be calculated using:

$$\hat{\theta} = \frac{\sum_{i=1}^{n} W_i Y_i}{\sum_{i=1}^{n} W_i} \tag{1.4}$$

The standard error of the summary effect estimate is calculated as:

$$s.e.(\hat{\theta}) = \sqrt{\frac{1}{\sum_{i=1}^{n} W_i}} \tag{1.5}$$

which is the square root of the reciprocal of the summation over all n studies of the study weights, Equation (1.3).

The 95% confidence interval for the summary effect for the fixed-effect meta-analysis can be calculated as follows:

$$CI_{\hat{\theta}} = \hat{\theta} \pm 1.96 \times s.e.(\hat{\theta}) \tag{1.6}$$

where $s.e.(\hat{\theta})$ is the standard error of the summary effect, Equation (1.5), and $\hat{\theta}$ is the summary effect estimate, Equation (1.4).

Forest plots

The results from a univariate meta-analysis are displayed in graphical form using forest plots. In a forest plot, the results (effect estimate and confidence interval) from each individual study are visually displayed using circles (effect estimate) and a line (confidence interval). The weights for each study can be displayed using a square, the size of the square represents the magnitude of the weight for that particular study. The summary effect estimate and respective confidence interval is displayed using a diamond (found at the bottom of the forest plot), where the centre of the diamond represents the summary effect estimate and the width of the diamond represents the summary confidence interval. Statistical packages, like metan (Bradburn et al. 1999, Harris et al. 2008) and metaan (Kontopantelis & Reeves 2010) in Stata, will produce forest plots upon request. In the examples following, the results are presented in forest plots.

1.3.2 Continuous outcome example - Ten hypertension trials

A meta-analysis of 10 randomised control trials investigated the use of an anti-hypertensive treatment to lower blood pressure (Riley et al. 2013, Wang et al. 2005). The randomised control trials measured the systolic blood pressure, a continuous outcome, at baseline and at follow-up. The treatment effect can be estimated in each study using three different model approaches: a change score model, a final score model or an analysis

of covariance (ANCOVA) model. The details for each model are now provided.

Change score model

The change score models the difference between the follow-up measurement and the baseline measurement. The change score model for the treatment effect for an individual trial i can be written as:

$$y_{F_{ij}} - y_{B_{ij}} = \phi_i + \theta_{Ci} x_{ij} + e_{ij} \qquad e_{ij} \sim N(0, \sigma_i^2) \tag{1.7}$$

where F represents the final score, B the baseline and C the change score. The follow-up and baseline values are $y_{F_{ij}}$ and $y_{B_{ij}}$, respectively for participant j in study i and x_{ij} is the indicator variable for the treatment/control group (where 0 is the control group and 1 is the treatment group) a participant belongs to. The residual error, e_{ij}, is assumed to be normally distributed with a residual variance, σ_i^2. The intercept, ϕ_i, represents the control group's mean change score and θ_{Ci} is the change score model's treatment effect in study i (the measure of interest).

Final score model

The final score models the follow-up measurements only and does not include the baseline measurements. The final score model can be written as:

$$y_{F_{ij}} = \phi_i + \theta_{Fi} x_{ij} + e_{ij} \qquad e_{ij} \sim N(0, \sigma_i^2) \tag{1.8}$$

where F represents the final score and $y_{F_{ij}}$ are the follow-up values for participant j in study i. The treatment effect from the final score model for study i is θ_{Fi} and the intercept, ϕ_i, is the mean final score for the control group in study i. The residual error, e_{ij}, and the treatment group indicator, x_{ij}, are the same as the change score model.

ANCOVA model

For the Analysis of Covariance (ANCOVA), the follow-up value is regressed and the baseline score is adjusted for. The ANCOVA model can be written as:

$$y_{F_{ij}} = \phi_i + \beta_i y_{B_{ij}} + \theta_{Ai} x_{ij} + e_{ij} \qquad e_{ij} \sim N(0, \sigma_i^2) \tag{1.9}$$

where F represents the final score, B the baseline and A the ANCOVA model. The follow-up and baseline values are $y_{F_{ij}}$ and $y_{B_{ij}}$, respectively, for participant j in study i and β_i is the effect of a one-unit increase in the baseline value on the follow-up value for study i. The treatment effect from the ANCOVA model for study i is θ_{Ai} and the intercept is ϕ_i. The residual error, e_{ij}, and the treatment group indicator, x_{ij}, are defined as in the previous models.

Fixed-effect meta-analysis

For each model estimating the treatment effect, a fixed-effect univariate meta-analysis was applied to obtain summary treatment effects from the three different models. The results for each meta-analysis are displayed in forest plots (Figure 1.1 and, Figures A.1 and A.2 in Appendix A.1) and summarised together in Table 1.1.

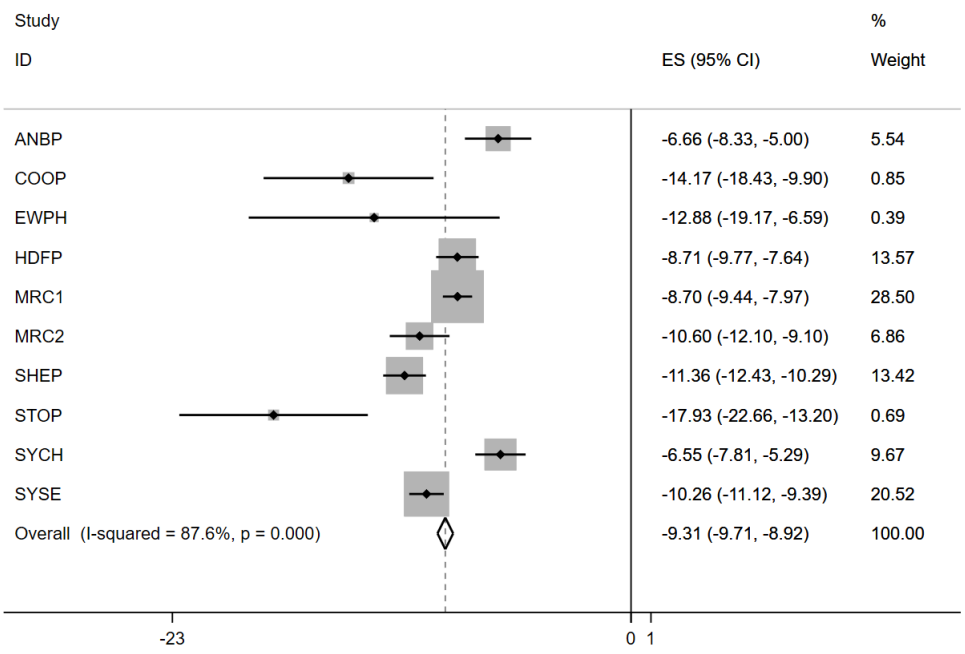


Figure 1.1: Fixed-effect meta-analysis of treatment effect estimates from the ANCOVA model

Table 1.1: Results from the fixed-effect univariate meta-analyses for the Hypertension data

Model	Fixed-effect	
	Estimate	95% CI
Change	-9.47	-9.91 to -9.04
Final	-9.29	-9.71 to -8.87
ANCOVA	-9.31	-9.71 to -8.92

The summary treatment effect estimates and confidence intervals from the three different analytical models (change score, final score and ANCOVA models) are all negative (Table 1.1 and Figure 1.1 and, Figures A.1 and A.2 in Appendix A.1). For example, the summary treatment effect estimate from the ANCOVA model was -9.31 with a confidence interval of -9.71 to -8.92. Therefore, from the summary treatment effect estimates and confidence intervals, there is statistically significant evidence that the anti-hypertensive treatment reduces the systolic blood pressure compared to control.

Due to the differences in the analytical models, the summary treatment effect estimates and confidence intervals differ. It is recommended, where possible, that treatment effects are estimated using the ANCOVA model, since the ANCOVA model adjusts for baseline values and thus accounts for the presence of baseline imbalance (Vickers & Altman 2001).

1.3.3 Binary outcome example - Bolus thrombolytic therapy

An example with a binary outcome is a meta-analysis of nine phase II trials, by Eikelboom et al. (2001), that investigated the use of Bolus thrombolytic therapy with standard infusion therapy for acute myocardial infarction by studying the number of events of death (Smalling et al. 1995, Bode et al. 1996, Kawai et al. 1997, Vanderschueren et al. 1997, Bär et al. 1998, Bleich et al. 1998, den Heijer et al. 1998, Cannon et al. 1998, Park 1998).

From the 2x2 contingency table, odds ratios for mortality were calculated for each trial. A univariate fixed-effect meta-analysis was used to analyse all nine trials together and the results are displayed in a forest plot.

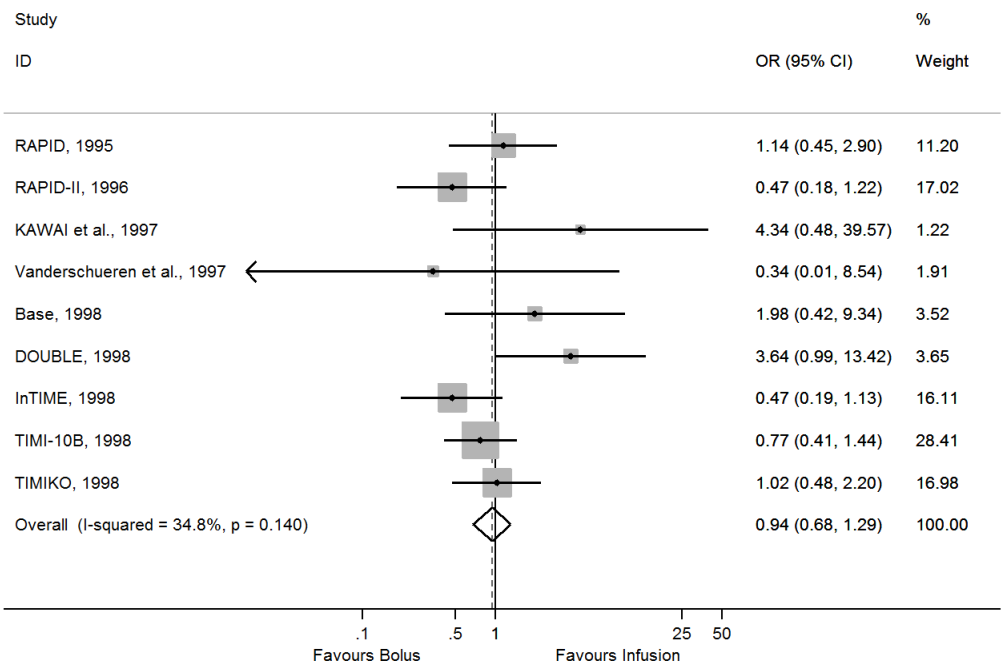


Figure 1.2: Fixed-effect meta-analysis of mortality from the Bolus thrombolytic therapy studies

From the fixed-effect univariate meta-analysis results, the odds of dying were 6% lower for those on the Bolus thrombolytic therapy than those on the standard infusion therapy (Figure 1.2). However, this was not statistically significant since the confidence interval was 0.68 to 1.29, which gave odds of dying between 32% lower and 29% higher for those given the Bolus thrombolytic therapy compared to those given the standard infusion.

1.3.4 Statistical heterogeneity

Heterogeneity is an umbrella term for different types of variability that arises between studies (Higgins & Green 2008). Statistical heterogeneity is the variation in the observed effect estimates that is beyond chance (i.e. over and above sampling error). Measures for quantifying statistical heterogeneity include Cochran's Q statistic, the between-study variance, τ^2, and the I^2 statistic.

Cochran's Q Statistic

The Cochran's Q Statistic is given as (Cochran 1954):

$$Q = \sum_{i=1}^{n} W_i (Y_i - \hat{\theta})^2 \tag{1.10}$$

where W_i is the study weight for study i, Y_i is the study effect in study i, $\hat{\theta}$ is the summary effect from the univariate fixed-effect meta-analysis and n is the total number of studies in the meta-analysis. Cochran's Q Statistic is the weighted sum of squares, a measure of variation (Higgins & Green 2008).

From the fixed-effect assumption that the true effect size is the same across all the studies, the expected value of Q is given by the degrees of freedom ($df = n - 1$). To calculate the excess variability among studies in the meta-analysis, the degrees of freedom is subtracted from Q. The Cochran's Q statistic can be used as a statistical test for heterogeneity where the null hypothesis is that all studies share the same true effect size and therefore there is no between-study variation. To test this null hypothesis, the Cochran's Q statistic is compared against a chi-square distribution with degrees of freedom $df = n - 1$.

I^2 statistic

The I^2 statistic provides a percentage of the total observed variation across studies that is due to the between-study variation (Higgins & Thompson 2002, Higgins et al. 2003). The greater the percentage for I^2, the stronger the evidence for heterogeneity having an important impact on the meta-analysis. The I^2 statistic is calculated using the following equation:

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\% \tag{1.11}$$

where Q is the Cochran's Q statistic, calculated from Equation 1.10, and it's degrees of freedom is df.

Between-study variance

The between-study variance, τ^2, is the actual variance of the true effect sizes across studies (Borenstein et al. 2009). The between-study variance can be estimated using:

$$\tau^2 = \frac{Q - df}{C} \tag{1.12}$$

where

$$C = \sum_{i=1}^{n} W_i - \frac{\sum_{i=1}^{n} W_i^2}{\sum_{i=1}^{n} W_i} \tag{1.13}$$

This is also known as the Dersimonian and Laird estimator for τ^2 (DerSimonian & Laird 1986). The Cochran's Q statistic (Equation 1.10) and it's degrees of freedom are denoted by Q and df, respectively, and C is calculated from the study weights, W_i, as shown in Equation 1.13. The between-study variance is required to be estimated (estimation methods are discussed in Section 1.3.9) to be included in a random-effects meta-analysis (described in the next section, Section 1.3.5).

1.3.5 Random-effects meta-analysis

Recall from Section 1.3.1, for a fixed-effect meta-analysis, it is assumed that all the studies are estimating the same true effect size and thus any difference in the estimated effect size is assumed to be due to random sampling error. However, for a random-effects meta-analysis, an alternative assumption is made that each study is estimating its own true effect size (Borenstein et al. 2009, Riley et al. 2011). To be included in the meta-analysis, the studies must be considered similar enough to be analysed together in a meta-analysis.

The random-effects meta-analysis can be expressed mathematically as:

$$Y_i = \underbrace{\mu + \eta_i}_{=\theta_i} + \epsilon_i \tag{1.14}$$

where Y_i is the observed effect size in study i and μ is the mean effect size for all studies, η_i is the true variation in effect size from the mean for study i, ϵ_i is the random sampling error in study i and θ_i is the true effect size for study i, which is most commonly modelled as normally distributed.

The random-effects meta-analysis model can alternatively be expressed using normal distributions, as follows (Kulinskaya et al. 2008, van Houwelingen et al. 2002):

$$Y_i \sim N(\theta_i, s_i^2) \tag{1.15}$$

$$\theta_i \sim N(\theta, \tau^2) \tag{1.16}$$

where Y_i is the observed effect size for study i which is modelled as normally distributed with mean θ_i, the true effect size for study i (which is also normally distributed in a random-effects meta-analysis model and this varies from study to study) and s_i^2 is the within study variance for study i. The true study effects are also normally distributed, with θ, the mean effect size and τ^2 the between-study variance.

In the same way as the studies in a fixed-effect meta-analysis were assigned a weight, so too are the studies in a random-effects meta-analysis (Borenstein et al. 2009). However, the calculation for the weight of a study in a random-effects meta-analysis is different to a fixed-effect. The calculation for the weight of a study in a random-effects meta-analysis is as follows:

$$W_i^* = \frac{1}{s_i^2 + \tau^2} \tag{1.17}$$

where s_i^2 is the within-study variance and τ^2 is the between-study variance. The DerSimonian and Laird equation for estimating the between-study variance, τ^2, was provided by Equation 1.12 in Section 1.3.4.

The summary effect estimate from the random-effects meta-analysis is estimated by:

$$\hat{\theta} = \frac{\sum_{i=1}^{n} W_i^* Y_i}{\sum_{i=1}^{n} W_i^*} \tag{1.18}$$

and the standard error of the summary effect estimate is calculated by:

$$s.e.(\hat{\theta}) = \sqrt{\frac{1}{\sum_{i=1}^{n} W_i^*}} \tag{1.19}$$

A confidence interval for the summary effect estimate can be calculated, such that the approximate 95% confidence interval for the summary effect from the random-effects meta-analysis is given as:

$$CI_{\hat{\theta}} = \hat{\theta} \pm 1.96 s.e.(\hat{\theta}) \tag{1.20}$$

where $\hat{\theta}$ is the summary effect estimate, Equation 1.18 and s.e.($\hat{\theta}$) is the standard error of the summary effect estimate, Equation 1.19. Alternatives to this standard confidence interval derivation have been proposed, such as the Hartung-Knapp correction, provided in Section 1.3.10.

1.3.6 Prediction intervals

Following a random-effects meta-analysis, it is possible to calculate a prediction interval (Higgins et al. 2009, Riley et al. 2011). From a summary effect estimate, a researcher only has information concerning the average effect estimate and its respective confidence interval; this does not provide information for the potential effect size in individual studies and their populations. Prediction intervals are used to quantify what the potential treatment effect might be for an individual study. The prediction interval can be calculated by:

$$\hat{\theta} \pm t_{n-2} \sqrt{\hat{\tau}^2 + \sigma^2} \tag{1.21}$$

where $\hat{\theta}$ is the estimate of the summary effect from the random-effects meta-analysis, n is the total number of studies in the meta-analysis and t_{n-2} is the $100(1-\alpha/2)$ percentile of the t-distribution with $n-2$ degrees of freedom. $\hat{\tau}^2$ is the estimate of between-study variation and σ^2 is the variance of the summary effect estimate.

The calculated prediction interval can be depicted on the forest plot for the random-effects meta-analysis (Riley et al. 2011). It is represented by two horizontal lines that extend out from the vertices of the diamond that represent the extremities of the confidence interval of the summary effect estimate. The width of the line represents the width of the prediction interval and extremities of the line are the prediction interval upper and lower bounds.

A couple of examples of a random-effects meta-analysis and calculation of prediction intervals are now provided.

1.3.7 Ten hypertension trials example revisited

Recall the continuous outcome example from 10 hypertension trials (Riley et al. 2013, Wang et al. 2005). The systolic blood pressure was measured at baseline and follow-up for those in the anti-hypertensive treatment group and those in the control group. The treatment effect estimates are derived from the final score, change score or ANCOVA models.

For the three different models, univariate random-effects meta-analyses were applied to analyse the treatment effect estimates from 10 hypertension trials, with the DerSimonian and Laird (DL) method used for estimation. The results from each meta-analysis were displayed in forest plots (Figure 1.3 and, Figures A.3 and A.4 in Appendix A.2) and the results from each meta-analysis (including the fixed-effect) were summarised in Table 1.2. The forest plots for the random-effects include the prediction intervals which are not included in forest plots from fixed-effect meta-analyses.

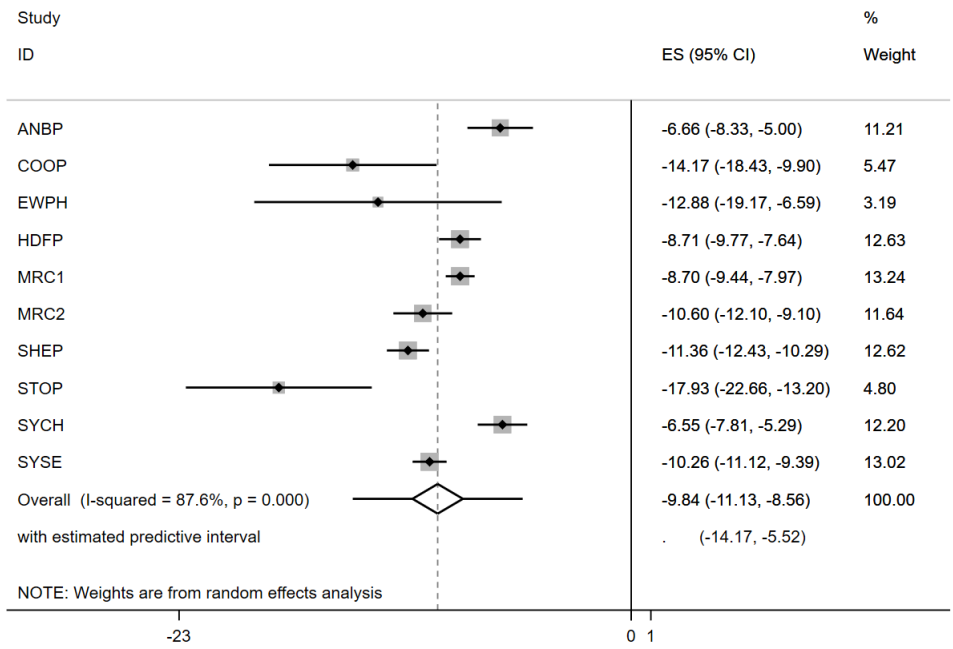


Figure 1.3: Random-effects meta-analysis of treatment effect estimates from the ANCOVA model

Table 1.2: Results from the fixed-effect and random-effects univariate meta-analyses for the Hypertension example

	Fixed-effect[a]		Random-effects		
Model	Estimate	95% CI	Estimate	95% CI	95% Prediction Interval
Change	-9.47	-9.91 to -9.04	-9.83	-11.15 to -8.52	-14.22 to -5.45
Final	-9.29	-9.71 to -8.87	-9.81	-11.12 to -8.50	-14.20 to -5.42
ANCOVA	-9.31	-9.71 to -8.92	-9.84	-11.13 to -8.56	-14.17 to -5.52

[a] Fixed-effect results as in Table 1.1

There is statistically significant evidence that, on average, the anti-hypertensive treatment reduces the systolic blood pressure compared to control, from the random-effects meta-analysis results from all models. From Table 1.2, the summary results from the random-effects meta-analyses differ slightly from the fixed-effect meta-analyses, with pooled estimates slightly larger. Additionally, the 95% confidence intervals are

wider from the random-effects meta-analyses than the fixed-effect meta-analyses. This reflects the additional between-study variance that is included in the weight and consequently the variance of the summary treatment effect estimate.

1.3.8 Bolus thrombolytic therapy example revisited

The binary outcome example is the meta-analysis of nine phase II trials that investigated the number of deaths in patients who were treated with Bolus thrombolytic therapy following an acute myocardial infarction compared to treated with standard infusion therapy (Eikelboom et al. 2001, Smalling et al. 1995, Bode et al. 1996, Kawai et al. 1997, Vanderschueren et al. 1997, Bär et al. 1998, Bleich et al. 1998, den Heijer et al. 1998, Cannon et al. 1998, Park 1998). The odds ratios from each trial were analysed in a random-effects meta-analysis via DL estimation method and the results are presented in Figure 1.4 and Table 1.3 (with the fixed-effect meta-analysis results).

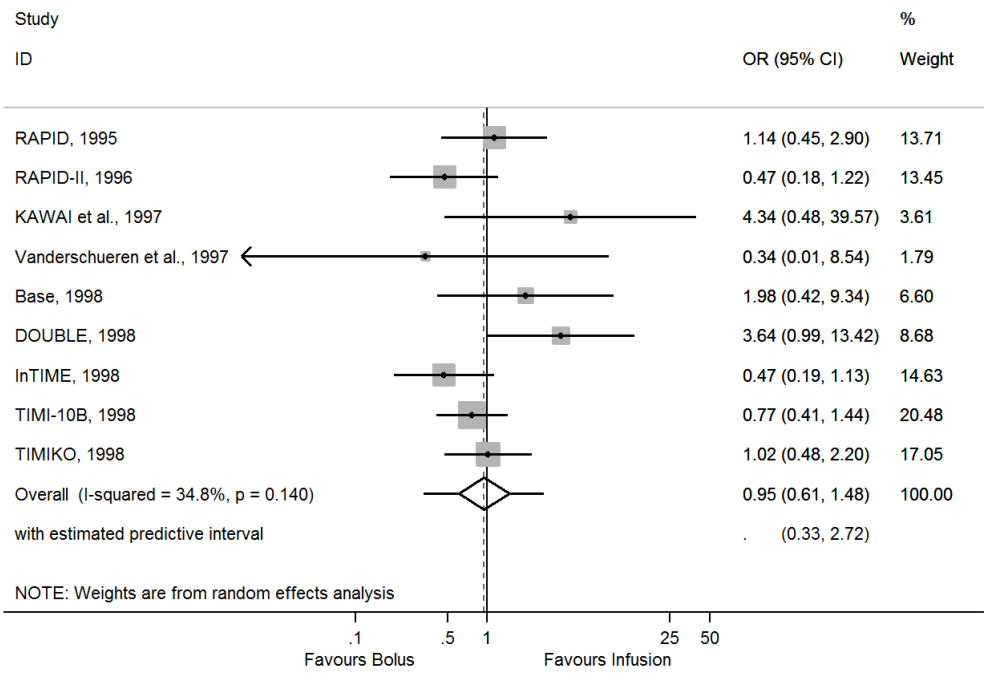


Figure 1.4: Random-effects meta-analysis of mortality from the Bolus thrombolytic therapy studies

Table 1.3: Results from the fixed-effect and random-effects univariate meta-analyses for the Bolus example

Assumption	Estimate	95% C.I.	95% Prediction Interval
Fixed-effect	0.94	0.64 to 1.29	
Random-effects	0.95	0.61 to 1.48	0.33 to 2.72

The odds of dying for patients treated with Bolus thrombolytic therapy was 6% lower (95% C.I.: 32% lower to 29% higher) than patients treated with standard infusion therapy following an acute myocardial infarction, from the fixed-effect meta-analysis. However, from the random-effects meta-analysis, the odds of dying were, on average, 5% lower (95% C.I.: 39% lower to 48% higher) for patients treated with Bolus thrombolytic therapy compared to standard infusion therapy following an acute myocardial infarction. From the random-effects meta-analysis (only), the prediction interval was 0.33 to 2.72 , suggesting large uncertainty in the actual treatment effect across different settings.

1.3.9 Estimation methods for random-effects meta-analyses

There are many different estimation methods that have been developed for the estimation of the value of τ^2 in random-effects meta-analyses and there have been many research papers that have compared estimation methods (Brockwell & Gordon 2001, Jackson et al. 2010, Thompson & Sharp 1999, Kontopantelis & Reeves 2012*a*,*b*). This section will provide details about some of the most common estimation methods: Dersimonian-Laird (DL), maximum likelihood (ML) and restricted maximum likelihood (REML) (DerSimonian & Laird 1986, Hardy & Thompson 1996, Normand 1999).

Dersimonian-Laird (DL)

The Dersimonian-Laird (DL) is a non-iterative method and therefore is less computationally intensive compared to some alternative methods (DerSimonian & Laird 1986, Kontopantelis & Reeves 2012*b*). To estimate the value of the between-study variance,

τ^2, the DL method uses a method of moments estimator, based on the Cochran's Q statistic (Kontopantelis & Reeves 2012*b*). The Cochran's Q statistic is described in Equation 1.12.

Maximum likelihood (ML)

The maximum likelihood (ML) method is an iterative, parametric method for estimating the between-study variance, τ^2 (Hardy & Thompson 1996, Kontopantelis & Reeves 2012*b*). The assumption for ML is that the random effects are normally distributed (Kontopantelis & Reeves 2012*b*, Ma & Mazumdar 2011). Additionally, the ML method is considered to be downwardly biased (Thompson & Sharp 1999).

Restricted maximum likelihood (REML)

Restricted maximum likelihood (REML) is an iterative and parametric method that assumes that the random effects from the study are normally distributed (Kontopantelis & Reeves 2012*b*,*a*). REML uses a penalised version of the likelihood to avoid the downward bias in variance parameters issue from ML estimation (Brown & Kempton 1994, Higgins et al. 2001).

Example to compare different estimation methods

Using the continuous data example with the systolic blood pressure from the antihypertensive treatment group and the control group, Section 1.3.2, the different estimation methods were examined.

Table 1.4: Examining the results from different estimation methods for the Hypertension data

Model	Estimation Method	Estimate	95% CI	I^2	Between-study variance
Change	ML	-10.01	-11.64 to -8.38	90.90	5.33
	REML	-10.08	-11.87 to -8.30	92.49	6.57
	DL	-9.83	-11.15 to -8.52	85.57	3.17
Final	ML	-10.05	-11.73 to -8.37	92.14	5.77
	REML	-10.14	-11.98 to -8.30	93.53	7.11
	DL	-9.81	-11.12 to -8.50	86.58	3.17
ANCOVA	ML	-10.09	-11.78 to -8.41	93.09	5.91
	REML	-10.17	-12.00 to -8.34	94.25	7.18
	DL	-9.84	-11.13 to -8.56	87.57	3.09

The I^2 statistic from each method was large, over 85%, and the estimate of the between-study variance, τ^2, was large, over 3 (Table 1.4). Therefore, the studies cannot be assumed to estimate the same effect size. The DL method had the greatest summary estimate and REML had the least summary effect for all the treatment effect estimation models.

1.3.10 Hartung-Knapp correction for confidence intervals

The Hartung-Knapp Correction was developed to address the problem of underestimating the true variance of the summary effect estimate (Hartung & Knapp 2001*a*,*b*). The Hartung-Knapp correction inflates the variance of the summary effect estimate and uses the t-distribution, with the number of degrees of freedom equal to the number of studies less one, to calculate the 95% confidence interval (Knapp & Hartung 2003).

Example - Continuous Outcomes

The Hartung-Knapp correction was used on the Hypertension continuous data example with the REML method to observe the difference the Hartung-Knapp correction has upon the width of the confidence intervals:

$$\hat{\theta} \pm t \times s.e._{HK}(\hat{\theta}) \tag{1.22}$$

Table 1.5: The results for the Hypertension data using the estimation method REML, with and without Hartung-Knapp (H-K) correction

Model	Confidence interval derivation	Estimate	95% CI
Change	With H-K	-10.08	-12.32 to -7.85
	Without H-K	-10.08	-11.87 to -8.30
ANCOVA	With H-K	-10.17	-12.45, -7.89
	Without H-K	-10.17	-12.00 to -8.34
Final	With H-K	-10.14	-12.44 to -7.83
	Without H-K	-10.14	-11.98 to -8.30

The summary estimates remained the same between the use and absence of the Hartung-Knapp correction, as expected (Table 1.5). The confidence intervals calculated using the Hartung-Knapp correction are wider than those calculated without the Hartung-Knapp correction. For example, for the summary treatment effect estimates from the final score model, the confidence interval calculated with the Hartung-Knapp was -12.44 to -7.83, compared to -11.98 to -8.30 without the Hartung-Knapp correction.

Example - Binary Outcomes

Similarly, the Hartung-Knapp correction was applied to the Bolus binary outcome example. The Hartung-Knapp correction was used with the REML estimation method.

Table 1.6: The results for the Bolus data using the estimation method REML, with and without Hartung-Knapp (H-K) correction

Confidence interval derivation	Estimate	95% CI
With H-K	0.937	0.549 to 1.598
Without H-K	0.937	0.616 to 1.425

The Hartung-Knapp corrected 95% confidence interval is wider than the 95% confidence interval without the Hartung-Knapp correction (Table 1.6).

1.4 Multivariate meta-analysis methods

Multivariate meta-analysis is an extension of the univariate meta-analysis, described in Section 1.3. Univariate meta-analysis considers one outcome in the analysis, whereas multivariate meta-analysis considers two or more outcomes in a single analysis (Raudenbush et al. 1988, Becker 2000, van Houwelingen et al. 2002).

1.4.1 Bivariate models

A bivariate meta-analysis is an approach for two outcomes from studies that have been identified to aid in answering the posed research question (Riley et al. 2007*a*,*b*). The bivariate meta-analysis is the simplest multivariate meta-analysis, however this is still more complicated than a univariate due to the inclusion of correlations in the model. In the same way as the univariate meta-analysis, the bivariate meta-analysis can either use a fixed-effect or a random-effects assumption.

Bivariate fixed-effect meta-analysis

The bivariate fixed-effect meta-analysis is modelled using the normal distribution, as follows (van Houwelingen et al. 2002, Riley et al. 2007*b*):

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \delta_i \right) \quad , \qquad \delta_i = \begin{pmatrix} s_{i1}^2 & \lambda_i \\ \lambda_i & s_{i2}^2 \end{pmatrix} \tag{1.23}$$

where Y_{i1} and Y_{i2} are the observed effect sizes from study i for outcomes 1 and 2, respectively. θ_1 and θ_2 are the true effect sizes for outcome 1 and 2 in each study in the meta-analysis. The within-study covariance, λ_i, is calculated for each study i as $\lambda_i = \rho_{W_i} s_{i1} s_{i2}$, where ρ_{W_i} is the within-study correlation, and s_{i1}^2 and s_{i2}^2 are the within-study variances for Y_{i1} and Y_{i2}, respectively.

For meta-analyses with within-study correlations of zero, the bivariate fixed-effect meta-analysis reduces down to a univariate fixed-effect meta-analysis:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \delta_i \right) \quad , \qquad \delta_i = \begin{pmatrix} s_{i1}^2 & 0 \\ 0 & s_{i2}^2 \end{pmatrix} \tag{1.24}$$

Bivariate random-effects meta-analysis

Recall that the random-effects meta-analysis does not assume that each study is estimating the same true effect size, and the observed effect sizes from each study and the true effect sizes for each study are assumed to be normally distributed. The bivariate random-effects meta-analysis is written as:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \delta_i \right) \quad , \qquad \delta_i = \begin{pmatrix} s_{i1}^2 & \lambda_i \\ \lambda_i & s_{i2}^2 \end{pmatrix} \tag{1.25}$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Omega \right) \quad , \qquad \Omega = \begin{pmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{pmatrix} \tag{1.26}$$

where Y_{i1} and Y_{i2} are the observed effect sizes from study i for outcomes 1 and 2, respectively. θ_{i1} and θ_{i2} are the true effect sizes for study i for outcomes 1 and 2. The within-study covariance, λ_i, is calculated by $\lambda_i = \rho_{W_i} s_{i1} s_{i2}$ where ρ_{W_i} is the within-study correlation, and s_{i1} and s_{i2}^2 are the within-study variances for outcome 1 and 2, respectively. The mean effect sizes across all studies is denoted by μ_1 and μ_2 for outcome 1 and outcome 2, respectively. The between-study covariance, τ_{12}, is calculated by $\tau_{12} = \rho_\mu \tau_1 \tau_2$ where ρ_μ is the between-study correlation, and τ_1^2 and τ_2^2 are the between-study variances for outcome 1 and outcome 2, respectively.

In the same way as the bivariate fixed-effect meta-analysis, the bivariate random-effects meta-analysis is reduced to a univariate random-effects meta-analysis, with zero within-study correlations and zero between-study correlation (Riley et al. 2007*b*, van Houwelingen et al. 2002).

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \delta_i \right) \quad , \qquad \delta_i = \begin{pmatrix} s_{i1}^2 & 0 \\ 0 & s_{i2}^2 \end{pmatrix} \tag{1.27}$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Omega \right) \quad , \qquad \Omega = \begin{pmatrix} \tau_1^2 & 0 \\ 0 & \tau_2^2 \end{pmatrix} \tag{1.28}$$

1.4.2 Multivariate meta-analysis

The bivariate meta-analysis method, Section 1.4.1, can be extended for the multivariate meta-analysis method (Raudenbush et al. 1988, Becker 2000, van Houwelingen et al. 2002, Jackson et al. 2011). It can be extended to generalised matrices for both fixed-effect and random-effects meta-analysis.

Multivariate fixed-effect meta-analysis

The generalised model for the multivariate fixed-effect meta-analysis is written as:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}, \delta_i \right) \quad , \quad \delta_i = \begin{pmatrix} s_{i1}^2 & \lambda_{i(1,2)} & \dots & \lambda_{i(1,n-1)} & \lambda_{i(1,n)} \\ \lambda_{i(1,2)} & s_{i2}^2 & \dots & \lambda_{i(2,n-1)} & \lambda_{i(2,n)} \\ \dots & \dots & \ddots & \dots & \dots \\ \lambda_{i(1,n)} & \lambda_{i(2,n)} & \dots & \lambda_{i(n-1,n)} & s_{in}^2 \end{pmatrix} \tag{1.29}$$

where the vector of $(Y_{i1}, ..., Y_{in})$ contains the observed values for each variable from study i and is normally distributed where the mean is the vector of true effect sizes, $(\theta_1, ..., \theta_n)$, for each outcome, 1 to n, and the variance is the within-study covariance matrix (a square matrix), δ_i. The elements of the lead diagonal of the within-study covariance matrix, δ_i, are the within-study variances, s_{ij}^2, is the variance of variable j in study i and $\lambda_{i(j_1,j_2)}$ is the within-study covariance for outcomes j_1 and j_2 in study i.

Multivariate random-effects meta-analysis

For the random-effects meta-analysis, the generalised multivariate model is written as:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \\ \vdots \\ \theta_{in} \end{pmatrix}, \delta_i \right) \quad , \quad \delta_i = \begin{pmatrix} s_{i1}^2 & \lambda_{i(1,2)} & \dots & \lambda_{i(1,n-1)} & \lambda_{i(1,n)} \\ \lambda_{i(1,2)} & s_{i2}^2 & \dots & \lambda_{i(2,n-1)} & \lambda_{i(2,n)} \\ \dots & \dots & \ddots & \dots & \dots \\ \lambda_{i(1,n)} & \lambda_{i(2,n)} & \dots & \lambda_{i(n-1,n)} & s_{in}^2 \end{pmatrix} \tag{1.30}$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \\ \vdots \\ \theta_{in} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \Omega \right) \quad , \quad \Omega = \begin{pmatrix} \tau_1^2 & \tau_{(1,2)} & \dots & \tau_{(1,n-1)} & \tau_{(1,n)} \\ \tau_{i(1,2)} & \tau_2^2 & \dots & \tau_{(2,n-1)} & \tau_{(2,n)} \\ \dots & \dots & \ddots & \dots & \dots \\ \tau_{(1,n)} & \tau_{(2,n)} & \dots & \tau_{(n-1,n)} & \tau_n^2 \end{pmatrix} \tag{1.31}$$

where the vector of $(Y_{i1}, ..., Y_{in})$ contains the observed values for each outcome from study i. This vector is normally distributed where the mean is the vector $(\theta_{i1}, ..., \theta_{in})$, that contains the true effect sizes for each outcome from study i, and the variance

is the matrix δ_i, the within-study variance-covariance matrix. The elements of the within-study variance-covariance matrix on the lead diagonal are s_{ij}^2, the variance of outcome j in study i and the remaining elements are $\lambda_{i(j_1,j_2)}$, the within-study covariance of outcome j_1 and j_2 in study i. For the random-effects meta-analysis the vector $(\theta_{i1}, ..., \theta_{in})$ is normally distributed where the mean is the vector of $(\mu_1, ..., \mu_n)$, the mean effect sizes across all studies for outcomes 1 to n and the variance is the between-study variance-covariance matrix, Ω. The between-study variance-covariance matrix contains on the lead diagonal the elements, τ_j^2, the between-study variance for outcome j and the remaining elements are the between-study covariances between the outcomes j_1 and j_2 , denoted by $\tau_{(j_1,j_2)}$.

1.4.3 Within-study and between-study correlations

Within-study correlations

The within-study correlation, ρ_{W_i}, is the association between the observed values for the effect estimates for two outcomes, Y_{i1} and Y_{i2}, within the same study (Riley et al. 2007*b*). It is assumed that the within-study correlation for each study is known, however, these are rarely reported in studies (Riley 2009, Jackson et al. 2011, Mavridis & Salanti 2013, Riley et al. 2007*b*, Wei & Higgins 2013*b*).

Between-study correlations

The between-studies correlation, ρ_μ, is the association between the true values across studies for two outcomes (Riley et al. 2007*b*, Riley 2009). Within a multivariate meta-analysis the between-study correlation is not assumed to be known and it is required that the between-study correlation is estimated (Riley et al. 2007*b*).

1.4.4 Within-study correlations using individual participant data

In a multivariate meta-analysis, the within-study correlations are assumed to be known (Riley 2009, Riley et al. 2007*b*). However, the within-study correlations are rarely reported in study/trial publications (Riley 2009, Jackson et al. 2011, Mavridis & Salanti 2013, Riley et al. 2007*b*, Wei & Higgins 2013*b*). Without within-study correlations the results from the meta-analysis are less accurate and the standard errors of the summary effect estimates are increased (Riley 2009). However, with the individual participant data (IPD) from the study/trial, the within-study correlation can be estimated (Riley et al. 2014).

There are two methods for estimating the within-study correlations between outcomes from IPD (Riley et al. 2014). The first method is joint linear regression and the second method is bootstrapping. Joint linear regression is used for the estimation of within-study correlations between continuous outcomes. For continuous outcomes with a baseline measure and a follow-up measure, the following model can be used in each trial to fit two regressions jointly:

$$\begin{gathered} y_{Fijk} = \alpha_{ik} + \beta_{ik} y_{Bijk} + \theta_{ik} x_{ij} + e_{ijk} \\ e_{ij1} \sim N(0, \sigma_{i1}^2) \quad e_{ij2} \sim N(0, \sigma_{i2}^2) \\ \operatorname{cov}(e_{ij1}, e_{ij2}) = \sigma_{i12} \end{gathered} \tag{1.32}$$

where y_{Fijk} and y_{Bijk} are the follow-up and baseline observed values for outcome k from participant j in trial i, respectively. β_{ik} denotes the mean increase in the follow-up measure for a unit increase in the baseline measure in trial i for outcome k. θ_{ik} is the effect size for outcome k in trial i and x_{ij} is the indicator variable for the treatment group. The residual error, e_{ijk}, is normally distributed with mean 0 and variance σ_{ik}^2. The inverse of Fisher's Information matrix provides the variances and the within-study correlation (Riley et al. 2014).

The second method, bootstrapping, can be used for both discrete and continuous outcomes (Riley et al. 2014). The method of bootstrapping randomly samples with replacement participants from the original dataset until the generated dataset is the same size as the original. The relevant model for each outcome is fitted to obtain the effect estimates of interest. This process is repeated, say 1000 times, for 1000 bootstrap samples and the analysis model is applied to each bootstrap dataset separately, to provide 1000 values of the effect estimates for each outcome. For each pair of outcome the observed correlation provides the within-study correlation.

1.4.5 Estimation methods for multivariate meta-analysis

The estimation methods for multivariate meta-analysis are extended from the estimation methods used for univariate random-effects meta-analysis, described in Section 1.3.9. The methods for multivariate meta-analysis described in this section are the restricted maximum likelihood (REML), maximum likelihood (ML) and two methods utilising the method of moments (MM) approach. This is not an exhaustive list as further estimation methods have been developed however this section only covers the most common methods (Ma & Mazumdar 2011).

Restricted maximum likelihood (REML) and maximum likelihood (ML)

Restricted maximum likelihood (REML) is a parametric, iterative method that can be extended from the univariate meta-analysis approach to a multivariate meta-analysis (van Houwelingen et al. 2002, Ma & Mazumdar 2011, Jackson et al. 2011, DerSimonian & Laird 1986). REML's multivariate random-effects model assumes a multivariate normal distribution (Ma & Mazumdar 2011). An advantage of REML is that it corrects downwardly biased variance estimates using a penalised version of the likelihood that otherwise would arise from maximum likelihood (ML) estimation (Jackson et al. 2011). The maximum likelihood (ML) can also be applied to multivariate meta-analyses (van Houwelingen et al. 2002).

Method of Moments (MM)

Jackson et al. (2010) suggested an extension of the Dersimonian-Laid (DL) method for multivariate meta-analysis. The DL method is a non-iterative method that does not require a normality assumption. Therefore, this method of moments approach is less computationally intensive than the previous likelihood approaches.

Another method of moments method was proposed by Jackson et al. (2013) and reduces to the method proposed by Chen et al. (2012). The method of moments method developed is non-iterative and it does not take into account the uncertainty in the estimated between-study covariance matrix.

1.4.6 Confidence intervals and prediction intervals for multivariate meta-analysis

The confidence intervals and prediction intervals can be derived in a similar way to the methods described for the univariate meta-analysis (Sections 1.3.1, 1.3.5 and 1.3.6). In particular, Jackson & Riley (2014) suggest how the Hartung-Knapp correction might be extended to the multivariate meta-analysis approach. The confidence intervals for the refined method proposed by Jackson & Riley (2014) can be calculated using:

$$\hat{\theta}_j \pm t_{(N-p;1-\frac{\alpha}{2})}\sqrt{H^2 \mathrm{Var}(\hat{\theta}_j} \tag{1.33}$$

where $t_{(N-p;1-\alpha/2)}$ represents the univariate t distribution at the $(1-\alpha/2)$ level with $(N-p)$ degrees of freedom , where N is the total number of estimates and p is the number of outcomes included in the meta-analysis. The H^2 statistic is a calculated scaling factor. Jackson & Riley (2014) suggest that for meta-analyses with many studies there are limited benefits. The scaling factor H^2 can be constrained to be greater than or equal to one, to ensure that narrower confidence intervals are not calculated.

1.4.7 Joint inferences to describe relationships between multiple outcomes

Following estimation of a multivariate meta-analysis model, joint inferences can be performed to inform about relationships between the effects of treatment on two or more outcomes (Riley et al. 2015). Examples of this are confidence regions and prediction regions. The confidence region is the area where there is, say, 95% confidence that the true pooled effects for the outcomes lie. The prediction interval is the area where there is, say, 95% confidence that for a new study the true pair of effects for the outcomes lie.

Joint inferences can be plotted on a graph using the paired treatment effects and within-study correlations (Riley et al. 2015). The graphs can visually display the joint probability statements that can be calculated using Bayesian statistics.

1.4.8 Borrowing of Strength (BoS)

Borrowing of strength in multivariate meta-analysis refers to using data from other available correlated effect sizes to gain extra information on the main effect size of interest (Jackson et al. 2017). For example, consider a meta-analysis with outcome overall survival, in a multivariate meta-analysis, strength can be borrowed from another outcome that is correlated with overall survival, such as disease-free survival. Borrowing of strength requires a multivariate meta-analysis framework, as this allows correlations between multiple effect sizes to be incorporated and used to inform the summary effect estimates. In contrast, univariate meta-analysis does not utilise such correlation and therefore cannot borrow strength between correlated effect sizes. An advantage of borrowing strength is that through the utilisation of additional data via the correlations, the summary effect estimates from a multivariate meta-analysis are often more precise than their respective univariate summary effect estimates (Riley et al. 2007*b*).

It therefore follows that the borrowing of strength can be quantified by comparing uni-

variate and multivariate analyses, as follows. Consider a univariate meta-analysis with one outcome, j, where the summary effect estimate from the univariate meta-analysis is $\hat{\theta}_j$. The precision of $\hat{\theta}_j$ from the univariate meta-analysis defines the information that contributed to the summary effect estimate $\hat{\theta}_j$ from all the studies with outcome j. This is also referred to as the *direct* information since the information is directly from outcome j. Now consider a multivariate meta-analysis with multiple outcomes, where one of these outcomes is outcome j (the same outcome in the univariate). The summary effect estimate for outcome j from the multivariate meta-analysis is $\hat{\theta}_{j,mv}$. The precision of $\hat{\theta}_{j,mv}$ defines the information that contributed to the summary effect estimate $\hat{\theta}_{j,mv}$ from all studies with outcome j and all the studies in the meta-analysis without outcome j, but with information from the other outcomes. This is referred to as the *total* information that can be attributed to the summary effect estimate for outcome j. Subsequently, the *indirect* information can be defined as the information gained in the meta-analysis from the other outcomes and the relationships between the other outcomes and outcome j, the outcome of interest. Note, *indirect* information would not be included in the univariate meta-analysis but the multivariate meta-analysis includes the *total* information, both *direct* and *indirect* information. The BoS statistic is defined mathematically as:

$$\text{BoS}_j = 100\% \times \left[1 - \frac{prec_{direct}(\hat{\theta}_j)}{prec_{total}(\hat{\theta}_j)}\right] \tag{1.34}$$

The precision is the reciprocal of the variance and in terms of the univariate and multivariate meta-analysis, the BoS can be written as:

$$\text{BoS}_j = 100\% \times \left[1 - \frac{\text{Var}_{mv}(\hat{\theta}_j)}{\text{Var}_{uv}(\hat{\theta}_j)}\right] \tag{1.35}$$

The definition of BoS is the percentage gain in information towards the summary result for outcome j from utilising a multivariate meta-analysis over a univariate meta-analysis.

Alternatively, Copas et al. (2018) proposed the BoS statistic is written in terms of E, the efficiency (Riley et al. 2017). The efficiency, E, is defined as the ratio between the variance of the summary effect estimate from the direct and related information and the variance of the summary estimate from the direct information only. The BoS statistic is defined in terms of E, as:

$$\text{BoS} = 100\% \times (1 - E) \tag{1.36}$$

1.4.9 Study weights

The derivation of the borrowing of strength statistic leads to a definition of study weights for multivariate meta-analysis (Jackson et al. 2017). The weight of study i for outcome j is the ratio of the *total* information from study i for outcome j and the *total* information for outcome j in the analysis. Therefore the study weights, W_{ij}, for outcome j and study i, can be described as the sum of the *direct* information for outcome j from study i, DI_{ij}, and the borrowing of strength of outcome j from study i, BoS_{ij}, as follows:

$$W_{ij} = DI_{ij} + \text{BoS}_{ij} \tag{1.37}$$

1.5 Benefits of multivariate meta-analysis over univariate meta-analysis

In this section, the key advantages of multivariate meta-analysis over univariate meta-analysis are described.

1.5.1 Single analysis for all outcomes

Multivariate meta-analysis allows for all the estimates to be obtained from one analysis (Jackson et al. 2011). This is advantageous for researchers as all outcomes can be investigated in a single analysis over multiple individual analyses.

1.5.2 Describes the relationships between multiple outcomes and joint inferences

The relationship between outcomes can be described by a multivariate meta-analysis (Riley et al. 2007*a*,*b*, Jackson et al. 2011). In a random-effects multivariate meta-analysis the between-study correlations (the correlation between the true effect sizes) between outcomes is estimated, which may itself be of interest, for example in surrogate outcomes research.

An additional method to describe the relationships between multiple outcomes, that is possible with multivariate meta-analysis, is the application of joint inferences, described in further detail in Section 1.4.7 (Riley et al. 2015).

1.5.3 Better statistical properties

This benefit of using a multivariate meta-analysis over a univariate meta-analysis is particularly relevant when there are studies with missing outcomes included (Riley et al. 2007*a*,*b*, 2008, Jackson et al. 2011, 2017, Copas et al. 2018). A multivariate meta-analysis can borrow strength across multiple outcomes to include further information in the analysis that otherwise would be absent in a univariate meta-analysis, as the studies without the outcome of interest would be excluded.

An example of a multivariate meta-analysis borrowing strength across multiple outcomes will now be discussed, using data from The Fibrinogen Studies Collaboration (The Fibrinogen Studies Collaboration 2009, Jackson et al. 2017). The data contained 31 observational studies and each study had available the Individual Participant Data (IPD). In each observational study, Cox proportional hazards models were used to obtain hazard ratios (Cox 1972). For each of the 31 observational studies, there were different confounders recorded and the main confounders were used to adjust the results for all 31 observational studies, these are referred to as the partially adjusted hazard ratios. For 14 of the observational studies, further confounders, in addition

to the main confounders, were also used to fully adjust the results, referred to as the fully adjusted hazard ratios. For these 14 studies, both the fully and partially adjusted hazard ratios were obtained, along with their variances. For the remaining 17 studies, only the partially adjusted hazard ratios and variances were obtained. Therefore, our two correlated outcomes are 'fully' and 'partially' adjusted estimates. Indeed, the within-study and between-study correlations are expected to be very large.

The summary hazard ratios were the same from the univariate and multivariate meta-analyses for the partially and fully adjusted hazard ratios, 1.41 and 1.31, respectively (Table 1.7). For the partially adjusted hazard ratios (outcome with complete data) the confidence intervals were similar between the univariate and multivariate results. However, for the fully adjusted hazard ratios the confidence intervals were narrower, and therefore more precise, from the multivariate results (95% C.I.: 1.25 to 1.38) than the univariate results (95% C.I.: 1.22 to 1.42).

Table 1.7: Meta-analysis results and borrowing of strength for the partially and fully adjusted results from the Fibrinogen studies

Meta-analysis	Adjusted Results	Hazard Ratio	Standard Error	95% Confidence Interval	BoS Statistic
Univariate	Partially	1.41	0.040	1.33 to 1.49	
	Fully	1.31	0.051	1.22 to 1.42	
Multivariate	Partially	1.41	0.042	1.33 to 1.50	1.8%
	Fully	1.31	0.035	1.25 to 1.38	53.3%

The borrowing of strength statistic was calculated as 53.3% for the fully adjusted results and 1.8% for the partially adjusted results (Table 1.7). There were fewer studies for the fully adjusted hazard ratios and consequently information from the relationship between the fully and partially adjusted hazard ratios was borrowed to inform the summary hazard ratio for the fully adjusted hazard ratio.

1.5.4 Reducing bias due to partial reporting

Bias due to partial reporting can be reduced by analysing the data using a multivariate meta-analysis (Jackson et al. 2011, Kirkham et al. 2012). An example of this is a meta-analysis that investigated the prognostic ability of marker p53 for patients with squamous cell carcinoma (Table 1.8) (Tandon et al. 2010, Jackson et al. 2011). There were six studies that were included in the multivariate meta-analysis. The outcomes were recorded as the log hazard ratio of disease-free survival and the log hazard ratio for overall survival. All six studies recorded an effect size for the overall survival but only three studies recorded an effect size for the disease-free survival.

Table 1.8: The data for the meta-analysis that investigated the prognostic ability of marker p53 for patients with squamous cell carcinoma

Study	Log hazard ratios		Standard errors	
	disease-free survival	overall survival	disease-free survival	overall survival
1	-0.58	-0.18	0.56	0.56
2		0.79		0.24
3		0.21		0.66
4	-1.02	-0.63	0.39	0.29
5		1.01		0.48
6	-0.69	-0.64	0.40	0.40

The studies that reported the disease-free survival log hazard ratios recorded negative log hazard ratios for both the disease-free survival and the overall survival log hazard ratios (Table 1.8). However, the overall log hazard ratios were positive for those studies that did not report the disease-free survival. A univariate meta-analysis for disease-free survival will assume that all the evidence provides only negative effect estimates (i.e. log HRs< 0). However with the correlation between the outcomes expected to be large, it is likely that the missing disease-free survival log hazard ratios are not negative as those studies provide positive estimates for the reported overall survival outcome.

Due to the partial reporting of the disease-free survival, the results from the univariate meta-analysis are at risk of bias towards the negative log hazard ratios. The bias in results can be reduced through the use of a multivariate meta-analysis.

From the example, the summary hazard ratio for the disease-free survival from the multivariate meta-analysis was 0.72 (95% C.I.: 0.41 to 1.29) when the correlation was assumed to be 0.7 and 0.84 (95% C.I.: 0.45 to 1.25) when the correlation was assumed to be 0.95 (Table 1.9). Thus, there was no statistically significant evidence that there was any difference in the survival between those with mutant p53 and normal p53. However, from the univariate meta-analysis there was statistically significant evidence that the patients with the mutant p53 were less likely to die or have a recurrence than patients without the mutant p53 gene (HR: 0.45, 95% C.I.: 0.27 to 0.73). So multivariate meta-analysis has changed our conclusions.

Table 1.9: The results for the disease-free survival for the p53 marker using REML

Meta-analysis	Correlation	Hazard Ratio	Standard Error	95% Confidence Interval
Univariate		0.45	0.11	0.27 to 0.73
Multivariate	0.7	0.72	0.21	0.41 to 1.29
	0.95	0.75	0.20	0.45 to 1.25

1.6 Limitations of multivariate meta-analysis

In this section, the key limitations of multivariate meta-analysis are described.

1.6.1 Complexity of multivariate meta-analysis

Multivariate meta-analysis is more complicated for researchers to understand and implement than univariate meta-analysis. This is partly due to the further assumptions that are required for multivariate meta-analysis. Similarly to the univariate, the multi-

variate approach assumes that the random-effects are normally distributed. However, this assumption is more difficult to verify in multivariate meta-analyses than univariate meta-analyses for small number of studies. Additionally, the multivariate meta-analysis assumes a linear relationship between pairs of outcomes effects across studies, since there is usually not enough studies for the effects to model a non-linear relationship. In addition to the further assumptions, additional information is required for a multivariate meta-analysis; within-study correlations are required to be known and for the between-study correlations to be estimable.

1.6.2 Estimation problems

Another limitation of applying multivariate meta-analysis is difficulties can arise during the estimation of the between-study variances and correlations. There can be estimation problems particularly for the correlations which can be estimated as either 1 or -1 (the extreme boundaries for the correlation). Additionally, the estimation of the between-study correlation varies depending on the estimation methods used to estimate the correlation. An example of this is the meta-analysis example that investigated the prognostic ability of marker p53 for patients with squamous cell carcinoma in Table 1.8 (Tandon et al. 2010, Jackson et al. 2011). The data was analysed using the estimation methods REML and MM. The results are provided in Table 1.10. The between-study correlation was estimated using the REML estimation method as 1 and using the MM estimation method as -1.

Table 1.10: The results for the disease-free survival and overall survival for the p53 marker using REML and MM estimation methods for univariate and multivariate meta-analyses

Method of Estimation	Meta-analysis	Correlation	Survival	Hazard Ratio	Standard Error	95% Confidence Interval	Between-study Correlation
REML	Multivariate	0.7	disease-free	0.72	0.21	0.41 to 1.29	1
		0.95		0.75	0.20	0.45 to 1.25	1
		0.7	overall	1.10	0.34	0.60 to 2.02	1
		0.95		1.11	0.34	0.61 to 2.03	1
	Univariate		disease-free	0.45	0.11	0.27 to 0.73	
			overall	1.09	0.34	0.59 to 2.02	
MM	Multivariate	0.7	disease-free	0.46	0.12	0.28 to 0.77	-1
		0.95		0.47	0.12	0.28 to 0.78	-1
		0.7	overall	1.07	0.36	0.55 to 2.07	-1
		0.95		1.06	0.36	0.55 to 2.06	-1
	Univariate		disease-free	0.45	0.11	0.27 to 0.73	
			overall	1.09	0.37	0.56 to 2.12	

1.6.3 Publication bias

Multivariate meta-analysis does not correct for publication bias in the meta-analysis; the results from the multivariate meta-analysis may still be biased if studies are unavailable because they are not reported at all. This might arise in cases where the missing data in the meta-analysis is missing not at random.

1.6.4 Benefits may be limited

It was previously discussed in Section 1.5.3 that the results from a multivariate meta-analysis had better statistical properties than the results from a univariate meta-analysis (Jackson et al. 2011). However, the statistical properties for the multivariate meta-analysis may only be marginally improved from the univariate meta-analysis (Trikalinos et al. 2013, 2014). That is, the gain in precision of summary results may only be small and may not change the clinical conclusions.

1.7 Aims of the thesis

The key aim for this thesis is to ascertain the benefits of multivariate meta-analysis over univariate meta-analysis through empirical reviews, statistical theory and simulation studies. To achieve this broad objective, I aim to facilitate the application of multivariate meta-analysis through the exploration and further understanding of the BoS statistic. In particular, to understand what settings lead to large BoS statistics, such that multivariate meta-analysis results may differ importantly from univariate meta-analysis results. Additionally, I aim to apply multivariate meta-analysis to novel applications in medical research and compare the results to current commonly used univariate methods.

1.8 Thesis outline

The benefits of multivariate meta-analysis have been discussed here and in published papers (Riley et al. 2007*a*,*b*, Hamza et al. 2009, Jones et al. 2009, Jackson et al. 2011, Kirkham et al. 2012, Riley et al. 2015, Frosi et al. 2015). A Health Technology Assessment (HTA) report that studied maternal and foetal outcomes following exercise and diet interventions during pregnancy is a meta-analysis study that may benefit from multivariate methods (Rogozińska et al. 2017). In Chapter 2, multivariate meta-analysis methods are applied to the HTA report data. The benefits of multivariate meta-analysis are discussed in the context of the HTA report, and compared to separate univariate meta-analysis and composite outcome meta-analyses.

In Chapter 3, the borrowing of strength (BoS) statistic is compared to the multivariate and univariate results from 43 Cochrane reviews (Trikalinos et al. 2014). This chapter investigates whether BoS can be used to identify situations when a multivariate approach is most beneficial. The meta-analysis level characteristics are investigated for their relationship with the magnitude of BoS. Chapter 4 follows with the development of prediction models for the magnitude of the BoS statistic given particular meta-analysis characteristics. The prediction models are applied to examples to demonstrate how accurately the predicted BoS statistic agrees with the observed BoS statistic. This work, comparing the univariate and multivariate meta-analysis results with the magnitude of BoS was presented at ISCB (International Society for Clinical Biostatistics) 2016 and the development of the prediction models was presented at YSM (Young Statisticians' Meeting) 2017 (Appendix A.3).

The BoS statistic is further investigated in Chapter 5 using interactive graphical tools developed in R shiny. The prediction models developed in Chapter 4 are embedded in the interactive graphs in Chapter 5. Additionally, in Chapter 5, the mathematical equation for the calculation of the BoS statistic is embedded in an interactive graphical tool. It is observed through the interactive tool for the equation of BoS that there

may exist a relationship between the magnitude of BoS and the percentage of missing data for the outcome of interest. Through mathematical reasoning, the relationship is proved for both fixed-effect and random-effects settings (Chapter 6). This work was presented at RSS (The Royal Statistical Society) conference 2018 (Appendix A.3).

In Chapter 7, there is an investigation as to whether the benefits of multivariate meta-analysis can be extended to a novel application in medical research. In randomised control trials with a continuous outcome, there are three different models to obtain the treatment effects: ANCOVA, change score and final score, as was illustrated earlier in this chapter. This leads to complexities for a meta-analysis, since randomised control trials (RCTs) may use different models to estimate the treatment effect. In Chapter 7 a multivariate meta-analysis approach is proposed to overcome this problem. An example individual participant data (IPD) meta-analysis of 10 hypertension trials is used to investigate the multivariate meta-analysis method against current univariate meta-analysis methods. Chapter 8 follows with a simulation study to assess the performance of the multivariate method against the performances of the univariate methods.

Finally, in Chapter 9 the key findings and recommendations are discussed. The practical implications of this thesis, as well as the needs for further research are presented.

Chapter 2

Multivariate meta-analysis for the effect of diet and exercise interventions on maternal and foetal outcomes

This chapter involves the re-analysis of an individual participant data (IPD) meta-analysis from a recent Health Technology Assessment (HTA) report on the effect of diet and exercise interventions during pregnancy on maternal and foetal outcomes (Rogozińska et al. 2017). The analysis in the HTA report included separate univariate meta-analyses of composite outcomes for maternal and foetal measures, as well as separate univariate meta-analysis for each individual outcome in the composite outcomes. This chapter seeks to extend the HTA work using novel multivariate meta-analysis methods; a joint model that analyses all outcomes together whilst accounting for their correlation. The intention is to produce meta-analysis results for each of the outcomes contributing toward the composite outcome, which are more clinically relevant than their respective composite meta-analysis results. The impact of missing data for individual outcomes is also reduced, which is an issue with separate univariate meta-analyses. This chapter starts by introducing the HTA report with a brief background, followed by the methods used. The alternative multivariate meta-analysis approach is then discussed and the multivariate meta-analysis results compared to the respective findings from the HTA report, which only used univariate meta-analysis of a composite outcome.

2.1 Summary of the HTA report

2.1.1 Background of the HTA report

Maternal obesity has been shown to increase the risk of adverse outcomes for the foetus as well as the mother (Cantwell et al. 2011, Thangaratinam & Jolly 2010, Rogozińska et al. 2017). Women who gain weight in excess during pregnancy are at risk of retaining excess weight and carrying this into subsequent pregnancies (Rogozińska et al. 2017). Diet and exercise interventions have been investigated as a means to address excessive weight gain. A Health Technology Assessment (HTA) report, conducted by the International Weight Management in Pregnancy (iWiP) collaborative group, synthesised existing evidence from independent studies to investigate the effect of diet and exercise interventions on maternal and foetal outcomes during pregnancy (Rogozińska et al. 2017).

The HTA report aimed to investigate the effect of diet and exercise interventions on gestational weight gain, maternal composite and foetal composite outcomes. A composite outcome is a combination of multiple individual outcomes that might happen, where any one of the events could count as part of the composite outcome. For the maternal composite outcome, the individual outcomes were pre-eclampsia or pregnancy induced hypertension, gestational diabetes, preterm delivery and any caesarean section. Similarly for the foetal composite outcome, the individual outcomes were intrauterine death, small for gestational age, large for gestational age and admission to a neonatal intensive care unit (NICU).

2.1.2 Methods from the HTA report

Inclusion and exclusion criteria

The HTA report identified randomised control trials (RCT) from a literature search that investigated the effects of diet and/or exercise interventions against no intervention. Cluster randomised control trials were included, as well as RCTs without

clustering. The RCTs were required to include an intervention of diet, exercise or a mixture of diet and exercise compared to a control of either no intervention or routine antenatal care. Additionally, the trials should have analysed gestational weight gain as well as other clinical outcomes relating to the mother and/or the foetus. Studies were excluded if the study was published before 1990, the study was in animals, the study only considered non-clinical outcomes (e.g. behaviour change) or the study was aiming to increase gestational weight gain.

Collection and checking individual participant data

The researchers from the HTA report contacted the authors of the RCTs to request the individual participant data (IPD). IPD was sought for 58 studies and was provided for 36 of these. Participants from the IPD were excluded from the analysis if the women were underweight (identified by a BMI of less than 18.5kg/m^2) or if the women had a multiple pregnancy.

Summary of IPD

There were 36 RCTs that provided IPD, 34 of which were randomised using individual participant allocation and two which were cluster RCTs. The 36 RCTs were conducted in six different continents; 22 in Europe, four in North America (three in USA and one in Canada), four in South America (Brazil), four in Australia, one in Africa (Egypt) and one in Asia (Iran). Four trials investigated diet as the intervention, 16 trials investigated exercise and 15 trials assessed an intervention of both diet and exercise. In the remaining trial, an exercise intervention arm was compared to a diet and exercise intervention arm. The number of women in each study ranged from 12 to 2212.

Analysis approach in the HTA report: Univariate meta-analysis of composite outcomes

To obtain summary effect estimates and 95% confidence intervals for the maternal composite and the foetal composite outcomes, the HTA report analysed the IPD us-

ing a two-stage IPD meta-analysis. A two-stage IPD meta-analysis analyses the IPD and obtains summary treatment effect estimates from the meta-analysis model in two stages. In the first stage, the IPD for each study is analysed to obtain the treatment effect estimates and within-study variances. In the second stage of the IPD meta-analysis, the univariate meta-analysis models described in Chapter 1 are applied to the aggregate data from the first stage.

In the first stage of the IPD meta-analysis from the HTA report, the composite outcomes were generated (Rogozińska et al. 2016) and their treatment effect estimates and standard errors for each study were obtained. Each of the composite outcomes were binary outcomes. For each study, logistic regression was applied to each outcome to obtain the log odds effect estimate (Cox 1958):

$$\text{logit}(p) = b_0 + b_1 x \tag{2.1}$$

where x is the intervention group. For a rare binary outcome with zero events in one treatment group, the Sweeting continuity correction (Sweeting et al. 2004) was applied to obtain the effect estimate. The Sweeting correction adds the reciprocal of the alternative treatment group sample size to each value in the 2x2 contingency table.

In the second-stage of the IPD meta-analysis, the treatment effect estimates for the composite outcome of interest were pooled using a univariate random-effects meta-analysis using REML (restricted maximum likelihood) estimation as described in Chapter 1. Additionally, a Hartung-Knapp correction (Knapp & Hartung 2003) was applied to the confidence intervals to account for the uncertainty in the estimate of the between-study variance (Section 1.3.10).

As well as a meta-analysis for each composite outcome, the association between the interventions and each individual outcome was investigated using separate univariate meta-analyses, using the same method as described for the composite outcomes.

2.2 Alternative approach: multivariate meta-analysis of individual outcomes

Rather than performing separate univariate meta-analyses or a univariate meta-analysis of a composite outcome, in this chapter a multivariate meta-analysis is considered. As described in the introduction chapter, multivariate meta-analysis jointly synthesises summary estimates for each outcome of interest in one analysis (Raudenbush et al. 1988, Becker 2000, van Houwelingen et al. 2002, Jackson et al. 2011). It utilises the within-study correlations and between-study correlations to provide more precise estimates and allows for the inclusion of studies that do not provide information for all the outcomes (Riley et al. 2007*a*,*b*), as explained in Chapter 1.

There were four maternal outcomes: pre-eclampsia or pregnancy induced hypertension (PE or PIH), gestational diabetes, preterm birth and caesarean section. Not all outcomes were reported in each trial and therefore, in this setting a multivariate meta-analysis will be beneficial; utilising the correlations between the outcomes, strength might be borrowed across outcomes. Additionally, there were four foetal outcomes: intrauterine death, small for gestational age, large for gestational age and admission to Neonatal Intensive Care Unit (NICU). Similarly, for the four foetal outcomes a multivariate meta-analysis will be beneficial. Hence, in the remainder of this chapter, a multivariate meta-analysis approach is outlined and applied to the HTA report.

2.3 Methods for the re-analysis of the HTA report

2.3.1 Obtaining relevant data for meta-analysis

The data for the application of a multivariate meta-analysis was requested from the iWiP collaborative group. It was requested that the within-study correlations between each pair of treatment effect estimates for the maternal and foetal outcomes be estimated using bootstrapping (Section 1.4.4). The iWiP collaborative group kindly

provided the relevant estimates for each outcome, that is, aggregate data in the form of treatment effect estimates, standard errors and the within-study correlations for each pair of maternal and foetal outcomes estimated using bootstrapping. The iWiP collaborative group indicated in the data which outcomes were calculated using the Sweeting correction and for these outcomes within-study correlations were unavailable. Thus the re-analysis in this chapter will be two-stage meta-analyses, with multivariate meta-analysis applied in the second stage.

2.3.2 Analysis 1: replicating HTA report results

For each of the maternal and foetal outcomes, a univariate meta-analysis model (Equation 1.15) was applied and estimated using REML. The confidence intervals were calculated using a Hartung-Knapp correction (Knapp & Hartung 2003). The results from the univariate meta-analyses for the data provided by the iWiP collaborators were compared to the univariate results from the HTA report for each outcome.

2.3.3 Analysis 2: application of multivariate meta-analysis model

In addition to the univariate meta-analyses for each outcome, the four maternal outcomes were analysed jointly and the four foetal outcomes were analysed jointly using multivariate meta-analysis models (Equation 1.30). For the outcomes that were calculated using the Sweeting correction, the within-study correlations for these studies were imputed using the mean within-study correlation for the two corresponding outcomes (Riley 2009). Following, the covariance for each pair of outcomes was calculated from the within-study correlation and the standard errors for the outcomes using $\sigma_{i(j_1,j_2)} = \sigma_{ij_1}\sigma_{ij_2}\rho_{W_{i(j_1,j_2)}}$ where $\sigma_{i(j_1,j_2)}$ is the covariance and $\rho_{W_{i(j_1,j_2)}}$ is the within-study correlation between outcome j_1 and j_2 in study i. The standard errors in study i for outcomes j_1 and j_2 are σ_{ij_1} and σ_{ij_2}, respectively.

The multivariate meta-analyses were estimated using REML. Their respective confi-

dence intervals for the summary treatment effects were calculated using a multivariate refined method similar to the Hartung-Knapp method in univariate meta-analysis, described in Section 1.4.6 (Jackson et al. 2013, Knapp & Hartung 2003). The calculated scaling factor, H^2 was restrained to one, to ensure the confidence intervals were not narrower. The Borrowing of Strength (BoS) was also calculated for each outcome.

2.3.4 Analysis 3: sensitivity analysis

Sensitivity analyses were conducted following the analysis to explore the effect the imputed within-study correlations may have had upon the meta-analyses' results. Extreme values for the within-study correlations were imputed for the within-study correlations which were originally missing. The extreme values were selected based upon the maximum/minimum correlation that did not provide estimation problems and a correlation of zero (Riley 2009). The within-study correlations for the sensitivity analysis for the foetal multivariate meta-analysis were zero, 0.55 and -0.39 and for the maternal multivariate meta-analysis were zero, 0.69 and -0.69.

2.4 Results of re-analysis

2.4.1 Summary of data

The number of studies that provided each outcome varied for each outcome (Table 2.1). The foetal outcome, intrauterine death was the outcome with the lowest number of studies. It also was the outcome with the largest number of treatment effects that were calculated using the Sweeting correction (Sweeting et al. 2004), since intrauterine death is a rare outcome.

Table 2.1: Number of studies reporting each outcome for the foetal and maternal outcomes

		Number of Studies	
	Outcome	Aggregate data provided	HTA report
Maternal	PE[a] or PIH[b]	22	22
	Gestational diabetes	27	27
	Preterm birth	33	32
	Caesarean section	32	32
Foetal	Intrauterine death	12	NA[c]
	Small for gestational age	34	33
	Large for gestational age	35	34
	Admission to NICU[d]	16	16

[a] Pre-eclampsia

[b] Pregnancy induced hypertension

[c] Insufficient data

[d] Neonatal Intensive Care Unit

The HTA report did not state how many studies provided intrauterine death but stated that there was insufficient data to perform a univariate meta-analysis. Additionally, there were slight differences in the number of studies that reported the outcomes preterm birth, small for gestational age and large for gestational age between the HTA report and the data. For example, in the HTA report, the number of studies with the outcome preterm birth was 32 studies, however there were 33 studies with the outcome preterm birth in the data set provided by the iWiP collaborators.

The mean within-study correlations were small in magnitude (<0.1) for each pair of outcomes (Table 2.2). The maximum magnitude of the within-study correlations between the pairs of outcomes was 0.25. Therefore, there was only a small association between the outcomes in each study.

Table 2.2: Mean and range of the within-study correlations between the maternal outcomes and the foetal outcomes

Outcome pairs	Number of studies	Within-study correlation	
		Mean	Range
Maternal Outcomes			
PE[a] or PIH[b] and gestational diabetes	19	0.023	-0.228 to 0.196
PE[a] or PIH[b] and preterm birth	21	0.053	-0.100 to 0.213
PE[a] or PIH[b] and caesarean section	21	0.024	-0.085 to 0.172
Gestational diabetes and preterm birth	24	0.021	-0.118 to 0.192
Gestational diabetes and caesarean section	25	0.041	-0.121 to 0.250
Preterm birth and caesarean section	30	0.040	-0.070 to 0.172
Foetal Outcomes			
intrauterine death and SGA[c]	4	0.034	0.003 to 0.087
intrauterine death and LGA[d]	4	0.043	-0.033 to 0.161
intrauterine death and admission to NICU[e]	4	-0.007	-0.024 to 0.011
SGA and LGA	31	-0.081	-0.228 to 0.024
SGA and admission to NICU	15	0.072	-0.028 to 0.210
LGA and admission to NICU	15	0.027	-0.068 to 0.139

[a] Pre-eclampsia

[b] Pregnancy induced hypertension

[c] Small for gestational age

[d] Large for gestational age

[e] Neonatal Intensive Care Unit

2.4.2 Comparison between univariate meta-analysis from HTA report and the data

The results between the HTA report and the analysis with the data provided by the iWiP collaboration were similar; for the majority of outcomes for both the maternal and foetal outcomes the results were equal (Table 2.3). Any differences in the results

from the HTA report and the analysis with the data in this chapter were slight and were likely to have occurred due to slight estimation variations and rounding.

Table 2.3: Univariate meta-analysis results from the HTA report and the data provided by the iWiP collaboration

Outcome	HTA[a] report		Using aggregate data provided		p-value
	Odds ratio	95% Confidence Interval	Odds ratio	95% Confidence Interval	
Maternal Outcomes					
PE[b] or PIH[c]	0.95	0.78 to 1.16	0.95	0.78 to 1.16	0.600
Gestational diabetes	0.89	0.72 to 1.10	0.89	0.72 to 1.09	0.243
Preterm birth	0.94	0.78 to 1.13	0.94	0.78 to 1.13	0.498
Caesarean section	0.91	0.83 to 0.99	0.91	0.83 to 0.99	0.034
Foetal Outcomes					
Intrauterine death	Insufficient data		0.92	0.36 to 2.33	0.851
Small for gestational age	1.06	0.94 to 1.20	1.06	0.94 to 1.20	0.311
Large for gestational age	0.90	0.76 to 1.07	0.90	0.76 to 1.07	0.209
Admission to NICU[d]	1.01	0.84 to 1.23	1.02	0.84 to 1.23	0.870

[a] Health Technology Assessment

[b] Pre-eclampsia

[c] Pregnancy induced hypertension

[d] Neonatal Intensive Care Unit

2.4.3 Multivariate meta-analysis results

The multivariate meta-analysis results from the IPD are provided in Table 2.4. Conclusions are very similar to those from the univariate meta-analyses. From the multivariate results, it is concluded that the odds of a caesarean section were reduced by 9% (95% C.I.:17% to 1%) for those on a diet and physical activity based interventions compared to those with no intervention. This was the only outcome, from the maternal and foetal outcomes, with a statistically significant result at the 5% significance level.

Table 2.4: Multivariate and univariate meta-analysis results for the maternal and foetal outcomes

Outcome	Multivariate meta-analysis				Univariate meta-analysis[a]	
	Odds Ratio	95% Confidence Interval	Between-study Variance	BoS (%)	Odds Ratio	95% Confidence Interval
Maternal Outcomes						
PE[b] or PIH[c]	0.85	0.67 to 1.08	0.270	17.1	0.95	0.78 to 1.16
Gestational diabetes	0.84	0.68 to 1.04	0.303	10.0	0.89	0.72 to 1.09
Preterm birth	0.94	0.76 to 1.15	0.176	2.6	0.94	0.78 to 1.13
Caesarean section	0.89	0.80 to 0.99	0.092	8.0	0.91	0.83 to 0.99
Foetal Outcomes						
Intrauterine death	0.98	0.38 to 2.50	0.966	5.5	0.92	0.36 to 2.33
Small for gestational age	1.06	0.93 to 1.20	0.093	3.5	1.06	0.94 to 1.20
Large for gestational age	0.89	0.77 to 1.03	0.188	2.3	0.90	0.76 to 1.07
Admission to NICU[d]	0.99	0.81 to 1.20	0.071	3.6	1.02	0.84 to 1.23

[a] Univariate results as in Table 2.3

[b] Pre-eclampsia

[c] Pregnancy induced hypertension

[d] Neonatal Intensive Care Unit

Borrowing of Strength (BoS)

The BoS statistic was small for all the Foetal outcomes (range: 2.3% to 5.5%) (Table 2.4). The largest BoS statistic was for the maternal outcome, pre-eclampsia or pregnancy induced hypertension, at 17.1%. The BoS statistics for gestational diabetes and caesarean section were not as large, 10% and 8%, respectively. For the preterm birth outcome, the BoS statistic was small at 2.6%. The low BoS values reflect that, in general, the within-study correlations were small to moderate at best.

Between-study correlations and variances

For the maternal outcomes the between-study correlations were all estimated to be positive, ranging from 0.173 to 0.971 (Table 2.5). The majority of the between-study correlations for the maternal outcomes were large. The between-study correlations for the foetal outcomes were all estimated to be +1 or -1, the extreme boundaries of the field for correlations, suggesting potential estimation difficulties, most likely because the magnitude of between-study variability was small for all foetal outcomes.

Table 2.5: Estimates of the between-study variance-covariance matrix, showing the between-study correlations in brackets.

Maternal Outcomes				
	PE[a] or PIH[b]	Gestational Diabetes	Preterm birth	Caesarean section
PE[a] or PIH[b]	0.073			
Gestational diabetes	0.079 (0.971)	0.092		
Preterm birth	0.019 (0.403)	0.009 (0.173)	0.031	
Caesarean section	0.022 (0.888)	0.021 (0.753)	0.012 (0.779)	0.009
Foetal Outcomes				
	Intrauterine death	SGA[c]	LGA[d]	Admission to NICU[e]
Intrauterine death	0.053			
SGA [c]	-0.022 (-1)	0.009		
LGA[d]	-0.043 (-1)	0.017 (1)	0.035	
Admission to NICU[e]	-0.017 (-1)	0.007 (1)	0.013 (1)	0.005

[a] Pre-eclampsia

[b] Pregnancy induced hypertension

[c] Small for gestational age

[d] Large for gestational age

[e] Neonatal Intensive Care Unit

2.4.4 Comparison between univariate and multivariate meta-analysis

The results were generally similar between the univariate and multivariate meta-analyses and the statistical significance for each outcome was unchanged (based upon the confidence intervals) (Table 2.4). However, there were some potentially clinically important differences in the actual magnitude of summary effect sizes, most notably for the maternal outcomes pre-eclampsia or pregnancy induced hypertension and gestational

diabetes. For example, the summary odds ratio for pre-eclampsia or pregnancy induced hypertension from the univariate meta-analysis was 0.95 (95% C.I.: 0.78 to 1.16) and from the multivariate meta-analysis, the summary odds ratio was smaller at 0.85 with a 95% confidence interval of 0.67 to 1.08. Therefore, the best estimate of the summary effect for intervention of diet and exercise is that it reduces the odds of pre-eclampsia or pregnancy induced hypertension by 15% in the multivariate meta-analysis, but only 5% in the univariate meta-analysis. Additionally, for gestational diabetes the summary odds ratio from the univariate was 0.89 (95% C.I.: 0.72 to 1.09) and from the multivariate the summary odds ratio was 0.84 with a 95% confidence interval of 0.68 to 1.04. As a result of the diet and exercise intervention the odds of gestational diabetes is reduced by 16%. Therefore, there was stronger evidence after the multivariate meta-analysis that the diet and exercise intervention is beneficial for these outcomes, although 95% confidence intervals are still wide.

Sensitivity analysis results

For the foetal sensitivity analysis, results were very similar when assuming a within-study correlation of zero in studies where it was not provided. However, the BoS increased when larger magnitudes of correlations were assumed (Table B.1 in Appendix B.1). The conclusions were unchanged for different assumed values of the within-study correlations. Similarly, for the maternal sensitivity, the results were very similar across the different within-study correlation values (Table B.2 in Appendix B.1). Greater differences between the multivariate meta-analysis and the sensitivity analyses were observed for large negative within-study correlations and the magnitude of BoS were greater.

2.5 Discussion

This chapter performed secondary data analysis of a HTA report that investigated the effect of diet and exercise interventions during pregnancy to reduce weight gain on both

maternal and foetal outcomes (Rogozińska et al. 2017). The original analysis in the HTA report developed composite outcomes for the maternal and foetal outcomes from the IPD (Rogozińska et al. 2016). The composite outcomes were analysed using univariate meta-analyses, one for the maternal composite outcome and one for the foetal composite outcome. Additionally, each foetal and maternal outcome was analysed in separate univariate meta-analyses in the HTA report. However, it was hypothesised that the treatment effect estimates may be correlated for each pair of outcomes. This chapter investigated whether, through the inclusion of the correlations, applying multivariate meta-analysis methods would be beneficial in this setting. The key findings from the analysis are summarised in Figure 2.1.

Figure 2.1: Key findings from the secondary analysis of the HTA report for the effect of diet and exercise interventions on maternal and foetal outcomes

Key Findings:

- The results from the multivariate meta-analysis were similar to those from the univariate meta-analysis; there was no change in statistical significance. However, there were some differences between the multivariate and univariate meta-analyses in the magnitude of the summary effect size, that might have clinical significance, although confidence intervals are still wide.
- The borrowing of strength statistics were larger for outcomes where there were greater differences between the univariate and multivariate results.
- The mean within-study correlations were small in magnitude and thus this is likely to have contributed to a small BoS statistic in general (Jackson et al. 2017).
- The estimated between-study correlations were large for the majority of the maternal outcome pairs. However, the between-study correlations for the foetal outcomes had estimation problems and were estimated at the extreme boundaries for correlations.

There was no change in the statistical significance of the results between the univariate and the multivariate meta-analyses for the foetal and maternal outcomes. The conclusions, based on statistical significance, for the effect of diet and exercise interventions during pregnancy were unaltered between the univariate and multivariate meta-analyses. For example, for the outcome small for gestational age the odds ratios were 1.06 from both univariate and multivariate results. The 95% confidence intervals were 0.94 to 1.20 and 0.93 to 1.20 from the univariate and multivariate, respectively and thus, there was no change in the not statistically significant result. However, there were potentially important clinical differences between the univariate and multivariate summary results for two maternal outcomes, pre-eclampsia or pregnancy induced hypertension and gestational diabetes. That is, the magnitude of effect was larger in the multivariate meta-analysis than the univariate meta-analysis. For example, for gestational diabetes the univariate meta-analysis results were 0.89 (95% C.I.: 0.72 to 1.09) and the multivariate results were 0.84 (95% C.I.: 0.76 to 1.15). Although further research is needed, due to the wide confidence intervals.

The borrowing of strength statistic was small for most of the maternal and foetal outcomes. However, for the maternal outcomes, pre-eclampsia or pregnancy induced hypertension and gestational diabetes, there was a slight quantity of BoS, with values of 17.1% and 10%, respectively. It was observed that the larger BoS values were for outcomes that were observed to have greater differences between univariate and multivariate meta-analysis. This motivates further research of BoS in the next chapters.

In this study, the mean within-study correlations between the treatment effect estimates were small for all pairs of maternal or foetal outcomes. This means there is only a small association between any two treatment effect estimates in the same study for the maternal or foetal outcomes. This may be one reason the BoS statistics were small in general, since it has previously been concluded that when the within-study correlation is zero there is no borrowing of strength (Riley et al. 2007*a*, Jackson et al.

2017). Therefore, it might be that there was little difference between the univariate and multivariate results, since there was only a small association between the pairs of treatment effect estimates for the maternal and foetal outcomes in each study. However, the majority of the between-study correlations were estimated to be quite large for the maternal outcomes, which means that the underlying true values for the pair of two outcomes across the trials included in the meta-analysis were highly associated. It was expected that by including the between-study correlations there would be a gain in information which may have resulted in differences between the univariate and multivariate results (Riley 2004, Riley et al. 2007*a*). However, in general, differences between the univariate and multivariate results were not observed. Although there were large between-study correlations, large differences in results may not have occurred since the between-study variances were small and ranged from 0.005 to 0.092 for all maternal and foetal outcomes.

Riley et al. (2007*b*) investigated the estimation of the between-study correlations in a normal random-effects meta-analysis using a maximum likelihood estimation; they found that often the between-study correlations are truncated to the boundary of the field for correlations. This is especially likely to occur if the within-study variances are large compared to the between-study variances. Certain conditions need to be met in a normal random-effects meta-analysis; these include the between-study variance covariance matrix satisfying the conditions to be a non-negative definite matrix meaning that the $\tau_j^2 \geq 0$ and $-1 \leq \rho_B \leq 1$. To satisfy the conditions, the between-study correlations are truncated at the boundary when estimated. A consequence of the between-study correlations estimated at the boundaries is an upward bias in the between-study variance estimates. Since the between-study variances were so small in this multivariate meta-analysis, this should not be a concern.

2.5.1 Does the multivariate meta-analysis approach have any additional benefits compared to the univariate meta-analysis in this setting?

The multivariate meta-analysis approach required more information with regards to the quantity of data needed and further assumptions were made compared to the univariate meta-analysis. In this particular example, there was no clear evidence that the multivariate meta-analysis was beneficial over the univariate meta-analysis especially considering the extra requirements that were needed. However, there were potentially important clinical differences between the univariate and multivariate results for two maternal outcomes, pre-eclampsia or pregnancy induced hypertension and gestational diabetes. For these outcomes, the multivariate meta-analysis may have proven to be beneficial due to the clinical implications the results might affect, yet confidence intervals remained large. Thus, the benefits of using the multivariate meta-analysis do not outweigh the additional complications of a multivariate approach over a univariate in this study. The differences in meta-analysis results between a univariate and a multivariate approach have been studied in many research papers and they concluded that usually there is little difference between the univariate meta-analysis and the multivariate meta-analysis results (Sohn 2000, Simel & Bossuyt 2009, Trikalinos et al. 2013, 2014).

2.5.2 Further work

Although using multivariate rather than univariate meta-analysis in this particular example did not change clinical or statistical conclusions, this may not hold in other applications. Further, there was a sign that multivariate meta-analysis can change summary results, as seen in this chapter for outcomes pre-eclampsia and pregnancy induced hypertension. Additionally, for these outcomes the BoS statistic was observed to be larger. A further investigation is needed into the relationship between the BoS statistic and the differences between the univariate and multivariate results. Could

the BoS statistic be used to identify situations and outcomes where the multivariate approach is beneficial over the univariate approach?

Alternatively, could the characteristics of the meta-analysis be used to identify situations where the multivariate approach is beneficial over the univariate approach? Were there particular characteristics, such as missing data, that meant that the multivariate approach was more beneficial for some outcomes (for example, pre-eclampsia and pregnancy induced hypertension) but not others (for example, small for gestational age and preterm birth)? For two maternal outcomes (preterm birth and caesarean section) and two foetal outcomes (small for gestational age and large for gestational age), the majority of the trials reported the outcome of interest. However, for the remaining outcomes there was missing data for the outcome of interest; for example, for intrauterine death the outcome was reported in 12 studies, whereas the outcome, large for gestational age, was reported in 35 studies. The characteristics of the data (including percentage of missing data for the outcome of interest, within-study and between-study correlations) which a meta-analysis is to be applied will differ between outcomes and meta-analysis studies. It is currently unclear which characteristics are more likely to result in differences between univariate and multivariate results. These characteristics including the correlations, might influence whether there are benefits between the univariate and the multivariate meta-analysis. This idea forms the motivation for the next few chapters in this thesis.

2.5.3 Conclusion

The results from the original HTA report are barely affected by the use of multivariate meta-analysis over the univariate meta-analysis approach. Hence, the benefits of multivariate meta-analysis in this setting were limited and the BoS statistic was often small. In the next chapter, the relationship between the magnitude of BoS and the differences between univariate and multivariate meta-analysis is explored. Meta-analysis level characteristics, including percentage of missing data for the outcome of interest,

within-study and between-study correlations, are investigated for their relationship with the magnitude of BoS.

Chapter 3

An evaluation of the distribution and magnitude of BoS in Cochrane reviews

3.1 Background

The advantages of multivariate meta-analysis and univariate meta-analysis have been examined and detailed in the literature. Studies have been conducted to investigate the differences between the univariate and multivariate results. Trikalinos et al. (2013, 2014) examined 45 Cochrane reviews that contained univariate meta-analyses for two or three binary outcomes that alternatively could be analysed using a multivariate meta-analysis. They compared univariate and multivariate meta-analysis results and generally found that the summary effect estimates and confidence intervals were similar. They concluded that if the "focus is on the summary effects and the confidence intervals then the choice between the univariate and multivariate meta-analysis has limited practical importance" (Trikalinos et al. 2014, pg 1456). Despite this conclusion, there are isolated examples within the review with potentially important differences between univariate and multivariate meta-analysis. For example, for outcome one in study 35 there is a noticeable change in the summary estimates between the univariate and the multivariate meta-analyses, (Figure 3.1). From the univariate meta-analysis, there was (visually) not a statistically significant difference in the odds of the outcome between treatment groups, whereas (visually) there was a statistically significant difference for the multivariate meta-analysis. Another example is outcome two in study 38 where

the multivariate meta-analysis made a substantial improvement to the precision of the summary effect estimate (Figure 3.1), seen from the narrower width of the credible interval in the multivariate meta-analysis. The conclusions from the meta-analyses also differ as the univariate meta-analysis results are not statistically significant, whereas the multivariate meta-analysis results are statistically significant.

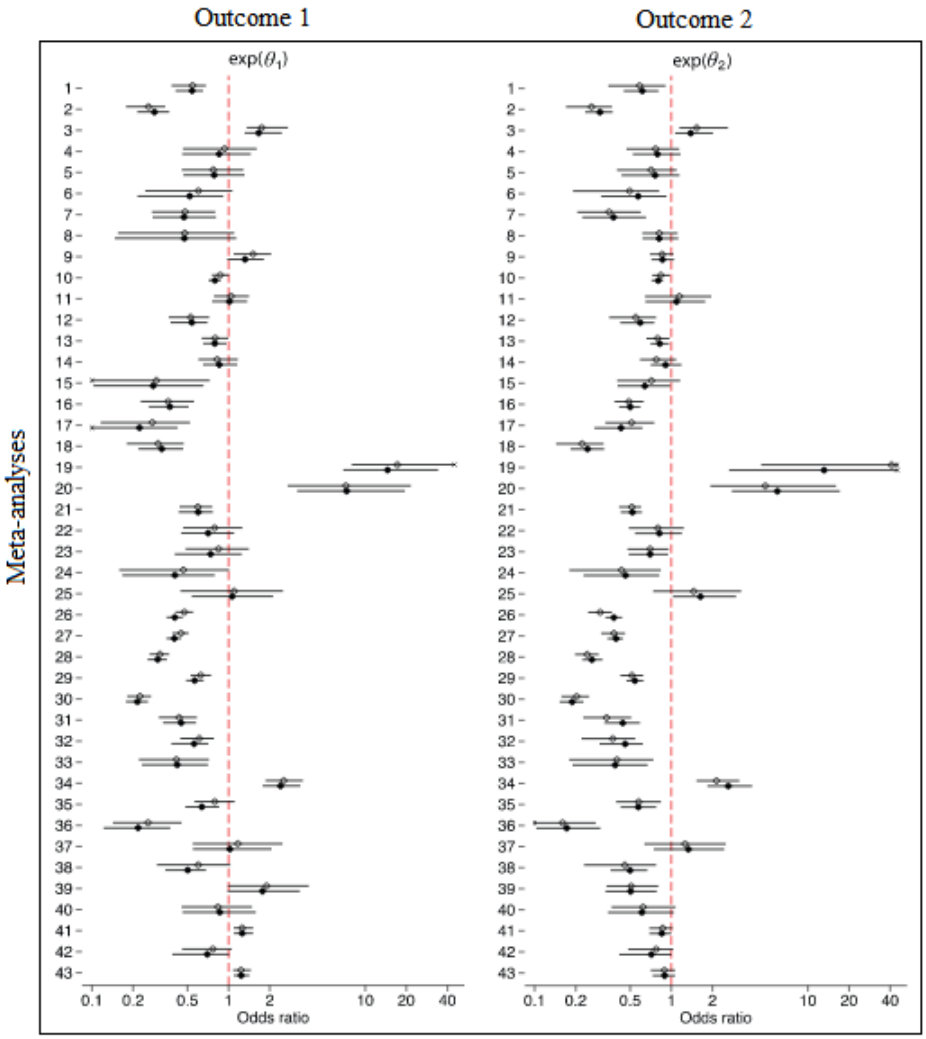


Figure 3.1: Comparison of univariate and bivariate meta-analyses of odds ratios (on log scale) for treatment versus comparator using the binomial or the multinomial distribution to model within-study variance (topics 1 through 43). Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1 : meta-analysis posterior median for the first outcome; θ_2 : meta-analysis posterior median for the second outcome. For topics 1 through 38, those experiencing the first outcome are a subset of those experiencing the second outcome. For topics 39 through 43, the two outcomes are mutually exclusive. Small 'x' markers denote truncated credible intervals. Figure used with permission from: Trikalinos, T. A., Hoaglin, D. C. & Schmid, C. H. (2014), 'An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes', Statistics in Medicine 33(9), 1441-1459.

Therefore, although generally Trikalinos et al. (2013, 2014) identifiy that multivariate

meta-analysis has no particular advantage, there are a few examples where the approach is important for the results and conclusions. This agrees with other examples (outside of the Trikalinos et al. (2013, 2014) review) where multivariate meta-analysis has shown to be beneficial (Riley et al. 2007*a*,*b*, Hamza et al. 2009, Jones et al. 2009, Jackson et al. 2011, Kirkham et al. 2012, Riley et al. 2015, Frosi et al. 2015). An example from The Fibrinogen Studies Collaboration (2009) that contained studies with missing fully adjusted results was shown in Chapter 1 that studies with missing outcomes can be included in the multivariate meta-analysis to gain information from partially adjusted results and thus gain precision in the summary result. In the same way, when selective outcome reporting is evident, multivariate meta-analysis can be beneficial to reduce bias in the summary result from the univariate meta-analysis. This was demonstrated for the prognostic ability of marker p53 for patients with squamous cell carcinoma in Chapter 1 (Jackson et al. 2011, Tandon et al. 2010). The inclusion of correlations in multivariate meta-analysis of longitudinal outcomes was important in an example from Jones et al. (2009) as it considerably changed the summary estimates and their precision. However, as Trikalinos et al. (2013, 2014) suggest, the concern is that such interesting examples are generally rare and thus recommending the use of the multivariate method seems premature. Indeed, multivariate meta-analysis is also more complicated than univariate meta-analysis due to the need to derive within-study correlations and estimate between-study correlations. Therefore, it would be useful to determine (preferably in advance) under which circumstances the utilisation of multivariate meta-analysis would be beneficial and hence worth the extra requirements and resources.

3.1.1 Aim of the chapter

Trikalinos et al. (2014) compared the summary effect estimates and their confidence intervals between the univariate and multivariate meta-analyses. The aim of this chapter is to empirically examine the magnitude of Jackson et al.(2017)'s BoS statistic in the examples from the Trikalinos et al. (2013, 2014) review, to ascertain whether the

general recommendation by Trikalinos et al. (2013, 2014) can be attributed to low BoS values in their cohort of examples, and whether those few examples where multivariate meta-analysis changes univariate meta-analysis conclusions can be attributed to a large BoS value. A secondary aim of this chapter is to explore which meta-analysis factors are associated with the magnitude of BoS values and therefore might highlight studies where it would be beneficial to utilise multivariate meta-analysis in the future.

This chapter follows with the methods used to compare the univariate and multivariate results, and the method for the calculation of the BoS statistic, in the next section. Then the potential factors associated with the magnitude of the borrowing of strength statistic are detailed with the method for determining the association.

3.2 Methods

3.2.1 Included meta-analyses for empirical evaluation

There were 45 reviews included in the empirical review by Trikalinos et al. (2013, 2014). Each review contained at least seven studies that reported both outcomes or at least half the studies with both outcomes if the total number of studies was greater than 14. Each of the studies satisfying the previous requirement (i.e. at least seven studies with both outcomes) must have at least 10 patients and at least two events in each treatment arm. There were two reviews examined by Trikalinos et al. (2013, 2014) that contained three outcomes and are not considered further here, since the focus in this chapter will be bivariate models. The remaining 43 reviews were included, and these contained two outcomes with cross classification tables for the treatment effect. The relationships between the pairs of outcomes were either mutually exclusive or an is-subset-of relationship. These two binary outcome relationship types were chosen, since formulae exist for the calculation of the within-study covariances with these binary outcome structures for each pair of treatment effect estimates in each study. The relationship between two binary outcomes named is-subset-of refer to when one

outcome is contained within the other. For example the number of patients that have survived with a particular condition at, say, 6 months and a year. The second binary outcome relationship type, the mutually exclusive relationship, is when the outcomes are independent of each other and therefore occur separately. An example is death from breast cancer and death from other causes, excluding breast cancer.

3.2.2 Derivation of treatment effects, within-study variances and within-study covariances

The data was kindly provided by the original authors (Trikalinos et al. 2013, 2014). The data for each outcome were contained in 2x2 contingency tables, containing the frequencies, for each trial. The treatment effect estimates, that is log odds ratio estimates ($\hat{\theta}_j$), for each outcome in each study were calculated from the contingency tables as follows:

$$\begin{aligned}\hat{\theta}_j =& \text{logit}(\hat{p}_t) - \text{logit}(\hat{p}_c) \\ =& \text{logit}\left(\frac{xt_j}{Nt}\right) - \text{logit}\left(\frac{xc_j}{Nc}\right)\end{aligned} \tag{3.1}$$

where xt_j is the number of patients with the event in outcome j, in the treatment group, xc_j is the number of patients with the event in the control group in outcome j, Nt is the number of patients in the treatment group in the study and Nc is the number of patients in the control group in the study. The variance of the log odds ratio estimate for each outcome was calculated (Bland & Altman 2000):

$$\sigma_i^2 = \frac{1}{xt_j} + \frac{1}{(Nt - xt_j)} + \frac{1}{xc_j} + \frac{1}{(Nc - xc_j)} \tag{3.2}$$

A fixed 0.5 continuity correction was required if any denominator in the equation for the variance was equal to zero, (Higgins & Green 2008, Sweeting et al. 2004); that is, if a study had a zero cell in the 2x2 contingency table then 0.5 was added to all cells for that study. This is a similar approach to that used by the authors in the Trikalinos

review adopted (Trikalinos et al. 2014).

The bivariate meta-analysis requires the within-study covariances between the outcomes to be calculated. The calculation for the within-study covariance in a study is dependent upon the relationship between the outcomes. For the is-subset-of relationship the equation was (Wei & Higgins 2013*b*):

$$\sigma_{1,2} = \frac{1}{Nt\left(\frac{xt_2}{Nt}\right)\left(1-\frac{xt_1}{Nt}\right)} - \frac{1}{Nc\left(\frac{xc_2}{Nc}\right)\left(1-\frac{xc_1}{Nc}\right)} \tag{3.3}$$

For the mutually exclusive relationship, the equation was (Trikalinos & Olkin 2008, Bagos 2012):

$$\sigma_{1,2} = -\frac{1}{Nt\left(1-\frac{xt_1}{Nt}\right)\left(1-\frac{xt_2}{Nt}\right)} - \frac{1}{Nc\left(1-\frac{xc_1}{Nc}\right)\left(1-\frac{xc_2}{Nc}\right)} \tag{3.4}$$

Then the within-study correlation can be calculated from the within-study variances (equation 3.2) and the within-study covariance (either equation 3.4 or 3.3) using the equation:

$$\rho_{ws} = \frac{\sigma_{1,2}}{\sqrt{\sigma_1^2}\sqrt{\sigma_2^2}} \tag{3.5}$$

In some studies, the within-study correlation was estimated as, 1 or -1, which can cause issues of singular variance matrices during the multivariate model estimation. To avoid this issue, Trikalinos et al. (2014) used a ridge-regression approach to shrink the magnitude of the correlations downwards, to avoid any issues of singular variance matrices in the model estimation. In this study, the within-study correlations of ± 1 are simply replaced with ± 0.99, in order to avoid the issue of singular variance matrices (Riley et al. 2014).

3.2.3 Derivation of univariate meta-analysis results

Univariate meta-analyses were applied to each meta-analysis study in the Trikalinos et al. (2013, 2014) dataset, using both the fixed-effect and the random-effects ap-

proaches. The fixed-effect univariate meta-analysis was modelled by applying the following to each outcome, j, separately:

$$\hat{\theta}_{ij} \sim N(\theta_j, \sigma_{ij}^2) \tag{3.6}$$

where $\hat{\theta}_{ij}$ is the observed log odds ratio estimate for outcome j in study i, θ_j is the true log odds ratio for outcome j across all studies in the meta-analysis and σ_{ij}^2 is the within-study variance for outcome j in study i. The random-effects meta-analysis was modelled using:

$$\hat{\theta}_{ij} \sim N(\theta_{ij}, \sigma_{ij}^2) \tag{3.7}$$

$$\theta_{ij} \sim N(\theta_j, \tau_j^2) \tag{3.8}$$

where $\hat{\theta}_{ij}$ is the observed log odds ratio estimate for outcome j in study i, θ_{ij} is the true log odds ratio for outcome j in study i, σ_{ij}^2 is the within-study variance for outcome j in study i, θ_j is the summary (average) log odds ratio for outcome j across all studies and τ_j^2 is the between-study variance for outcome j.

3.2.4 Derivation of bivariate meta-analysis results

Similarly, fixed-effect and a random-effects bivariate meta-analysis were applied to each meta-analysis in the Trikalinos et al. (2013, 2014) dataset. The bivariate fixed-effect meta-analysis is given as:

$$\begin{pmatrix} \hat{\theta}_{i1} \\ \hat{\theta}_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \delta_i \right) \quad , \quad \delta_i = \begin{pmatrix} \sigma_{i1}^2 & \sigma_{i(1,2)} \\ \sigma_{i(1,2)} & \sigma_{i2}^2 \end{pmatrix} \tag{3.9}$$

where $\hat{\theta}_{i1}$ and $\hat{\theta}_{i2}$ are the observed log odds ratio estimates for outcomes 1 and 2, respectively, θ_1 and θ_2 are the true log odds ratios for outcome 1 and 2, respectively, and δ_i is the within-study covariance matrix for study i. The elements of the lead diagonal of the within-study covariance matrix, δ_i, are the within-study variances, σ_{i1}^2 and σ_{i2}^2, for outcomes 1 and 2, respectively. The off diagonal element is the within-study

covariance which is calculated from the within-study variances and the within-study correlation, $\sigma_{i(1,2)} = \rho_{W_i}\sigma_{i1}\sigma_{i2}$.

The bivariate random-effects meta-analysis that was used is given as:

$$\begin{pmatrix} \hat{\theta}_{i1} \\ \hat{\theta}_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \delta_i\right) \quad , \quad \delta_i = \begin{pmatrix} \sigma_{i1}^2 & \sigma_{i(1,2)} \\ \sigma_{i(1,2)} & \sigma_{i2}^2 \end{pmatrix} \tag{3.10}$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \Omega\right) \quad , \quad \Omega = \begin{pmatrix} \tau_1^2 & \tau_{(1,2)} \\ \tau_{(1,2)} & \tau_2^2 \end{pmatrix} \tag{3.11}$$

where Y_{i1}, Y_{i2} and δ_i are defined as above, θ_{i1} and θ_{i2} are the true treatment effects for each outcome for each study i, θ_1 and θ_2 are the summary (average) treatment effects for each outcome and Ω is the between-study covariance matrix. The elements of Ω are the between-study variances and covariance. The between-study variances are the elements of the lead diagonal, τ_1^2 and τ_2^2, and $\tau_{(1,2)}$ is the between-study covariance. The between-study correlation is calculated from the between-study correlation; $\tau_{(1,2)} = \rho_\mu\tau_1\tau_2$.

For both univariate and multivariate models, the models were estimated using the inverse of fisher's information matrix, which estimates the standard error (s.e.) of the summary estimate for each outcome (Jackson & Riley 2014). After estimation the 95% C.I.s for the summary estimate were derived using the following formula:

$$\hat{\theta}_j \pm 1.96(\text{s.e.}) \tag{3.12}$$

For simplicity, in the univariate and multivariate random-effects no post-estimation inflation of the variance of the summary effect estimates was used, which is sometimes proposed to account for uncertainty in the estimated between-study variances (or estimated between-study covariance matrix) (Cornell et al. 2014, Knapp & Hartung 2003).

3.2.5 Derivation of BoS

The aim of this chapter is to evaluate the magnitude of the BoS statistic for each outcome in each of the 43 Cochrane reviews. The BoS quantifies the extra precision gained from using multivariate meta-analysis of both outcomes jointly (and thus utilising their correlations) rather than a separate univariate meta-analyses (for each outcome independently) (Jackson et al. 2017). As mentioned in Section 1.4.8, the BoS statistic for a particular outcome is given mathematically as:

$$\text{BoS} = 100\% \times \left[1 - \frac{\text{var}(\hat{\theta}_{mv,j})}{\text{var}(\hat{\theta}_{uv,j})}\right] \tag{3.13}$$

where $\hat{\theta}_{mv,j}$ is the summary effect estimate from the multivariate meta-analysis for outcome j and $\hat{\theta}_{mv,j}$ is the summary effect estimate from the univariate meta-analysis for outcome j.

The BoS statistic was calculated for each outcome in each meta-analysis in the Trikalinos et al. (2013, 2014) dataset, for each of the fixed-effect and random-effects approaches. The fixed-effect model was fitted using maximum likelihood (ML) estimation. For the random-effects, two models were fitted; the first used restricted maximum likelihood (REML) and the second used a Method of Moments (MM) approach (Jackson et al. 2010), a non-iterative method which is extended from the Dersimonian and Laird approach for univariate meta-analyses. This produced a total of three BoS statistics for each outcome in each meta-analysis (one from ML fixed-effect approach, one from REML random-effects and one from MM random-effects) in the Trikalinos et al. (2013, 2014) dataset. The distribution of BoS statistics was summarised using descriptive statistics and also graphically via histograms.

3.2.6 Factors associated with the magnitude of BoS

The second aim of this chapter is to identify meta-analysis level factors that are associated with the magnitude of the BoS. The rationale is that the identified factors that

are associated with affecting the magnitude of BoS might highlight when a multivariate meta-analysis approach should be considered. The potential factors that were considered are meta-analysis level factors that were decided upon *a priori* by the research team; the meta-analysis level factors were considered, by the research team, potentially associated with the magnitude of BoS, or were previously suggested to influence the magnitude of BoS (The Fibrinogen Studies Collaboration 2009, Jackson et al. 2011, Riley 2009).

- the percentage of studies with missing data for the outcome of interest
- the percentage of studies with missing data across both outcomes
- the number of studies in the meta-analysis
- the number of studies with only the outcome of interest
- the number of studies with both outcomes
- the average within-study correlation for that outcome with others
- the average absolute within-study correlation for that outcome with others
- the maximum within-study correlation for that outcome with others
- the maximum absolute within-study correlation for that outcome with others

Additionally for the random-effects meta-analyses the following meta-analysis level factors were also considered:

- the between-study correlation for that outcome with others
- the absolute between-study correlation for that outcome with others
- the multivariate between-study variance for both outcomes
- the univariate between-study variance for the outcome of interest

Two approaches that generalised Higgin's I^2 in the multivariate meta-analysis setting were also considered(Higgins & Thompson 2002, Higgins et al. 2003).

- White's I^2 statistics, as derived for each outcome using the approach of White (White 2011).
- Jackson's I^2 statistics, as derived for each outcome using the approach of (Jackson et al. 2012).

An additional factor, the difference between the univariate and multivariate summary estimates, was considered to investigate whether the magnitude of the BoS statistic identified a difference between summary effect estimates. Since the difference in the summary estimates depend upon clinical context, the difference was divided by the multivariate standard error for a standardised difference.

Univariable Regression

Univariable regression models were fitted for each of the factors of interest to investigate the association between the BoS statistic and the factor of interest. For each estimation type (fixed-effect, random-effects MM and REML), there were two BoS statistics available per meta-analysis (one for each outcome). Since there were two BoS statistics from each meta-analysis, there was concern that the BoS statistics from the same study may be correlated. To address this, the univariable regression model needed to account for clustering at the meta-analysis level. Therefore, a mixed-effect multilevel model was utilised to fit univariable models with BoS as the outcome response. A random effect was applied to the intercept and was assumed to be approximately normally distributed. The following model was fitted in Stata:

$$\begin{aligned} \text{BoS}_{ij} &= (\gamma_0 + u_i) + \gamma_1(X_{ij}) + e_{ij} \\ e_{ij} &\sim N(0, \sigma_e^2) \\ u_i &\sim N(0, \sigma_u^2) \end{aligned} \tag{3.14}$$

where $\gamma_0 + u_i$ is the random intercept, where γ_0 is the fixed mean intercept and u_i is the random component for study i, which is normally distributed with mean 0 and variance σ_u^2, the heterogeneity. γ_1 is the effect on the magnitude of BoS for a one unit increase

in the covariate, X_{ij}. The term, e_{ij}, is the random error for outcome j in study i. The error terms are normally distributed with mean 0 and variance σ_e^2, the residual variance.

Sensitivity analysis were undertaken which transformed BoS onto the log scale and similar findings were shown. Thus, the results are only discussed on the original BoS scale for ease of interpretation.

3.3 Results

3.3.1 Distribution of BoS statistic

The distribution of BoS statistics from the 43 meta-analyses for the three methods, (ML fixed-effect, REML random-effects, MM random-effects) are graphically displayed in histograms in Figures 3.2 and summarised in Table 3.1. For each estimation method, the distributions of BoS were positively skewed. A large proportion of the BoS statistics were small, with the largest frequency of BoS statistics in the 0-5% category. However, there were BoS statistics that were as large as 57.15% (Table 3.1), although as the magnitude of BoS increases, the frequency of results in each category decreases (Figure 3.2). The summary statistics for BoS were very similar for the three estimation methods. For example, the median BoS statistics and the inter-quartile ranges (IQR) for the estimation methods (ML, REML and MM) were 8.98% (IQR: 2.71% to 20.18%), 11.06% (IQR: 4.03% to 23.28%) and 8.58% (IQR: 3.83% to 21.98%), respectively. Additionally for the estimation methods (ML, REML and MM) the corresponding maximum BoS statistics were 57.15%, 51.40% and 45.78%.

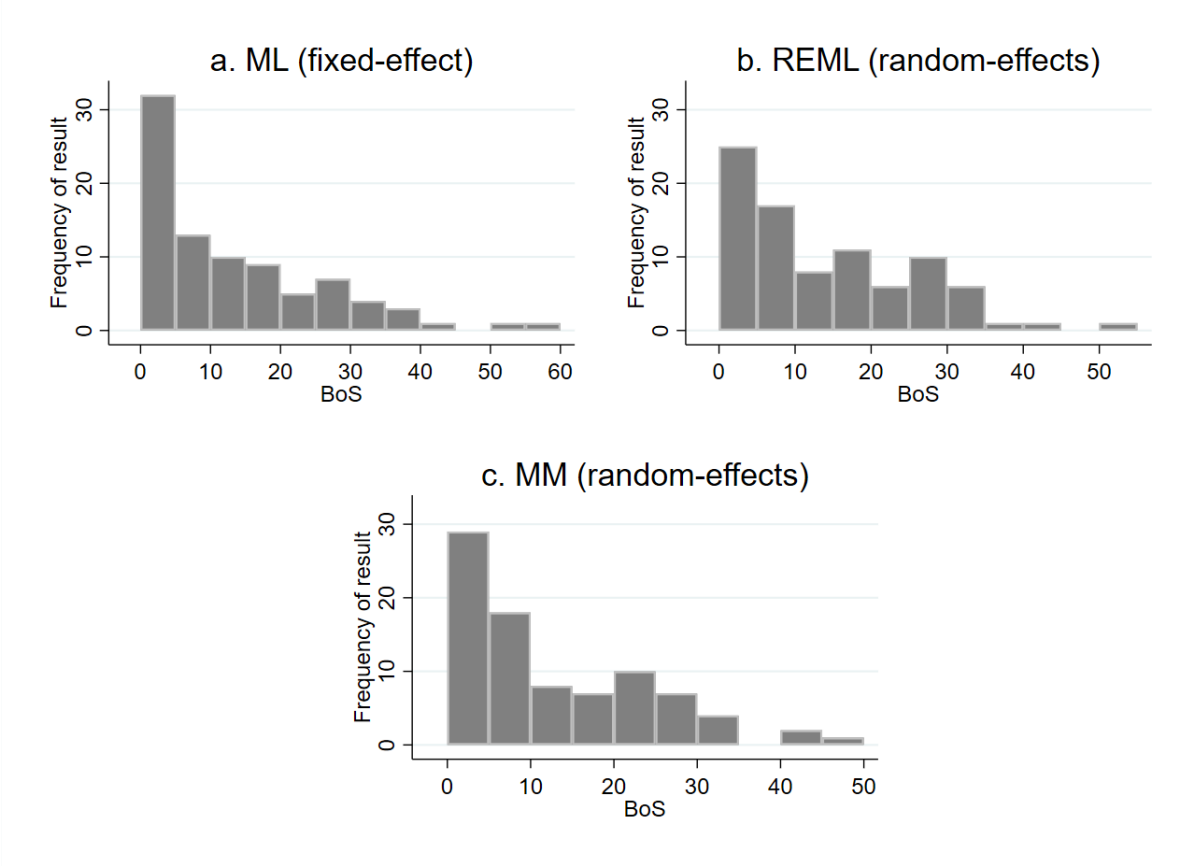


Figure 3.2: Histograms of the BoS statistics from both outcomes for each estimation method

Table 3.1: Summary statistics for BoS from each estimation method

	ML BoS	REML BoS	MM BoS
	%	%	%
Minimum	0.049	0.015	0.087
Maximum	57.150	51.402	45.782
Mean	13.193	14.130	12.927
25^{th} Percentile	2.706	4.025	3.826
Median	8.983	11.058	8.584
75^{th} Percentile	20.179	23.276	21.977

The agreement scatter plots (Figures 3.3, 3.4 and 3.5) show the majority of the BoS values from the meta-analyses agree between two estimation methods. The BoS values

were more similar between REML and MM (Figure 3.5), than for comparisons with the fixed-effect approach (Figures 3.3 and 3.4), which was expected given they are both random-effects estimation methods.

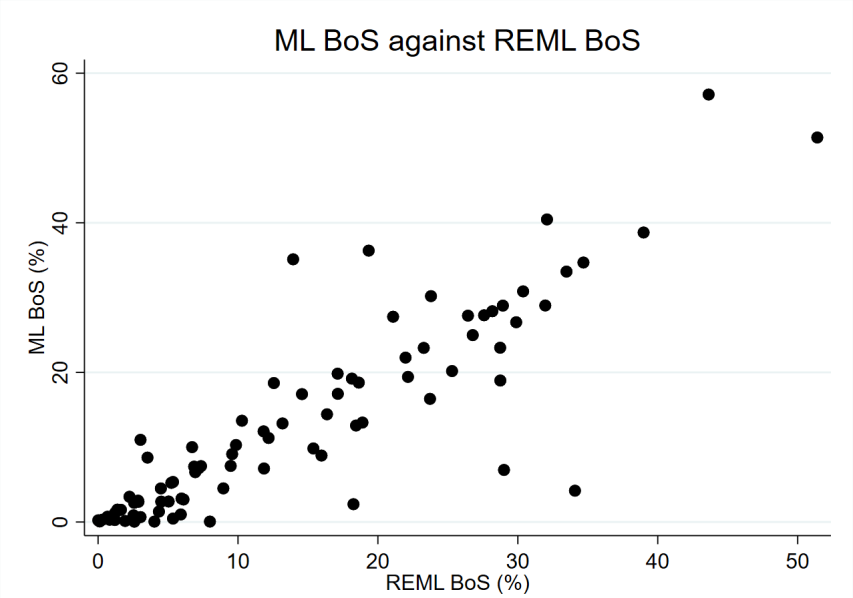


Figure 3.3: Scatter plot of the BoS statistics from REML random-effects meta-analyses against the BoS statistics from ML fixed-effect meta-analyses

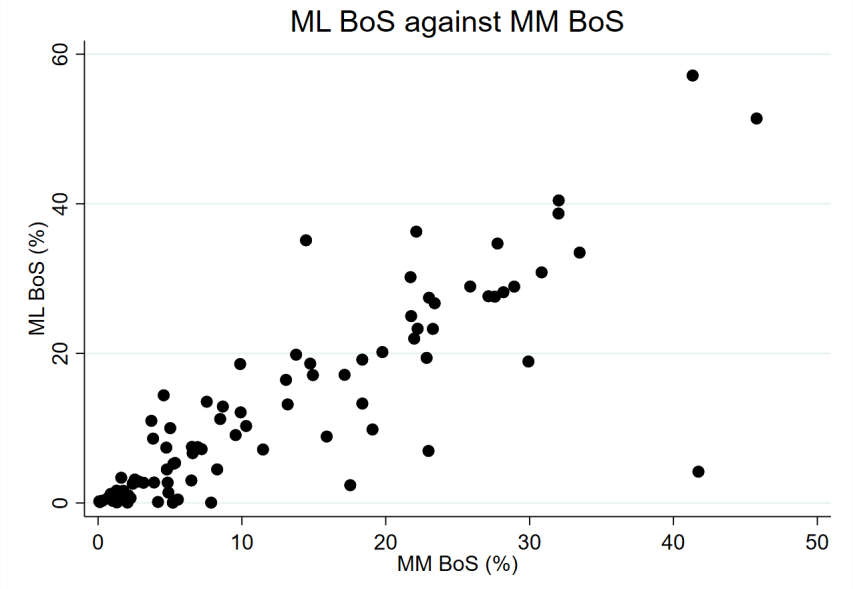


Figure 3.4: Scatter plot of the BoS statistics from MM random-effects meta-analyses against the BoS statistics from ML fixed-effect meta-analyses

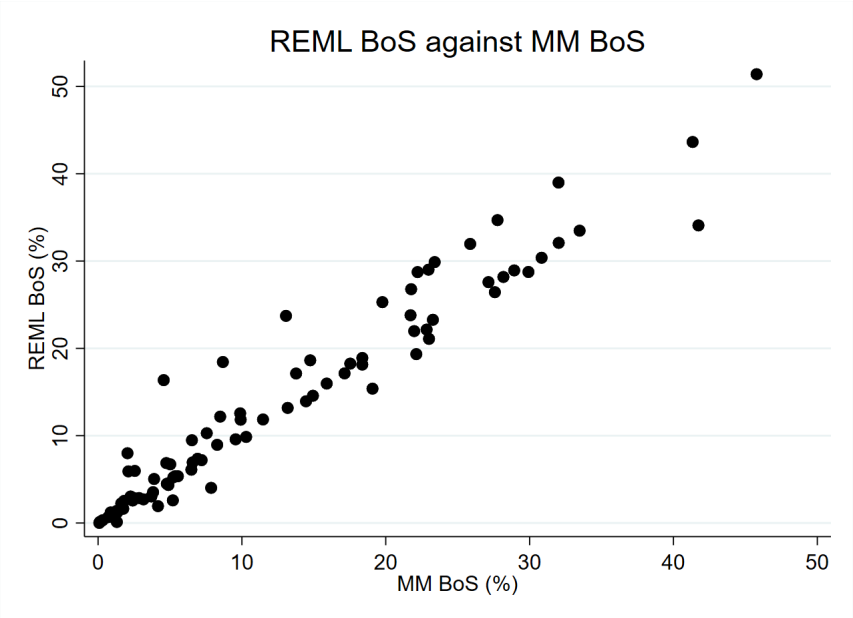


Figure 3.5: Scatter plot of the BoS statistics from REML random-effects meta-analyses against the BoS statistics from MM random-effects meta-analyses

3.3.2 Differences between the univariate and multivariate meta-analyses results

In the majority of meta-analyses, there was very little difference between the univariate meta-analysis and the multivariate meta-analysis summary estimates and the respective confidence intervals (Figures 3.6, 3.7, 3.8, 3.9, 3.10, 3.11 and Figures C.1, C.2, C.3, C.4, C.5, C.6 in Appendix C.1). This corresponds with the review by Trikalinos et al. (2013, 2014) who concluded that in the majority of cases any numerical differences were often small. This was irrespective of the estimation method used in the meta-analysis.

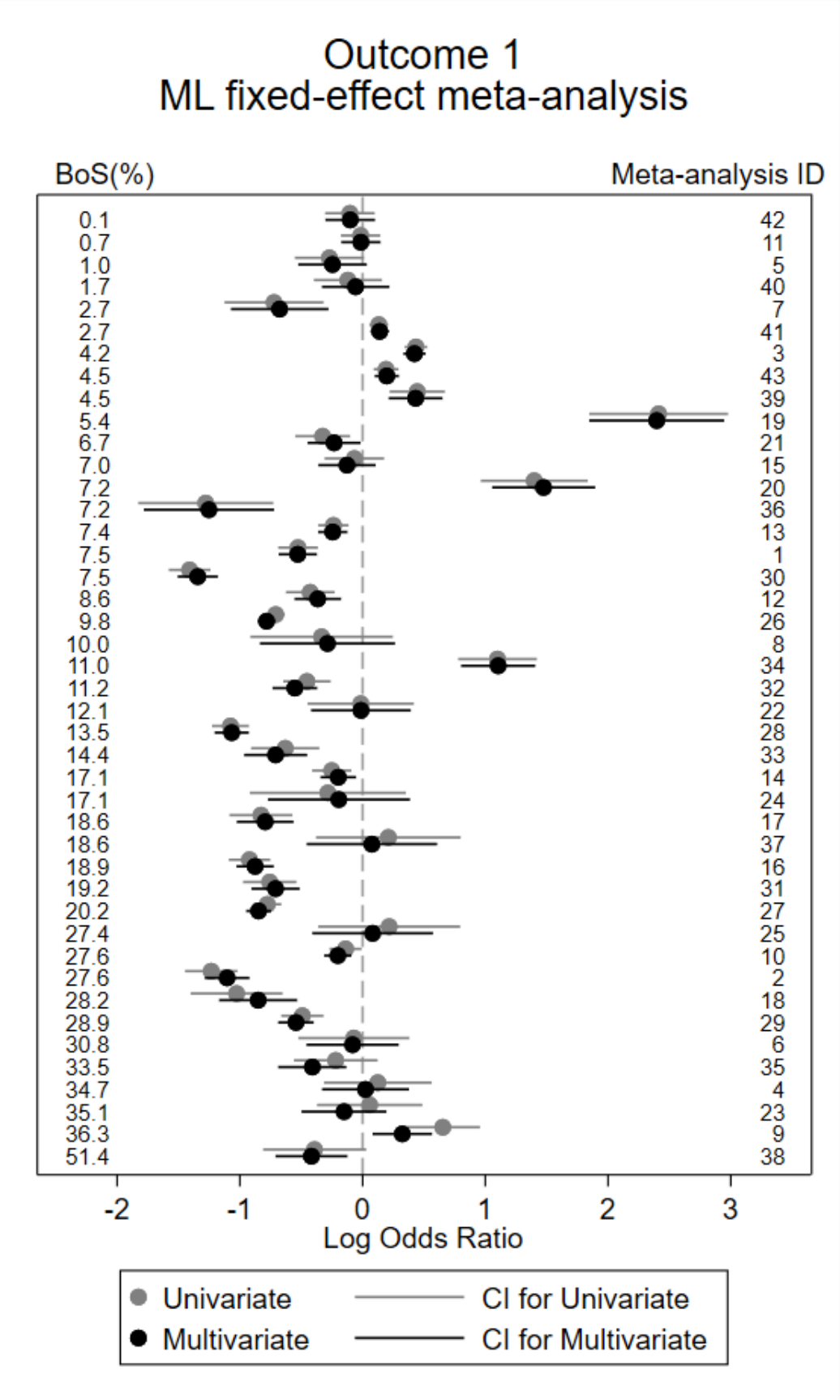


Figure 3.6: Comparison of the univariate and multivariate meta-analysis results on the log odds ratio scale for outcome one from the fixed-effect meta-analysis ordered by the BoS statistic

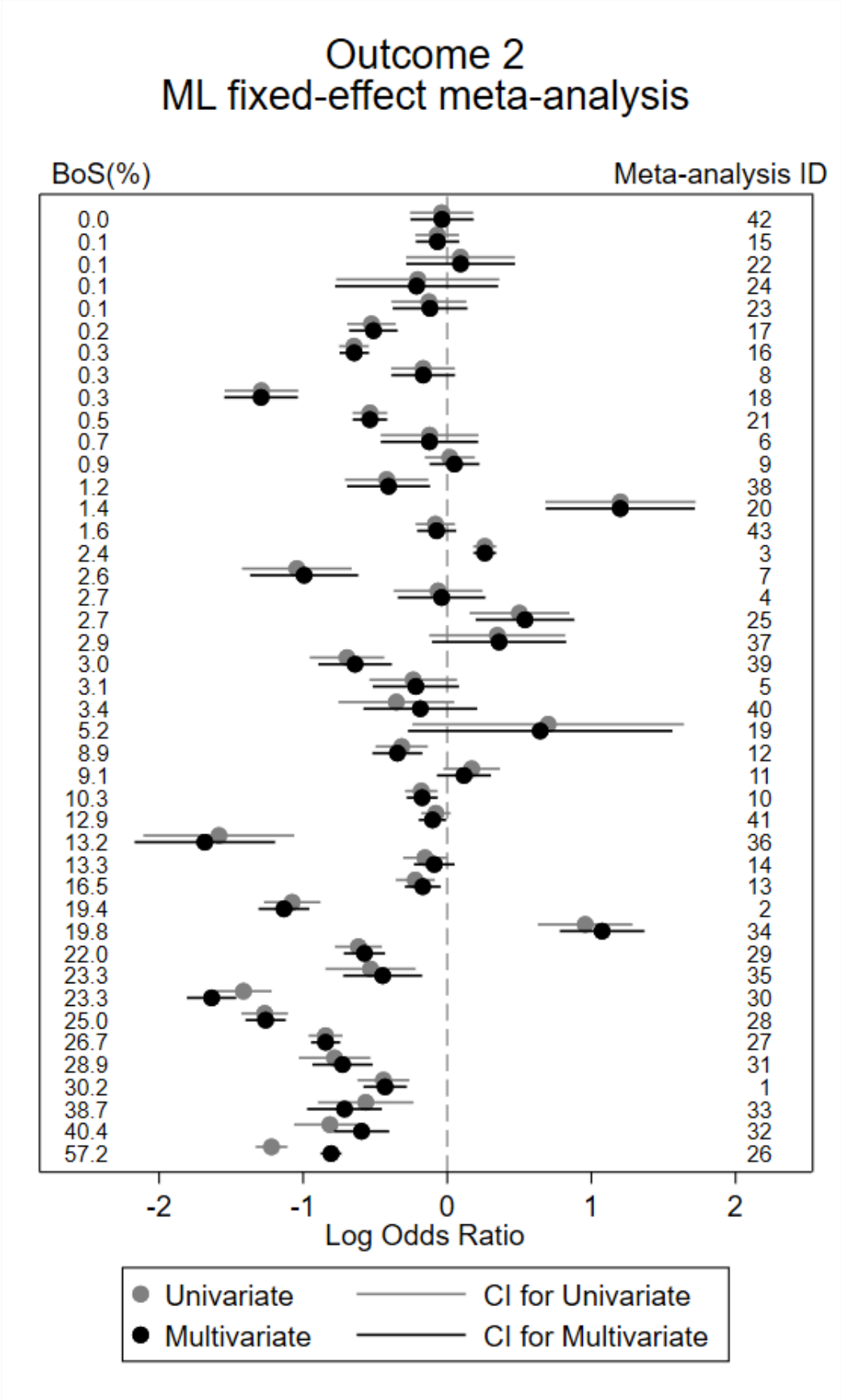


Figure 3.7: Comparison of the univariate and multivariate meta-analysis results on the log odds ratio scale for outcome two from the fixed-effect meta-analysis ordered by the BoS statistic

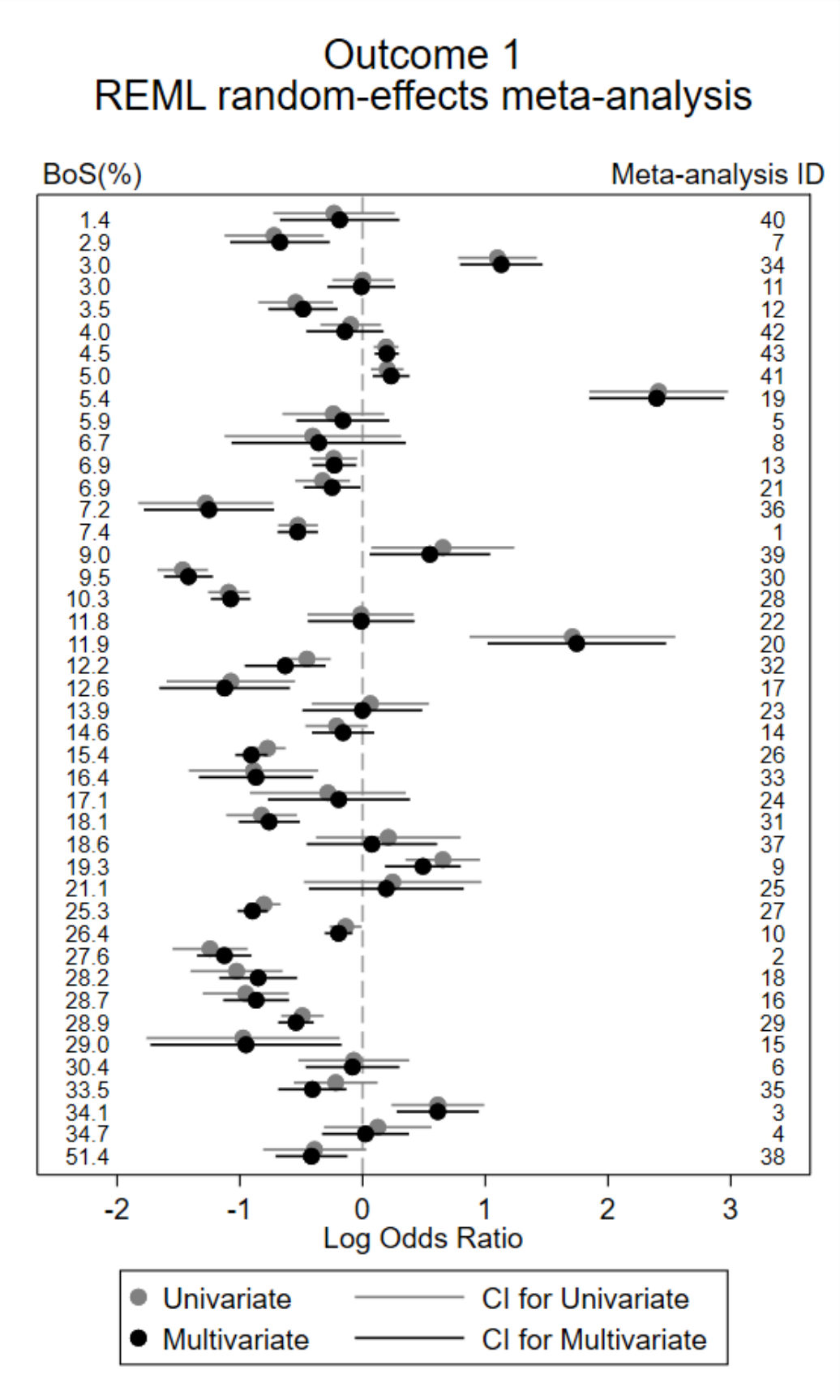


Figure 3.8: Comparison of the univariate and multivariate meta-analysis results on the log odds ratio scale for outcome one from the REML random-effects meta-analysis ordered by the BoS statistic

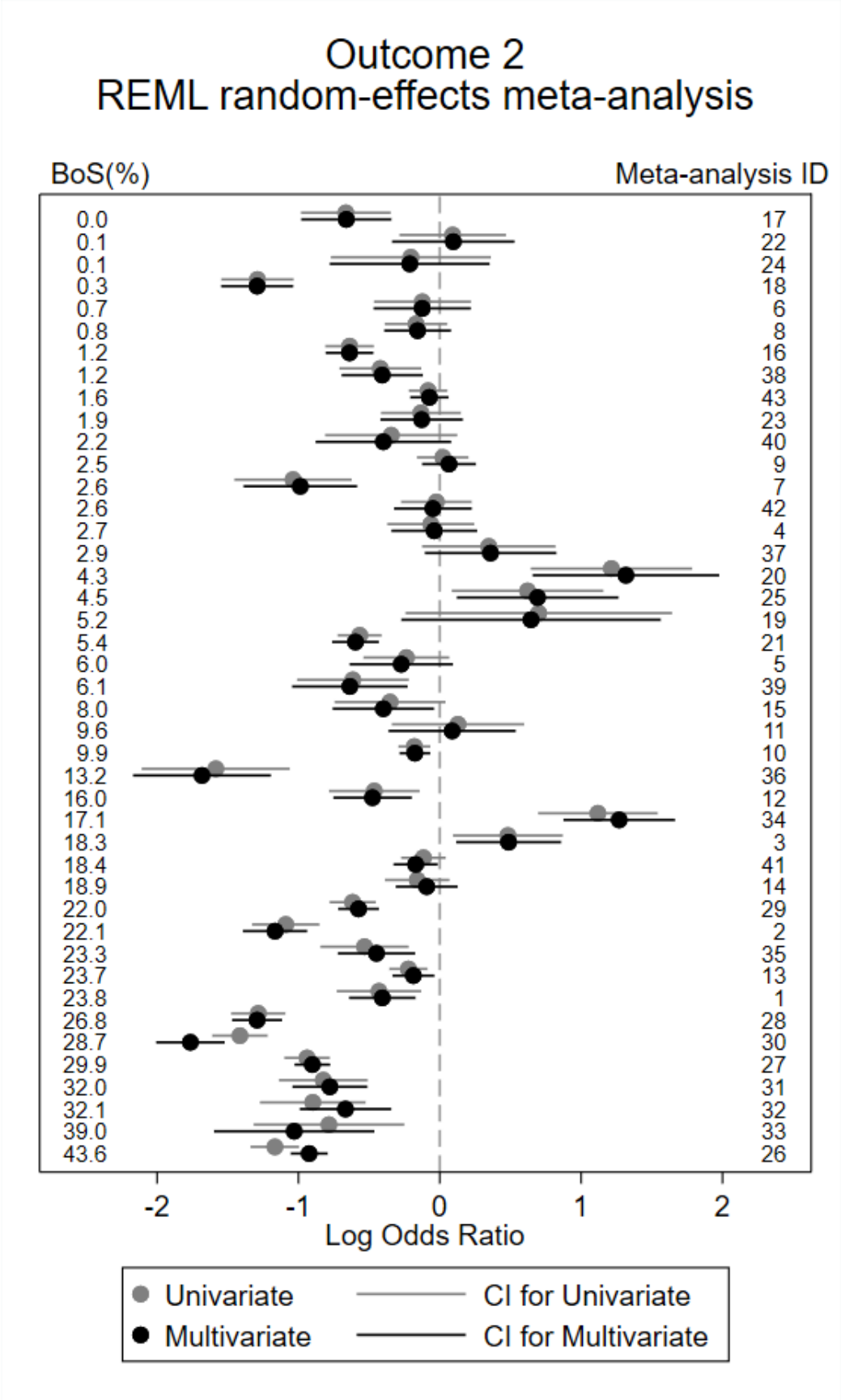


Figure 3.9: Comparison of the univariate and multivariate meta-analysis results on the log odds ratio scale for outcome two from the REML random-effects meta-analysis ordered by the BoS statistic

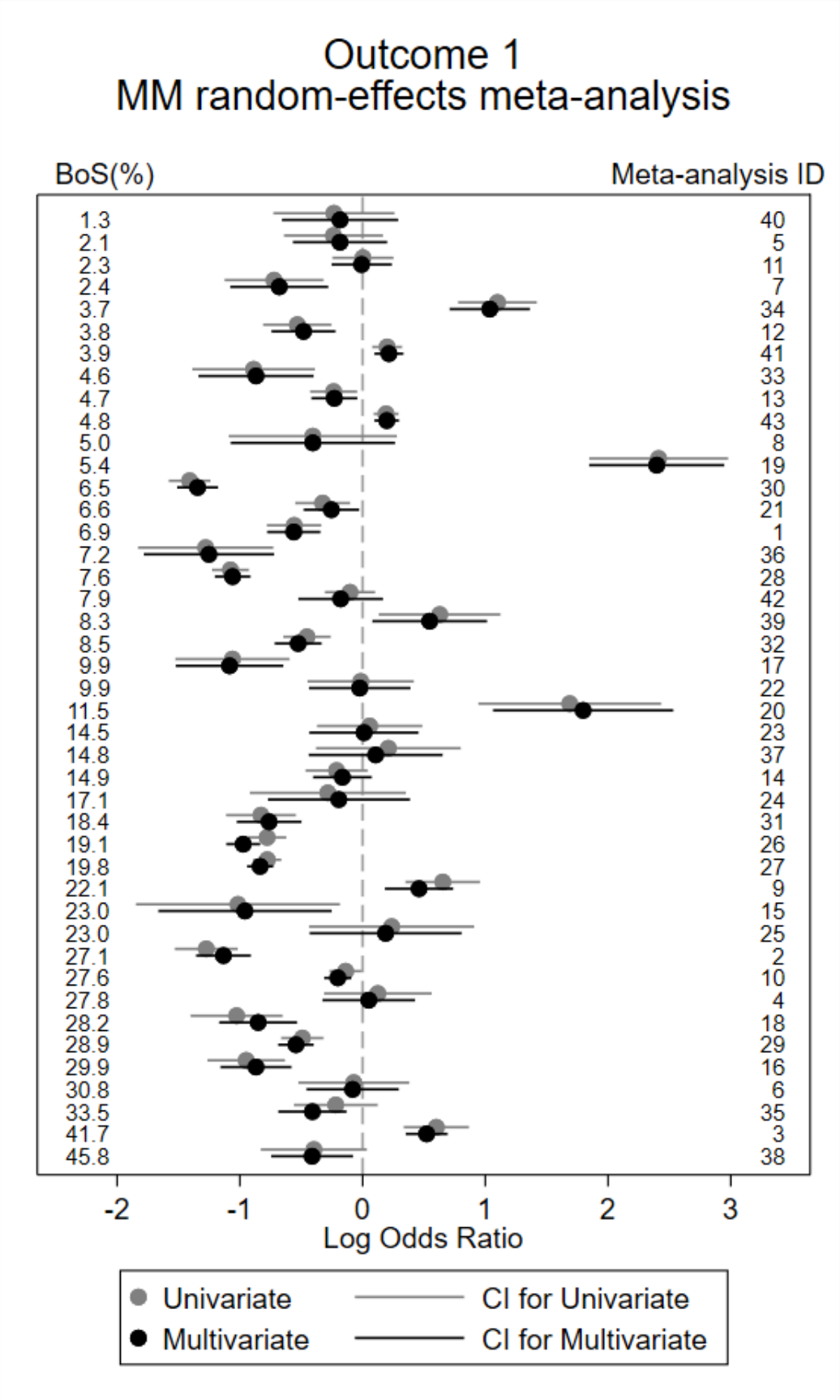


Figure 3.10: Comparison of the univariate and multivariate meta-analysis results on the log odds ratio scale for outcome one from the MM random-effects meta-analysis ordered by the BoS statistic

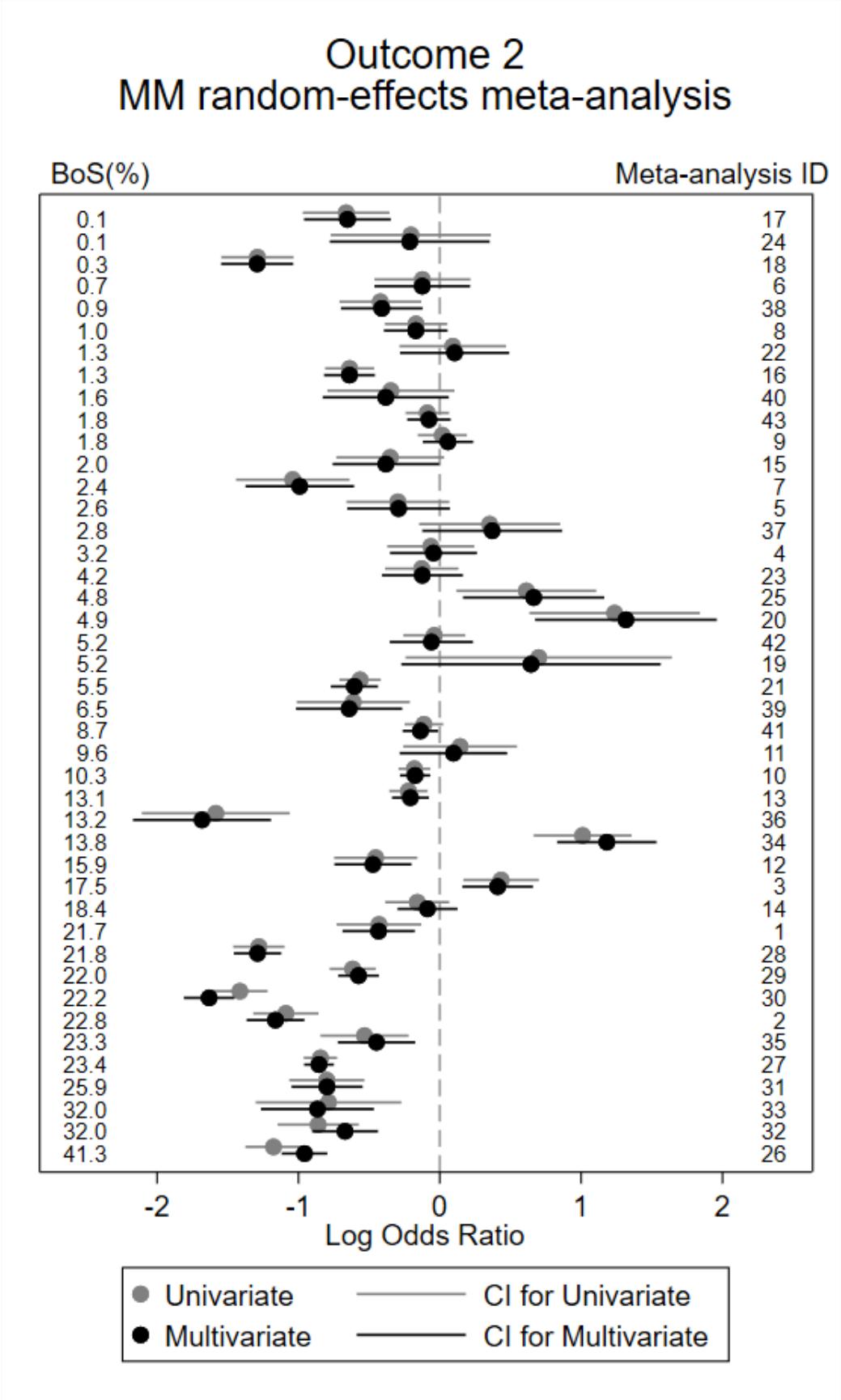


Figure 3.11: Comparison of the univariate and multivariate meta-analysis results on the log odds ratio scale for outcome two from the MM random-effects meta-analysis ordered by the BoS statistic

However, there are examples of meta-analyses where important differences between the univariate and multivariate meta-analysis results arose. In these meta-analyses changes in summary estimates and/or changes in confidence intervals were observed. An example of a change in summary estimate is from outcome one in meta-analysis 18 (Figures 3.6, 3.8, 3.10) regardless of estimation method. For example for the fixed-effect estimation, the summary log odds ratio was -1.03 from the univariate meta-analysis and -0.85 from the multivariate meta-analysis. In this example, the univariate summary log odds ratio for outcome one has a larger difference between the treatment groups compared to the multivariate summary log odds ratio.

For a noticeable change in the confidence interval, an example is meta-analysis six for outcome one. The summary log odds ratio was approximately the same for all estimation methods at -0.07 for the univariate meta-analysis and -0.08 for the multivariate meta-analysis. However, the confidence interval was narrower for the multivariate (95% C.I. -0.46 to 0.29 for MM) compared to the univariate meta-analysis (95% C.I.: -0.52 to 0.38).

Alongside the previous changes, a change in statistical significance sometimes occurs. An example is in meta-analysis 38 for outcome one, where the statistical significance changed between the univariate and the multivariate meta-analyses for all estimation methods. For example, for the REML estimation method the univariate result was -0.39 (95% C.I.: -0.81 to 0.03) with a p-value of 0.068 and the multivariate result was -0.42 (95% C.I.: -0.71 to -0.12) with a p-value of 0.005. In this example, there was only a statistically significant result in the multivariate meta-analysis and the conclusion of a beneficial treatment effect is much stronger from the multivariate meta-analysis.

3.3.3 Is the magnitude of BoS associated with difference between univariate and multivariate meta-analysis?

For each estimation method for small values of the BoS statistics, the results for the univariate and the multivariate meta-analyses were similar. For example, in meta-analysis 26 for outcome two, from the REML estimation method, the BoS was 1.2% and the univariate and multivariate results were -0.64 (95% C.I.: -0.81 to -0.47).

Additionally, generally when the BoS statistic was larger there were greater differences between univariate meta-analysis and multivariate meta-analysis results, than compared to when BoS was smaller. Meta-analyses 38 for outcome one (results in Section 3.3.2) and 35 for outcome one were examples where differences between the univariate and multivariate meta-analysis results were observed. The BoS statistics in these examples were large, at 33.5% (meta-analysis 35) and 51.4% (meta-analysis 38) from the REML estimation method. In both these meta-analyses, the statistical significance changed between the univariate and the multivariate results. For meta-analysis 35, the p-value from the univariate meta-analysis was 0.209 (summary log OR: -0.22, 95% C.I.: -0.56 to 0.12) which is not statistically significant, whereas the p-value for the multivariate meta-analysis was 0.004 (summary log OR: -0.41, 95% C.I.: -0.69 to -0.13), which is statistically significant.

When the BoS was large, the magnitude of the summary effect and the width of the confidence interval was also observed to change between the univariate and the multivariate results. For example, outcome two from meta-analysis 26 had a BoS value of 57.2% from the ML fixed-effect estimation method. The summary univariate and multivariate estimates were very different in magnitude at -1.22 and -0.81, respectively; this may be clinically important. There were also large differences in the width of their 95% confidence intervals from the univariate and multivariate meta-analyses, -1.33 to -1.11 and -0.88 to -0.73, respectively. Although in this example there was no change in the statistical significance between the univariate and multivariate results, there were

potentially important clinical differences relating to the summary effect estimates.

In some examples there were large changes in the magnitude of the summary estimates between the univariate and multivariate, although this did not necessarily result in a large BoS statistic as one might have expected. For example, outcome one from meta-analysis 32 from the REML estimation method there was a large difference in the magnitude of the summary estimates from the univariate and the multivariate results, -0.45 and -0.63, respectively; however the BoS statistic was only 12.2%.

3.3.4 Factors associated with the magnitude of BoS

The factors listed in Section 3.2.6 identified by the research team as potentially associated with the magnitude of the BoS statistic were examined using univariable regression and the results are shown in Tables 3.2, 3.3 and 3.4 for fixed-effect, REML and MM estimation methods, respectively. The results from the univariable regressions for every potential factor were similar between the three estimation methods. All the factors for fixed-effect were statistically significantly associated with the BoS statistic. For REML and MM, the between-study variance (τ^2) and I^2 were not statistically significantly associated with the BoS statistic. Additionally for REML, the between-study correlation was not statistically significant, which was surprising but perhaps reflects a lack of power, as the confidence interval was wide.

Table 3.2: Table of univariable results for the possible predictors of BoS statistic from the fixed-effect meta-analyses

Covariate	Effect of covariate on BoS			
	Coef.	S.E.	p-value	95% C.I.
Number of studies	0.192	0.050	<0.001	0.094 to 0.291
Percentage missing across both outcomes	0.490	0.129	<0.001	0.236 to 0.743
Percentage missing for outcome of interest	0.525	0.073	<0.001	0.381 to 0.668
Number of studies with outcome of interest	0.174	0.070	0.013	0.036 to 0.312
Number of studies with both outcomes	0.358	0.109	0.001	0.146 to 0.571
Average within-study correlation	12.559	3.577	<0.001	5.548 to 19.569
Average absolute within-study correlation	28.362	7.138	<0.001	14.371 to 42.352
Maximum within-study correlation	12.834	3.417	<0.001	6.138 to 19.531
Maximum absolute within-study correlation	29.076	6.516	<0.001	16.304 to 41.847

Table 3.3: Table of univariable results for the possible predictors of BoS statistic from the REML meta-analyses

Covariate	Effect of covariate on BoS			
	Coef.	S.E.	p-value	95% C.I.
Number of studies	0.169	0.047	<0.001	0.077 to 0.260
Percentage missing across both outcomes	0.537	0.114	<0.001	0.313 to 0.761
Percentage missing for outcome of interest	0.549	0.060	<0.001	0.430 to 0.667
Number of studies with outcome of interest	0.146	0.065	0.025	0.019 to 0.273
Number of studies with both outcomes	0.285	0.101	0.005	0.087 to 0.483
Average within-study correlation	10.776	3.308	0.001	4.292 to 17.260
Average absolute within-study correlation	23.135	6.668	0.001	10.066 to 36.203
Maximum within-study correlation	10.934	3.167	0.001	4.727 to 17.141
Maximum absolute within-study correlation	23.786	6.113	<0.001	11.805 to 35.766
Multivariate between-study variance	6.636	7.782	0.394	-8.616 to 21.888
Univariate between-study variance	2.502	7.865	0.750	-12.914 to 17.918
White's I^2	0.027	0.051	0.596	-0.073 to 0.127
Jackson's I^2	0.039	0.070	0.575	-0.098 to 0.176
Between-study correlation	2.721	1.784	0.127	-0.775 to 6.217
Absolute between-study correlation	8.852	11.467	0.440	-13.624 to 31.328

Table 3.4: Table of univariable results for the possible predictors of BoS statistic from the MM meta-analyses

Covariate	Effect of covariate on BoS			
	Coef.	S.E.	p-value	95% C.I.
Number of studies	0.159	0.044	<0.001	0.073 to 0.245
Percentage missing across both outcomes	0.501	0.109	<0.001	0.289 to 0.714
Percentage missing for outcome of interest	0.511	0.058	<0.001	0.398 to 0.625
Number of studies with outcome of interest	0.141	0.061	0.021	0.021 to 0.261
Number of studies with both outcomes	0.270	0.096	0.005	0.082 to 0.457
Average within-study correlation	10.770	3.108	0.001	4.679 to 16.861
Average absolute within-study correlation	22.072	6.302	<0.001	9.721 to 34.424
Maximum within-study correlation	10.568	2.988	<0.001	4.713 to 16.424
Maximum absolute within-study correlation	21.217	5.840	<0.001	9.770 to 32.664
Multivariate between-study variance	3.753	7.539	0.619	-11.022 to 18.528
Univariate between-study variance	3.041	7.595	0.689	-11.844 to 17.926
White's I^2	0.016	0.048	0.741	-0.079 to 0.111
Jackson's I^2	0.044	0.072	0.543	-0.098 to 0.185
Between-study correlation	2.776	1.375	0.044	0.081 to 5.470
Absolute between-study correlation	-1.346	3.031	0.657	-7.287 to 4.594

Number of studies

For the number of studies in the meta-analysis from the fixed-effect estimation method, there was an increase of 0.19% (95% C.I.: 0.09% to 0.29%) in the magnitude of BoS for each additional study included in the meta-analysis (Table 3.2). In other words, for every five additional studies the BoS value increases by 1%. In the REML estimation method, for approximately every six additional studies the BoS value increased by 1% (increase per study, REML: 0.17%, 95% C.I.: 0.08% to 0.26% (Table 3.3)). From Figure 3.12, it can be seen that the majority of meta-analysis studies had less than 50 studies, with the largest concentration around 20 studies. An upward trend such that as the number of studies in the meta-analysis increases the BoS also increases is apparent in Figure 3.12.

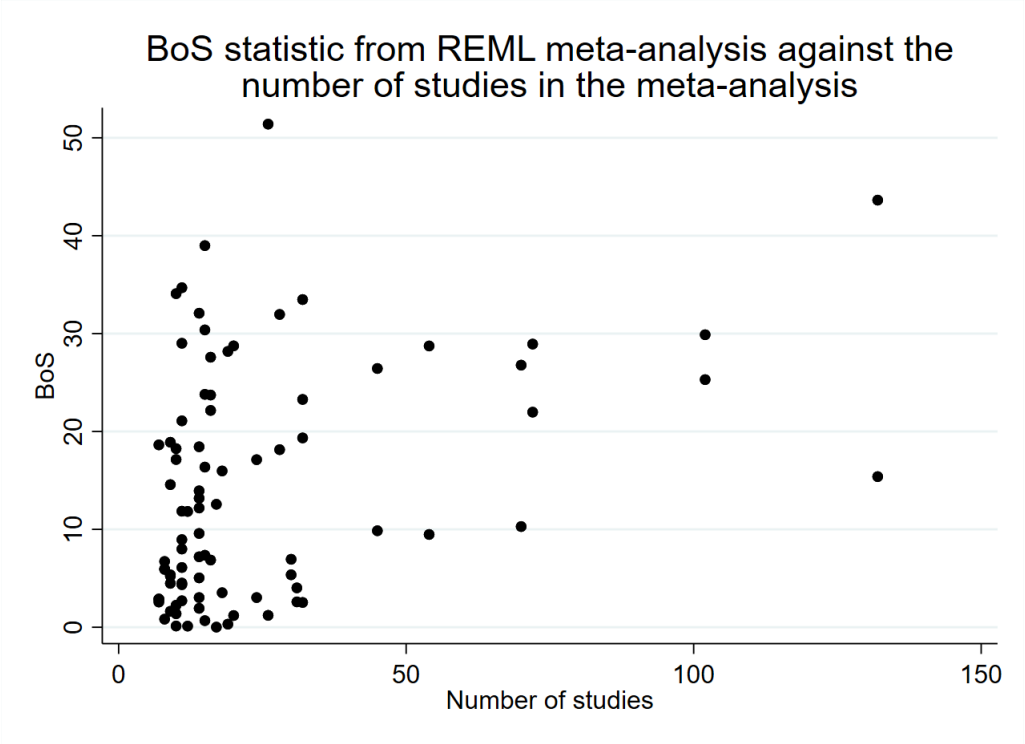


Figure 3.12: Scatter of BoS statistics from REML against the number of studies in the meta-analysis

Percentage of missing data

The percentage of missing data for the outcome of interest and the percentage of missing data across both outcomes were important factors associated with the magnitude of

BoS. Generalising across all estimation methods, for a 1% increase in missing data (for either percentage of missing data across both outcomes or for the outcome of interest) there was a 0.5% increase in the BoS. For example, considering the BoS statistics derived from the REML meta-analysis, the BoS statistic increased by 0.54% (95% C.I.: 0.31% to 0.76%) for an increase of 1% in the percentage of missing data for both outcomes (Table 3.3, Figure 3.13). A similar result was found for the percentage of missing data for the outcome of interest (0.55%, 95% C.I.: 0.43% to 0.67%)(Table 3.3, Figure 3.14). Upward trends are evident in both Figures 3.13 and 3.14, although a trend is more prominent in Figure 3.14 for the percentage of missing data for the outcome of interest, which is sensible as an outcome is more likely to gain information from other outcomes when it is missing in more studies.

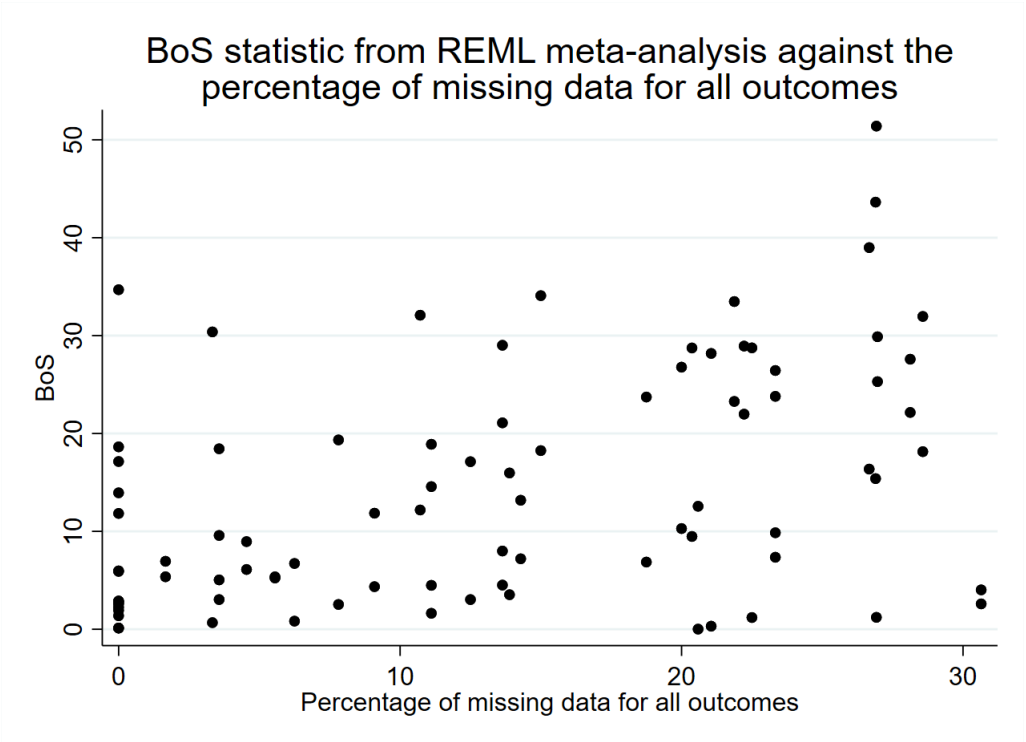


Figure 3.13: Scatter of BoS statistics from REML against the percentage of missing data for both outcomes

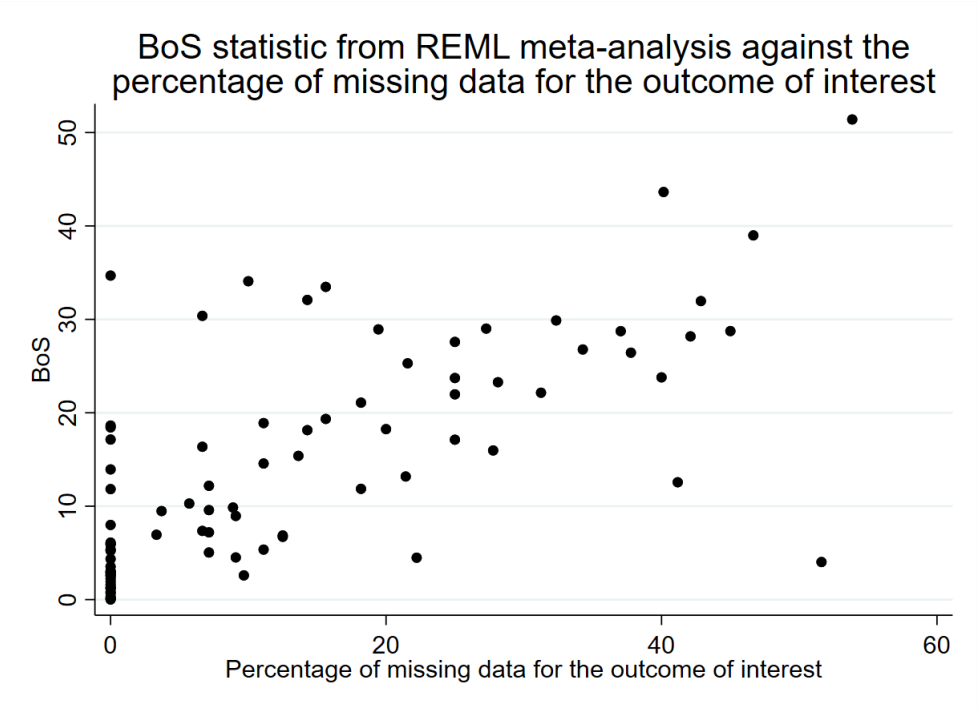


Figure 3.14: Scatter of BoS statistics from REML against the percentage of missing data for the outcome of interest

Within-study correlation

The magnitude of the within-study correlation was also associated with the magnitude of the BoS statistic. The magnitude of the association differed slightly between the fixed-effect and the random-effects meta-analyses (Tables 3.2, 3.3 and 3.4). For the fixed-effect meta-analysis, an increase of one in the average absolute within-study correlation resulted in an increase of 28.37% (95% C.I.: 14.37% to 42.35%) in the BoS statistic (Table 3.2). However, in the random-effects meta-analysis estimated using REML the BoS statistic only increases by 23.13% (95% C.I.: 10.07% to 36.20%) for an increase of one in the within-study correlation (Table 3.3). From Figure 3.15, there appears to be an association between the BoS statistic and the average absolute within-study correlation. It is unclear from the scatter graph whether the magnitude of the BoS statistic is bounded according to the magnitude of the average absolute within-study correlation. However, from the scatter plot it appears that it is unlikely for the BoS statistic to be large if the average absolute within-study correlation is small.

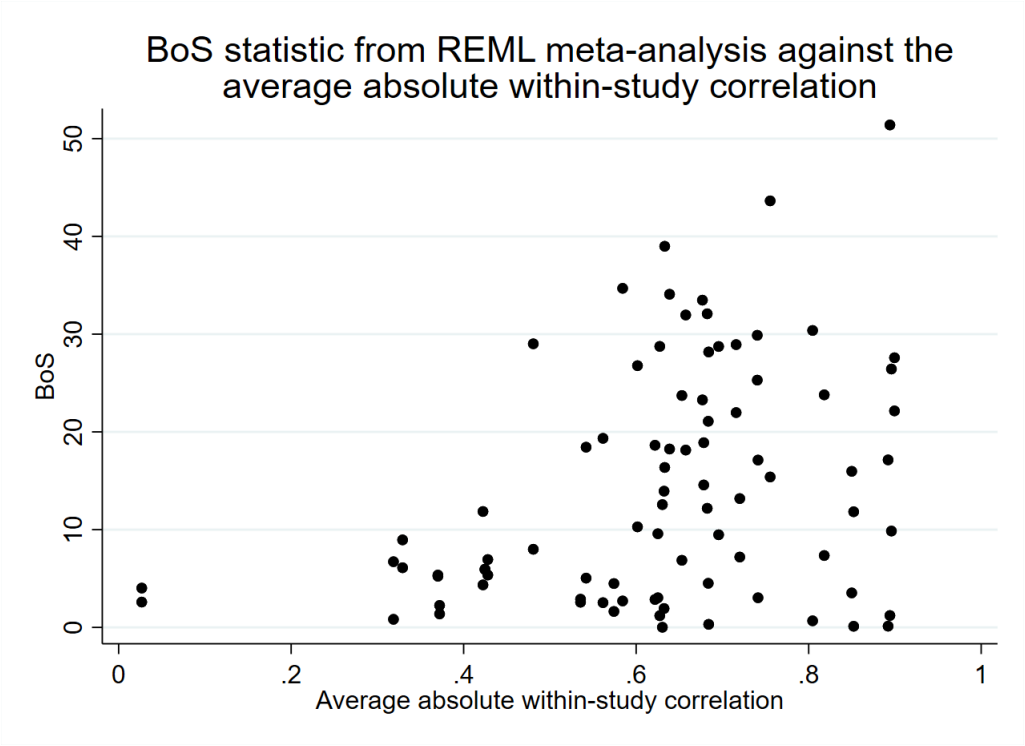


Figure 3.15: Scatter of BoS statistics from REML against the average absolute within-study correlation

Between-study correlation

The absolute between-study correlation was only calculated for the random-effects meta-analyses. For the REML and MM meta-analyses, there was no strong evidence of an association between the magnitude of the BoS statistic and the absolute between-study correlation. However for the MM meta-analyses, there was a statistically significant association between the magnitude of BoS and the between-study correlation (Table 3.4). For example, for an increase of 1 in the between-study correlation, the magnitude of BoS increases by 2.78% (95% C.I.: 0.8% to 5.47%), which is quite a small increase.

The lack of clear evidence for a relationship between BoS and the between-study correlation is likely due to a lack of power, due to the low variability of the estimate of between-study correlation. In Figure 3.16, the majority of absolute between-study correlations for the meta-analyses were estimated with magnitude one, which makes

it difficult to see an association between the absolute between-study correlation and the BoS statistic. Additionally, this shows that the between-study correlation often encounters estimation problems.

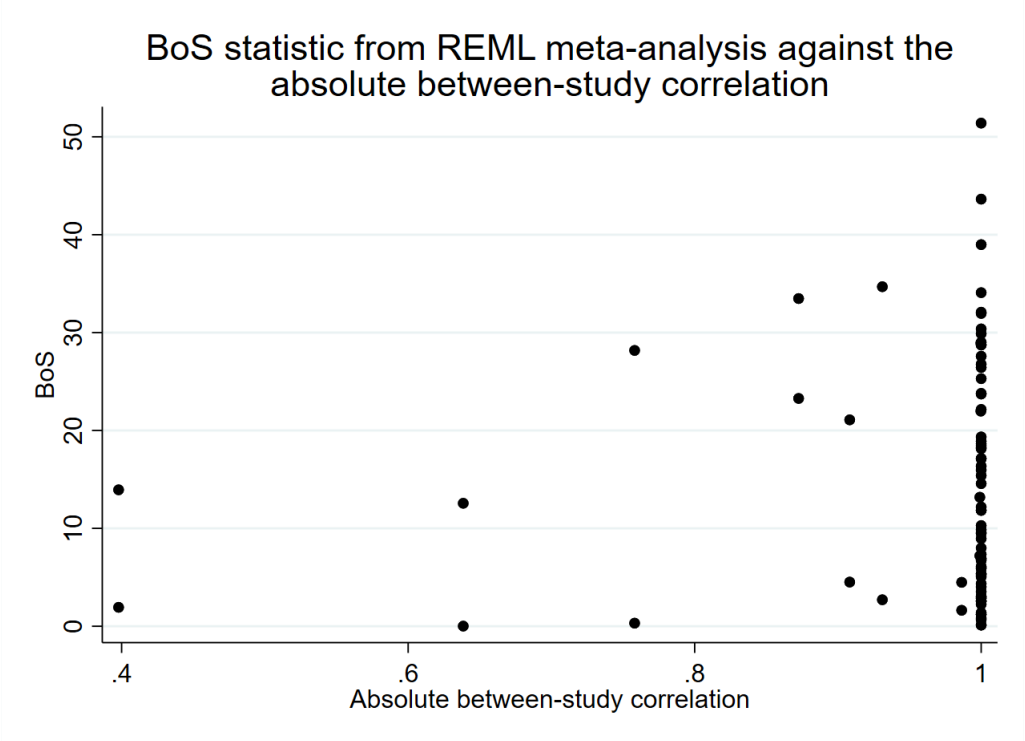


Figure 3.16: Scatter of BoS statistics from REML against the absolute between-study correlation

White's and Jackson's I^2

The distributions of White's I^2 and Jackson's I^2 differed between the meta-analyses estimated using REML and MM (Figure 3.17). For White's I^2 statistic, a large percentage of meta-analyses, for each estimation method, had very small I^2 statistics that were less than 5% (Figure 3.17a and b). For each 5% increment of I^2 values between 5% and 80% there were far fewer corresponding meta-analysis. The largest value of White's I^2 was 79.5% for REML and 75.1% for MM.

Jackson's I^2 also had its largest proportion of meta-analyses between 0% and 5% (Figure 3.17c and d). However, the proportion was not as large as that of White's I^2 (Figure 3.17a and b) and the largest values of Jackson's I^2 were not as large, 69.1% for REML

and 64.1% for MM.

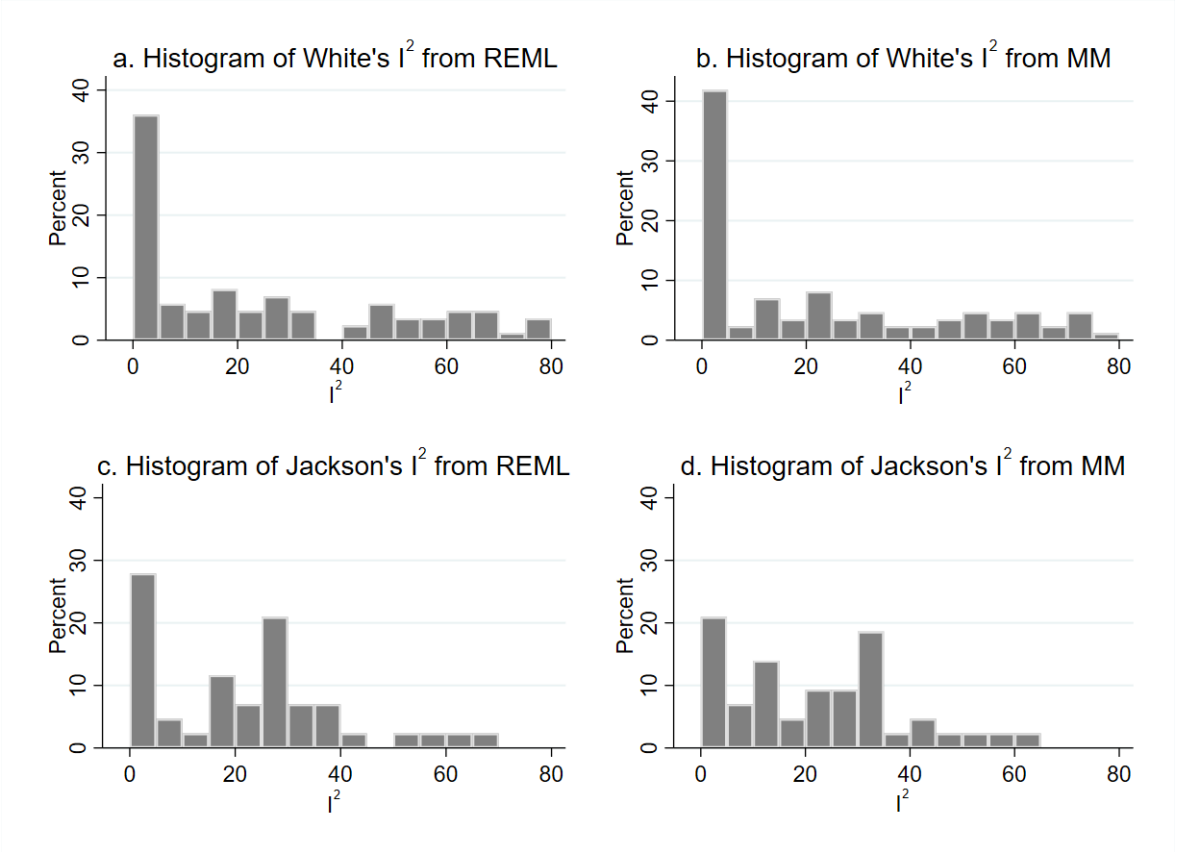


Figure 3.17: Histograms of the magnitude of the Jackson's and White's I^2 statistics for REML and MM meta-analyses

There was no statistically significant evidence to suggest an association between White's I^2 or Jackson's I^2 and the magnitude of BoS (Tables 3.3 and 3.4). For example, for Jackson's I^2 statistic from the REML meta-analyses for an increase of one in the I^2 statistic there was an increase of 0.04% (95% C.I.: -0.10% to 0.18%) in the BoS statistic. Visually, this can be explored through the scatter plots for White's I^2 (Figures 3.18 and 3.19) and Jackson's I^2 (Figures 3.20 and 3.21).

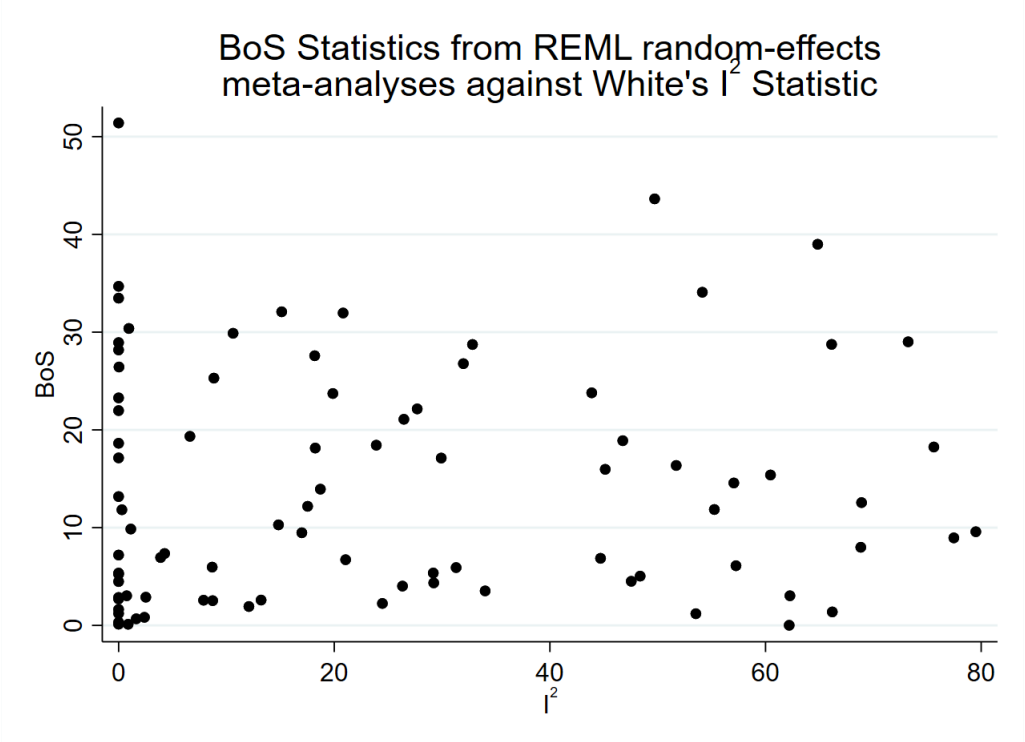


Figure 3.18: Scatter plot of the BoS statistics against White's I^2 for both outcomes from the REML random-effects meta-analyses

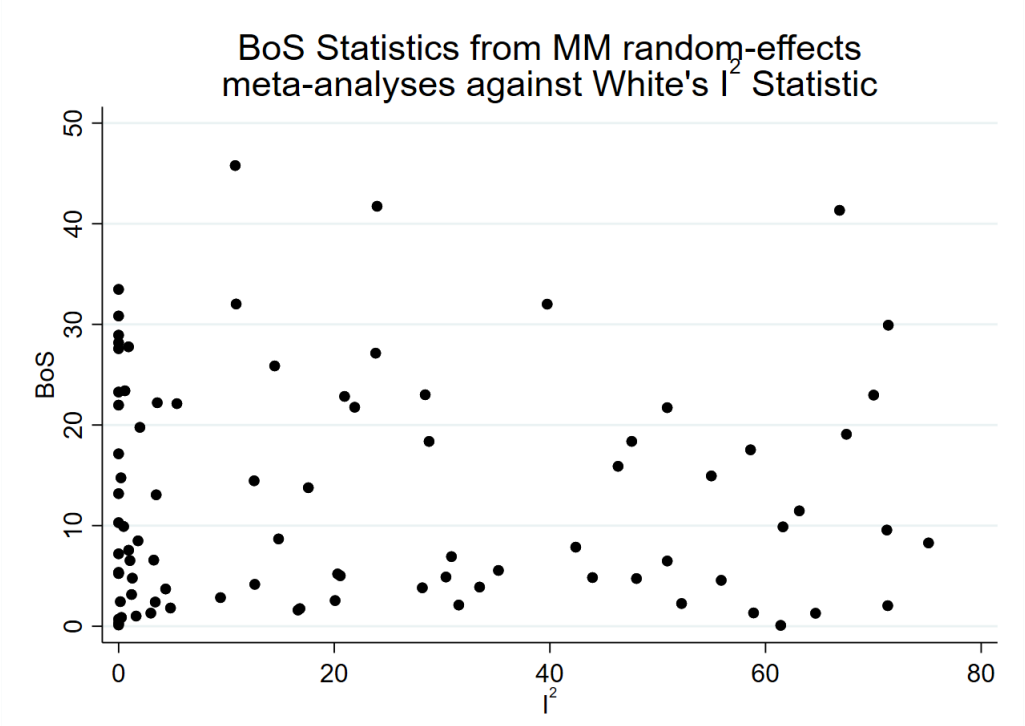


Figure 3.19: Scatter plot of the BoS statistics against White's I^2 for both outcomes from the MM random-effects meta-analyses

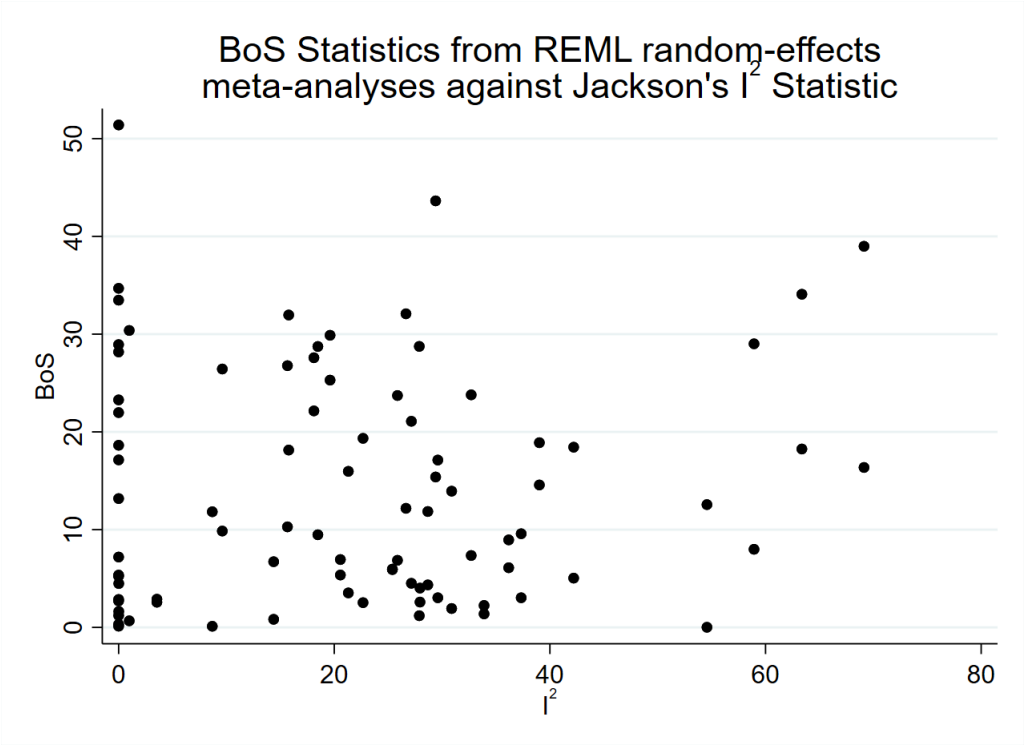


Figure 3.20: Scatter plot of the BoS statistics against Jackson's I^2 for both outcomes from the REML random-effects meta-analyses

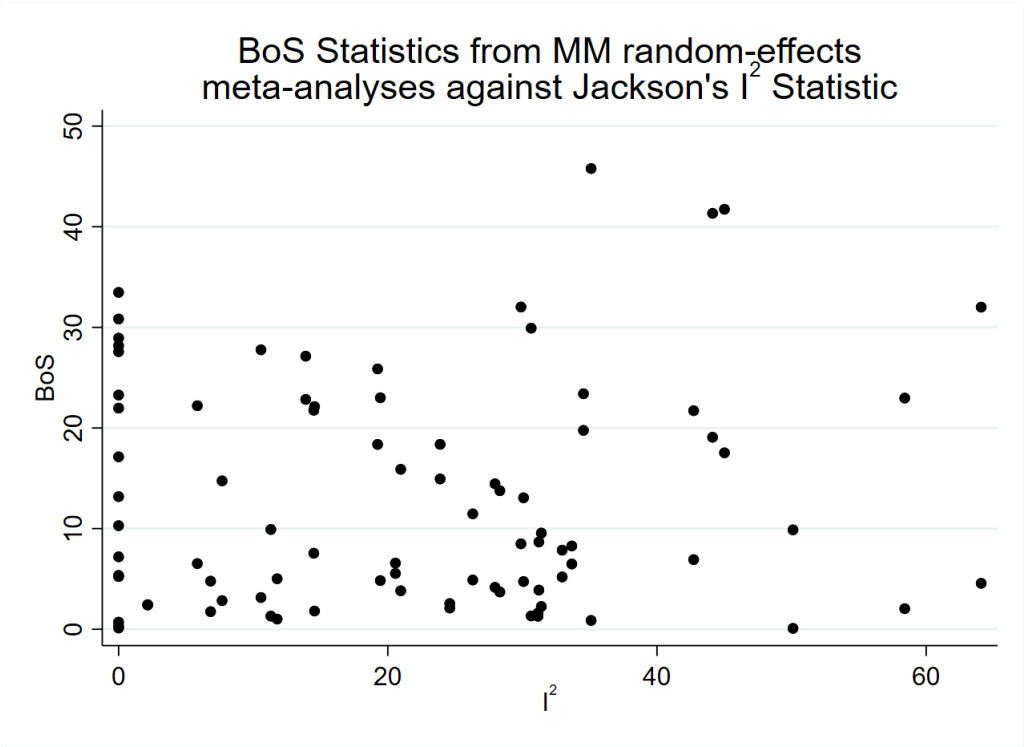


Figure 3.21: Scatter plot of the BoS statistics against Jackson's I^2 for both outcomes from the MM random-effects meta-analyses

Multivariate and Univariate between-study variance

The relationship between the BoS statistic and the estimated between-study variance was explored using scatter plots. Figures 3.22, 3.23, 3.24 and 3.25 suggest that, for either REML, MM, univariate or multivariate analyses, the estimated between-study variance did not appear to show any association with the BoS statistic. There was no evidence from the univariable regressions of an association between the univariate or the multivariate between-study variance and the magnitude of the BoS statistic, irrespective of random-effects estimation method (Tables 3.3 and 3.4). For example, for an increase of one in the multivariate between-study variance the BoS statistic from the REML meta-analyses increased by 6.64% (95% C.I.: -8.62 to 21.89). From the MM meta-analyses, an increase of one in the multivariate between-study variance the BoS statistic increased by 3.75% (95% C.I.: -11.02 to 18.53).

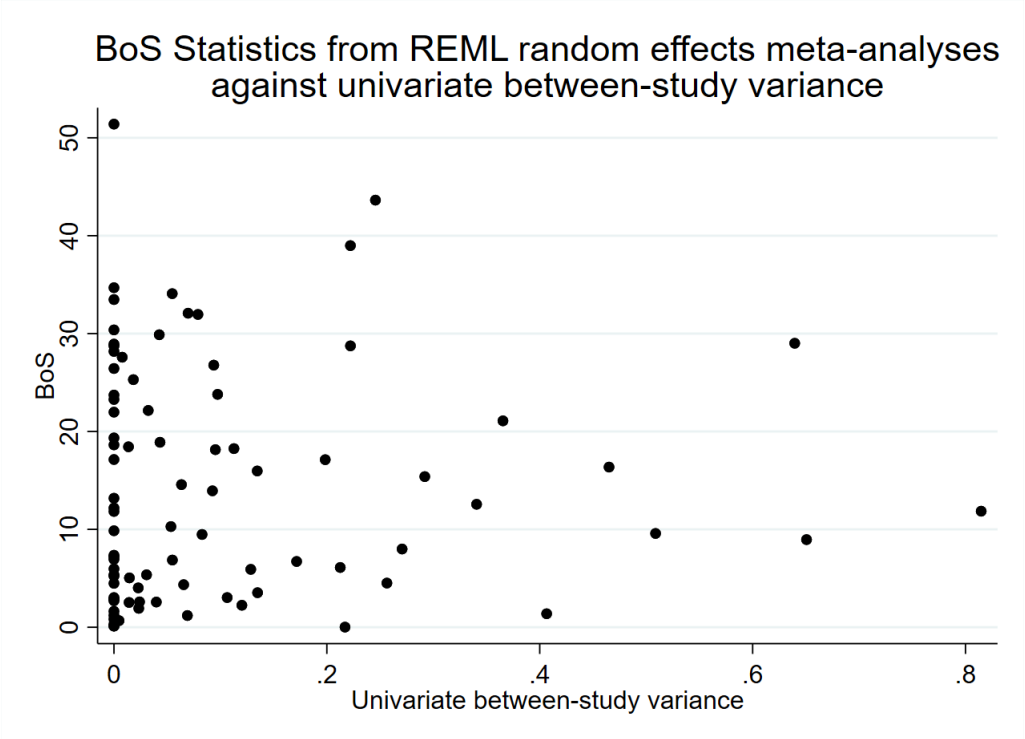


Figure 3.22: Scatter plot of the BoS statistics against the univariate between-study variance from the REML random-effects meta-analyses

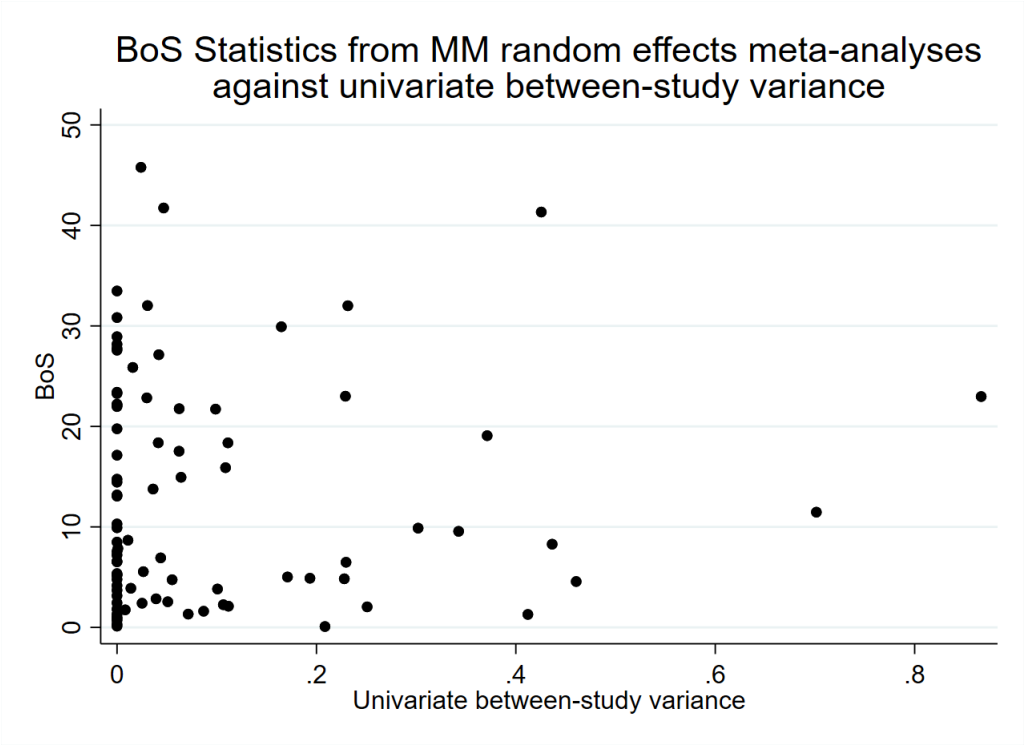


Figure 3.23: Scatter plot of the BoS statistics against the univariate between-study variance from the MM random-effects meta-analyses

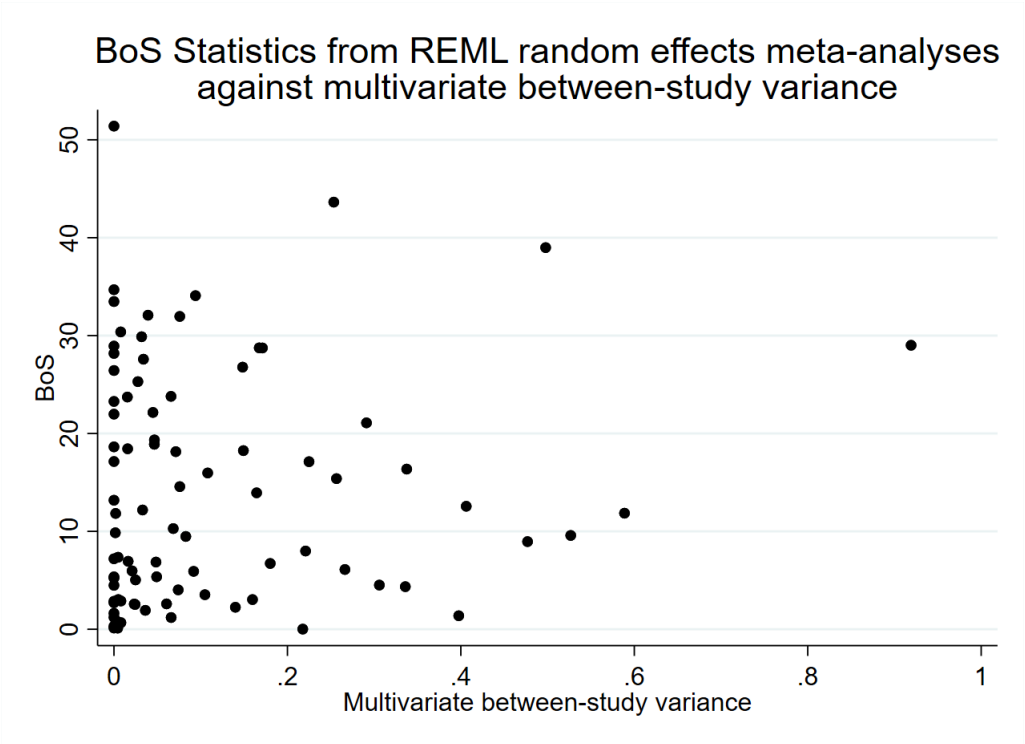


Figure 3.24: Scatter plot of the BoS statistics against the multivariate between-study variance from the REML random-effects meta-analyses

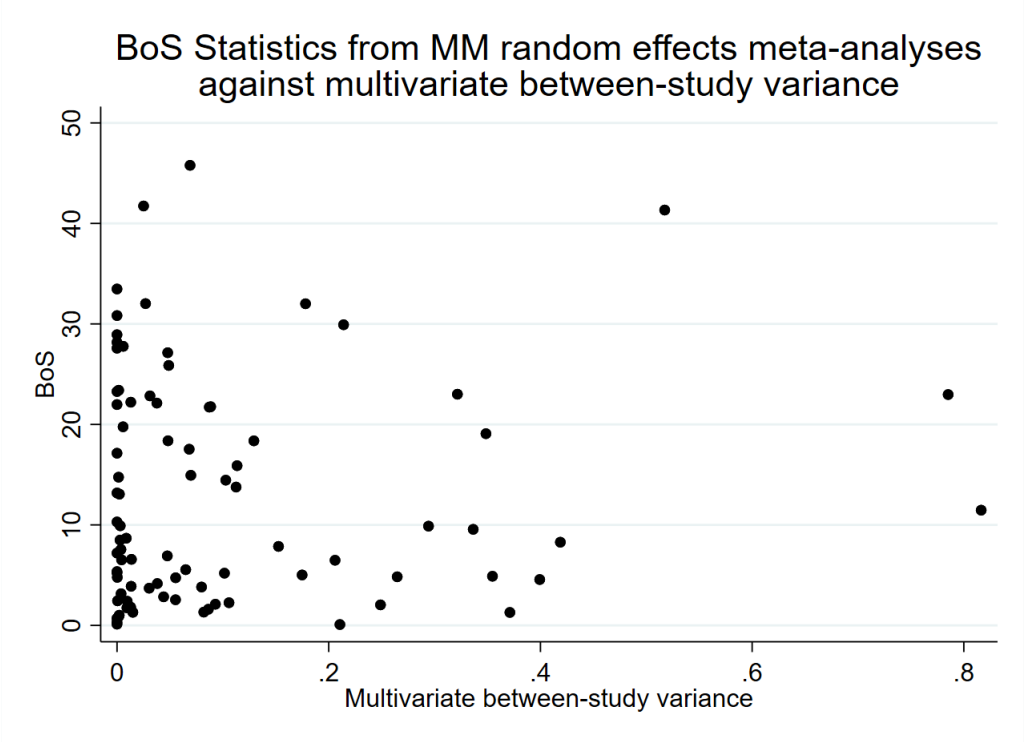


Figure 3.25: Scatter plot of the BoS statistics against the multivariate between-study variance from the MM random-effects meta-analyses

3.3.5 Example Revisited

In section 3.3.3, a few example meta-analyses from the Trikalinos et al. (2014) review were discussed due to their relationship between the univariate and multivariate results and the magnitude of BoS. In particular, outcome one from meta-analysis 38 was interesting, since there was a statistically important difference between the univariate and the multivariate meta-analysis results. For the REML meta-analysis, the univariate results were -0.39 (95% C.I.: -0.81 to 0.03) and the multivariate results were -0.42 (95% C.I.: -0.71 to -0.12). The point estimates are similar between the univariate and the multivariate, however different conclusions could be drawn from the confidence intervals. The confidence interval for the multivariate results is narrower due to the gain in precision and this corresponds to a large BoS statistic, 51.4%. Through the exploration of meta-analysis level characteristics that have been shown, in this section, to be associated with the magnitude of BoS, can the large BoS statistic in this example be understood further? Specifically from the number of studies, percentage of missing

data and the magnitude of the within-study correlation, the magnitude of the BoS can be understood since in this meta-analysis all these meta-analysis level characteristics were large. The number of studies included in the meta-analysis was 26. There was also 54% missing data for the outcome of interest and the average absolute within-study correlation was 0.9. Thus, it is perhaps not a surprise that, in this example, the BoS was large, given all the key factors that increase BoS were large.

3.3.6 Difference between the univariate and multivariate summary estimates divided by the standard error and its association with the BoS statistic

The borrowing of strength (BoS) statistic quantifies the gain in precision from a multivariate meta-analysis compared to a univariate meta-analysis for the outcome of interest. Therefore, the magnitude of the BoS statistic can assist in identifying when confidence intervals for multivariate meta-analyses are more precise than univariate meta-analyses. However, it is unclear whether the BoS statistic would also identify changes between the univariate and multivariate summary estimates. As a method of assessing this, the difference between the summary estimates for univariate and multivariate meta-analyses was calculated and divided by the multivariate standard error; this is referred to as the 'difference statistic' from this point.

The difference statistic was associated with the magnitude of BoS from the fixed-effect and random-effects meta-analyses. For example, an increase of one in the effect size (the difference divided by the standard error), the BoS statistic increased by 2.07% (95% C.I.: 0.18% to 3.96%) (Tables 3.5 and C.4). However, for the REML and MM meta-analyses, the difference statistic was not associated with the magnitude of the BoS statistic. For example, for an increase of one in the difference statistic the BoS increased by -0.44% (95% C.I.: -3.88% to 3.01%, p-value: 0.804) for the MM meta-analyses and by 0.40% (95% C.I.: -3.05% to 3.85%, p-value: 0.820) for the REML meta-analyses. This suggests that although BoS reflects gain in precision, it may not

relate to change in summary estimate.

Table 3.5: Table of univariable results for the difference between the univariate and multivariate summary estimates divided by the standard error for each estimation method

Estimation	Effect of difference statistic on BoS			
Method	Coef.	S.E.	p-value	95% C.I.
Fixed-effect	2.071	0.964	0.032	0.1825 to 3.960
REML	0.401	1.759	0.820	-3.046 to 3.848
MM	-0.435	1.757	0.804	-3.878 to 3.008

3.4 Discussion

This chapter empirically examined the differences between univariate and multivariate summary effect estimates and 95% confidence intervals from 43 Cochrane reviews containing two binary outcomes of interest, with the magnitude of the BoS statistic (Trikalinos et al. 2013, 2014). The aim of this chapter was to analyse whether the magnitude of BoS identified differences between the univariate and multivariate meta-analysis results. Following, meta-analysis level factors were analysed univariably to determine whether there was an unadjusted association with the magnitude of BoS. The key findings from this chapter are summarised in Figure 3.26.

Figure 3.26: Key findings

Key Findings:

- When BoS was small, the differences between the univariate and the multivariate summary estimates and confidence intervals tended to be small.
- When BoS was large, there was a gain in precision, thus narrower confidence intervals and sometimes a change in statistical significance.

- The majority of meta-analysis level factors analysed using univariable mixed-effect multilevel models were associated with the magnitude of BoS: number of studies in the meta-analysis, percentage of missing data across both outcomes, percentage of missing data for the outcome of interest, the number of studies with the outcome of interest and number of studies with both outcomes and the within-study correlation (average and maximum).
- The meta-analysis level factors that were not associated with the magnitude of BoS were the I^2 statistics, between-study variance values and often the measures of the between-study correlation were not associated.
- The difference in summary estimates between the univariate and the multivariate was not associated with the magnitude of the BoS. This suggests that although BoS reflects gain in precision, it may not reflect changes in the summary estimates.

In the majority of meta-analysis studies, the summary estimates and the confidence intervals did not differ between the univariate and the multivariate meta-analyses (Trikalinos et al. 2013, 2014). However, there were meta-analysis studies with clinically and/or statistically important differences between the univariate and multivariate meta-analysis results. This motivated research into the comparison between the results and the magnitude of the BoS statistic.

A key finding from this chapter was that in the majority of situations where the BoS statistic was large, greater differences between the univariate and multivariate results were observed compared to when BoS was small. The types of differences included differences in statistical significance, differences in width of confidence intervals and differences in the confidence intervals values. Note that these were only observed in particular meta-analyses with large BoS, not all meta-analyses with large BoS. Often when BoS was large, the confidence intervals were narrower for the multivariate than the univariate meta-analysis, which was expected due to the additional information

included in a multivariate meta-analysis.

There were many factors that were associated with the magnitude of the BoS statistic. The percentage of missing data across all outcomes was associated with the magnitude of BoS, since the greater the quantity of missing data the greater the opportunities for borrowing strength between outcomes will occur. Similarly, Riley et al. (2007*a*) discussed that borrowing strength when there is missing data is advantageous in providing smaller standard errors in the bivariate meta-analysis compared to the univariate meta-analysis. The percentage of missing data for the outcome of interest and the absolute within-study correlation were also highly associated with the magnitude of BoS. Likewise, Riley (2009) discussed ignoring the within-study correlation reduces the borrowing strength.

In contrast, there were meta-analysis level factors that were not associated with the magnitude of the BoS statistic. These included the absolute between-study correlation, I^2 statistics and the between-study variances. The between-study correlation is additional information that is utilised in the multivariate meta-analysis that is not utilised in the univariate meta-analysis, so it was expected that there would be a relationship between BoS and the between-study correlation. However, there was no evidence of any association with the BoS statistic. This may be due to low power in identifying an association since the absolute between-study correlation was one in the majority of meta-analyses in the review. Perhaps the key reason is that the between-study correlation is often poorly estimated at -1 or 1 and as a result the absolute between-study correlation does not vary importantly across meta-analyses.

In Section 3.3.6, the difference between the univariate and the multivariate summary estimates divided by the multivariate standard error was used to investigate the association between the difference in summary estimates and the magnitude of the BoS statistic. There was no evidence to suggest an association between the difference in

effect sizes and the magnitude of the BoS statistic for the random-effects estimation methods. In the fixed-effect setting there was evidence that there may be an association, although only a weak association. Therefore, it was concluded that the BoS statistic does not detect changes in the effect sizes well for the random-effects meta-analyses but it appears to be able to detect changes in the standard error and therefore the width of the confidence intervals.

3.4.1 Limitations and further work

The BoS statistic is calculated during a meta-analysis. If the BoS statistic is recommended to be used to identify settings where a multivariate meta-analysis is beneficial over a univariate meta-analysis, the calculation of BoS during analysis will be a limitation. Therefore, for further research it is a priority to investigate whether the BoS can be predicted accurately in advance of the application of meta-analysis (see Chapter 4).

A limitation of this chapter was the quantity of data used to compare the univariate and multivariate results was limited. There were only 43 Cochrane reviews included. The studies also only had two outcomes of interest, which were either mutually exclusive or one was a subset of the other. Therefore the settings reflect a narrow range of multivariate meta-analysis applications.

The standard errors were calculated without accounting for the uncertainty in the estimation of the between-study variances and covariances. This provided narrow confidence intervals which could be adjusted in practice using the Hartung-Knapp correction (Hartung & Knapp 2001*a*,*b*, Knapp & Hartung 2003). An additional limitation of this chapter is it is unclear whether or not the results are directly applicable to other settings, such as continuous or survival outcomes, and this requires further work to explore these findings in data with these outcomes.

The findings from this chapter should be investigated in different data. The BoS

statistic should be compared with differences in univariate and multivariate results from meta-analyses with greater than two outcomes. Additionally, further research should be undertaken to determine whether the between-study correlation is associated with the magnitude of BoS in data with more power to detect an association.

The BoS appeared to identify differences between the univariate and the multivariate results in the majority of meta-analyses. However, there were some studies where the BoS was found to be quite small and the differences between the univariate and the multivariate were great. This was especially noted for meta-analyses when the summary estimates differed between the univariate and the multivariate results, but the width of the confidence intervals did not differ. However, there was little evidence to suggest that BoS increases when there are larger differences in the summary estimates between the univariate and multivariate meta-analyses. There needs to be further work done to determine whether there exists a statistic/measure that is associated with changes in the summary estimate, not just precision.

Next steps

The next chapter investigates whether the BoS statistic can be predicted from meta-analysis level characteristics. Multivariable models are developed to predict the magnitude of BoS in bivariate meta-analysis.

Chapter 4

Development of a multivariable model to predict the magnitude of Borrowing of Strength in bivariate meta-analysis

4.1 Background and aims

In Chapter 3, the differences between the univariate and multivariate results from 43 bivariate meta-analyses were compared alongside their respective Borrowing of Strength (BoS) statistics. In the majority of the meta-analyses, the univariate and multivariate results were the same. However, in some examples statistically and/or potentially clinically important differences between the results were observed. Furthermore, the BoS statistic was often quite high in these situations.

Following this, the characteristics of the meta-analyses were used to investigate the association between a single characteristic and the magnitude of the BoS statistic. A large proportion of the characteristics were found to be highly associated with the magnitude of the BoS statistic. However, the question remains: when should researchers (e.g. for Cochrane) utilise a multivariate over a univariate approach to meta-analysis? It is not sensible to suggest that multivariate meta-analysis should be used routinely, since generally there is little difference in the results between the two approaches (Trikalinos et al. 2014). However, multivariate meta-analysis should also not be disregarded, in

some situations potentially important differences occur as shown in the previous chapter.

In this chapter, a multivariable model containing multiple predictors (characteristics of the meta-analyses) is developed, with the intention that it will predict the magnitude of BoS in a new meta-analysis dataset. The objective is to identify the most important predictors (from those considered in Chapter 3) of large BoS. These can then be used in combination to help identify situations in the future where multivariate meta-analyses are most likely to be beneficial to Cochrane reviews of binary outcomes over and above univariate meta-analyses. The focus is on predicting BoS as quantified by the BoS statistic proposed by Jackson et al. (2017), which corresponds to the relative reduction in the variance of the summary estimate for multivariate meta-analysis compared to univariate meta-analysis, as defined in the introduction (Section 1.4.8). Therefore, the developed model aims to identify situations where the gain in precision from a multivariate meta-analysis is large compared to the univariate. The methods for the chapter are described in Section 4.2 and the models derived for predicting BoS are described in Section 4.3, followed by illustration of their potential use to Cochrane reviews in Section 4.3.5.

4.2 Methods for the development of BoS prediction models

4.2.1 Data and choice of candidate predictors

To develop the prediction model, this chapter utilises the data from Chapter 3. Recall this data contained a BoS statistic for each outcome from 43 bivariate meta-analyses fitted using a particular estimation method (e.g. restricted maximum likelihood (REML)), as well as all the candidate predictors. Thus, there were 86 observations (two BoS statistics from each of the 43 meta-analyses) and these were used as the

response data in the linear regression models that included predictors of BoS as covariates. Separate models were considered for predicting BoS from bivariate fixed-effect meta-analysis estimated using ordinary maximum likelihood (ML) and from bivariate random-effects meta-analysis estimated using REML. Note, in this chapter the method of moments (MM) estimation method is not considered, since the univariable regression results were similar. Furthermore, REML is generally the preferred estimation method (Viechtbauer 2005, Novianti et al. 2014, Langan et al. 2019).

The following were the characteristics of bivariate meta-analyses that were investigated as the candidate predictors for BoS:

- the percentage of studies with missing data for the outcome of interest,
- the percentage of studies with missing data across both outcomes,
- the number of studies in the meta-analysis,
- the number of studies with the outcome of interest,
- the number of studies with both outcomes.
- the maximum absolute within-study correlation between the effect sizes for both outcomes,
- the average absolute within-study correlation between the effect sizes for both outcomes.

In addition, just for predicting BoS from a bivariate random-effects meta-analysis the following characteristics were, also, considered as candidate predictors:

- the absolute between-study correlation between the true effect sizes across trials for both outcomes,
- White's I^2 statistics, as derived for each outcome using the approach of White (2011),

- Jackson's I^2 statistics, as derived for each outcome using the approach of Jackson et al. (2012),
- the multivariate between-study variance for the outcome of interest,
- the univariate between-study variance for the outcome of interest.

In total, when the BoS was derived using a fixed-effect bivariate meta-analysis, there were seven candidate predictors and there were 12 candidate predictors when the BoS was derived from the bivariate random-effects meta-analysis (REML). Since there were 86 observations of BoS, there were 12.3 subjects per variable for fixed-effect and 7.2 subjects per variable for random-effects.

4.2.2 Procedure to obtain fitted prediction models for BoS

Full multivariable models

In Chapter 3, univariable multi-level linear regression models were used to investigate the unadjusted association between the candidate predictors (listed in previous section) and the magnitude of BoS. In the univariable models, a random intercept was used to account for the clustering of BoS observations in each bivariate meta-analysis. However, in almost all the univariable models the variability in the intercept was zero, thus the heterogeneity between the BoS observations was very small and unimportant to account for (Tables C.1, C.2 and C.3 in Appendix C.2). Therefore, in this chapter, it is assumed that the 86 BoS observations were independent and multivariable linear regression models were used to further investigate the candidate predictors.

In the first instance, multivariable linear regression models containing all the candidate predictors (forcing them all to be included, regardless of statistical significance) were fitted for the fixed-effect and random-effects BoS statistics. The dependent variable was modelled as BoS, although log BoS was considered to see if the normality of errors was more suitable. However, there was little benefit identified in transforming BoS, and BoS scale is more meaningful for those using the model. The equation for

the multivariable linear regression model for BoS is:

$$\text{BoS}_{ij} = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + ... + \gamma_p X_p \tag{4.1}$$

where γ_0 is the intercept, γ_i are the regression coefficients for the candidate predictors ($i = 1...p$ for p candidate predictors) and X_i are the candidate predictors. Only linear trends were considered for continuous predictors. The performance of the models was summarised using the R^2 statistic, which is the percentage of total variability between the observations (the BoS values) in the data that is described by the fitted model.

Backwards Selected Multivariable Model

For a model to be applied in practice (e.g. by Cochrane reviewers), a parsimonious model is more desirable than a full model. A parsimonious model should contain as few predictors as necessary whilst maintaining explanatory power (i.e. maintaining an acceptable R^2 value). Therefore, in addition to fitting full models, containing all of the candidate predictors, reduced multivariable models were obtained. A backwards selection/elimination procedure was utilised, which only kept covariates in the model if the association with the magnitude of BoS was statistically significant. The backwards selection procedure starts with the full model. The predictor with the largest p-value that is also greater than a predefined value (in this chapter p=0.1 was used) is removed from the model. The resulting model is refitted to the data and the process of identifying and removing the predictor with the largest p-value is repeated. This process is repeated until the model contains only covariates with p-values less than the predefined value ($p \leq 0.1$). The resulting model is the backwards selected multivariable model.

Observed BoS and Predicted BoS plots

The observed BoS values were plotted against the predicted BoS values, so that the prediction could be empirically examined. There were some BoS values that were predicted to be below 0%, since the models were not bounded; the values that were

predicted to be below 0% were truncated at zero.

Model assumptions checking

For all fitted models, the assumption of constant variance of residual errors and the assumption of normally distributed residual variances were checked (Appendix D.1). To check the assumption of constant variance of residual errors the studentised residuals and the fitted values of BoS were plotted against one another in a scatter plot (Figure D.1 in Appendix D.1). For the normality of the residuals, a normal probability plot was utilised with the standardised residuals (Figure D.2 in Appendix D.1).

Exploratory analysis

Further exploratory analysis was conducted to consider select interactions. To minimise the effect of overfitting due to large number of variables, interactions with the absolute between-study correlation and the average absolute within-study correlations were chosen. The methods and results for the exploratory analysis are provided in Appendix D.2).

4.2.3 Shrinkage and optimism

Shrinkage

A prediction model that was developed to fit a sample dataset will predict well for that sample data (Copas 1983, Steyerberg 2008). As a consequence, it is known that prediction models are often overly optimistic when fitted to a new dataset. Thus, the model predictions in this new dataset are unlikely to be as accurate. To improve the predictions from the developed models in this chapter a uniform shrinkage method was applied, which calculates a shrinkage factor and applies it to the estimated coefficients, $\hat{\gamma}_i$, of the included predictors, X_i, of the prediction model. The shrinkage factor can be calculated between zero and one, and thus the shrinkage factor shrinks the coefficients towards zero.

The uniform shrinkage factor was calculated and applied using the following steps:

1. Sample with replacement the observations from the original dataset, to generate a bootstrap sample with the same number of observations as the original data.

2. Develop a new model in the bootstrap data using the same model development strategy as the original model (e.g. linear regression, backwards selection, see Section 4.2.2)

3. Using the newly developed model (from step two), predict the BoS statistic for the original data.

4. Estimate the calibration slope, by modelling the observed BoS as a function of the predicted BoS statistic:

$$\text{observed BoS} = \alpha + \phi * \text{predicted BoS} \tag{4.2}$$

   where ϕ is the calibration slope.

5. Repeat steps one to four, 1000 times, to obtain 1000 estimates of the calibration slope, ϕ.

6. Calculate the average calibration slope, which is the uniform shrinkage factor, Φ.

7. Adjust the original model (i.e. as developed in the original data) using the shrinkage factor by multiplying the coefficients in the model by the uniform shrinkage factor, Φ.

8. Re-estimate the intercept whilst holding fixed the revised predictor effects, to provide the shrinkage adjusted model.

Optimism

The R^2 measure of model performance is often overly optimistic when the model is fitted to a new dataset (Steyerberg 2008). The percentage of variability the model

explains will likely be lower in a new dataset (compared to the dataset used to develop the model). For a more realistic R^2, the optimism can be calculated and the R^2 can be adjusted by the optimism.

The optimistic-adjusted R^2 was obtained through a bootstrap procedure, as follows:

1. The original data was sampled with replacement to generate a bootstrap sample which contained the same number of observations as the original data.

2. Develop a new model in the bootstrap sample, using the same model development strategy as the original model (e.g. linear regression, backwards selection, see Section 4.2.2)

3. Calculate the R^2 of the developed model in the bootstrap data. This is denoted as the Bootstrap R^2.

4. Using the original data fit the newly developed model (from step two), and calculate the R^2 (denoted as the Tested R^2).

5. Calculate the optimism as the difference between the R^2 from the bootstrap data (Bootstrap R^2) and the R^2 from the original data (Tested R^2).

6. Repeat steps one to five 1000 times, for 1000 Bootstrap R^2, 1000 Tested R^2 values and 1000 optimisms (the difference between the Bootstrap R^2 and the Tested R^2.

7. Calculate the average optimism.

8. Take the average optimism (calculated in the previous step) away from the apparent R^2 (the R^2 from the original model fitted to the original data) to obtain the optimistic-adjusted R^2.

4.2.4 Application of the developed model to a selection of examples

Cochrane reviews

The final models were applied to two Cochrane reviews, which were not included in the Trikalinos et al. (2013, 2014) dataset, but are similar in setting. The reviews were both published after the review by Trikalinos et al. (2013, 2014). The reviews were chosen using the same inclusion and exclusion criteria as the Trikalinos et al. (2013, 2014) review (Section 3.2.1). Additionally, reviews were excluded if they were included in the Trikalinos et al. (2013, 2014) review (even if they have since been updated). The Cochrane database was searched for published Cochrane reviews starting from Issue 5, 2017 and working backwards in time. The search was stopped when two Cochrane reviews which had two correlated outcomes and satisfied the inclusion criteria were identified.

The two correlated outcomes were chosen if they were either mutually exclusive or one outcome was a subset of the other. The reviews were required to contain at least seven studies with both outcome with at least 10 participants and at least two events in each treatment arm.

Further examples

In addition to applying the model to Cochrane reviews, further examples were chosen outside the context of Cochrane reviews that were used to derive the prediction models. These examples provide settings for us to study the performance of the models outside of the setting they were developed in.

The real studies from Chapter 1 that were considered were the 10 hypertension trials (Riley et al. 2015, Wang et al. 2005), the Fibrinogen study (Jackson et al. 2013) and the p53 study (Tandon et al. 2010, Jackson et al. 2011). There were four outcomes of

interest in the hypertension study, but only two outcomes of interest in the Fibrinogen and the p53 studies.

4.3 Results

4.3.1 Fitted models before adjustment for overfitting

The fitted full models and the backwards selected models for the BoS observations derived from the fixed-effect and random-effects bivariate meta-analyses are provided in Table 4.1. The most important statistically significant predictors in all the models were the maximum absolute within-study correlation, the percentage of missing data for the outcome of interest, and the number of studies in the meta-analysis. For example, from the full model for BoS estimated using the random-effects bivariate meta-analyses and after adjustment for other variables, there was an increase of 42.06% (95% C.I.: 21.90% to 62.21%, $p<0.001$) in BoS for an increase of one in the maximum absolute within-study correlation. A one percent increase in missing data for the outcome of interest resulted in a 0.47% (95% C.I.: 0.24% to 0.69%, $p<0.001$) increase in BoS. For the number of studies in a meta-analysis, an increase of one study increased BoS by 0.67% (95% C.I.: 0.19% to 1.14%, $p=0.007$).

Table 4.1: Table of multivariable results from the fitted full models and backwards selected models for BoS observations derived using the fixed-effect and random-effects meta-analyses

Covariate	Coef.	S.E.	p-value	95% C.I.
Fixed-effect Full Model: Apparent R^2= 0.58				
Number of studies	0.7012	0.2726	0.010	0.1669 to 1.2355
Percentage missing across both outcomes	-0.1211	0.1936	0.532	-0.5006 to 0.2584
Percentage missing for outcome of interest	0.3523	0.1253	0.005	0.1067 to 0.5978
Number of studies with outcome of interest	-0.6908	0.3335	0.038	-1.3445 to -0.0371
Number of studies with both outcomes	-0.2312	0.4631	0.618	-1.1389 to 0.6766
Average absolute within-study correlation	-9.9798	11.0829	0.368	-31.7019 to 11.7422
Maximum absolute within-study correlation	35.2347	10.8817	0.001	13.9070 to 56.5625
Intercept	-13.9029	4.4594	0.002	-22.6430 to -5.1627
Random-effects Full Model: Apparent R^2= 0.65				
Number of studies	0.6660	0.2390	0.007	0.1896 to 1.1425
Percentage missing across both outcomes	-0.2110	0.1768	0.237	-0.5634 to 0.1413
Percentage missing for outcome of interest	0.4662	0.1140	<0.001	0.2390 to 0.6934

Continued on next page

Table 4.1: continued

Covariate	Coef.	S.E.	p-value	95% C.I.
Number of studies with outcome of interest	-0.3074	0.2936	0.299	-0.8926 to 0.2778
Number of studies with both outcomes	-0.8307	0.4087	0.046	-1.6453 to -0.0160
Average absolute within-study correlation	-19.0690	10.2991	0.068	-39.5952 to 1.4571
Maximum absolute within-study correlation	42.0569	10.1124	<0.001	21.9029 to 62.2108
Multivariate between-study variance	12.4340	13.1972	0.349	-13.8680 to 38.7359
Univariate between-study variance	0.7774	12.7070	0.951	-24.5477 to 26.1024
White's I^2	-0.1282	0.0708	0.074	-0.2692 to 0.0128
Jackson's I^2	0.1079	0.0858	0.212	-0.0631 to 0.2789
Absolute between-study correlation	14.1154	7.3651	0.059	-0.5631 to 28.7940
Intercept	-25.3138	8.7458	0.005	-42.7441 to -7.8835
Fixed-effect Backwards selected Model: Apparent R^2= 0.57				
Number of studies	0.6546	0.2339	0.006	0.1892 to 1.1200
Percentage missing for outcome of interest	0.2936	0.0958	0.003	0.2029 to 0.4842
Number of studies with outcome of interest	-0.7852	0.2994	0.010	-1.3809 to -0.1894
Maximum absolute within-study correlation	26.5895	5.0580	<0.001	16.5257 to 36.6533

Continued on next page

Table 4.1: continued

Covariate	Coef.	S.E.	p-value	95% C.I.
Intercept	-14.0537	4.4186	0.002	-22.8453 to -5.2621
Random-effects Backwards selected Model: Apparent R^2= 0.64				
Number of studies	0.4774	0.1909	0.014	0.0974 to 0.8573
Percentage missing across both outcomes	-0.2738	0.1486	0.069	-0.5697 to 0.0220
Percentage missing for outcome of interest	0.5571	0.0708	<0.001	0.4161 to 0.6981
Number of studies with both outcomes	-0.9075	0.3824	0.020	-1.6687 to -0.1463
Average absolute within-study correlation	-19.1021	9.8098	0.055	-38.6279 to 0.4238
Maximum absolute within-study correlation	39.2969	9.6317	<0.001	20.1255 to 58.4684
Intercept	-9.3278	3.9471	0.021	-17.1843 to -1.4713

The apparent R^2 values for the models range from 0.57 to 0.65 (Table 4.1), thus a large proportion of the variation in the set of observations is accounted for by the models. The apparent R^2 values were similar between the full model and the backwards selected model for each meta-analysis estimation method. For example, for the random-effects meta-analysis, the R^2 for the full model and backwards selected model was 0.65 and 0.64, respectively. The full and backwards selected models for the random-effects meta-analyses accounted for a greater quantity of variability (apparent R^2: 0.65 and 0.64, respectively) than the models for the fixed-effect meta-analyses (apparent R^2: 0.58 and 0.57, respectively).

From Figure 4.1, which displays the plots of the observed BoS against the predicted BoS, it was observed that generally there was good agreement. This is expected, as the models are fitted to this data, and so calibration will be perfect (i.e. calibration slope of 1). The predicted BoS statistic increased as the observed BoS statistics increased. Although in each graph there are many observations that lie on or close to the line that represents perfect prediction, there are many that are at a distance from the line.

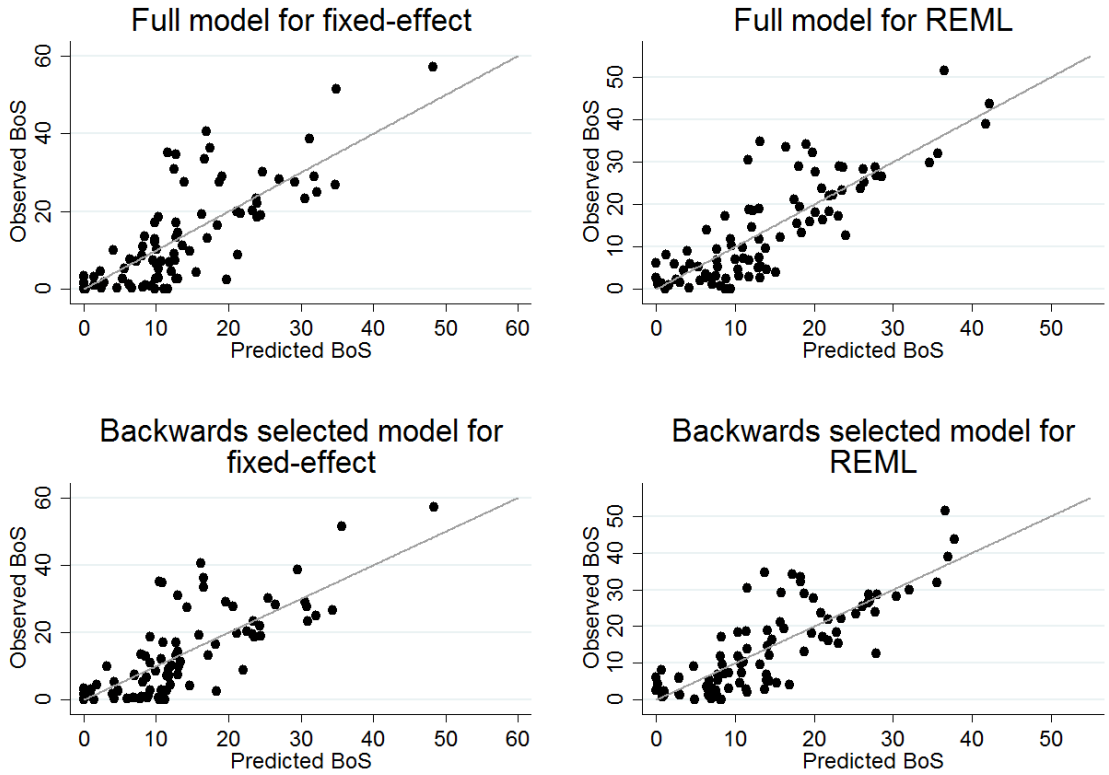


Figure 4.1: Scatter plots of the observed BoS against the predicted BoS from the full and the backwards selected models

4.3.2 Shrinkage

All the models in Table 4.1 can be re-written in the form of a mathematical equation:

$$\text{BoS} = \alpha + \text{LP} \tag{4.3}$$

where α is the intercept and LP is the linear predictor for the model. For example, for the backwards selected model for the fixed-effect meta-analyses, the intercept (α) was -14.05 and the linear predictor (LP) was:

$$\begin{aligned}\text{LP} =& 0.655 \times \text{number of studies} + 0.294 \times \text{percentage of missing data for outcome of interest} \\ & - 0.785 \times \text{number of studies with outcome of interest} \\ & + 26.590 \times \text{maximum absolute within-study correlation}\end{aligned} \tag{4.4}$$

For each model, a shrinkage factor was calculated as detailed in Section 4.2.3. The models adjusted by the shrinkage factor can be mathematically written as:

$$\text{BoS} = \beta + \Phi \times \text{LP} \tag{4.5}$$

where LP is as above , β is the new calculated intercept and Φ is the shrinkage factor.

For every model, the shrinkage factor was close to one (Table 4.2). Therefore, the coefficients for each model were only shrunk slightly towards zero. The impact on the predictions can be seen in Figure 4.2. For each model, the shrinkage factor shrinks prediction towards the mean BoS statistic. Therefore, greater differences between the predictions from the models before shrinkage and after shrinkage are observed when BoS is predicted to be large (i.e. further from the mean of BoS).

Table 4.2: Table of shrinkage factors for each model

Model	Meta-analysis approach	Shrinkage factor	Revised intercept
Full Model	Fixed-effect	0.95	-12.52
	Random-effects	0.92	-22.22
Backwards Selected Model	Fixed-effect	0.96	-13.02
	Random-effects	0.93	-7.56

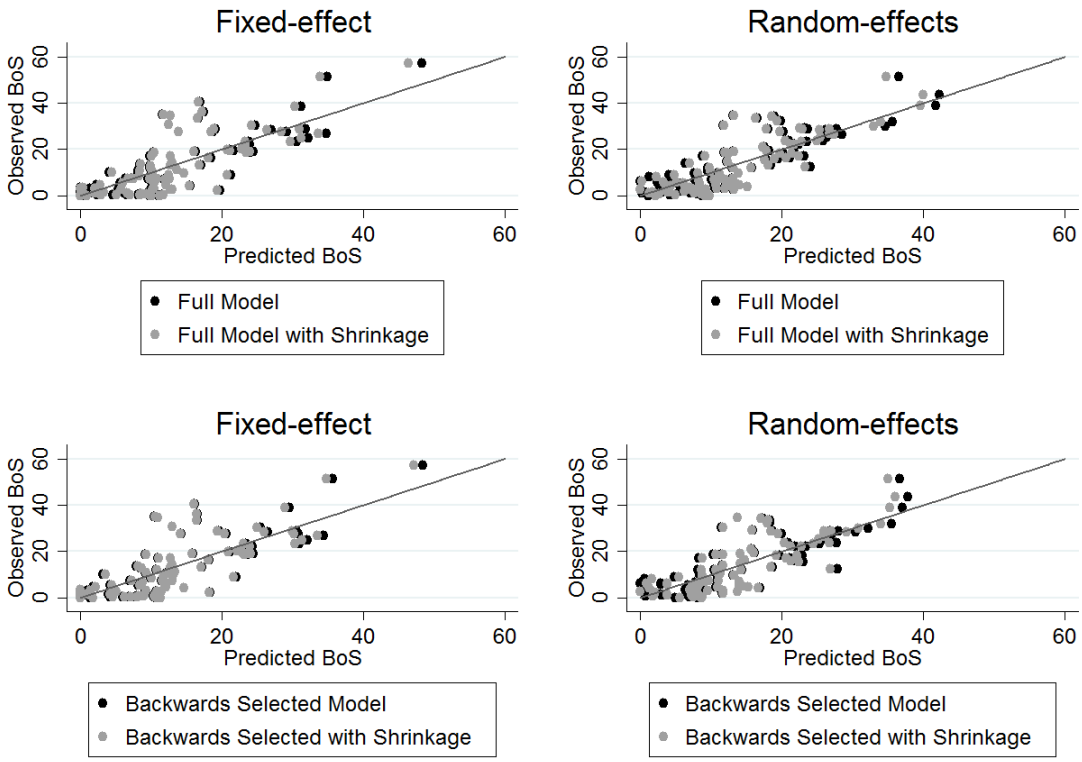


Figure 4.2: Scatter plots of the observed BoS against the predicted BoS from the models without and with shrinkage

4.3.3 Description of the final models after shrinkage

As there was little difference between the full and backwards selection models, the latter was chosen as the final models. The final models, derived after backwards selection and shrinkage, are shown for the fixed-effect and REML random-effects models below.

Final model for fixed-effect BoS

The final model for BoS from fixed-effect meta-analyses was:

$$\begin{aligned}\hat{\text{BoS}} = -13.020 + 0.962 * \Big(&0.655 * \text{number of studies} \\ &+ 0.294 * \text{percentage of missing data for outcome of interest} \\ &- 0.785 * \text{number of studies with outcome of interest} \\ &+ 26.590 * \text{maximum absolute within-study correlation}\Big)\end{aligned} \tag{4.6}$$

Final model for random-effects BoS

The final model for BoS from REML random-effects meta-analyses was:

$$\begin{aligned}\hat{\text{BoS}} = -7.561 + 0.925 * \Big(&0.477 * \text{number of studies} \\ &- 0.274 * \text{percentage of missing data across all outcomes} \\ &+ 0.557 * \text{percentage of missing data for outcome of interest} \\ &- 0.907 * \text{number of studies with both outcomes} \\ &- 19.102 * \text{average absolute within-study correlation} \\ &+ 39.297 * \text{maximum absolute within-study correlation}\Big)\end{aligned} \tag{4.7}$$

4.3.4 Optimism in R^2 of the developed models

In this section, the optimism and optimism adjusted R^2 for the models are presented (Table 4.3). The optimism corrected R^2's were smaller than the apparent R^2's, as was expected.

Table 4.3: The apparent, bootstrap, tested and optimism corrected R^2, and the mean calculated optimism for each model that had shrinkage and optimism applied

Models	Apparent R^2	Mean Bootstrap R^2	Mean Tested R^2	Mean Calculated Optimism	Optimism corrected R^2
Fixed-effect Full	0.581	0.610	0.527	0.083	0.498
Random-effects Full	0.678	0.716	0.613	0.103	0.565
Fixed-effect Final	0.574	0.599	0.521	0.078	0.496
Random-effects Final	0.641	0.700	0.593	0.107	0.534

The variability explained by the models for the fixed-effect meta-analyses were very similar between the full model and the final model, 0.498 and 0.496, respectively. The optimism corrected R^2's were large for the models for the random-effects meta-analyses compared to the those for the fixed-effect meta-analyses. For example, the optimism corrected R^2's from the final models for the fixed-effect meta-analyses and the random-effects meta-analyses were 0.496 and 0.534, respectively. The larger optimism is due to the larger number of predictors considered for the prediction model in the random-effects setting.

4.3.5 Application of the prediction models to examples

The predicted BoS from our models was compared in other datasets from the same setting (i.e. binary outcomes from Cochrane reviews), using the Buzzetti et al. (2017) and Feinberg et al. (2017) reviews.

Cochrane Examples

The Buzzetti et al. (2017) review and the Feinberg et al. (2017) review both contained an is a subset of relationship between the two outcomes. The predicted BoS statistics were predicted in these examples using the final models for the fixed-effect and the random-effects meta-analyses. The setting for the Buzzetti et al review was the

use of Glucocorticosteriods for alcoholic hepatitis compared with no intervention. The two correlated outcomes were mortality at maximal follow-up and mortality at 30 days. Feinberg et al. (2017) reviewed randomised control trials (RCTs) that compared experimental nutrition support with a control for disease-related malnutrition in Intensive Care Unit participants including trauma characterised as "at nutritional risk." The correlated outcomes of interest are all cause mortality at the end of intervention and at maximum follow-up.

Table 4.4: True and predicted BoS values for two Cochrane reviews, predicted using the final prediction models Equation 4.6 for the fixed-effect meta-analyses and Equation 4.7 for the random-effects meta-analyses. The results from their multivariate and univariate meta-analysis results are presented alongside.

Review	Outcome	Number of studies	Observed BoS	Predicted BoS	Multivariate			Univariate		
					Trt eff[a]	s.e.[b]	95% C.I.	Trt eff[a]	s.e.[b]	95% C.I.
Fixed-effect				from model 4.6						
Buzzetti et al.	1	12	3.2	10.8	-0.117	0.127	-0.365 to 0.131	-0.099	0.129	-0.351 to 0.153
(2017)	2	9	47.7	20.1	-0.118	0.131	-0.375 to 0.138	-0.375	0.181	-0.729 to -0.021
Feinberg et al.	1	11	32.8	21.0	-0.032	0.091	-0.210 to 0.145	-0.031	0.110	-0.247 to 0.186
(2017)	2	15	0.2	10.4	-0.042	0.081	-0.201 to 0.118	-0.041	0.081	-0.201 to 0.119
Random-effects				from model 4.7						
Buzzetti et al.	1	12	0.8	7.7	-0.433	0.293	-1.007 to 0.142	-0.459	0.322	-1.091 to 0.173
(2017)	2	9	29.4	20.6	-0.429	0.287	-0.991 to 0.133	-0.527	0.287	-1.089 to 0.034
Feinberg et al.	1	11	32.8	22.3	-0.032	0.091	-0.210 to 0.145	-0.100	0.158	-0.409 to 0.210
(2017)	2	15	0.2	8.5	-0.042	0.081	-0.201 to 0.118	-0.041	0.081	-0.201 to 0.119

[a] Treatment effect

[b] Standard error

The predictions of BoS, using the final models (Equation 4.6 for the fixed-effect meta-analyses and Equation 4.7 for the random-effects meta-analyses), were not perfect, however the outcomes with small values of BoS were predicted small values, although not as small (Table 4.4). For example, from Buzzetti et al. (2017) for outcome one from the random-effects, the observed BoS was 0.2%, and the predicted was 10.4%. Similarly for large values of BoS, the predictions were large, although not as large. For example, for Buzzetti et al. (2017) review, from the random-effects meta-analyses, the observed BoS for outcome two was 29.4%, and the predicted BoS was 20.6%. Nevertheless, they are of a reasonably similar magnitude.

For Buzzetti et al. (2017) and Feinberg et al. (2017) reviews the results were similar between the univariate and multivariate fixed and random-effects meta-analyses in the majority of comparisons. For example, from the Feinberg et al. (2017) review, outcome one from the fixed-effect meta-analysis the multivariate result was -0.032(95% C.I.: -0.210 to 0.145) and -0.031 (95% C.I.: -0.201 to 0.119) from the univariate, and the observed and predicted BoS were 32.8% and 21.0%, respectively. There was one example that contained a change in statistical significance between the univariate and multivariate result; from the Buzzetti et al. (2017) review for outcome two from the fixed-effect meta-analyses the results from the multivariate and univariate were -0.118 (95% C.I.: -0.375 to 0.138) and -0.375 (95%: -0.729 to -0.021), respectively. Additionally, the BoS statistic was large at 47.7% and was predicted to be 20.1%. There was a potentially important clinical difference between the univariate and multivariate results for outcome one from Feinberg et al. (2017) review from the random-effects meta-analysis. The univariate meta-analysis results were -0.100 (95% C.I.: -0.409 to 0.210) and the multivariate results were -0.032 (95% C.I.: -0.210 to 0.145). The observed and predicted BoS was 32.8% and 22.3%, respectively.

Further examples of the application of the prediction models

The predictions of BoS from the final models (Equation 4.6 and 4.7) were also checked in completely different settings (i.e. non binary outcomes not in Cochrane reviews). The values of the predicted BoS for the fixed-effect were very similar for each outcome in the hypertension, ranging from 5.7% to 5.9%. However, the true BoS values had a greater spread of values from 1.3% to 11.3% (Table 4.5). Similar results can be seen for the REML random-effects estimation method. Nevertheless, the predicted BoS values were a similar magnitude to the observed values.

Table 4.5: A table of the observed and predicted BoS values from real examples

Study	Outcome	Observed BoS	Predicted BoS
Fixed-effect			from model 4.6
Hypertension	SBP	5.3	5.9
	DBP	1.3	5.9
	CVD	1.9	5.7
	Stroke	11.3	5.7
Fibrinogen	Fully	57.9	36.8
	Partially	0.1	8.4
p53	Disease-free survival	52.3	26.9
	Overall survival	8.6	10.5
REML Random-effects			from model 4.7
Hypertension	SBP	0.9	13.1
	DBP	0.5	13.3
	CVD	4.2	13.3
	Stroke	20.8	13.5
Fibrinogen	Fully	53.3	34.7
	Partially	1.8	6.5
p53	Disease-free survival	44.7	29.7
	Overall survival	0.3	4.0

The predictions from the models were not accurate for the Fibrinogen and p53 studies. However, the models do predict BoS to be large or small appropriately, although the predictions are not as large as the observed BoS. For example, from p53 study from the fixed-effect meta-analysis the observed BoS was 52.3% whereas the predicted BoS was 26.9%. The predicted BoS statistics, although not as large as the observed, would still help to identify situations where BoS is likely to be large.

4.4 Discussion

The aim of this chapter was to develop prediction models for the BoS statistic for fixed-effect meta-analyses and random-effects meta-analyses, using 86 BoS statistics from 43 bivariate meta-analyses from the Cochrane library. Full models, containing all potential predictors, were developed for the fixed-effect and the random-effects meta-analyses. The characteristics of meta-analysis were included in the prediction models as linear covariates. These models were developed using backwards selection and shrinkage was applied to obtain the final models. The key findings from the prediction model are summarised in Figure 4.3.

Figure 4.3: Key findings

Key findings:

- Two prediction models were developed for the prediction of BoS, one for fixed-effect meta-analyses and one for random-effects meta-analyses.
- The important covariates for predicting the BoS statistic were the within-study correlation, the number of studies with both outcomes and the percentage of missing outcomes either for all the outcomes or the outcome of interest.
- BoS statistic predictions may differ between the models developed for the fixed-effect and the random-effects multivariate meta-analyses.

- Between-study correlation, between-study variance and I^2 statistics were not important to prediction of the magnitude of the BoS.

The backwards selected models with shrinkage applied contained the meta-analysis characteristics most important to the prediction of BoS, and were the final models used in the prediction of BoS examples. For these models the optimism adjusted R^2 was 0.49 and 0.53 for the fixed-effect and the random-effects multivariate meta-analysis BoS prediction models, respectively, which suggests the models explain 49% and 53% of the variability in the observations, respectively. The most important predictors contained in both the final models were the maximum absolute within-study correlation, the number of studies and the percentage of missing data for the outcome of interest. In each of the final models, these predictors were also the most statistically significant in both the multivariable models and the univariable regression models in Chapter 3. The association between borrowing strength and missing data and/or the within-study correlation was discussed in Chapter 3 and discussed in the literature in such papers by Riley et al. (2007*a*), Riley (2009), Jackson et al. (2017), Copas et al. (2018).

These models were applied to two further Cochrane bivariate examples, as well as some examples from Chapter 1, to examine how accurately the final models predicted (Section 4.3.5. For the Cochrane examples, the predictions for BoS were similar to the true BoS. Athough the predictions were not exact, but for very small BoS values the predicted BoS were small and for large BoS the predicted BoS were large (Table 4.4). The Cochrane examples were similar (two correlated binary outcomes) settings to the Trikalinos et al. (2013, 2014) review data the prediction models were developed for. Therefore, the predictions were likely to be similar to the observed.

For the examples from Chapter 1, the BoS predictions for the Hypertension example with four outcomes were similar to each other, however, the true BoS values varied in magnitude. It was unclear with so few examples as to whether this was due to the

number of outcomes in the example that were greater than two, or due to another reason that was unknown. The true BoS statistics for the hypertension study, with four outcomes, did not have a range of values as great for the BoS statistic as the bivariate real examples. The characteristics for each of the four outcomes in the hypertension study were very similar. This resulted in very similar BoS predictions, which could mean that the final models are missing an unknown factor that affects the magnitude of the BoS statistic. For the bivariate real examples, Fibrinogen and p53 studies, the predicted BoS statistics were predicted to be large, similarly the observed BoS were large. It is unclear how accurate the prediction of BoS in further examples would be.

For exploratory purposes, interactions were investigated in further models to determine the importance of the effects of the within-study and between-study correlations (Appendix D.2). The inclusion of interactions with the within-study and between-study correlations did not identify anything of importance.

4.4.1 Limitations and further research

A limitation of this chapter is the number of observations per variable. For the models for the prediction of the BoS from a fixed-effect meta-analysis and from a random-effects meta-analysis, the number of observations per variable was 12.3 and 7.2, respectively. Although Austin & Steyerberg (2015) discuss in their 2015 paper that two subjects per variable is adequate in linear regression for the estimation of coefficients, standard errors and confidence intervals, there is alternative research that suggests that problems can still occur when there are greater than 10 subjects (events) per variable (Courvoisier et al. 2011). Ten events per variable is a rule of thumb that has arisen from two simulation studies and is often cited as best practice (Peduzzi et al. 1995, 1996). The number of observations per variable is a limitation and this chapter has used shrinkage and optimism to adjust the model for overfitting.

The prediction models were developed to be applied in prospective meta-analyses to

predict the expected magnitude of BoS. However, the performance of the model outside of the original sample is unknown and has not been assessed through external validation, an important step in model development (Altman & Royston 2000, Altman et al. 2009, Debray et al. 2015*b*). Furthermore, a limitation of this chapter was there was no large data that the model could be externally validated with. The next step for the models would be to externally validate them with new data across a range of examples. This would involve Cochrane reviews where the within-study correlation is known or can be calculated. Unfortunately as seen by the number of observations in the dataset in this chapter, there are few reviews that report the within-study correlations or provide enough information from which the within-study correlations can be calculated.

The data the models were developed on was a set of meta-analysis studies with only two binary outcomes of interest. It is not known whether these models would be transferable to multivariate meta-analyses with continuous or survival outcomes, or with greater than two outcomes. It was not developed or studied in detail in this chapter as to whether these models could be used to predict BoS statistics in continuous outcomes in studies with more than two outcomes of interest.

The models were very different between the fixed-effect and random-effects meta-analyses. Therefore, predictions from these models for the same outcome were different and the models should not be transferred freely between the estimation methods. In Chapter 3, the true BoS statistics were sometimes different between the estimation methods for the same outcome of interest in the same meta-analysis.

4.4.2 Recommendations

For researchers who want to identify whether a multivariate meta-analysis may be important for borrowing strength, the models developed may be useful to predict the potential magnitude of BoS before data analysis. If BoS is predicted to be 'moderate' or

'large', then this provides more motivation to undertake a multivariate meta-analysis, and results may change from separate univariate meta-analyses. In addition it is recommended that the final model for the random-effects meta-analysis is used for predictions of BoS. This follows from the recommendation that random-effects meta-analysis should be used in preference to a fixed-effect meta-analysis.

Next steps

In the next chapter the use of interactive graphs developed in R shiny for the presentation of the prediction models is explored. Furthermore, the relationship between the meta-analysis level characteristics and the magnitude of BoS are also explored through an interactive tool developed from the equation for the BoS statistic.

Chapter 5

Development of interactive graphical tools for the prediction and exploration of the BoS statistic

5.1 Introduction

The Borrowing of Strength (BoS) statistic is calculated following a multivariate meta-analysis from the variances of the outcome of interest from the multivariate and univariate meta-analyses. For the BoS statistic to be used as a decision making tool to determine when a multivariate meta-analysis is most beneficial, its magnitude needs to be predicted in advance of any analysis. In Chapter 4, prediction models were developed for fixed-effect and random-effects bivariate meta-analyses from a dataset of 43 Cochrane reviews with two binary outcomes.

The use of interactive graphical tools is becoming increasingly used for presenting data and results, including clinical prediction models (Bonnett et al. 2019). An advantage of interactive graphical tools for clinical prediction models is that these can be embedded into websites, which can be made accessible to researchers. This allows patients and clinicians to predict and visually explore these predictions for themselves. Advantages of interactive graphical tools include; ease of use (particularly for complex prediction models), full models can be embedded in the background for the predication and thereby avoid approximations.

Additionally, the development of interactive graphs can be a useful tool for methodology research. Interactive graphical tools allow data and/or relationships to be explored visually in a reactive real time environment.

The relationship between the BoS statistic and some meta-analysis characteristics have previously been discussed in the literature (Jackson et al. 2017, Copas et al. 2018). Jackson et al. (2017) explored the relationship between BoS and the within-study correlation. Copas et al. (2018) reviewed the impact that additional studies have on the contribution made by the original studies in the meta-analysis on the BoS statistic. However the relationship between BoS and meta-analysis characteristics has yet to be explored through interactive tools.

The aim of this chapter is to present the prediction models from Chapter 4 on interactive graphs and embed these on a website accessible to the reader. A secondary aim of this chapter is to use interactive graphs for an exploratory purpose, to observe the relationship between meta-analysis level characteristics and the magnitude of BoS, from its mathematical calculation.

This chapter follows by firstly developing and presenting interactive graphical tools produced using R shiny for the prediction of the BoS statistic for the fixed-effect and random-effects bivariate meta-analyses. The formula for the BoS statistic is programmed into an interactive graphical tool to visually explore the relationship between the magnitude of BoS and characteristics of meta-analyses. This tool will allow researchers without a mathematical background to visualise BoS from its formula and gain a greater understanding of how the magnitude of BoS changes according to characteristics of meta-analyses.

5.2 Interactive tools for the prediction of BoS from developed prediction models

In this section interactive graphs are developed following the development of prediction models in Chapter 4. In Chapter 4, two prediction models were developed, one for bivariate fixed-effect meta-analyses and one for bivariate random-effect meta-analyses.

There are two advantages with displaying the developed prediction models on interactive graphs. Firstly, it allows researcher a tool to easily and visually predict the magnitude of BoS in their setting. Secondly, it allows for the predictions to be visually explored across a range of different percentages of missing data for the outcome of interest.

5.2.1 Method to develop interactive tools for the prediction of the BoS

In this section, the interactive graphical tools for the prediction of the BoS statistic for a bivariate meta-analysis are presented. Two graphical tools were developed; one for fixed-effect meta-analyses and one for random-effects meta-analyses. The interactive graphs were developed in R shiny and the code for the two graphs can be found in Appendices E.2 and E.3.

Recall from Chapter 4, the prediction model for the BoS for a bivariate fixed-effect meta-analyses was (Equation 4.6):

$$\begin{aligned}\hat{\text{BoS}} = -13.020 + 0.962 * \Big(&0.655 * \text{number of studies} \\ &+ 0.294 * \text{percentage of missing data for outcome of interest} \\ &- 0.785 * \text{number of studies with outcome of interest} \\ &+ 26.590 * \text{maximum absolute within-study correlation}\Big)\end{aligned}$$

The prediction model for the BoS for a bivariate random-effects meta-analysis was (Equation 4.7):

$$\begin{aligned}\hat{\text{BoS}} = -7.561 + 0.925 * \Big(&0.477 * \text{number of studies} \\ &- 0.274 * \text{percentage of missing data across all outcomes} \\ &+ 0.557 * \text{percentage of missing data for outcome of interest} \\ &- 0.907 * \text{number of studies with both outcomes} \\ &- 19.102 * \text{average absolute within-study correlation} \\ &+ 39.297 * \text{maximum absolute within-study correlation}\Big)\end{aligned}$$

The interactive graphical tools were developed such that the BoS statistic (y axis) was provided against the percentage of missing data for the outcome of interest (x axis). For the prediction of the BoS statistic for a fixed-effect meta-analysis, the required predictors were the number of studies expected to be included in the meta-analysis and the expected maximum absolute within-study correlation. Both the values for the number of studies and the maximum absolute within-study correlation were able to be input into the interactive graph using sliders. The percentage of missing data for the outcome of interest is the x axis of the interactive graph and the number of studies with the outcome of interest can be obtained from the number of studies included in the meta-analysis and the percentage of missing data for the outcome of interest.

The prediction model was plotted by calculating the co-ordinates where the value of the BoS statistic was 0% and 100%, and the coordinates where the percentage of missing data for the outcome of interest was 0% and 100%.

There were more predictors included in the prediction model for the BoS statistic for the random-effects meta-analysis. As well as the number of studies expected to be included in the meta-analysis and the maximum absolute within-study correlation, the average absolute within-study correlation and the percentage of missing data for the

alternative outcome were important predictors of the BoS statistic and required for the interactive tool. These values were able to be input into the interactive graphical tool using sliders. The remaining predictors, percentage of missing data across all outcomes and the number of studies with both outcomes can be obtained from the number of studies, the percentage of missing data for the outcome of interest and the percentage of missing data for the alternative outcome.

Bivariate fixed-effect meta-analysis BoS prediction model

The only predictors in the prediction model for the BoS statistic for a bivariate fixed-effect are the number of studies in the meta-analysis and the maximum absolute within-study correlation. The interactive graphical tool for the prediction of the BoS for the fixed-effect bivariate meta-analysis is shown below (Figure 5.1):

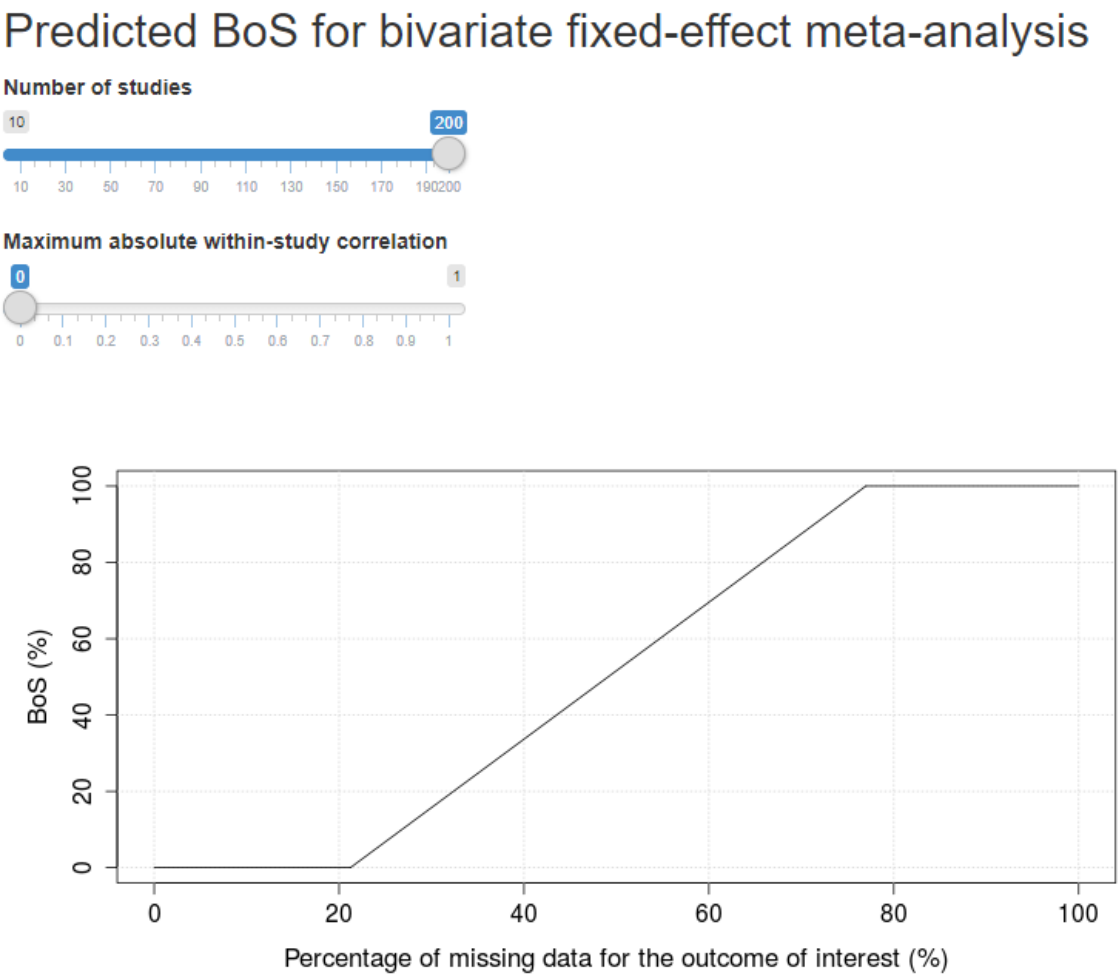


Figure 5.1: A single frame of the interactive graph for the prediction model for the prediction of BoS for a fixed-effect meta-analysis accessed from `https://mhattle.shinyapps.io/PredictingFEBoS/`

Bivariate random-effects meta-analysis BoS prediction model

For the prediction of the BoS for a bivariate random-effects meta-analysis the predictors are the number of studies in the meta-analysis, the percentage of missing data for the alternative outcome, the average absolute within-study correlation and the maximum average within-study correlation. The average absolute within-study correlation provides the minimum value for the maximum absolute within-study correlation. The slider for the maximum absolute within-study is automatically updated given the average absolute within-study correlation. The interactive graphical tool for the prediction of the BoS for the random-effects bivariate meta-analysis is shown below (Figure 5.2):

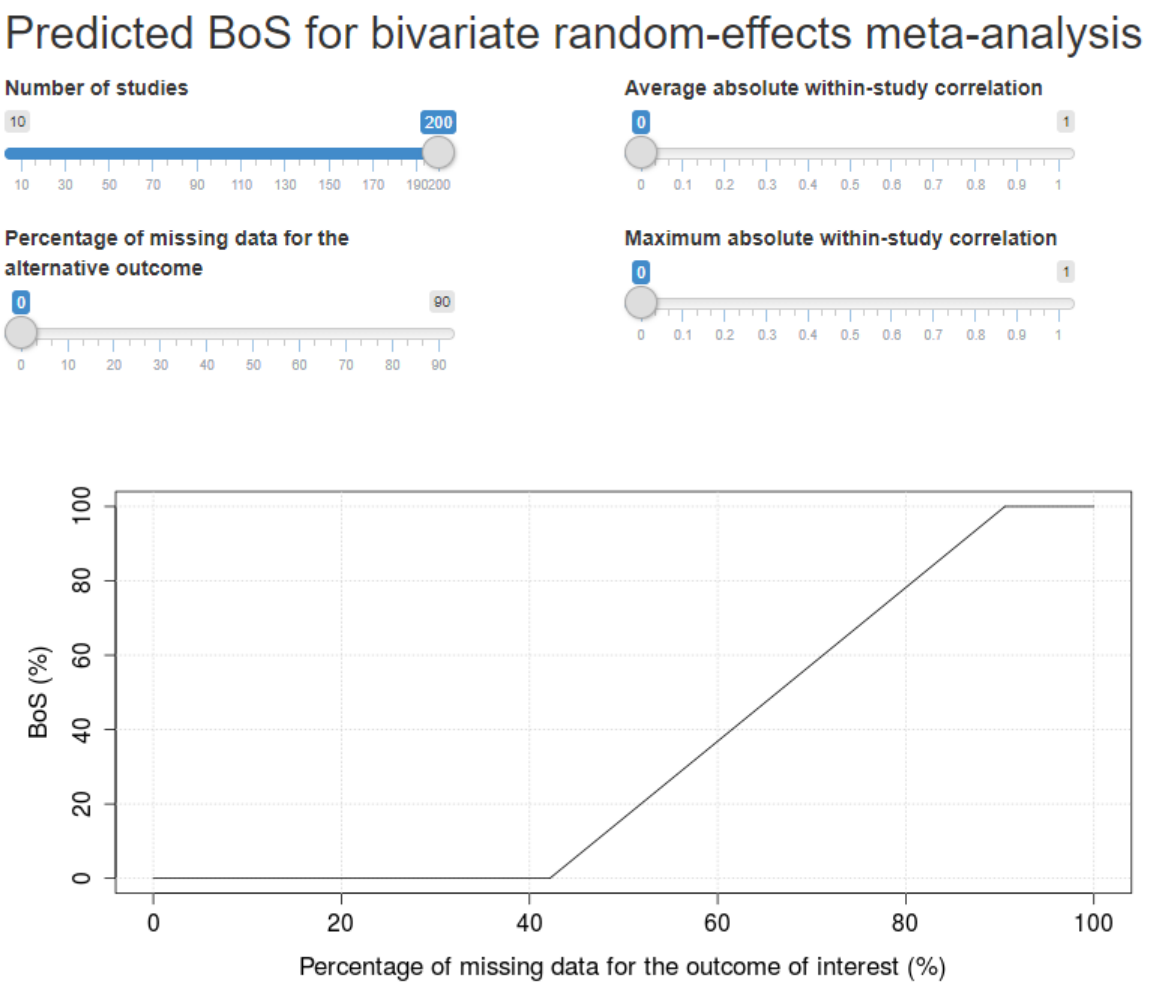


Figure 5.2: A single frame of the interactive graph for the prediction model for the prediction of BoS for a random-effects meta-analysis accessed from `https://mhattle.shinyapps.io/PredictingREBoS/`

5.2.2 Application of the interactive tools to an example

Recall the Buzzetti et al. (2017) review from Chapter 4, a Cochrane review that investigated the use of Glucocorticosteriods for alcoholic hepatitis compared with no intervention. The two binary outcomes had an is a subset of relationship between the two outcomes, which were mortality at maximal follow-up and mortality at 30 days. The BoS statistic was predicted using the interactive graphical tools for the Buzzetti et al. (2017) review in an example in Section 4.3.5. In this section, the Buzzetti et al. (2017) review is used as an example for the interactive graphs for the prediction of BoS.

The Buzzetti et al. (2017) review contained 12 studies and the maximum absolute within-study correlation was 0.99. There was no missing data for the first outcome (mortality at maximal follow-up) and there was 25% (3 studies) missing data for the secondary outcome (mortality at 30 days). The interactive graph for the prediction of the BoS statistic for the bivariate fixed-effect meta-analysis had the sliders set to reflect this information, with the number of studies equal to ten and the maximum absolute within-study correlation 1 (Figure 5.3).

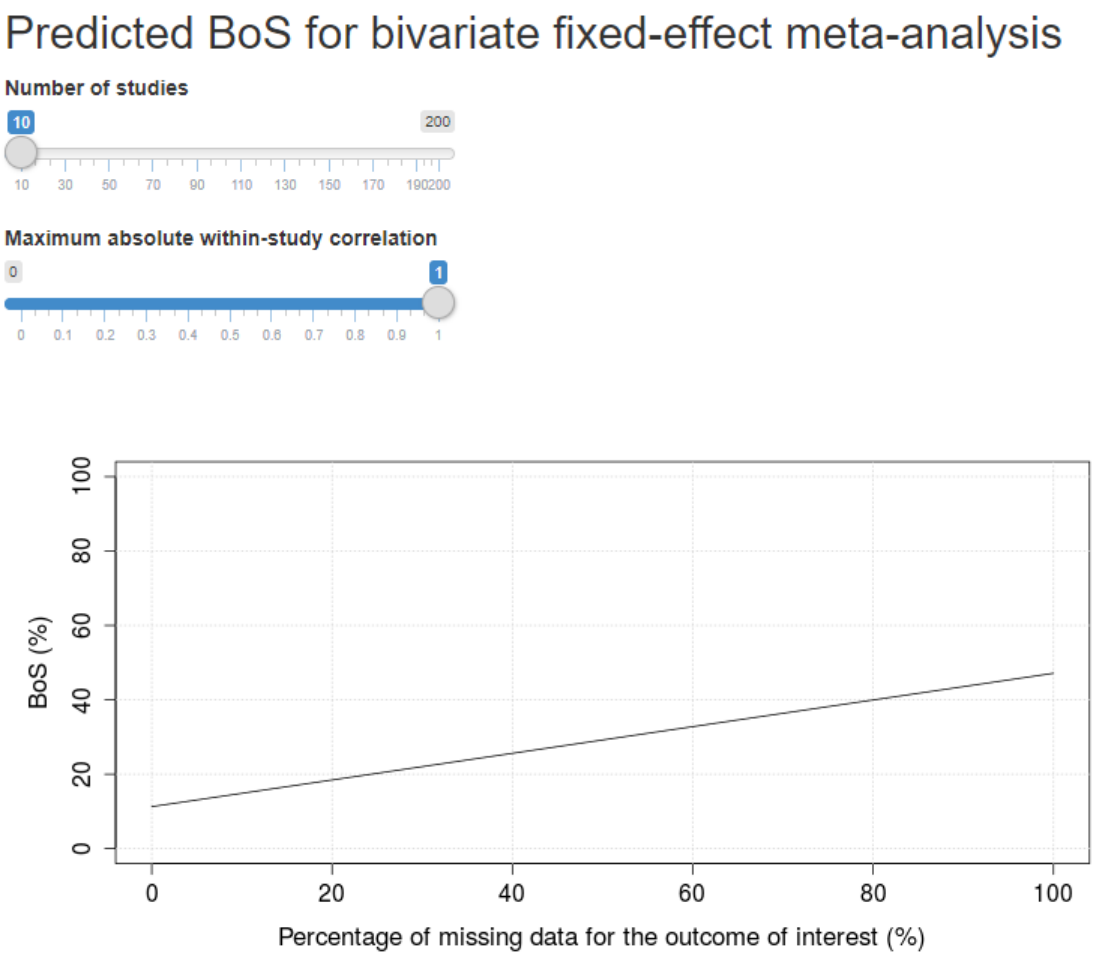


Figure 5.3: A single frame of the interactive graph for the prediction of the Borrowing of Strength (BoS) statistic for a fixed-effect bivariate meta-analysis for both outcomes from the Buzzetti et al. (2017) review example accessed from `https://mhattle.shinyapps.io/PredictingFEBoS/`

The BoS statistic for both outcomes (mortality at maximal follow-up and mortality at 30 days) can be predicted from the same interactive graph. For the mortality at maximal follow-up (outcome one) there was no missing data for the outcome, therefore from the interactive graph the predicted BoS was about 10%. For the mortality at 30 days (outcome two) there was 25% missing data, therefore from the interactive graph the predicted BoS was about 20%. Recall, from Table 4.4 in Chapter 4 the observed BoS statistics were 3.2% and 47.7%, respectively.

The interactive graph for the the prediction of the BoS statistic for the bivariate random-effects meta-analysis required two single frames of the graph to predict the BoS statistics in both outcomes, due to the different values for the percentage of missing data for the alternative outcome. For the mortality at maximal follow-up (outcome one), there was 25% missing data for the alternative outcome. However the slider was developed with increasing increments of 10%, so the percentage of missing data for the alternative outcome was input as 20%, such that the BoS statistic was not overestimated (Figure 5.4). For the remaining meta-analysis level characteristics, the number of studies in the Buzzetti et al. (2017) was 12, the average absolute within-study correlation was 0.866 (3dp) and the maximum absolute within-study correlation was 0.99. Therefore, the values that were input into the interactive graphs for the random-effects meta-analysis were 10 studies, average absolute within-study correlation of 0.85 and a maximum absolute within-study correlation of 1 (Figures 5.4 and 5.5).

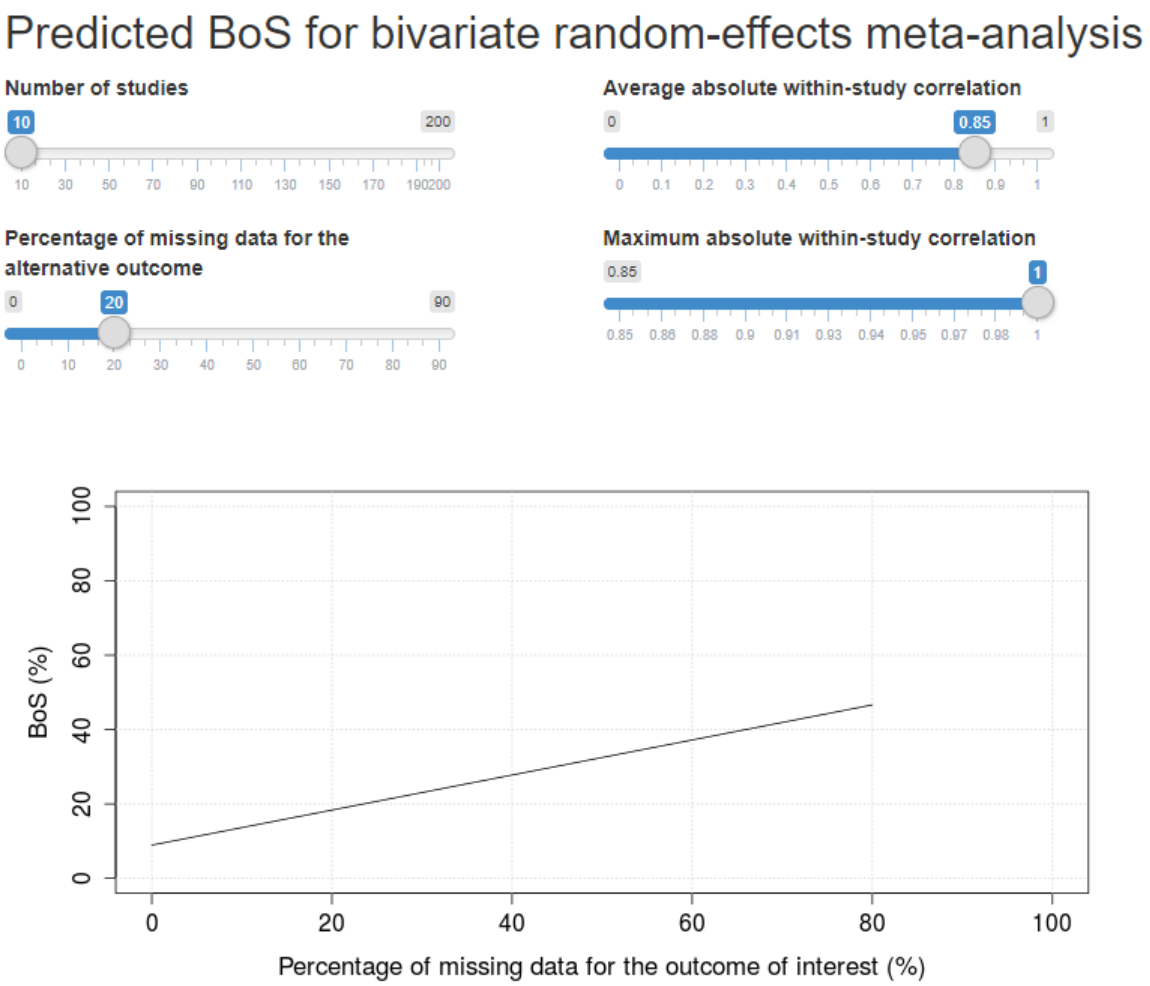


Figure 5.4: A single frame of the interactive graph for the prediction of the Borrowing of Strength (BoS) statistic for a random-effects bivariate meta-analysis for the outcome one, mortality at maximal follow-up, from the Buzzetti et al. (2017) review example accessed from `https://mhattle.shinyapps.io/PredictingREBoS/`

There was no missing data for the outcome, mortality at follow-up (outcome one) and therefore the predicted BoS statistic from Figure 5.4 was just under 10%. For the outcome mortality at 30 days, the predicted BoS from Figure 5.5 was just over 20%. Recall from Table 4.4 in Chapter 4 the observed BoS statistics were 0.8% and 29.4%, respectively.

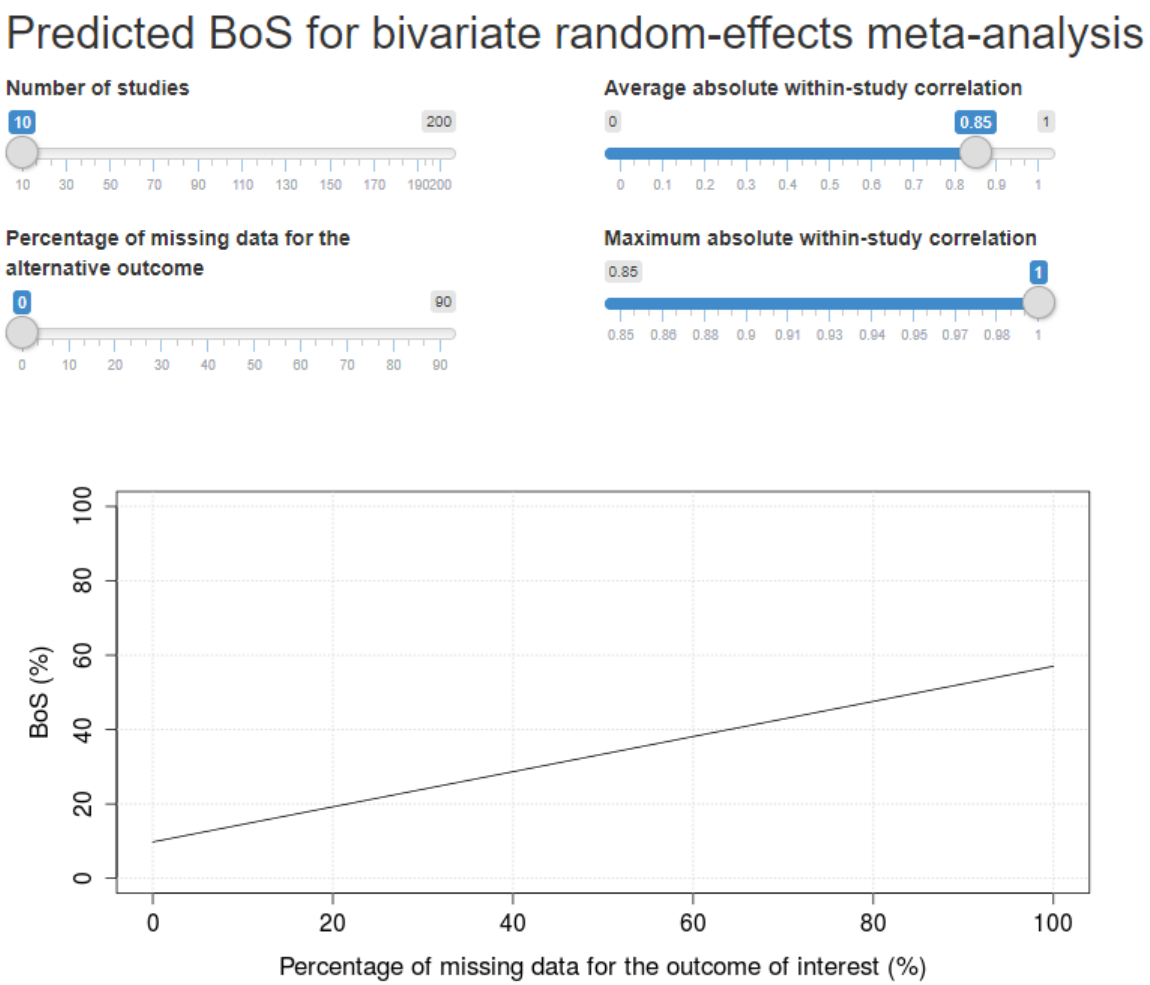


Figure 5.5: A single frame of the interactive graph for the prediction of the Borrowing of Strength (BoS) statistic for a random-effects bivariate meta-analysis for the outcome two, mortality at 30 days follow-up, from the Buzzetti et al. (2017) review example accessed from `https://mhattle.shinyapps.io/PredictingREBoS/`

5.3 Interactive tool for investigating the relationship of meta-analysis characteristics with BoS

An interactive graph is now developed to help visualise the magnitude of BoS and the influence meta-analysis characteristics have upon the value of BoS in a bivariate meta-analysis. The graph allows a user to plot the BoS against the percentage of missing data for the outcome of interest, conditional on the chosen values for meta-analysis

characteristics:

- Within-study correlation
- Within-study variances for the outcome of interest and alternative outcome
- Number of studies
- Between-study correlation
- Between-study variances for the outcome of interest and alternative outcome
- Percentage of missing data for the alternative outcome

The value of BoS is obtained based on mathematical theory, as now described.

5.3.1 Method to develop the interactive graph

Developing an alternative formula for BoS in a bivariate meta-analysis

The Borrowing of Strength (BoS) statistic (Jackson et al. 2017) has been denoted in previous chapters in its simplest form:

$$\text{BoS} = 100\% \times \left[1 - \frac{\text{var}(\hat{\theta}_{mvmeta})}{\text{var}(\hat{\theta}_{uvmeta})}\right] \tag{5.1}$$

To express BoS conditional on the values of the meta-analysis characteristics listed in this section, Equation 5.1 requires expanding. The variances of $\hat{\theta}$ from the univariate and bivariate meta-analyses can be expressed in terms of the within-study variances, s_{i1}^2 and s_{i2}^2, the within-study covariances, λ_i, the between-study variances, $\hat{\tau}_1^2$ and $\hat{\tau}_2^2$, and the between study covariances, $\hat{\tau}_{12}^2$. This is shown by Riley (2004), for the variance of $\hat{\theta}$ from a bivariate meta-analysis:

$$\text{var}(\hat{\theta}_{mvmeta}) = \frac{\sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\hat{\tau}_{12}+\lambda_i)^2}}{\sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\hat{\tau}_{12}+\lambda_i)^2} \sum_{i=1}^{n} \frac{\hat{\tau}_2^2+s_{i2}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\hat{\tau}_{12}+\lambda_i)^2} - \left[\sum_{i=1}^{n} \frac{\hat{\tau}_{12}+\lambda_i}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\hat{\tau}_{12}+\lambda_i)^2}\right]^2} \tag{5.2}$$

where s_{i1}^2 and s_{i2}^2 are within-study variances for study i for outcome 1 and 2, respectively and the within-study covariance is denoted as λ_i. The total number of studies in the meta-analysis is represented by n and i denotes a single study in the meta-analysis. The between-study variances are $\hat{\tau}_1^2$ and $\hat{\tau}_2^2$, and $\hat{\tau}_{12}$ is the between-study covariance.

Similarly, the variance for $\hat{\theta}$ from the univariate meta-analysis can be written as:

$$\text{var}(\hat{\theta}_{uvmeta}) = \frac{1}{\sum_{i=1}^{n} \frac{1}{\hat{\tau}_1^2+s_{i1}^2}} \tag{5.3}$$

where $\hat{\tau}_1^2$ is the between-study variance for outcome 1 (the outcome of interest) and s_1^2 is the within-study variance for outcome 1. The BoS statistic (Equation 5.1) can, therefore be re-expressed by substituting in Equations 5.2 and 5.3 as follows:

$$\text{BoS} = 100\% \times \left\{ 1 - \frac{\sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\hat{\tau}_{12}+\lambda_i)^2}}{\frac{\sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\hat{\tau}_{12}+\lambda_i)^2} \sum_{i=1}^{n} \frac{\hat{\tau}_2^2+s_{i2}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\hat{\tau}_{12}+\lambda_i)^2} - \left[\sum_{i=1}^{n} \frac{\hat{\tau}_{12}+\lambda_i}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\hat{\tau}_{12}+\lambda_i)^2}\right]^2}{\frac{1}{\sum_{i=1}^{n} \frac{1}{\hat{\tau}_1^2+s_{i1}^2}}} \right\} \tag{5.4}$$

The BoS statistic is now expressed in terms of the within-study variances (s_{i1} and s_{i2}) and covariances (λ_i), between-study variances ($\hat{\tau}_1^2$ and $\hat{\tau}_1^2$) and covariance ($\hat{\tau}_{12}^2$) for the total number of studies (n) in the meta-analysis. To express the BoS statistic in terms of the within-study correlations (ρ_{ws_i}) and the between-study correlation (ρ_{bs}), substitute in $\lambda_i = \rho_{ws_i} s_{i1} s_{i2}$ and $\hat{\tau}_{12} = \rho_{bs} \hat{\tau}_1 \hat{\tau}_2$. The BoS statistic in terms of the correlation is then:

$$\text{BoS} = 100\% \times \left\{ 1 - \frac{\sum_{i=1}^{n} \frac{1}{\hat{\tau}_1^2 + s_{i1}^2} \sum_{i=1}^{n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs} \hat{\tau}_1 \hat{\tau}_2 + \rho_{ws_i} s_{i1} s_{i2})^2}}{\sum_{i=1}^{n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs} \hat{\tau}_1 \hat{\tau}_2 + \rho_{ws_i} s_{i1} s_{i2})^2} \sum_{i=1}^{n} \frac{\hat{\tau}_2^2 + s_{i2}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs} \hat{\tau}_1 \hat{\tau}_2 + \rho_{ws_i} s_{i1} s_{i2})^2} - \left[\sum_{i=1}^{n} \frac{\rho_{bs} \hat{\tau}_1 \hat{\tau}_2 + \rho_{ws_i} s_{i1} s_{i2}}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs} \hat{\tau}_1 \hat{\tau}_2 + \rho_{ws_i} s_{i1} s_{i2})^2} \right]^2} \right\} \quad (5.5)$$

To express the BoS statistic in terms of the missing data, the information provided by the n studies in the meta-analysis need to be considered. Consider three mutually exclusive subsets of the n studies. The first subset contains mn studies where m is the proportion of n studies with missing data for the outcome of interest. The second subset contains kn studies where k is the proportion of the n studies with missing data for the alternative outcome. The final subset contains the studies with complete data for both outcomes and contains $(1 - k - m)n$ studies, such that, $n = mn + kn + (1 - k - m)n$, where $0 \leq m, k < 1$ and mn, kn and $(1 - k - m)n$ are integers $(mn, kn, (1 - m - k)n \in \mathbb{Z})$. Following, the equation for the BoS statistic is then:

$$\text{BoS} = 100\% \times \left\{ 1 - \frac{\left[\sum_{i=1}^{mn} \frac{1}{\hat{\tau}_1^2+s_{i1}^2} + \sum_{i=mn+1}^{(m+k)n} \frac{1}{\hat{\tau}_1^2+s_{i1}^2} + \sum_{i=(m+k)n+1}^{n} \frac{1}{\hat{\tau}_1^2+s_{i1}^2}\right]\left[\sum_{i=1}^{mn} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} + \sum_{i=mn+1}^{(m+k)n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} + \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2}\right]}{\left[\sum_{i=1}^{mn} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} + \sum_{i=mn+1}^{(m+k)n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} + \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2}\right]\left[\sum_{i=1}^{mn} \frac{\hat{\tau}_2^2+s_{i2}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} + \sum_{i=mn+1}^{(m+k)n} \frac{\hat{\tau}_2^2+s_{i2}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} + \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_2^2+s_{i2}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2}\right] - \left[\sum_{i=1}^{mn} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2}}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} + \sum_{i=mn+1}^{(m+k)n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2}}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} + \sum_{i=(m+k)n+1}^{n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2}}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2}\right]^2 \right\} \tag{5.6}$$

Therefore, for a particular bivariate meta-analysis dataset, the user could calculate BoS using Equation 5.6 and inputting the values for the within-study variances and correlations, and between-study variances and correlations. However, Equation 5.6 is complex, with large quantities of values to be entered, especially as the number of studies in the meta-analysis increases. Some assumptions help to simplify the mathematical complexities:

- The within-study variances (where available) are assumed to be equal across all the studies in the meta-analysis for the outcome of interest and similarly, for the alternative outcome.

- The within-study correlations (where available) are assumed to be equal in all studies, between +1 and −1.

- The proportions of missing data must not exceed 100% across both outcomes

This leads to the simplified formula for BoS:

$$\text{BoS} = 100\% \times \left\{ 1 - \frac{\left[\frac{mn}{\hat{\tau}_1^2+s_L^2} + \frac{(1-m)n}{\hat{\tau}_1^2+s_1^2}\right]\left[\frac{mn(\hat{\tau}_1^2+s_L^2)}{(\hat{\tau}_1^2+s_L^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \frac{kn(\hat{\tau}_1^2+s_1^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_L^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \frac{(1-m-k)n(\hat{\tau}_1^2+s_1^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{ws}s_1s_2+\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}\right]}{\left[\frac{mn(\hat{\tau}_1^2+s_L^2)}{(\hat{\tau}_1^2+s_L^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \frac{kn(\hat{\tau}_1^2+s_1^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_L^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \frac{(1-m-k)n(\hat{\tau}_1^2+s_1^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{ws}s_1s_2+\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}\right]\left[\frac{mn(\hat{\tau}_2^2+s_2^2)}{(\hat{\tau}_1^2+s_L^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \frac{kn(\hat{\tau}_2^2+s_L^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_L^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \frac{(1-m-k)n(\hat{\tau}_2^2+s_2^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{ws}s_1s_2+\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}\right] - \left[\frac{mn(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)}{(\hat{\tau}_1^2+s_L^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \frac{kn(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_L^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \frac{(1-m-k)n(\rho_{ws}s_1s_2+\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{ws}s_1s_2+\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}\right]^2} \right\} \tag{5.7}$$

where m is the proportion of missing data for the outcome of interest and k is the proportion of missing data for the alternative outcome, such that for n, the total number of studies in the meta-analysis, $n = mn + kn + (1-k-m)n$. The available within-study variances are s_1^2 and s_2^2 for outcomes one and two, respectively and the within-study variances for the missing outcomes, s_L^2, are large. The within-study correlation is ρ_{ws} and the between-study correlation is ρ_{bs}. The between-study variances for outcome one and outcome two are $\hat{\tau}_1^2$ and $\hat{\tau}_2^2$, respectively.

R shiny tool

This formula, Equation 5.7, was embedded in the R package, Shiny, which allows for web applications to be built, and an interactive graph built using the package. A snapshot of the interactive graph is shown in Figure 5.6, and can be accessed by the reader at `https://mhattle.shinyapps.io/BoSstatistic/`. The code for the interactive graph can be found in Appendix E.4.

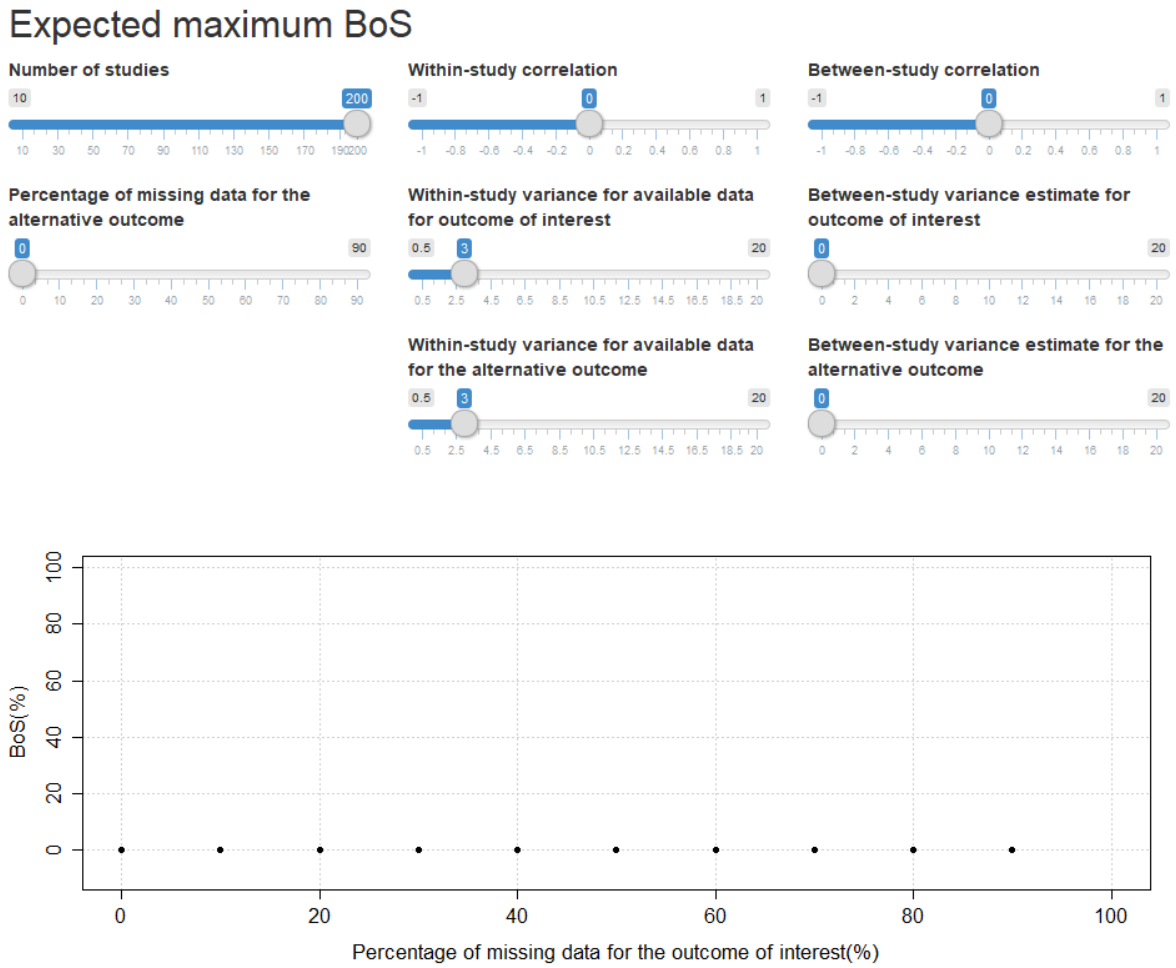


Figure 5.6: A single frame of the interactive graph for the expected maximum BoS calculated from the eqution for BoS statistic. Accessed from `https://mhattle.shinyapps.io/BoSstatistic/`

The tool allows the user to obtain BoS values for a bivariate meta-analysis by simply inputting the number of studies in the meta-analysis, the proportion of studies with missing data for the alternative outcome (outcome two), the within-study variances

for both outcomes, the within-study correlation (assumed equal across all trials), the between-study variances for both outcomes and the between-study correlation. Furthermore, the values for these characteristics could be varied using the 'sliders' above the graph. The slider for the within-study correlation varied in increments of 0.05 and the minimum and maximum values were -0.999 and 0.999, respectively. The within-study variances for each outcome (the outcome of interest and the alternative outcome) were able to vary from 0.5 to 20, increasing in increments of 0.1 on the slider. The number of studies included in the meta-analysis were varied using a slider from 10 to 200 in increments of 10. Increments of 10 were decided upon to make sure that the proportion of studies missing was always an integer.

The percentages of missing data for the alternative outcome ranged from 0% to 90% and increased in increments of 10%. The percentage of missing data for the outcome of interest was represented on the x axis on the interactive graph; the values on the axis ranged from 0% to 90% and increased in increments of 10%. The percentages of missing data for both outcomes were coded such that if the percentages summed to 100% or greater the interactive graph would not plot these points, since a multivariate meta-analysis would not exist in these cases.

For a random-effects meta-analysis, the between-study variances and correlation were able to be varied. The between-study variances for both outcomes varied from 0 to 20 in increments of 0.1. The between-study correlation increased in increments of 0.05 from -1 to 1.

5.3.2 Interactive graph in operation

The influence of the within-study correlation and the within-study variances under a fixed-effect assumption

The magnitude of BoS increases as the absolute value of the within-study correlation increases, under a fixed-effect assumption (Figure 5.7). This is particularly prominent

when the percentage of missing data for the outcome of interest is large. For example, for 20% missing data and 60% missing data the BoS statistics for a within-study correlation of 0.25 were 1.25% and 3.75%, respectively and for a within-study correlation of 0.55 were 6.05% and 18.15%, respectively (Figure 5.7). Nevertheless, the magnitude of BoS is zero when the within-study correlation is zero, irrespective of the percentage of missing data for outcome of interest.

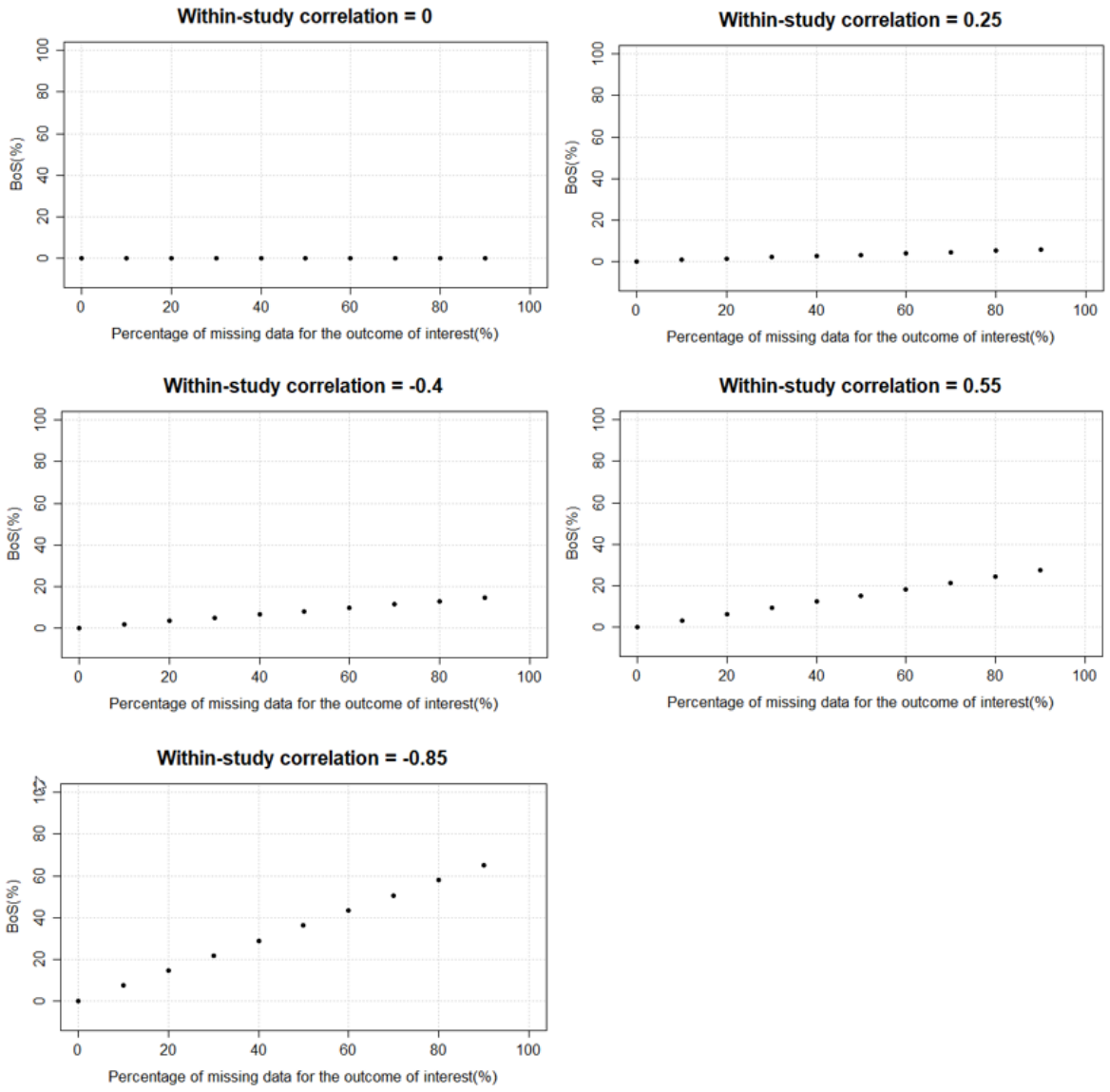


Figure 5.7: Single frames of the interactive graph for the effect of the percentage of missing data for the outcome of interest on the magnitude of BoS for different within-study correlations with the following characteristics: 200 studies, 0% missing data for the alternative outcome and variances of 3 for both outcomes, under a fixed-effect assumption.

The magnitude of the within-study variances for the outcome of interest and the alternative outcome had no effect on the magnitude of BoS (Figure E.1 In Appendix E.1). The magnitude of BoS was unaffected even when the within-study variances were different for the outcome of interest and the alternative outcome.

The influence of the quantity of missing data for the alternative outcome

The percentage of missing data for the alternative outcome appears to influence the magnitude of the BoS in the interactive graph (Figure 5.8), irrespective of a fixed or random-effects assumption. In general, as the percentage of missing data for the alternative outcome increases, the magnitude of BoS decreases. For example, for 40% missing data for the outcome of interest, the magnitude of BoS decreased from 25.60% to 23.88% to 14.88% as the percentage of missing data for the alternative outcome increased from 0% to 20% to 50%, respectively.

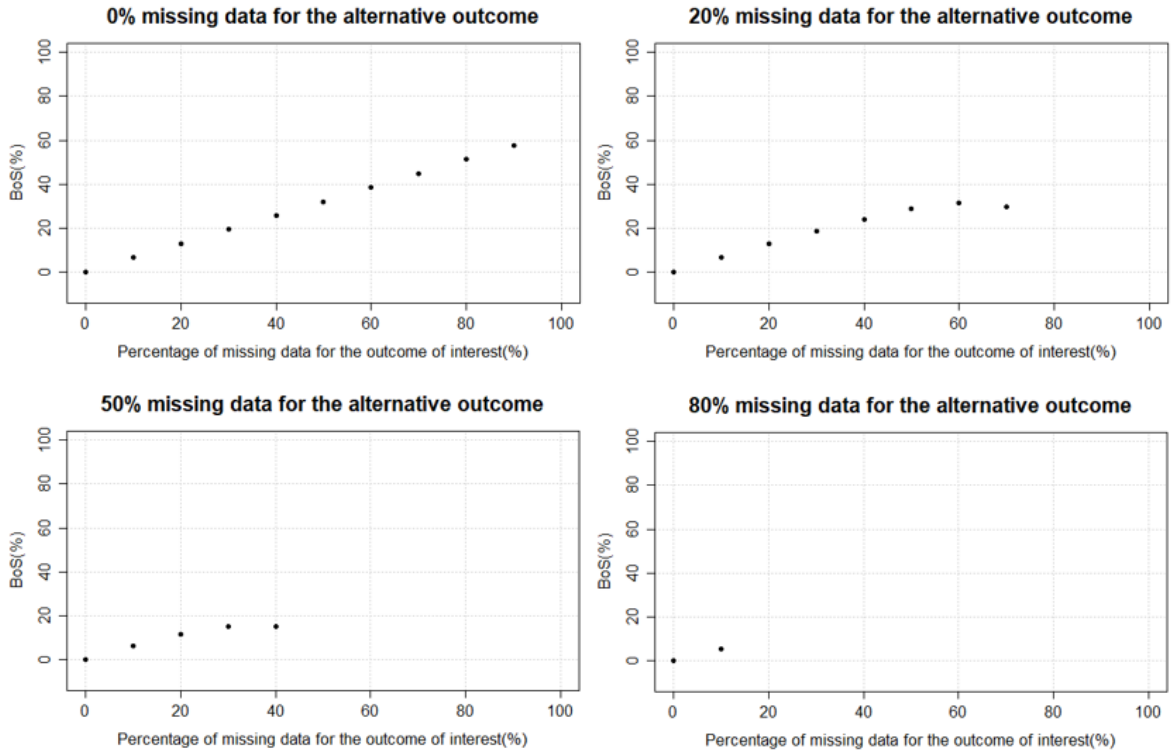


Figure 5.8: Single frames of the interactive graph for the effect of the percentage of missing data for the outcome of interest on the magnitude of BoS for different percentages of missing data for the alternative outcome with the following characteristics; 200 studies, within-study correlation of 0.8 and variances of 3 for both outcomes, under a fixed-effect assumption

The influence of the total number of studies

There was no effect on the magnitude of BoS for different total number of studies in the meta-analysis, irrespective of a fixed or random-effects assumption (Figure E.2 in Appendix E.1).

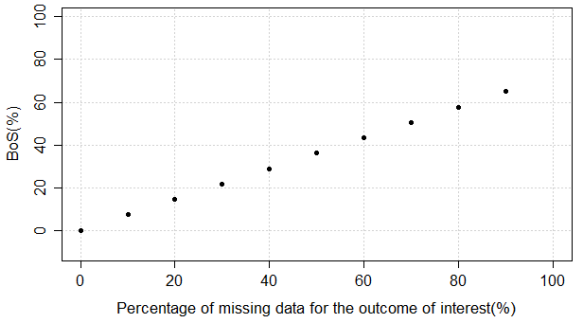
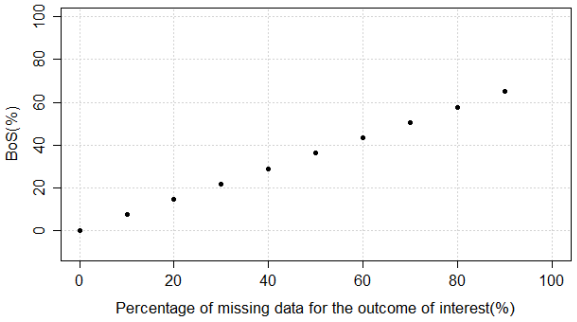
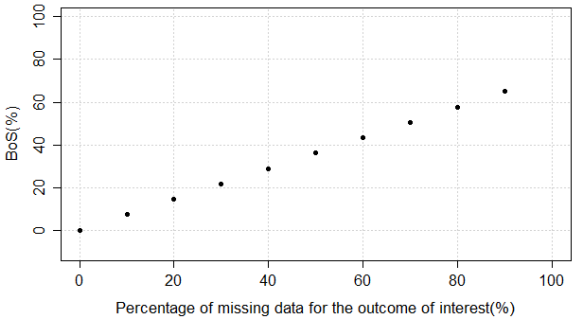
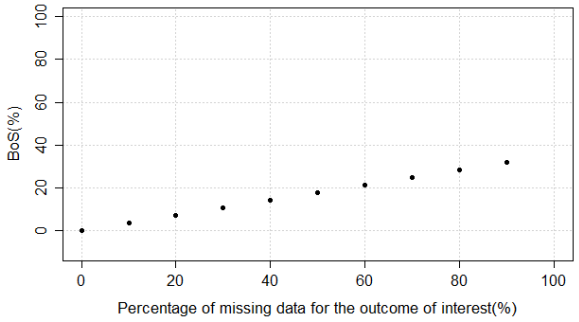
The influence of the between-study correlation and variances (random-effects assumption)

Under a random-effects assumption, the within-study and between-study correlations and the within-study and between-study variances influence the magnitude of BoS jointly. Firstly, the relationship between the correlations is explored here. The relationship is determined by the sign of the within-study correlation. In general, when

the within-study correlation is negative, the BoS increases as the between-study correlation decreases. Alternatively, when the within-study correlation is positive, the BoS increases as the between-study correlation increases.

The within-study and between-study variances interact together to influence the magnitude of BoS from a random-effects meta-analysis. The magnitude of BoS is greatest when there are pairs of variances that are approximately equal, provided the correlations are equal. The pairs of variances that are required to be approximately equal for this statement to hold can be either the within-study variances, the between-study variances, the variances for the outcome of interest or the variances for the alternative outcome, as well as all four variances. Note, this does not hold for pairs of complete opposite variances e.g. the within-study variance for the outcome of interest and the between-study variance for the alternative outcome. For example, in Figure 5.9 the BoS values are greatest when the magnitude of the variances are paired either by the variance (Figure 5.9 i. and ii.), i.e. within-study or between-study, or by the outcome they correspond to (Figure 5.9 iii.). In these scenarios, for 60% missing data for the outcome of interest the BoS statistic ranged between 43.33% to 43.34%. The BoS statistic decreased to 21.22% for 60% missing data for the outcome of interest when the magnitude of the variances were contained in unrelated pairs (Figure 5.9 iv.) e.g. the variances with magnitude of 3 are the within-study variance for the outcome of interest and the between-study variance for the alternative outcome. When all the variances were equal, with magnitude of 3, apart from the between-study variance for the outcome of interest, which had a variance of 18, the BoS statistic was 36.84% for 60% missing data for the outcome of interest (Figure 5.9 v.). Although the magnitude of BoS was greater than the scenario of unrelated pairs, the magnitude of BoS was greatest when the magnitude of the variances were in related pairs.

Figure 5.9

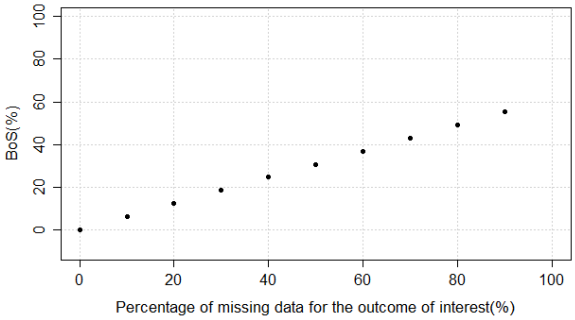
	Variances				BoS graph
	Within-study		Between-study		
	1[a]	2[b]	1[a]	2[b]	
i.	3	3	18	18	
ii.	18	18	3	3	
iii.	18	3	18	3	
iv.	3	18	18	3	

Continued on next page

[a] Outcome of interest

[b] Alternative outcome

Figure 5.9: continued

	Variances				BoS graph
	Within-study		Between-study		
	1[a]	2[b]	1[a]	2[b]	
v.	3	3	18	3	

[a] Outcome of interest

[b] Alternative outcome

Figure 5.9: Single frames of the interactive graph for the effect of the percentage of missing data for the outcome of interest on the magnitude of BoS for different combinations of values for the within-study and between-study variances with the following characteristics; 200 studies, within-study and between-study correlations of 0.85

Maximising BoS

The aim of the interactive graph was to visually explore the relationship between the magnitude of BoS and the percentage of missing data for the outcome of interest. Through the exploration of the interactive tool the value of BoS was often observed to be less than the percentage of missing data for the outcome of interest. It will now be explored whether the value of BoS can exceed the percentage of missing data for the outcome of interest in the interactive tool, by maximising BoS using the conditions that have been observed to increase the magnitude of BoS.

The number of studies was not found to influence the magnitude of BoS in the interactive graph. To maximise the magnitude of BoS large absolute values of the within-study and between-study correlations with equal signs is required with no

missing data for the alternative outcome. Figures 5.10 and 5.11 display the maximised magnitude of BoS; it can be seen that the magnitude of BoS does not exceed the percentage of missing data for the outcome of interest.

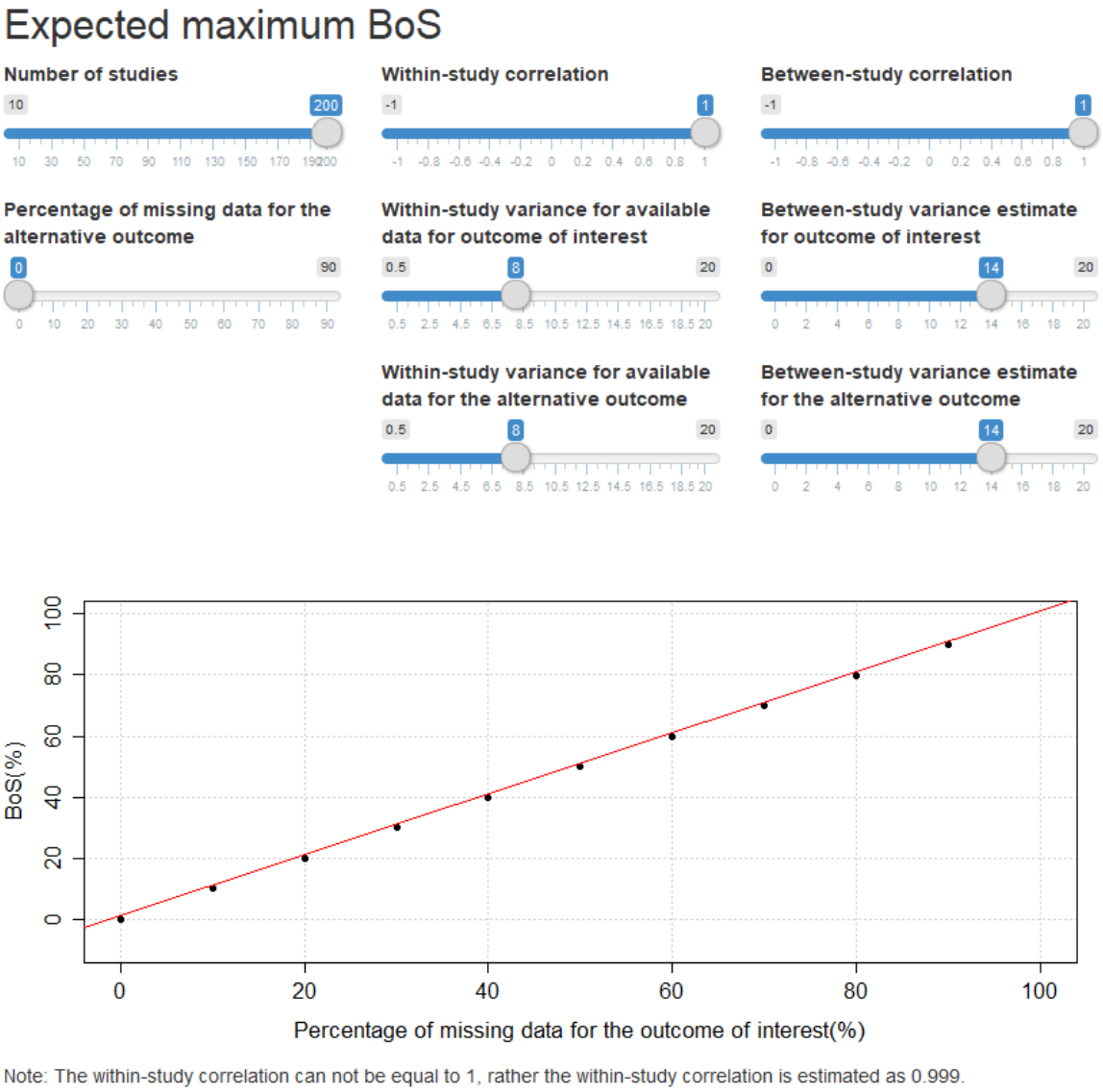


Figure 5.10: A single frame of the interactive graph when BoS statistic was maximised using large positive correlations. The red line represents y=x, when BoS statistic is equal to the percentage of missing data for the outcome of interest.

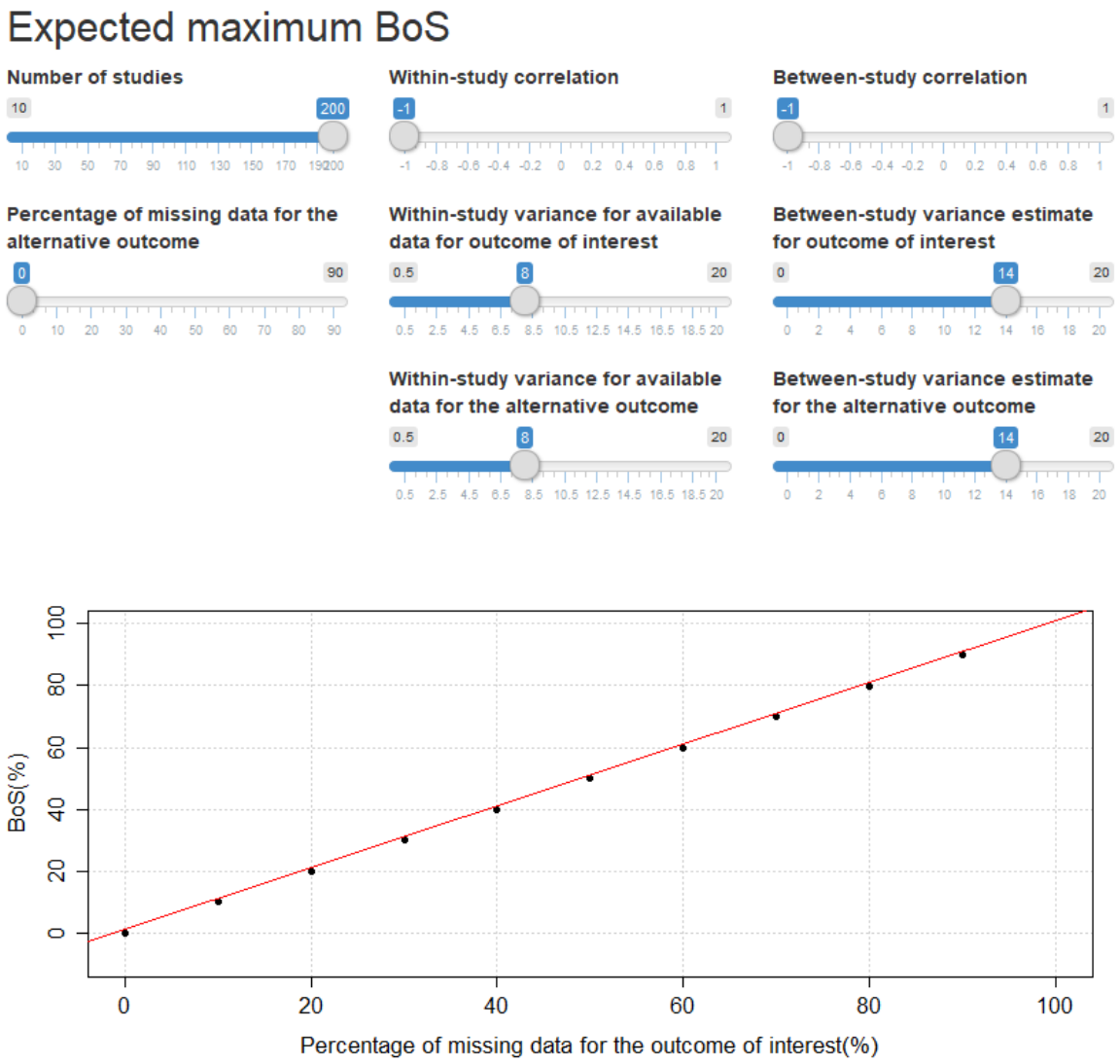


Figure 5.11: A single frame of the interactive graph when BoS statistic was maximised using large negative correlations. The red line represents y=x, when BoS statistic is equal to the percentage of missing data for the outcome of interest.

5.4 Discussion

In this chapter the development and use of interactive graphical tools has been explored. The prediction models from Chapter 4, were inbuilt into interactive graphical tools; one for the fixed-effect and one for the random-effects bivariate meta-analyses. The equation for the BoS statistic was also encoded into an interactive graphical tool for the exploration of the relationship between the BoS statistic and meta-analysis

characteristics. The interactive tools were developed using R shiny and allow a user to visually explore the predicted value of BoS or the expected maximum of BoS for different characteristics of a meta-analysis. In this section, the key findings (summarised in Figure 6.1) are discussed and areas for future research are described.

Figure 5.12: Key findings from the interactive tool and mathematical proofs

Key Findings:

- Two interactive tools were produced from the developed prediction models (Chapter 4) to aid researchers to visualise their predicted BoS.
- An interactive tool was developed to aid researchers to visualise the magnitude of BoS and what characteristics influence the magnitude, using the mathematical formula of BoS.
- The within-study correlation and the percentage of missing data were identified from the potential maximum BoS interactive graph as highly influential over the magnitude of BoS.
- By maximising the BoS statistic using the within-study correlation and the percentage of missing data for the alternative outcome, the magnitude of BoS was found to not exceed the percentage of missing data for the outcome of interest.

There were two interactive graphs that were developed for the prediction models from Chapter 4 in Section 5.2. The use of the interactive graphs for the prediction models was demonstrated using an example, Buzzetti et al. (2017), a Cochrane review.

A third interactive tool was developed in R shiny that visually displayed the magnitude of BoS calculated from the BoS equations for different quantities of missing data for the outcome of interest using provided values of meta-analysis characteristics. The interactive graph was explored to investigate which combination

of characteristics increased the magnitude of BoS. The within-study correlation was highly influential in the magnitude of BoS. This is not a surprising result; Jackson et al. (2017) discussed that for a complete data scenario the BoS statistic is bounded by the square of the within-study correlation. Additionally the percentage of missing data for the outcome of interest was highly influential to the magnitude of BoS.

The BoS statistic was maximised using meta-analysis characteristic values that were observed to increase the magnitude of BoS. Under these conditions the interactive graph showed that the BoS may be bounded by the percentage of missing data for the outcome of interest (since the values of BoS did not exceed the percentages of missing data).

The prediction models and the mathematical formula for BoS were not displayed on the same graph due to each graph requiring and presenting different information. The interactive tools for the prediction models present the predicted BoS statistic based on the meta-analysis characteristics. The interactive graph for the mathematical formula presents the expected maximum based on the meta-analysis characteristics. Therefore, if these two approaches were presented on the same tool there is potential to mislead or cause confusion for researchers, which negates the aim for these tools to be accessible and easy for researchers to use. Additionally by presenting the two approaches on the same interactive tool it leads to comparisons to be made between the predicted BoS and the expected maximum BoS, two very different measures.

5.4.1 Strengths and limitations

The prediction models have been discussed in the previous chapter (Chapter 4) and therefore the prediction models specifics will not be discussed here. In this section the strengths and limitations of the interactive graphical tools will be discussed.

Strengths

This chapter has many strengths including providing a greater understanding of which characteristics of meta-analyses affect the magnitude of BoS. This chapter builds upon the findings and investigations into the BoS statistic by Jackson et al. (2017) and Copas et al. (2018). A greater understanding of BoS will help to identify situations in which the BoS is considerably large. This will help researchers predict what magnitude of BoS they might expect from their meta-analysis and in turn could help them to decide if a multivariate approach would be beneficial over a univariate approach.

This chapter is the first time the BoS has been displayed on interactive graphs using prediction models for BoS and the mathematical equation for BoS. The interactive graphs allow researchers to easily visually explore the BoS and the affects of certain characteristics on its magnitude.

The interactive graphical tools can be easily accessed by researchers through a website. Additionally, the interactive tools are usable with the input of values aided through embedded sliders.

Limitations

The mathematical equation for the BoS statistic was simplified using assumptions that included equal within-study correlations across all studies in the meta-analysis. These conditions were made to aid usability. However, they are unlikely to hold for real data analysed in a meta-analysis. It might be expected that provided the variances and within-study correlations are similar in each study the BoS statistic will be similar to those in the interactive graph.

Another limitation of the interactive graphical tools is it is possible for researchers to manipulate the information provided to the graph or from the graph to fit a desirable narrative or analysis plan.

5.4.2 Further work

This chapter presented the prediction models from Chapter 4 using interactive graphical which allow researchers to visually predict the BoS statistic in their application. In addition the mathematical equation for the calculation of the BoS statistic was explored for its relationship to meta-analysis level characteristics. Through this exploration it was observed that there is possibly a relationship between the magnitude of the BoS statistic and the percentage of missing data for the outcome of interest. If there is a relationship then it is important for future research to explore this further.

For the use of the interactive graphs, it would be beneficial to speak to researchers about whether they would use such a tool for predicting the BoS statistic.

Next steps

In the next chapter, the relationship between the BoS statistic and the percentage of missing data for the outcome of interest is explored further through mathematical reasoning and examples.

Chapter 6

Mathematical investigation of the magnitude of the BoS statistic

6.1 Introduction

In previous chapters, BoS was studied extensively within the framework of multivariate meta-analysis. In Chapter 3, the BoS statistic was observed to identify differences between the univariate and multivariate meta-analysis results. In Chapters 3 and 4, statistical models for predicting the value of BoS were developed using characteristics of existing multivariate meta-analyses as predictors. The characteristics of meta-analyses that were the most important predictors included; the number of studies, the maximum absolute within-study correlation and the percentage of missing data for the outcome of interest.

In the previous chapter, the BoS statistic was explored using developed interactive graphical tools. The prediction models developed in Chapter 4 were embedded into interactive graphical tools, to enable researchers to predict the magnitude of BoS in their setting (Chapter 5). In particular, the potential maximum BoS calculated from the equation for the BoS statistic was also included in an interactive graph. This enabled the visual exploration of the affect the characteristics of multivariate meta-analysis had on the magnitude of BoS.

The Borrowing of Strength (BoS) statistic was proposed by Jackson et al. (2017) and

was re-examined by Copas et al. (2018) to provide a greater understanding of how the borrowing of strength depends on individual study characteristics. Jackson et al. (2017) discussed the relationship between BoS and the within-study correlation; when there is no within-study correlation the BoS is zero in a fixed-effect meta-analysis. The magnitude of BoS is bound, due to the within-study correlations when there is no missing data.This chapter seeks to build upon the findings of previous chapters, and the work by Jackson et al. (2017) and Copas et al. (2018), by further investigating the relationship between the characteristics of meta-analyses, in particular the percentage of missing data for the outcome of interest, and the magnitude of BoS. The percentage of missing data for the outcome of interest, along with the within-study correlation and the number of studies were the most important predictors of the magnitude of BoS. However within-study correlations are rarely available (Riley 2009) and therefore in these settings the BoS cannot be predicted using the prediction models in Chapter 4. Through the development and use of a novel visual display tool in R, the magnitude of BoS in relation to the within-study correlation, the percentage of missing data for the outcome of interest and other meta-analysis characteristics were explored in Chapter 5. Through the exploration of this interactive graphical tool, it was observed, while maximising BoS, that the value of BoS was observed to not exceed the percentage of missing data for the outcome of interest.

This chapter follows by firstly discussing whether it is plausible to consider that the BoS might be bounded by the percentage of missing data for the outcome of interest (Section 6.2). Following, using mathematical theory, Section 6.3 will provide an informal rule of thumb for when multivariate meta-analysis may be most important to use.

6.2 Is it plausible that BoS might be bounded by percentage of missing data?

In the previous chapter (Section 5.3), the magnitude of BoS was shown to not exceed the value of the percentage of missing data for the outcome of interest in the interactive graph. Therefore it could be considered that the magnitude of BoS is bounded by the percentage of missing data for the outcome of interest but is it logical to consider this to be the relationship? In this section, a simplistic explanation containing a univariate meta-analysis with no heterogeneity is used to outline the plausibility of the boundedness of BoS.

In this explanation, consider a univariate meta-analysis of a particular outcome with five studies available. Let the within-study variance for the effect estimate of the outcome in each study, i, be s_{i1}^2. Then, the variance from the univariate meta-analysis with no heterogeneity for outcome one is

$$\left(\sum_{i=1}^{5} \frac{1}{s_{i1}^2}\right)^{-1} = \frac{s_1^2}{5}, \tag{6.1}$$

assuming that the within-study variances are equal across all studies.

Consider the situation where the univariate meta-analysis was updated and assume that a further five studies were included in the meta-analysis, each having variances of equal value to the original five studies. Furthermore, assume that there is no heterogeneity in treatment effect across all 10 studies. The variance for the summary statistic from the updated univariate meta-analysis with 10 studies is:

$$\begin{aligned}\left(\sum_{i=1}^{10} \frac{1}{s_{i1}^2}\right)^{-1} &= \frac{s_1^2}{10} \\ &= \frac{1}{2}\left(\frac{s_1^2}{5}\right),\end{aligned} \tag{6.2}$$

assuming that the within-study variance is equal across all studies. Therefore, the variance from the meta-analysis with 10 studies is smaller than the variance from the meta-analysis with five studies.

An alternative way to consider the original univariate meta-analysis with five studies is that there were missing studies, with 50% of trials missing (i.e. 5 out of 10). Therefore, a potential maximum BoS that a multivariate meta-analysis could achieve for this outcome is to completely recover the missing information in the five trials where it is not available, i.e. to get back to the univariate meta-analysis result where 10 studies truly had been available. In this situation, the BoS statistic, which is written as (Jackson et al. 2017)

$$\text{BoS} = 100\% \times \left(1 - \frac{\text{var}_{\text{mvmeta}}}{\text{var}_{\text{uvmeta}}}\right) \tag{6.3}$$

can be calculated from the variances from the meta-analysis with 10 studies and the meta-analysis with five studies. The univariate meta-analysis with 10 studies contains the total information contribution to the summary effect estimate and the univariate meta-analysis with five studies contains the direct information contribution to the summary effect estimate. Therefore, the BoS statistic equation can be rewritten as:

$$\begin{aligned}
\text{BoS} &= 100\% \times \left(1 - \frac{\frac{s_{i1}^2}{10}}{\frac{s_{i1}^2}{5}}\right) && (6.4)\\
&= 100\% \times \left(1 - \frac{s_{i1}^2}{10}\frac{5}{s_{i1}^2}\right) && (6.5)\\
&= 100\% \times \left(1 - \frac{1}{2}\right) && (6.6)\\
&= 50\% && (6.7)
\end{aligned}$$

Therefore, the BoS in this example is 50% which is equal to the percentage of missing data in the original meta-analysis, which had five out of ten trials. This is a simplistic explanation to illustrate why we might expect the BoS statistic to be bounded by the percentage of missing data for the outcome of interest. However, neither this example

nor the graphical tool formally prove that the BoS statistic is bounded and, if so, under what conditions the boundedness holds.

6.3 Mathematical proofs for the relationship between the percentage of missing data and the magnitude of BoS

In this section the mathematical relationship between the percentage of missing data and the magnitude of BoS is examined further. The aim of this section is to formally prove whether the percentage of missing data restricts BoS to a maximum value, under certain conditions.

First, recall the BoS equation (Equation 5.5) that was expanded using the equations for the variances for $\hat{\theta}$ from the univariate and multivariate meta-analyses and written in terms of the correlations:

$$\text{BoS} = 100\% \times \left\{ 1 - \frac{\sum_{i=1}^{n} \frac{1}{\hat{\tau}_1^2+s_{i1}^2} \sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2}}{\sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} \sum_{i=1}^{n} \frac{\hat{\tau}_2^2+s_{i2}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} - \left[\sum_{i=1}^{n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2}}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} \right]^2} \right\} \quad (5.5)$$

This can also be written as a proportion as opposed to a percentage as follows:

$$\text{BoS} = 1 - \frac{\sum_{i=1}^{n} \frac{1}{\hat{\tau}_1^2+s_{i1}^2} \sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2}}{\sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} \sum_{i=1}^{n} \frac{\hat{\tau}_2^2+s_{i2}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} - \left[\sum_{i=1}^{n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2}}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} \right]^2} \quad (6.8)$$

Theorem 6.3.1. *The Borrowing of Strength (BoS) statistic from a bivariate fixed-effect meta-analysis is bounded by the percentage of missing data for the outcome of interest when there is complete data for the alternative outcome; if and only if the within-study variances and covariances are known to be equal across all studies within the bivariate fixed-effect meta-analysis.*

Proof. For a bivariate fixed-effect meta-analysis, let the between-study correlation and the between-study variance equal zero.

$$\begin{aligned}
\text{BoS} &= 1 - \frac{\sum_{i=1}^{n} \frac{1}{s_{i1}^2} \sum_{i=1}^{n} \frac{s_{i1}^2}{s_{i1}^2 s_{i2}^2 - (\rho_{ws_i} s_{i1} s_{i2})^2}}{\sum_{i=1}^{n} \frac{s_{i1}^2}{s_{i1}^2 s_{i2}^2 - (\rho_{ws_i} s_{i1} s_{i2})^2} \sum_{i=1}^{n} \frac{s_{i2}^2}{s_{i1}^2 s_{i2}^2 - (\rho_{ws_i} s_{i1} s_{i2})^2} - \left[\sum_{i=1}^{n} \frac{\rho_{ws_i} s_{i1} s_{i2}}{s_{i1}^2 s_{i2}^2 - (\rho_{ws_i} s_{i1} s_{i2})^2} \right]^2} \\
&= 1 - \frac{\sum_{i=1}^{n} \frac{1}{s_{i1}^2} \sum_{i=1}^{n} \frac{1}{s_{i2}^2 - s_{i2}^2 \rho_{ws_i}^2}}{\sum_{i=1}^{n} \frac{1}{s_{i2}^2 - s_{i2}^2 \rho_{ws_i}^2} \sum_{i=1}^{n} \frac{1}{s_{i1}^2 - s_{i1}^2 \rho_{ws_i}^2} - \left(\sum_{i=1}^{n} \frac{\rho_{ws_i}}{s_{i1} s_{i2} - s_{i1} s_{i2} \rho_{ws_i}^2} \right)^2} \\
&= 1 - \frac{\sum_{i=1}^{n} \frac{1}{s_{i1}^2} \sum_{i=1}^{n} \frac{1}{s_{i2}^2 (1 - \rho_{ws_i}^2)}}{\sum_{i=1}^{n} \frac{1}{s_{i2}^2 (1 - \rho_{ws_i}^2)} \sum_{i=1}^{n} \frac{1}{s_{i1}^2 (1 - \rho_{ws_i}^2)} - \left[\sum_{i=1}^{n} \frac{\rho_{ws_i}}{s_{i1} s_{i2} (1 - \rho_{ws_i}^2)} \right]^2}
\end{aligned} \tag{6.9}$$

To understand the BoS statistic further, the summations need to be explored. Each summation in the BoS statistic (Equation 6.9) is summed for every study in the meta-analysis where there are n studies in the meta-analysis. For each study contained within the bivariate meta-analysis, there are two potential missing data scenarios the study can belong to, since the alternative outcome has complete data. Therefore, either a study has complete data, where both the outcome of interest and the alternative outcome are available or for a study the outcome of interest is unavailable (therefore missing) and the alternative outcome is available. It should be noted that these two

scenarios are mutually exclusive.

To mathematically model these scenarios notation assigned to each scenario is required. Therefore, let m be the proportion of studies with missing data for the outcome of interest and following the proportion of studies with complete data can be calculated as $(1-m)$. Therefore the total number of studies can be expressed as $n = mn + (1-m)n$, where $0 \leq m < 1$ and mn is an integer ($mn \in \mathbb{Z}$) and $(1-k-m)n$ is an integer ($(1-k-m)n \in \mathbb{Z}$). In this situation, mn is the number of studies that have missing information for the outcome of interest and $(1-m)n$ is the number of studies with complete data.

Consider the situation where the studies in the meta-analysis are ordered by scenario, with all the studies with a missing outcome of interest first followed by studies with complete data and subsequently assigned a number. Therefore, the summations in Equation 6.9 can be expanded out in the following way:

$$\text{BoS} = 1 - \frac{\left(\sum_{i=1}^{mn} \frac{1}{s_{i1}^2} + \sum_{i=mn+1}^{n} \frac{1}{s_{i1}^2}\right)\left[\sum_{i=1}^{mn} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)} + \sum_{i=mn+1}^{n} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}\right]}{\left[\sum_{i=1}^{mn} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2]} + \sum_{i=mn+1}^{n} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}\right]\left[\sum_{i=1}^{mn} \frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)} + \sum_{i=mn+1}^{n} \frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)}\right] - \left[\sum_{i=1}^{mn} \frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)} + \sum_{i=mn+1}^{n} \frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)}\right]^2} \tag{6.10}$$

For the studies that have missing information for the outcome of interest (studies 1 to mn), the within-study correlation, ρ_{ws_i} is zero and the within-study variance, s_{i1}^2 is modelled as infinity (Riley 2009, Jackson et al. 2011), furthermore $\frac{1}{s_{i1}^2}$ tends to zero as s_{i1}^2 tends to infinity.

$$\text{BoS} = 1 - \frac{\left(\sum_{i=mn+1}^{n} \frac{1}{s_{i1}^2}\right)\left[\sum_{i=1}^{mn} \frac{1}{s_{i2}^2} + \sum_{i=mn+1}^{n} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}\right]}{\left[\sum_{i=1}^{mn} \frac{1}{s_{i2}^2} + \sum_{i=mn+1}^{n} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}\right]\left[\sum_{i=mn+1}^{n} \frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)}\right] - \left[\sum_{i=mn+1}^{n} \frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)}\right]^2} \tag{6.11}$$

Using the assumption that the within-study correlations, ρ_{ws_i}, the variance for the outcome of interest, s_{i1}^2, and the variance for the alternative outcome, s_{i2}^2, are equal across each study in the bivariate meta-analysis, the summation can be written as follows.

$$\text{BoS} = 1 - \frac{\left[\frac{nm}{s_2^2} + \frac{n(1-m)}{s_2^2(1-\rho_{ws}^2)}\right]\left[\frac{n(1-m)}{s_1^2}\right]}{\left[\frac{nm}{s_2^2} + \frac{n(1-m)}{s_2^2(1-\rho_{ws}^2)}\right]\left[\frac{n(1-m)}{s_1^2(1-\rho_{ws}^2)}\right] - \left[\frac{n(1-m)\rho_{ws}}{s_1 s_2(1-\rho_{ws}^2)}\right]^2} \tag{6.12}$$

The brackets that contain a sum of two fractions can be simplified by generating common denominators in the two fractions and combining into one fraction. Following the equation can be simplified further.

$$\text{BoS} = 1 - \frac{\left[\frac{nm(1-\rho_{ws}^2)}{s_2^2(1-\rho_{ws}^2)} + \frac{n(1-m)}{s_2^2(1-\rho_{ws}^2)}\right]\left[\frac{n(1-m)}{s_1^2}\right]}{\left[\frac{nm(1-\rho_{ws}^2)}{s_2^2(1-\rho_{ws}^2)} + \frac{n(1-m)}{s_2^2(1-\rho_{ws}^2)}\right]\left[\frac{n(1-m)}{s_1^2(1-\rho_{ws}^2)}\right] - \left[\frac{n^2(1-m)^2\rho_{ws}^2}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}\right]}$$

$$= 1 - \frac{\left[\frac{nm(1-\rho_{ws}^2)+n(1-m)}{s_2^2(1-\rho_{ws}^2)}\right]\left[\frac{n(1-m)}{s_1^2}\right]}{\left[\frac{nm(1-\rho_{ws}^2)+n(1-m)}{s_2^2(1-\rho_{ws}^2)}\right]\left[\frac{n(1-m)}{s_1^2(1-\rho_{ws}^2)}\right] - \left[\frac{n^2(1-m)^2\rho_{ws}^2}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}\right]}$$

$$= 1 - \frac{\left[\frac{n^2 m(1-m)(1-\rho_{ws}^2)+n^2(1-m)^2}{s_1^2 s_2^2(1-\rho_{ws}^2)}\right]}{\left[\frac{n^2 m(1-m)(1-\rho_{ws}^2)+n^2(1-m)^2}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}\right] - \left[\frac{n^2(1-m)^2\rho_{ws}^2}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}\right]} \tag{6.13}$$

Generate common denominators in both the numerator fraction and the denominator fraction, such that the common denominators will cancel out.

$$\text{BoS} = 1 - \frac{\left[\frac{n^2 m(1-m)(1-\rho_{ws}^2)^2+n^2(1-m)^2(1-\rho_{ws}^2)}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}\right]}{\left[\frac{n^2 m(1-m)(1-\rho_{ws}^2)+n^2(1-m)^2-n^2(1-m)^2\rho_{ws}^2}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}\right]}$$

$$= 1 - \frac{n^2 m(1-m)(1-\rho_{ws}^2)^2 + n^2(1-m)^2(1-\rho_{ws}^2)}{n^2 m(1-m)(1-\rho_{ws}^2) + n^2(1-m)^2 - n^2(1-m)^2\rho_{ws}^2}$$

$$= 1 - \frac{n^2 m(1-m)(1-\rho_{ws}^2)^2 + n^2(1-m)^2(1-\rho_{ws}^2)}{n^2 m(1-m)(1-\rho_{ws}^2) + n^2(1-m)^2(1-\rho_{ws}^2)}$$

$$= 1 - \frac{n^2(1-m)(1-\rho_{ws}^2)[m(1-\rho_{ws}^2) + (1-m)]}{n^2(1-m)(1-\rho_{ws}^2)[m + (1-m)]}$$

$$= 1 - [m(1-\rho_{ws}^2) + (1-m)]$$

$$= 1 - (m - m\rho_{ws}^2 + 1 - m)$$

$$= 1 - (1 - m\rho_{ws}^2)$$

$$= m\rho_{ws}^2 \tag{6.14}$$

Since $\rho_{ws_i}^2$ is the square of the within study correlation which has a range of between -1 and 1, $\rho_{ws_i}^2$ has a range of between 0 and 1. Therefore, BoS can not be greater than m, the proportion of missing data for the outcome of interest, although BoS can be less than m when it is scaled by a within-study correlation that is not equal to 1 or -1. Since the within-study correlation is squared and m is a proportion between 0 and 1, the BoS statistic can not take a value that is negative.

Corollary 6.3.2. *The BoS statistic is bounded between 0% and 100% (proportion: 0 and 1).*

Therefore, under the conditions of equal within-study variances and covariances across studies, when there is complete data for the alternative outcome BoS is bounded by the percentage of missing data for the outcome of interest. ■

Theorem 6.3.3. *The borrowing of strength statistic from a bivariate fixed-effect meta-analysis is bounded by the percentage of missing data for the outcome of interest when there is missing data for the alternative outcome; if and only if the within-study variances and covariances are known to be equal across all studies.*

Proof. Recall Equation 6.9;

$$\text{BoS} = 1 - \frac{\sum_{i=1}^{n} \frac{1}{s_{i1}^2} \sum_{i=1}^{n} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}}{\sum_{i=1}^{n} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)} \sum_{i=1}^{n} \frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)} - \left[\sum_{i=1}^{n} \frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)}\right]^2} \tag{6.9}$$

was the equation for the BoS statistic from a bivariate fixed-effect meta-analysis written as a proportion in terms of the within-study variances and the within-study correlations. Similarly to the proof for Theorem 6.3.1, consider each study as belonging to a missing data scenario. For this proof there are three possible scenarios for each study. For a study a possible scenario is complete data, or an alternative scenario is there is missing data for the outcome of interest and available information for the alternative outcome, as discussed in the previous proof. A third possible scenario is for a study to have the outcome of interest available and the alternative outcome missing. This scenario will also be included to extend the previous proof.

Some further notation is required, so let m be the proportion of missing data for the outcome of interest and let k be the proportion of missing data for the alternative outcome in the bivariate meta-analysis. The proportion of studies with complete data can be calculated as $(1-k-m)$, such that, the total number of studies is $n = mn + kn + (1-k-m)n$, $0 \leq m, k, < 1$, where mn is an integer ($mn \in \mathbb{Z}$), kn is an integer ($kn \in \mathbb{Z}$) and $(1-k-m)n$ is an integer ($(1-k-m)n \in \mathbb{Z}$). The summation in Equation 6.9, can then be expanded using the above notation.

First, assume that the studies in the meta-analysis are ordered by the scenario in which they belong to. Therefore, the first mn studies have missing information for the outcome of interest, the studies numbered between $mn+1$ and $n(m+k)$ (inclusive) are studies

with missing information for the alternative outcome and for the studies numbered between $n(m+k)+1)$ to n are the studies with complete data.

$$\text{BoS} = 1 - \frac{\left[\sum_{i=1}^{mn}\frac{1}{s_{i1}^2} + \sum_{i=mn+1}^{n(m+k)}\frac{1}{s_{i1}^2} + \sum_{i=n(m+k)+1}^{n}\frac{1}{s_{i1}^2}\right]\left[\sum_{i=1}^{mn}\frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)} + \sum_{i=mn+1}^{n(m+k)}\frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)} + \sum_{i=n(m+k)+1}^{n}\frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}\right]}{\left\{\left[\sum_{i=1}^{mn}\frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)} + \sum_{i=mn+1}^{n(m+k)}\frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)} + \sum_{i=n(m+k)+1}^{n}\frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}\right]\left[\sum_{i=1}^{mn}\frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)} + \sum_{i=mn+1}^{n(m+k)}\frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)} + \sum_{i=n(m+k)+1}^{n}\frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)}\right]\right.} \\ \left. - \left[\sum_{i=1}^{mn}\frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)} + \sum_{i=mn+1}^{n(m+k)}\frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)} + \sum_{i=n(m+k)+1}^{n}\frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)}\right]^2\right\} \tag{6.15}$$

The studies with a missing outcome, either the outcome of interest or the alternative outcome, have a within-study correlation, ρ_{ws_i}, equal to zero. The missing within-study variances, either s_{i1}^2 or s_{i2}^2, are modelled as infinity and thus $\frac{1}{s_{i1}^2}$ tends to zero as s_{i1}^2 tends to infinity (and $\frac{1}{s_{i1}^2}$ tends to zero as s_{i1}^2 tends to infinity) (Riley 2009, Jackson et al. 2011). Using these values and rules the summations can be simplified:

$$\text{BoS} = 1 - \frac{\left[\sum_{i=mn+1}^{n(m+k)}\frac{1}{s_{i1}^2} + \sum_{i=n(m+k)+1}^{n}\frac{1}{s_{i1}^2}\right]\left[\sum_{i=1}^{mn}\frac{1}{s_{i2}^2} + \sum_{i=n(m+k)+1}^{n}\frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}\right]}{\left\{\left[\sum_{i=1}^{mn}\frac{1}{s_{i2}^2} + \sum_{i=n(m+k)+1}^{n}\frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}\right]\left[\sum_{i=mn+1}^{n(m+k)}\frac{1}{s_{i1}^2} + \sum_{i=n(m+k)+1}^{n}\frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)}\right] - \left[\sum_{i=n(m+k)+1}^{n}\frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)}\right]^2\right\}} \tag{6.16}$$

Using the assumption that the variances for available outcome and the available within-study correlations are equal across all studies in the bivariate meta-analysis, the equation for BoS can be simplified to:

$$\text{BoS} = 1 - \frac{\left[\frac{nk}{s_1^2} + \frac{(1-m-k)n}{s_1^2}\right]\left[\frac{nm}{s_2^2} + \frac{(1-m-k)n}{s_2^2(1-\rho_{ws}^2)}\right]}{\left[\frac{nm}{s_2^2} + \frac{(1-m-k)n}{s_2^2(1-\rho_{ws}^2)}\right]\left[\frac{nk}{s_1^2} + \frac{(1-m-k)n}{s_1^2(1-\rho_{ws}^2)}\right] - \left[\frac{(1-m-k)n\rho_{ws}}{s_1 s_2(1-\rho_{ws}^2)}\right]^2} \tag{6.17}$$

To simplify the contents of the brackets, generate common denominators and combine these into a single fraction. Following this, the brackets in the numerator and denominator can be expanded.

$$\text{BoS} = 1 - \frac{\left[\frac{nk+(1-m-k)n}{s_1^2}\right]\left[\frac{nm(1-\rho_{ws}^2)+(1-m-k)n}{s_2^2(1-\rho_{ws}^2)}\right]}{\left[\frac{nm(1-\rho_{ws}^2)+(1-m-k)n}{s_2^2(1-\rho_{ws}^2)}\right]\left[\frac{nk(1-\rho_{ws}^2)+(1-m-k)n}{s_1^2(1-\rho_{ws}^2)}\right] - \left[\frac{(1-m-k)^2n^2\rho_{ws}^2}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}\right]}$$

$$= 1 - \frac{\frac{[nm(1-\rho_{ws}^2)+(1-m-k)n][nk+(1-m-k)n]}{s_1^2 s_2^2(1-\rho_{ws}^2)}}{\frac{[nm(1-\rho_{ws}^2)+(1-m-k)n][nk(1-\rho_{ws}^2)+(1-m-k)n]-(1-m-k)^2n^2\rho_{ws}^2}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}} \tag{6.18}$$

To simplify the BoS statistic further, a common denominator in both the numerator fraction and the denominator fraction are required. To achieve this multiply the numerator and denominator in the numerator fraction by $(1-\rho_{ws})$.

$$\text{BoS} = 1 - \frac{\frac{[nm(1-\rho_{ws}^2)+(1-m-k)n][nk+(1-m-k)n](1-\rho_{ws}^2)}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}}{\frac{[nm(1-\rho_{ws}^2)+(1-m-k)n][nk(1-\rho_{ws}^2)+(1-m-k)n]-(1-m-k)^2n^2\rho_{ws}^2}{s_1^2 s_2^2(1-\rho_{ws}^2)^2}}$$

$$= 1 - \frac{\Big[nm(1-\rho_{ws}^2) + (1-m-k)n\Big]\Big[nk + (1-m-k)n\Big](1-\rho_{ws}^2)}{\Big[nm(1-\rho_{ws}^2) + (1-m-k)n\Big]\Big[nk(1-\rho_{ws}^2) + (1-m-k)n\Big] - (1-m-k)^2 n^2 \rho_{ws}^2} \tag{6.19}$$

Simplify the BoS equation further by removing common factors and expanding out the brackets.

$$\text{BoS} = 1 - \frac{n^2(1-\rho_{ws}^2)\Big[m(1-\rho_{ws}^2) + (1-m-k)\Big]\Big[k + (1-m-k)\Big]}{n^2[m(1-\rho_{ws}^2) + (1-m-k)][k(1-\rho_{ws}^2) + (1-m-k)] - (1-m-k)^2 n^2 \rho_{ws}^2}$$

$$= 1 - \frac{n^2(1-\rho_{ws}^2)\Big[km(1-\rho_{ws}^2) + (1-m-k)m(1-\rho_{ws}^2) + (1-m-k)k + (1-m-k)^2\Big]}{n^2\{[m(1-\rho_{ws}^2) + (1-m-k)][k(1-\rho_{ws}^2 + (1-m-k)] - (1-m-k)^2\rho_{ws}^2\}}$$

$$= 1 - \frac{n^2(1-\rho_{ws}^2)\Big\{km(1-\rho_{ws}^2) + (1-m-k)[m(1-\rho_{ws}^2) + k + (1-m-k)]\Big\}}{n^2(1-k-m\rho_{ws}^2)(1-m-k\rho_{ws}^2) - (1-m-k)^2\rho_{ws}^2} \tag{6.20}$$

Cancel like terms from the numerator and denominator. The numerator will be in the simplest form. To simplify the denominator further begin to expand out the brackets.

$$\text{BoS} = 1 - \frac{(1-\rho_{ws}^2)\Big[km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)\Big]}{(1-k-m\rho_{ws}^2)(1-m-k\rho_{ws}^2) - (1-m-k-m+m^2+mk-k+mk+k^2)\rho_{ws}^2}$$

$$= 1 - \frac{(1-\rho_{ws}^2)\Big[km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)\Big]}{\Big[1 - m - k\rho_{ws}^2 - k + km + k^2\rho_{ws}^2 - m\rho_{ws}^2 + m^2\rho_{ws}^2 + km(\rho_{ws}^2)^2 - \rho_{ws}^2 + m\rho_{ws}^2 + k\rho_{ws}^2 + m\rho_{ws}^2 - m^2\rho_{ws}^2 - mk\rho_{ws}^2 + k\rho_{ws}^2 - mk\rho_{ws}^2 - k^2\rho_{ws}^2\Big]}$$

$$= 1 - \frac{(1-\rho_{ws}^2)\Big[km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)\Big]}{\Big[1 - m - k + km + km(\rho_{ws}^2)^2 - \rho_{ws}^2 + k\rho_{ws}^2 - mk\rho_{ws}^2 - mk\rho_{ws}^2 + m\rho_{ws}^2\Big]} \tag{6.21}$$

In the denominator, group like terms together and factorise.

$$\text{BoS} = 1 - \frac{(1-\rho_{ws}^2)\Big[km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)\Big]}{\Big[1 - \rho_{ws}^2 - m + m\rho_{ws}^2 - k + k\rho_{ws}^2 + km - 2mk\rho_{ws}^2 + km(\rho_{ws}^2)^2\Big]}$$

$$= 1 - \frac{(1-\rho_{ws}^2)\Big[km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)\Big]}{(1-\rho_{ws}^2) - m(1-\rho_{ws}^2) - k(1-\rho_{ws}^2) + km(1-\rho_{ws}^2)^2}$$

$$= 1 - \frac{(1-\rho_{ws}^2)\Big[km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)\Big]}{(1-\rho_{ws}^2)\Big[1 - m - k + km(1-\rho_{ws}^2)\Big]}$$

$$= 1 - \frac{km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)}{km(1-\rho_{ws}^2) + 1 - m - k} \tag{6.22}$$

To prove that the BoS statistic is bounded by the percentage of missing data for the outcome of interest, the method proof by contradiction will be used. A proof by contraction assumes that the converse statement is true and if this statement can be proven to be false/a contradiction then the original statement must be true. To prove that the BoS statistic is bounded by the percentage of missing data, assume that the converse is true. Therefore assume that:

$$\text{BoS} > m$$

$$1 - \frac{km(1-\rho_{ws}^2) + (1-m-k)[1-m\rho_{ws}^2]}{1-m-k+km(1-\rho_{ws}^2)} > m$$

$$-\frac{km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)}{1-m-k+km(1-\rho_{ws}^2)} > m-1$$

Multiply through by -1 and consequently change the direction of the inequality.

$$\frac{km(1-\rho_{ws}^2) + (1-m-k)(1-m\rho_{ws}^2)}{1-m-k+km(1-\rho_{ws}^2)} < 1-m$$

$$km(1-\rho_{ws}^2)+(1-m-k)(1-m\rho_{ws}^2)<(1-m)[1-m-k+km(1-\rho_{ws}^2)]$$

Expand the brackets on both sides of the inequality.

$$km(1-\rho_{ws}^2)+(1-m-k)(1-m\rho_{ws}^2)<km(1-\rho_{ws}^2)+(1-m-k)-km^2(1-\rho_{ws}^2)-m(1-m-k)$$

$$1-m-k-m\rho_{ws}^2+m^2\rho_{ws}^2+km\rho_{ws}^2<(1-m-k)-km^2(1-\rho_{ws}^2)-m(1-m-k)$$

$$-m\rho_{ws}^2+m^2\rho_{ws}^2+km\rho_{ws}^2<-km^2(1-\rho_{ws}^2)-m+m^2+mk$$

$$m-m\rho_{ws}^2+m^2\rho_{ws}^2-m^2+mk\rho_{ws}^2-mk+km(1-\rho_{ws}^2)<0$$

$$m(1-\rho_{ws}^2)-m^2(1-\rho_{ws}^2)-mk(1-\rho_{ws}^2)+km^2(1-\rho_{ws}^2)<0$$

$$m(1-\rho_{ws}^2)(1-m-k+mk)<0$$

$$m(1-\rho_{ws}^2)(1-m)(1-k)<0 \tag{6.23}$$

For this statement to be true, either one of the components (m, $(1-\rho_{ws}^2)$ $(1-m)$ and $(1-k)$) or three must be negative. Firstly, m is a proportion and ranges between $0 \leq m < 1$, therefore m can not be negative. Following, $(1-m)$ can not be negative. Recall, that k is the proportion of missing data for the alternative outcome (outcome 2) and ranges between $0 \leq k < 1$, therefore $(1-k)$ can not be negative. The square of the within-study correlation, ρ_{ws}^2, ranges between $0 \leq \rho_{ws}^2 \leq 1$ and therefore $1-\rho_{ws}^2$ is not negative. Therefore, there is a contraction to the truth of the statement that the BoS statistic can be greater than the proportion of missing data for the outcome of interest, BoS $> m$ (Equation 6.3). Consequently, the BoS statistic has been proved to be bounded by the proportion of missing data for the outcome of interest in a bivariate fixed-effect meta-analysis, if and only if the within-study variances and correlations are equal across all studies. ■

Theorem 6.3.4. *The borrowing of strength statistic from a bivariate random-effects meta-analysis is bounded by the percentage of missing data for the outcome of interest, if and only if the within-study variances and correlations are known to be equal across all studies.*

Proof. Recall the BoS statistic equation (Equation 5.5) containing the within-study and between-study variances and correlations:

$$\text{BoS} = 100\% \times \left\{ 1 - \frac{\sum_{i=1}^{n} \frac{1}{\hat{\tau}_1^2+s_{i1}^2} \sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2}}{\sum_{i=1}^{n} \frac{\hat{\tau}_1^2+s_{i1}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} \sum_{i=1}^{n} \frac{\hat{\tau}_2^2+s_{i2}^2}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} - \left[\sum_{i=1}^{n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2}}{(\hat{\tau}_1^2+s_{i1}^2)(\hat{\tau}_2^2+s_{i2}^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws_i}s_{i1}s_{i2})^2} \right]^2} \right\} \quad (5.5)$$

Due to the complexity of the BoS equation (Equation 5.5), consider each summation individually. Note, the second summation in the numerator is identical to the first summation in the denominator. In order to simplify the summations, the assumptions of equal

within-study variances and correlations across studies will be applied. Additionally, the simplification of the summations will utilise the limit as x tends to infinity , $\frac{1}{x}$ tends to zero.

The first summation simplified is the first summation of the numerator of BoS (Equation 5.5):

$$\sum_{i=1}^{n} \frac{1}{\hat{\tau}_1 + s_{i1}^2} = \sum_{i=1}^{mn} \frac{1}{\hat{\tau}_1 + s_{i1}^2} + \sum_{i=mn+1}^{(m+k)n} \frac{1}{\hat{\tau}_1 + s_{i1}^2} + \sum_{i=(m+k)n+1}^{n} \frac{1}{\hat{\tau}_1 + s_{i1}^2}$$

$$= \sum_{\cancel{i=1}}^{mn} \cancelto{0}{\frac{1}{\hat{\tau}_1 + s_{i1}^2}} + \frac{kn}{\hat{\tau}_1^2 + s_1^2} + \frac{(1-m-k)n}{\hat{\tau}_1^2 + s_1^2}$$

$$= \frac{(1-m)n}{\hat{\tau}_1^2 + s_1^2} \tag{6.24}$$

Secondly, the second summation in the numerator and the first summation in the denominator of BoS (Equation 5.5) is simplified, as follows:

$$\sum_{i=1}^{n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_1s_2)^2} = \sum_{i=1}^{mn} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \sum_{i=mn+1}^{(m+k)n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}$$
$$+ \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_1s_2)^2}$$

$$= \sum_{i=1}^{mn} \frac{1}{(\hat{\tau}_2^2 + s_{i2}^2) - \frac{(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}{(\hat{\tau}_1^2 + s_{i1}^2)}} + \sum_{i=mn+1}^{(m+k)n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} \to 0$$
$$+ \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_1s_2)^2}$$

$$= \sum_{i=1}^{mn} \frac{1}{(\hat{\tau}_2^2 + s_{i2}^2) - \frac{(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}{(\hat{\tau}_1^2 + s_{i1}^2)}} \to 0 + \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_1s_2)^2}$$

$$= \sum_{i=1}^{mn} \frac{1}{\hat{\tau}_2^2 + s_{i2}^2} + \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_1^2 + s_{i1}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_1s_2)^2}$$

$$= \frac{mn}{(\hat{\tau}_2^2 + s_2^2)} + \frac{(1-m-k)n(\hat{\tau}_1^2 + s_1^2)}{(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2} \tag{6.25}$$

The second summation of the denominator of Equation 5.5 is simplified as follows:

$$\sum_{i=1}^{n} \frac{\hat{\tau}_2^2 + s_{i2}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_1s_2)^2} = \sum_{i=1}^{mn} \frac{\hat{\tau}_2^2 + s_{i2}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \sum_{i=mn+1}^{(m+k)n} \frac{\hat{\tau}_2^2 + s_{i2}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}$$

$$+ \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_2^2 + s_{i2}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_1 s_2)^2}$$

$$= \sum_{i=1}^{mn} \frac{\hat{\tau}_2^2 + s_{i2}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \sum_{i=mn+1}^{(m+k)n} \frac{1}{(\hat{\tau}_1^2 + s_{i1}^2) - \frac{(\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}{(\hat{\tau}_2^2 + s_i 2^2)}}$$

$$+ \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_2^2 + s_{i2}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_{i1}s_{i2})^2}$$

$$= \sum_{i=mn+1}^{(m+k)n} \frac{1}{(\hat{\tau}_1^2 + s_{i1}^2)} + \sum_{i=(m+k)n+1}^{n} \frac{\hat{\tau}_2^2 + s_{i2}^2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_{i1}s_{i2})^2}$$

$$= \frac{kn}{(\hat{\tau}_1^2 + s_1^2)} + \frac{(1-m-k)n(\hat{\tau}_2^2 + s_2^2)}{(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1 s_2)^2} \tag{6.26}$$

Finally, the remaining summation of the denominator of Equation 5.5 is simplified:

$$\sum_{i=1}^{n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_{i1}s_{i2}}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_{i1}s_{i2})^2} = \sum_{i=1}^{mn} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \sum_{i=mn+1}^{(m+k)n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}$$

$$+ \sum_{i=(m+k)n+1}^{n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_{i1}s_{i2}}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_{i1}s_{i2})^2}$$

$$= \sum_{i=1}^{mn} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2} + \sum_{i=mn+1}^{(m+k)n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2)^2}$$
$$+ \sum_{i=(m+k)n+1}^{n} \frac{\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_{i1}s_{i2}}{(\hat{\tau}_1^2 + s_{i1}^2)(\hat{\tau}_2^2 + s_{i2}^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws_i}s_{i1}s_{i2})^2}$$

$$= \frac{(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)(1-m-k)n}{(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2} \tag{6.27}$$

All the summations from Equation 5.5, have been simplified. The simplified summations (Equations 6.24, 6.25, 6.26 and 6.27), are substituted back into the BoS statistic equation (Equation 5.5) and the equation is written as a proportion as opposed to a percentage, as follows:

$$\text{BoS} = 1 - \frac{\left[\frac{(1-m)n}{\hat{\tau}_1^2+s_1^2}\right]\left[\frac{mn}{\hat{\tau}_2^2+s_2^2} + \frac{(1-m-k)n(\hat{\tau}_1^2+s_1^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}\right]}{\left[\frac{mn}{\hat{\tau}_2^2+s_2^2} + \frac{(1-m-k)n(\hat{\tau}_1^2+s_1^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}\right]\left[\frac{kn}{(\hat{\tau}_1^2+s_1^2)} + \frac{(1-m-k)n(\hat{\tau}_2^2+s_2^2)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}\right] - \left[\frac{(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)(1-m-k)n}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}\right]^2} \tag{6.28}$$

Expand and simplify equation 6.28 using algebra:

$$\text{BoS} = 1 - \frac{\frac{mn^2(1-m)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)} + \frac{(1-m)n^2(1-m-k)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}}{\frac{mkn^2}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)} + \frac{(1-m-k)mn^2}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2} + \frac{(1-m-k)kn^2}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2} + \frac{(1-m-k)^2n^2(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)}{[(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2]^2} - \frac{(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2(1-m-k)^2n^2}{[(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2]^2}}$$

$$= 1 - \frac{\frac{mn^2(1-m)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)} + \frac{(1-m)n^2(1-m-k)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}}{\frac{mkn^2}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)} + \frac{(1-m-k)n^2(m+k)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2} + \frac{(1-m-k)^2n^2[(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2]}{[(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2]^2}}$$

$$= 1 - \frac{\frac{mn^2(1-m)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)} + \frac{(1-m)n^2(1-m-k)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}}{\frac{mkn^2}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)} + \frac{(1-m-k)n^2[(m+k)+(1-m-k)]}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}}$$

$$= 1 - \frac{\frac{mn^2(1-m)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)} + \frac{(1-m)n^2(1-m-k)}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}}{\frac{mkn^2}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)} + \frac{(1-m-k)n^2}{(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2}} \tag{6.29}$$

To simplify, Equation 6.29, generate a common denominator in both the numerator and denominator fractions, which can then be cancelled out. The equation can then be simplified further:

$$\text{BoS} = 1 - \frac{\frac{mn^2(1-m)[(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2]+(1-m)n^2(1-m-k)(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)}{[(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2](\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)}}{\frac{mkn^2[(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2]+(1-m-k)n^2(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)}{[(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)-(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2](\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)}}$$

$$= 1 - \frac{n^2\{m(1-m)[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2] + (1-m)(1-m-k)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)\}}{n^2\{mk[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2] + (1-m-k)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)\}}$$

$$= 1 - \frac{m(1-m)[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2] + (1-m-k-m+m^2+mk)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)}{mk[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2] + (1-m-k)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)}$$

$$= 1 - \frac{\begin{array}{r} m(1-m)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - m(1-m)(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2 - m(1-m)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) \\ + (1-m-k+mk)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) \end{array}}{(1-m-k+mk)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - mk(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2}$$

$$= 1 - \frac{(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)(1-m)(1-k) - m(1-m)(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2}{(1-m)(1-k)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - mk(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2} \tag{6.30}$$

To prove the theorem the BoS statistic is bounded by the percentage of missing data for the outcome of interest (Theorem 6.3.4), the method proof by contradiction will be applied here. To set up the proof by contradiction, assume that the converse of the theorem statement is true. Therefore, assume that:

$$\text{BoS} > m$$

$$1 - \frac{(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)(1-m)(1-k) - m(1-m)(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2}{(1-m)(1-k)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - mk(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2} > m$$

$$\frac{(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)(1-m)(1-k) - m(1-m)(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2}{(1-m)(1-k)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - mk(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2} < 1 - m$$

$$(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)(1-m)(1-k) - m(1-m)(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2 < (1-m)[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2)(1-m)(1-k) - mk(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2]$$

$$(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)(1-m)(1-k)-(1-m)(\hat{\tau}_1^2+s_1^2)(\hat{\tau}_2^2+s_2^2)(1-m)(1-k) < m(1-m)(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2-(1-m)mk(\rho_{bs}\hat{\tau}_1\hat{\tau}_2+\rho_{ws}s_1s_2)^2$$

$$m(1-m)(1-k)(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) < m(1-m)(1-k)(\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2$$

$$m(1-m)(1-k)[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2] < 0 \tag{6.31}$$

The statement "the BoS statistic is greater than the percentage of missing data for the outcome of interest" (BoS $> m$) is true if and only if Equation 6.31 is true. For this equation to hold one or three of m, $(1-m)$, $(1-k)$ and $[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2]$ must be negative.

The proportion of missing data for the outcome of interest, m ranges between $0 \leq m < 1$ and therefore m and $(1-m)$ can not be negative. Similarly, k is the proportion of missing data for the alternative outcome and ranges between $0 \leq k < 1$, thus $(1-k)$ can not be negative. The sign of $[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2]$ will be investigated by assuming that it is negative and investigating whether this assumption holds. Following, assume that:

$$(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2 < 0 \tag{6.32}$$

$$\hat{\tau}_1^2\hat{\tau}_2^2 + s_1^2s_2^2 + \hat{\tau}_2^2s_1^2 + \hat{\tau}_1^2s_2^2 - \rho_{bs}^2\hat{\tau}_1^2\hat{\tau}_2^2 - \rho_{ws}^2s_1^2s_2^2 - 2\rho_{ws}\rho_{bs}s_1s_2\hat{\tau}_1\hat{\tau}_2 < 0$$

$$\hat{\tau}_1^2\hat{\tau}_2^2(1-\rho_{bs}^2) + s_1^2s_2^2(1-\rho_{ws}^2) + s_1^2\hat{\tau}_2^2 + s_2^2\hat{\tau}_1^2 - 2\rho_{ws}\rho_{bs}s_1s_2\hat{\tau}_1\hat{\tau}_2 < 0 \tag{6.33}$$

The between-study and the within-study variances, $\hat{\tau}_1^2$, $\hat{\tau}_2^2$, s_1^2 and s_2^2 are positive and the square of the within-study and between-study correlations are positive and range between $0 \leq \rho_{bs}^2, \rho_{ws}^2 \leq 1$. Therefore, $\hat{\tau}_1^2\hat{\tau}_2^2(1-\rho_{bs}^2) + s_1^2s_2^2(1-\rho_{ws}^2)$ are positive. The sign of $s_1^2\hat{\tau}_2^2 + s_2^2\hat{\tau}_1^2 - 2\rho_{ws}\rho_{bs}s_1s_2\hat{\tau}_1\hat{\tau}_2$ needs to be determined, since if it is positive the assumption in equation 6.32 is false. Thus, there will be a contradiction to the statement that the BoS statistic is greater than the percentage of missing data for the outcome of interest. Following

a proof by contradiction set up, we assume that it is negative:

$$s_1^2\hat{\tau}_2^2 + s_2^2\hat{\tau}_1^2 - 2\rho_{ws}\rho_{bs}s_1s_2\hat{\tau}_1\hat{\tau}_2 < 0 \tag{6.34}$$

Alternatively, Equation 6.34 can be written by considering the signs and magnitudes of the between-study and within-study correlations; either

$$s_1^2\hat{\tau}_2^2 + s_2^2\hat{\tau}_1^2 + 2|\rho_{ws}||\rho_{bs}|s_1s_2\hat{\tau}_1\hat{\tau}_2 < 0 \tag{6.35}$$

when either ρ_{ws} or ρ_{bs} is negative (but not both), or

$$s_1^2\hat{\tau}_2^2 + s_2^2\hat{\tau}_1^2 - 2|\rho_{ws}||\rho_{bs}|s_1s_2\hat{\tau}_1\hat{\tau}_2 < 0 \tag{6.36}$$

when both ρ_{ws} and ρ_{bs} have the same sign either positive or negative.

The first equation (Equation 6.35) when either the between-study correlation or within-study correlation is negative can not be a true statement since the right hand side can not be negative. Therefore, Equation 6.34 can only be true if and only if Equation 6.36 is true.

To further investigate whether these equations are true, consider the following quadratic equation:

$$(s_1\hat{\tau}_2 - s_2\hat{\tau}_1)^2 = s_1^2\hat{\tau}_2^2 + s_2^2\hat{\tau}_1^2 - 2s_1s_2\hat{\tau}_1\hat{\tau}_2 \tag{6.37}$$

This can be rearranged, such that

$$s_1^2\hat{\tau}_2^2 + s_2^2\hat{\tau}_1^2 = (s_1\hat{\tau}_2 - s_2\hat{\tau}_1)^2 + 2s_1\hat{\tau}_2^2 s_2\hat{\tau}_1 \tag{6.38}$$

which can be substituted into Equation 6.36:

$$(s_1\hat{\tau}_2 - s_2\hat{\tau}_1)^2 + 2s_1s_2\hat{\tau}_1\hat{\tau}_2 - 2|\rho_{ws}||\rho_{bs}|s_1s_2\hat{\tau}_1\hat{\tau}_2 < 0 \tag{6.39}$$

$$(s_1\hat{\tau}_2 - s_2\hat{\tau}_1)^2 + 2s_1s_2\hat{\tau}_1\hat{\tau}_2(1 - |\rho_{bs}||\rho_{ws}|) < 0 \tag{6.40}$$

It will now be investigated whether this equation holds. The quadratic $(s_1\hat{\tau}_2 - s_2\hat{\tau}_1)^2$ can not be negative. For the remaining component of Equation 6.40, $2s_1s_2\hat{\tau}_1\hat{\tau}_2(1 - |\rho_{bs}||\rho_{ws}|)$, it is known that the within-study and between-study standard deviations (s_1, s_2, $\hat{\tau}_1$ and $\hat{\tau}_2$) are positive and therefore, the sign is the component is determined by the sign of $(1 - |\rho_{bs}||\rho_{ws}|)$. The moduli of the within-study and between-study correlations, $|\rho_{bs}|$ and $|\rho_{ws}|$ respectively, range between $0 \leq |\rho_{bs}|, |\rho_{ws}| \leq 1$. Therefore, the sign of $(1 - |\rho_{bs}||\rho_{ws}|)$ is not negative since $0 \leq (1 - |\rho_{bs}|, |\rho_{ws}|) \leq 1$. As a result, Equation 6.40,

$$(s_1\hat{\tau}_2 - s_2\hat{\tau}_1)^2 + 2s_1s_2\hat{\tau}_1\hat{\tau}_2(1 - |\rho_{bs}||\rho_{ws}|) < 0 \tag{6.41}$$

is a contradiction, since it can not be negative. Therefore, Equation 6.32,

$$(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2 < 0 \tag{6.42}$$

is also a contradiction, since it can not be negative. Furthermore, all the components of Equation 6.31, m, $(1-m)$, $(1-k)$ and $[(\hat{\tau}_1^2 + s_1^2)(\hat{\tau}_2^2 + s_2^2) - (\rho_{bs}\hat{\tau}_1\hat{\tau}_2 + \rho_{ws}s_1s_2)^2]$ are positive, which results in a contradiction in the statement BoS $> m$. Therefore, for a bivariate random-effects meta-analysis the BoS statistic is bounded by the percentage of missing data for the outcome of interest, if and only if, the within-study correlations and variances are equal across all studies. ■

6.4 Discussion

In this chapter, the relationship between the BoS and the percentage of missing data for the outcome of interest was investigated using mathematical proofs. During the exploration of the meta-analysis characteristics in the interactive tool developed using the equation for BoS in Chapter 5, the magnitude of BoS was observed to never exceed the percentage of missing data for the outcome of interest. Through a series of mathematical proofs the BoS statistic was proved to be bounded by the percentage of missing data for the outcome of interest, under certain conditions. In this section, the key findings (summarised in Figure 6.1) are discussed and areas for future research are described.

Figure 6.1: Key findings from the interactive tool and mathematical proofs

Key Findings:

- Through a simplified example, of a univariate meta-analysis and its update, the plausibility of the boundedness of the statistic was discussed.
- The BoS statistic was proved to be bounded by the percentage of missing data for the outcome of interest under a fixed-effect meta-analysis assumption, under the conditions of equal within-study variances and correlations across all studies for the outcomes that are reported (Theorems 6.3.1 and 6.3.3).
- The BoS statistic was proved to be bounded by the percentage of missing data for the outcome of interest under a random-effects assumption, in addition to the conditions of equal within-study variances and correlations across all studies for the outcomes reported (Theorem 6.3.4).

In the first instance, the BoS statistic from a fixed-effect meta-analysis was proved to be bounded by the percentage of missing data for the outcome of interest when there was no missing data for the alternative outcome and secondly when there was missing

data for the alternative outcome. The proof was extended for the random-effects meta-analysis and included the between-study correlation and the between-study variances. For all the proofs some assumptions were made about the within-study correlations and the variances. The within-study correlations were assumed to be equal across all studies in the meta-analysis. The variances for the outcome of interest in each study were assumed to be equal and the variances for the alternative outcome were also assumed to be equal. For any variances that were missing, these were assumed to be infinity (Riley 2009, Jackson et al. 2011). It is under these conditions that BoS was proved to be bounded by the percentage of missing data for the outcome of interest under a fixed or random-effects assumption.

6.4.1 Limitations

These conditions were made to aid simplicity. However, they are unlikely to hold for real data analysed in a meta-analysis. It might be expected that provided the variances and within-study correlations are similar in each study the BoS statistic might still be bounded by the percentage of missing data for the outcome of interest. A contrived example can be used to show that it is possible for the BoS to exceed the percentage of missing data when the condition of equal variances does not hold. Consider a fixed-effect bivariate meta-analysis with three studies, two containing complete data and one with missing data for the outcome of interest. The within-study variances and within-study correlations for the three studies are shown in Table 6.1.

Table 6.1: Contrived example, for demonstration purposes, of a bivariate fixed-effect meta-analysis with three studies

study number	Within-study variances		Within-study
	s_1^2	s_2^2	correlation (ρ_{ws})
1	5	18	0.9
2	20	5	0.9
3	missing $\approx$ 10000	9	0

For the missing data in study three, the missing variance is assumed to be large, say 10000, and the within-study correlation is zero. In this example, the percentage of missing data for the outcome of interest is 33.3% and the BoS statistic is calculated as follows:

$$\text{BoS} = 100\% \times \left\{ 1 - \frac{\sum_{i=1}^{n} \frac{1}{s_{i1}^2} \sum_{i=1}^{n} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)}}{\sum_{i=1}^{n} \frac{1}{s_{i2}^2(1-\rho_{ws_i}^2)} \sum_{i=1}^{n} \frac{1}{s_{i1}^2(1-\rho_{ws_i}^2)} - \left[\sum_{i=1}^{n} \frac{\rho_{ws_i}}{s_{i1}s_{i2}(1-\rho_{ws_i}^2)} \right]^2} \right\}$$

$$= 100\% \times \left\{ 1 - \frac{\left[\frac{1}{5} + \frac{1}{20} + \frac{1}{10000}\right]\left[\frac{1}{18(1-0.9^2)} + \frac{1}{5(1-0.9^2)} + \frac{1}{9}\right]}{\left[\frac{1}{18(1-0.9^2)} + \frac{1}{5(1-0.9^2)} + \frac{1}{9}\right]\left[\frac{1}{5(1-0.9^2)} + \frac{1}{20(1-0.9^2)} + \frac{1}{10000}\right] - \left[\frac{0.9}{\sqrt{5}\sqrt{18}(1-0.9^2)} + \frac{0.9}{\sqrt{20}\sqrt{5}(1-0.9^2)}\right]^2} \right\}$$

$$= 62.43\% \tag{6.43}$$

It can be seen in this example that the BoS (62.43%) is greater than the percentage of missing data for the outcome of interest (33.3%). This is only one example of a violation of one condition resulting in the BoS statistic no longer being bounded by the percentage of missing data for the outcome of interest. The boundedness may also not hold when other conditions are violated. Nevertheless, the findings in the chapter demonstrate useful results that hold when the conditions assumed are true.

6.4.2 The relationship between the borrowing of strength and missing data

This chapter proved under certain conditions the BoS statistic in a bivariate meta-analysis is bounded by the percentage of missing data for the outcome of interest. Alternatively, Copas et al. (2018) investigated the effect missing data has upon the borrowing of strength attributed to an individual study in a bivariate meta-analysis. Copas et al. (2018) discussed that the BoS given to a particular study is dependent upon

how similar its characteristics are to the other studies in the meta-analysis. Consider a bivariate meta-analysis that contains studies with complete data and studies with a missing outcome. The studies with complete data are not altered but the given BoS alters compared to the same meta-analysis with complete data for all studies of the same size. Copas et al. (2018) showed that the sum of the given BoS for the studies with complete data is affected by the remaining studies in the meta-analysis. The given BoS increases when the remaining studies are missing the outcome of interest and decreases when the alternative outcome is missing, compared to complete data in the studies. Similarly, in this chapter the interactive tool showed that the BoS statistic in the meta-analysis increased with increasing proportions of studies with the outcome of interest missing and decreased when the proportion of studies with the alternative outcome missing increases.

6.4.3 Strengths

This chapter has many strengths including providing a greater understanding of which characteristics of meta-analyses affect the magnitude of BoS. This chapter builds upon the findings and investigations into the BoS statistic by Jackson et al. (2017) and Copas et al. (2018). A greater understanding of BoS will help to identify situations in which the BoS is considerably large. This will help researchers predict what magnitude of BoS they might expect from their meta-analysis and in turn could help them to decide if a multivariate approach would be beneficial over a univariate approach.

A focus of this chapter was the relationship between the borrowing of strength statistic and the percentage of missing data for the outcome of interest, which complements the models developed in Chapter 4. The relationship has been detailed in a simplistic usable manner; the BoS is bounded by the percentage of missing data for the outcome of interest under particular conditions. This provides a very simple 'rule of thumb' that some researchers may find useful, i.e. consider multivariate meta-analysis when the percentage of missing data for the outcome of interest is large.

Researchers might consider both the simple rule of thumb that BoS is bounded by the percentage of missing data for the outcome of interest and also the value predicted from the models in Chapter 4. Conversely, for complete data, it may be unlikely to be worthwhile.

6.4.4 Further work

This chapter proved under certain conditions (equal within-study variances and correlations across studies) the BoS statistic is bounded by the percentage of missing data for the outcome of interest. However this chapter did not investigate the consequences of violating the conditions for the boundedness. It is important to understand the impacts violating the conditions have on the boundedness relationship. This is particularly important if it is recommended that the magnitude of BoS is considered in determining whether a multivariate meta-analysis is beneficial in that setting.

Next steps

In the next chapter, a novel application of multivariate meta-analysis is considered, where missing data for the key outcome is often a problem, such that, multivariate meta-analysis may have considerable BoS.

Chapter 7

A new bivariate approach for meta-analysis of continuous outcomes from randomised trials providing a mixture of final score and analysis of covariance treatment effect estimates

In the previous chapters the benefits of a multivariate meta-analysis over multiple univariate meta-analyses have been explored. The results from multivariate meta-analyses have been compared to the results from univariate meta-analyses. In Chapter 2, the multivariate meta-analysis results and the univariate meta-analysis results were often similar in an applied example that extended a previous HTA report. However, there were incidences in which potential clinical differences arose between the results and the Borrowing of Strength (BoS) statistic was larger for these outcomes. Following in Chapter 3, the magnitude of BoS was found to identify meta-analysis studies where the multivariate meta-analysis is beneficial over the univariate in terms of gain in precision and potentially stronger/different conclusions. The focus of the thesis in Chapter 4 and 6 was the further understanding of the BoS statistic and meta-analysis level characteristics that influence the magnitude of BoS. In the next two chapters, the thesis returns to investigate further potential benefits of the multivariate meta-analysis approach in a novel application.

7.1 Context and aims

A randomised control trial (RCT) is considered to be the best design for the comparison of treatments (Aiello et al. 2011), with a meta-analysis of RCTs considered to be the "gold standard" for summarising the evidence about treatment effects (Trowman et al. 2007). In RCTs, continuous outcomes (e.g. blood pressure, weight) are usually measured at baseline (before the treatment period) and at follow up (during or at the end of the treatment period). The treatment effect measure of interest for a continuous outcome is usually the mean difference (i.e. treatment versus control) in the outcome at follow-up. However, this can be estimated using a variety of different analytical approaches, depending on if and how the baseline variable (i.e. the continuous outcome at baseline) is adjusted for. The most commonly used methods are the final score model, which does not adjust for the baseline variable; the change score model, which accounts for the baseline values by modelling the change in the outcome between follow up and baseline; and an ANCOVA (analysis of covariance) model, which models the follow-up outcome value in a linear regression that adjusts for the baseline variable using covariate adjustment. These three methods were introduced in Chapter 1 in Section 1.3.2. The ANCOVA model is considered to be the method of choice for the estimation of the mean treatment difference (Vickers & Altman 2001, Van Breukelen 2006, Zhang et al. 2014).

When multiple RCTs are available that evaluate the same treatment in terms of a particular continuous outcome, then different studies may have used different methods for the analysis, potentially. Where possible, it is recommended that the ANCOVA treatment effect estimates should be combined in the meta-analysis (Riley et al. 2013, Fu et al. 2013). However, the researcher is restricted by the treatment effect estimates reported unless they can be derived using individual participant data (IPD), which is unlikely to be available for all trials. The question for the researcher is therefore: what is the best way to combine treatment effect estimates for continuous outcomes in a meta-analysis, given that the method to derive them

will vary across trials? A previously proposed meta-analysis approach suggests a univariate meta-analysis model where a single treatment effect from each trial is obtained regardless of the continuous outcome modelling method (Higgins & Green 2008). This allows, for example, for treatment effect estimates from trials using ANCOVA models to be combined with treatment effect estimates from trials using the final score approach.

An interesting further issue arises when some trials report more than one treatment effect estimate based on two or more of the different estimation methods (e.g. ANCOVA and final score). For example, a trial may report the estimate of mean difference based on a final score model, and also report the estimate from an ANCOVA model. Furthermore, if IPD are available for a particular trial then the meta-analysts themselves can calculate treatment effect estimates for each of the different models using the trial's IPD. These estimates will be correlated with each other and therefore may provide an opportunity to gain more information for the meta-analysis results.

Therefore, the aim of this chapter is to consider a new multivariate meta-analysis approach for synthesising treatment effect estimates from RCTs of continuous outcomes. The multivariate approach allows for the joint synthesis of multiple treatment effect estimates from the same trial, to borrow strength for each effect estimate by utilising additional information from the correlation between the effect estimates (Raudenbush et al. 1988, Becker 2000, van Houwelingen et al. 2002, Jackson et al. 2011). Particular interest lies with the premise that the ultimate aim is to produce a summary result based on ANCOVA effect estimates, and that in trials where ANCOVA results are not available, information could be borrowed from other effect estimates (e.g. final score) that are available.

The structure of this chapter hereby follows with an introduction to the analysis approach of a single RCT with a continuous outcome, using final score, change

score and ANCOVA. The current univariate meta-analysis approaches are then described and the new multivariate meta-analysis approach introduced. The chapter will follow with an example dataset to illustrate the use of the methods as well as facilitate comparison between the approaches. A detailed simulation study will be described in the subsequent chapter.

7.2 Methods

7.2.1 Models for analysis of continuous data in a randomised control trial

The setting for the models described in this section is a single parallel-group RCT where the study participants are randomly allocated to either a treatment group or a control group. A baseline value for each participant is recorded prior to the treatment period and similarly, a follow up value is recorded following the treatment period for each participant. The models are fitted to this data to obtain the estimated mean treatment effect and its standard error, which could then contribute to a subsequent meta-analysis.

The three potential models (change score, final score and ANCOVA) are described below. Each model uses a linear regression framework including the treatment group as a covariate, but they model the baseline and final values of the outcome in different ways. Although these were introduced in the Introduction (Chapter 1), they are shown again below for completeness for this chapter.

Change score model

The change score approach models the difference between the final value and the baseline value. For a simple, unadjusted analysis of the treatment effect for an individual

trial, i, the model can be written as:

$$y_{F_{ij}} - y_{B_{ij}} = \phi_i + \theta_{Ci} x_{ij} + e_{ij} \qquad e_{ij} \sim \mathrm{N}(0, \sigma_i^2) \tag{7.1}$$

where F represents the final score, B the baseline and C is the change score. $y_{F_{ij}}$ and $y_{B_{ij}}$ are the follow up and baseline values, respectively, for participant j in trial i and x_{ij} is 0/1 for participants in the control/treatment group. θ_{Ci} is the change score model's treatment effect (mean difference in the change score between the treatment and control) for study i, and ϕ_i is the fixed trial effect, which represents the control group's mean change score in this case. The residual error, e_{ij}, is assumed to be normally distributed with mean 0 and variance σ_i^2.

Final score model

The final score model is similar to the change score model, except the final score model does not account for the baseline values. The final score model can be written as:

$$y_{F_{ij}} = \phi_i + \theta_{Fi} x_{ij} + e_{ij} \qquad e_{ij} \sim \mathrm{N}(0, \sigma_i^2) \tag{7.2}$$

where F represents the final score and $y_{F_{ij}}$ is the final measurement value from participant j in trial i. The intercept, ϕ_i, is the mean final score for the control group in trial i. The treatment effect for the final score model for trial i is θ_{Fi}. The residual error, e_{ij} and the treatment group indicator, x_{ij} are the same as the change score model.

ANCOVA model

An ANCOVA model adjusts for the baseline measurement values while regressing the final value.

$$y_{F_{ij}} = \phi_i + \beta_i y_{B_{ij}} + \theta_{Ai} x_{ij} + e_{ij} \qquad e_{ij} \sim \mathrm{N}(0, \sigma_i^2) \tag{7.3}$$

where F represents the final score, B is the baseline and A represents the ANCOVA model. The interpretation of β_i is the effect of a 1-unit increase in the baseline value on the final value for study i. The final measurement value and baseline measurement

value for participant j in study i are denoted as $y_{F_{ij}}$ and $y_{B_{ij}}$, respectively. The treatment effect for the ANCOVA model is θ_{Ai} and ϕ_i is the intercept for study i. The residual error, e_{ij}, and the treatment group indicator, x_{ij}, are as defined previously.

7.2.2 Choice of model for continuous data and baseline imbalance

The statistical properties of the three approaches have been extensively reviewed (Senn 1993, Wei & Zhang 2001). The ANCOVA is considered to be the best approach for analysing continuous outcomes from RCTs compared to the final score and the change score models since it has greater statistical power to detect a treatment effect (Vickers & Altman 2001) whilst accounting for the correlation between baseline and follow-up values (unlike final score) and adjusting for the starting value (unlike change score and final score). It yields unbiased treatment effect estimates and maintains the nominal type I error rate when there is baseline imbalance that occurs by chance (Senn 1993, Wei & Zhang 2001, Fu & Holmer 2016).

Baseline imbalance occurs when the mean of the baseline values for treatment and control groups is different for one or more covariates in a trial (Fu & Holmer 2016). When there is imbalance in the continuous outcome at baseline, it can lead to biased treatment effect estimates for both the final score and change score analyses. The final score model ignores the baseline values, consequently when there is baseline imbalance the model does not account for the imbalance, which results in biased treatment effect estimates. The change score model is often considered to account for baseline imbalance, however this is incorrect (Vickers & Altman 2001). The change score model contains the baseline values, but it does not account for baseline imbalance as it simply calculates the change score from the baseline and final values. Thus the magnitude of the baseline value is lost when the change value is calculated and similarly the correlation between the change value and baseline value is also lost. The ANCOVA model is the only model that adjusts for the baseline values and

thereby accounts for baseline imbalances across treatment groups.

When multiple RCTs are to be synthesised, and there is little or no baseline imbalance in all RCTs, the choice of analytical approach in each trial becomes less important and the impact on the meta-analysis summary estimates may be minor (Riley et al. 2013). However, when there is baseline imbalance in some of the trials included, the results from the meta-analysis may be more likely to produce treatment effects that are biased (Trowman et al. 2007). Some argue that baseline imbalance is not a concern since its impact is likely to be greater within an individual trial than across trials (McKenzie et al. 2016, Wei & Zhang 2001); for example, baseline imbalances may vary randomly around zero when averaged over similar trials. However, this may not always be the case, since there may be systematic differences in the direction of imbalance across trials, or the trials with a particular direction of imbalance may have larger study weights. Therefore, in situations with baseline imbalance it is recommended to synthesise ANCOVA derived treatment effect estimates rather than change score or final score treatment effect estimates, wherever possible. This forms the premise for the new meta-analysis method to summarise ANCOVA results, to be introduced in Section 7.2.6.

7.2.3 Mathematical relationship between the ANCOVA, change score and final score treatment effects

The relationship between the ANCOVA, change score and the final score treatment effects in a single trial can be expressed mathematically as (Senn 1993, McKenzie et al. 2016):

$$\theta_A = (1 - \lambda) * \theta_F + \lambda * \theta_C \tag{7.4}$$

where θ_A, θ_F and θ_C are the treatment effects from the ANCOVA (A), final score (F) and change score (C) models, respectively. The term λ is the slope coefficient from the regression of the final measurement values from participants on the baseline

measurement values from participants, written as:

$$y_{F_j} = \alpha + \lambda y_{B_j} \tag{7.5}$$

where α is the intercept, y_{F_j} is the final measurement value and y_{B_j} is the baseline measurement value for participant j. This can be derived when a trial's IPD is available. Alternatively, λ can be calculated using the following equation:

$$\lambda = \rho_{FB} \frac{sd_F}{sd_B} \tag{7.6}$$

where ρ_{FB} is the correlation between the final and baseline measurement values for the participants in the trial, sd_F and sd_B, are the standard deviations of the final measurements and baseline measurements, respectively. When the variances for the final and baseline are equal, λ is equal to ρ_{FB} (Senn 1993).

To provide further insight into the relationship of ANCOVA, final score and change score treatment effect estimates, some theorems and lemmas are now introduced and proved.

Theorem 7.2.1. *The value of λ influences the relationship between the ANCOVA, final score and change score treatment effects.*

For example, when the value of λ is between 0 and 1, the value of ANCOVA lies between the final score and change score values. This can be proved, from Equation 7.4.

Proof of Theorem 7.2.1

Recall Equation 7.4, which describes the relationship between the ANCOVA, change score and final score treatment effects.

$$\theta_A = (1 - \lambda) * \theta_F + \lambda * \theta_C$$

Let θ_A, θ_F and θ_C be the treatment effects from the ANCOVA, final score and change score models, respectively. Also, let $x = 1 - \lambda$ and $y = \lambda$.

Lemma 7.2.2. *The treatment effects from the ANCOVA and final score models are equal when* $\lambda = 0$.

Consider the situation where $\lambda = 0$, then $x = 1$ and $y = 0$, therefore $\theta_A = \theta_F$.

Lemma 7.2.3. *The treatment effects from the ANCOVA and change score models are equal when* $\lambda = 1$.

Consider the situation where $\lambda = 1$, then $x = 0$ and $y = 1$, therefore $\theta_A = \theta_C$.

Lemma 7.2.4. *The value of the ANCOVA treatment effect lies between the final score and change score treatment effect values when* λ *is between zero and one.*

Consider the situation where $0 < \lambda < 1$, then $0 < x, y < 1$.
If $\theta_F > \theta_C$,

$$\begin{aligned} &\theta_A = x\theta_F + y\theta_C && \text{and} && \theta_A = x\theta_F + y\theta_C \\ &\theta_A > x\theta_C + y\theta_C && && \theta_A < x\theta_F + y\theta_F \\ &\theta_A > (x+y)\theta_C && && \theta_A < (x+y)\theta_F \\ &\theta_A > \theta_C && && \theta_A < \theta_F \end{aligned}$$

$$\therefore \theta_C < \theta_A < \theta_F \tag{7.7}$$

If $\theta_F < \theta_C$,

$$\begin{aligned} &\theta_A = x\theta_F + y\theta_C && \text{and} && \theta_A = x\theta_F + y\theta_C \\ &\theta_A < x\theta_C + y\theta_C && && \theta_A > x\theta_F + y\theta_F \\ &\theta_A < (x+y)\theta_C && && \theta_A > (x+y)\theta_F \\ &\theta_A > \theta_C && && \theta_A > \theta_F \end{aligned}$$

$$\therefore \theta_C > \theta_A > \theta_F \tag{7.8}$$

Therefore, the ANCOVA treatment effect lies between the treatment effect for the final score and the change score, when the regression coefficient (λ) from the final values regressed on the baseline values is between 0 and 1.

Lemma 7.2.5. *The value of the change score treatment effect lies between the final score and ANCOVA treatment effect values when λ is greater than one.*

Consider the situation where $\lambda > 1$, then x is negative and y is positive.
If $\theta_F > \theta_C$,

$$\begin{aligned} \theta_A &= x\theta_F + y\theta_C && \text{and} && \theta_A = x\theta_F + y\theta_C \\ \theta_A &< x\theta_C + y\theta_C && && \theta_A < x\theta_F + y\theta_F \\ \theta_A &< (x+y)\theta_C && && \theta_A < (x+y)\theta_F \\ \theta_A &< \theta_C && && \theta_A < \theta_F \end{aligned}$$

$$\therefore \theta_A < \theta_C < \theta_F \tag{7.9}$$

If $\theta_F < \theta_C$,

$$\begin{aligned} \theta_A &= x\theta_F + y\theta_C && \text{and} && \theta_A = x\theta_F + y\theta_C \\ \theta_A &> x\theta_C + y\theta_C && && \theta_A > x\theta_F + y\theta_F \\ \theta_A &> (x+y)\theta_C && && \theta_A > (x+y)\theta_F \\ \theta_A &> \theta_C && && \theta_A > \theta_F \end{aligned}$$

$$\therefore \theta_A > \theta_C > \theta_F \tag{7.10}$$

Lemma 7.2.6. *The value of the final score treatment effect lies between the change score and ANCOVA treatment effect values when λ is less than zero.*

Consider the situation where $\lambda < 0$, then x is positive and y is negative.

If $\theta_F > \theta_C$,

$$\begin{aligned} \theta_A &= x\theta_F + y\theta_C \qquad \text{and} \qquad \theta_A = x\theta_F + y\theta_C \\ \theta_A &> x\theta_C + y\theta_C \qquad\qquad\quad \theta_A > x\theta_F + y\theta_F \\ \theta_A &> (x+y)\theta_C \qquad\qquad\quad \theta_A > (x+y)\theta_F \\ \theta_A &> \theta_C \qquad\qquad\qquad\quad \theta_A > \theta_F \end{aligned}$$

$$\therefore \theta_A > \theta_F > \theta_C \tag{7.11}$$

If $\theta_F < \theta_C$,

$$\begin{aligned} \theta_A &= x\theta_F + y\theta_C \qquad \text{and} \qquad \theta_A = x\theta_F + y\theta_C \\ \theta_A &< x\theta_C + y\theta_C \qquad\qquad\quad \theta_A < x\theta_F + y\theta_F \\ \theta_A &< (x+y)\theta_C \qquad\qquad\quad \theta_A < (x+y)\theta_F \\ \theta_A &< \theta_C \qquad\qquad\qquad\quad \theta_A < \theta_F \end{aligned}$$

$$\therefore \theta_A < \theta_F < \theta_C \tag{7.12}$$

Therefore, the ANCOVA treatment effect does not lie in between the treatment effects for final score and the change score when $\lambda < 0$ or $\lambda > 1$, as it is either the largest or smallest in value of the three results. This situation is likely to occur in trials if the correlation between baseline and final is large, and the variation in final outcome values is considerably greater than the variation in baseline outcomes values (as recall Equation 7.6 says that $\lambda = \rho_{FB}\frac{sd_F}{sd_B}$). It may also occur if there is a negative correlation between baseline and follow-up values. However, in most RCTs $0 < \lambda < 1$.

7.2.4 Univariate meta-analysis: a current method for meta-analysis

The current methods for a meta-analysis of treatment effect estimates for a continuous outcome involve a univariate meta-analysis approach; that is one estimate per trial

is used in the analysis. These approaches are now introduced using a random-effects specification, since heterogeneity is usually expected in a meta-analysis.

Separate Univariate meta-analysis

The first current option is separate univariate meta-analyses for each set of treatment effect estimates using an ANCOVA, change score and final score model. For the ANCOVA model, the separate univariate meta-analysis can be expressed as:

$$\begin{aligned} \hat{\theta}_{Ai} &\sim \mathrm{N}(\theta_{Ai}, s^2_{Ai}) \\ \theta_{Ai} &\sim \mathrm{N}(\theta_A, \tau^2_A) \end{aligned} \tag{7.13}$$

where, for an ANCOVA model in trial i, $\hat{\theta}_{Ai}$ is the treatment effect estimate, θ_{Ai} is the true treatment effect and s^2_{Ai} is the estimated variance of the treatment effect estimate in trial i. The average treatment effect from the ANCOVA model is denoted by θ_A and τ^2_A is the between-study variance for the true treatment effect across trials.

For the final score and the change score models, the univariate random-effects meta-analysis model can be re-written by replacing the estimates from ANCOVA with either those from the final score model (Equation 7.14) or the change score model (Equation 7.15).

$$\begin{aligned} \hat{\theta}_{Fi} &\sim \mathrm{N}(\theta_{Fi}, s^2_{Fi}) \\ \theta_{Fi} &\sim \mathrm{N}(\theta_F, \tau^2_F) \end{aligned} \tag{7.14}$$

$$\begin{aligned} \hat{\theta}_{Ci} &\sim \mathrm{N}(\theta_{Ci}, s^2_{Ci}) \\ \theta_{Ci} &\sim \mathrm{N}(\theta_C, \tau^2_C) \end{aligned} \tag{7.15}$$

7.2.5 A meta-analysis method recommended by the Cochrane Collaboration

Previously in this chapter, it was discussed that the preferred method for analysing continuous outcomes in a single RCT is with an ANCOVA model. However, although this method is known to have the best statistical properties compared to the final and change score models, it is often the least reported (Higgins & Green 2008). More commonly, a researcher will have studies that reported the final score and/or the change score. Unlike the separate univariate meta-analysis approach given above, the Cochrane Handbook (Higgins & Green 2008) states that it is possible to include a mixture of treatment effect estimates from different analytical approaches in a single univariate meta-analysis, under the assumption that all the models are estimating the same underlying treatment effects. In this way, only a single treatment effect estimate from each study will be included in the meta-analysis. Priority should be given to that from the ANCOVA model due to the reasons discussed earlier in this chapter. The Cochrane combined approach therefore is a random-effects meta-analysis model that can be expressed generally as:

$$\begin{aligned} \hat{\theta}_i &\sim \mathrm{N}(\theta_i, s_i^2) \\ \theta_i &\sim \mathrm{N}(\theta, \tau^2) \end{aligned} \tag{7.16}$$

where, for trial i, $\hat{\theta}_i$ is the treatment effect estimate derived from ***any*** analytical approach, θ_i is the true treatment effect in trial i and s_i^2 is the within-study variance (assumed known). θ denotes the average treatment effect and τ^2 is the between-study variance.

7.2.6 A new application of bivariate meta-analysis for continuous outcomes

An alternative meta-analysis approach for combining effect estimates from RCTs with a continuous outcome is a bivariate meta-analysis model, to simultaneously synthesise treatment effect estimates from a pair of analytical method (e.g. ANCOVA and final score). As shown in previous chapters, an advantage of multivariate meta-analysis over univariate meta-analysis is borrowing of strength, which is defined as the gain in precision (=1/variance) of summary meta-analysis results when using a multivariate meta-analysis over a univariate meta-analysis (Jackson et al. 2017). This gain in precision arises from the inclusion of correlated effect estimates in the same analysis. Therefore, the idea is that we may learn about ANCOVA estimates by borrowing information from the final score (or change score) estimates, when ANCOVA estimates are not reported.

In section 7.2.1, three models for continuous outcomes (ANCOVA, final score and change score) and additionally the relationship between the treatment effect estimates from the three models were described. The three treatment effect estimates from the three models are structurally related. Due to this, a meta-analysis containing all three is likely to encounter the potential problem of non-convergence. Therefore, a bivariate meta-analysis was decided as being the most appropriate for avoiding this problem. The bivariate meta-analysis of final score and ANCOVA treatment effect estimates was decided to be the most realistic and useful situation. The ANCOVA model is the ideal model for the analysis of continuous data, whilst the final score treatment effect is often the most reported (or can be derived from published information).

In summary, the aim is to produce a summary ANCOVA estimate of the treatment effect, based on available ANCOVA estimates, but borrowing information from final score estimates in trials that do not provide ANCOVA estimates. The proposed

bivariate random-effects meta-analysis model for ANCOVA and final score treatment effect estimates can be mathematically expressed as:

$$\begin{pmatrix} \hat{\theta}_{Ai} \\ \hat{\theta}_{Fi} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \theta_{Ai} \\ \theta_{Fi} \end{pmatrix}, \mathbf{\Gamma_i} \right), \qquad \mathbf{\Gamma_i} = \begin{pmatrix} s^2_{Ai} & s_{(A,F)i} \\ s_{(A,F)i} & s^2_{Fi} \end{pmatrix}$$
$$\begin{pmatrix} \theta_{Ai} \\ \theta_{Fi} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \theta_{A} \\ \theta_{F} \end{pmatrix}, \mathbf{\Omega} \right), \qquad \mathbf{\Omega} = \begin{pmatrix} \tau^2_{A} & \tau_{(A,F)} \\ \tau_{(A,F)} & \tau^2_{F} \end{pmatrix} \tag{7.17}$$

where A and F represent the ANCOVA and the final score models, respectively. For trial i, $\hat{\theta}_{Ai}$ and $\hat{\theta}_{Fi}$ represent the treatment effect estimate from the ANCOVA and final score models respectively and θ_{Ai} and θ_{Fi} are the true treatment effect for the ANCOVA and final score models, respectively. The average treatment effect for the ANCOVA and final score models are θ_A and θ_F, respectively. The within-study covariance matrix and between-study covariance matrix are represented as $\mathbf{\Gamma_i}$ and $\mathbf{\Omega}$, respectively. The elements of the lead diagonal of $\mathbf{\Omega}$, τ^2_A and τ^2_F, are the between-study variances for the ANCOVA and final score models, respectively. The remaining off-diagonal elements of $\mathbf{\Omega}$ are the between-study covariances, which are calculated using the between-study correlations. For example, for the ANCOVA and the final score models:

$$\tau_{(A,F)} = \rho_B \tau_A \tau_F \tag{7.18}$$

where ρ_B is the between-study correlation, which is the correlation between the true treatment effects, θ_{Ai} and θ_{Fi}. The lead diagonal of $\mathbf{\Gamma_i}$, s^2_{Ai} and s^2_{Fi} are the variances of the summary treatment effect estimates from the respective analytical models. The remaining off-diagonal elements of matrix $\mathbf{\Gamma_i}$ are the within-study covariances, which are calculated using within-study correlations. For example, for the ANCOVA and final score models:

$$s_{(A,F)i} = \rho_{W_{(A,F)i}} s_{Ai} s_{Fi} \tag{7.19}$$

where $\rho_{W_{(A,F)i}}$ is the within-study correlation between the treatment effect estimates from the ANCOVA and the final score models for trial i.

Calculating within-study correlations in a trial providing IPD

The within-study correlation between the treatment effect estimates from the analytical models is required for a bivariate meta-analysis and can be calculated from the IPD. The above equation (Equation 7.19) can be rearranged to give the equation for the within-study correlation:

$$\rho_{W_{(A,F)i}} = \frac{s_{(A,F)i}}{s_{Ai}s_{Fi}} \tag{7.20}$$

where $s_{(A,F)i}$, s_{Ai} and s_{Fi} are as above.

There are two methods to calculate the within-study correlations from IPD: seemingly unrelated regression and bootstrapping (Riley et al. 2015). These were explained in Section 1.4.4 and here the bootstrapping approach is used.

Potential advantages of a bivariate meta-analysis in this setting

If a trial only reports the treatment effect estimate using one of either an ANCOVA or a final score approach, the other effect estimate will be missing. These missing values can be handled in a bivariate meta-analysis under the missing at random assumption. In fact, this is the reason for proposing this meta-analysis approach in this setting; the model can borrow strength between effect estimates, to improve summary results.

In this setting, missing data refers to an unreported treatment effect estimate and standard error from a particular model (ANCOVA or final score). The treatment effect estimates from the ANCOVA are the desired outcome for pooling in a meta-analysis, however the ANCOVA estimates are not usually available for all trials. Therefore, the premise is that for the trials with missing ANCOVA treatment effects, the final score treatment effects are available.

Additionally, for some trials both the ANCOVA and the final score treatment effects are reported or are both available since they have been calculated from available IPD. For these trials with available IPD the within-study correlation can

be calculated as previously described. So the bivariate method may be most suitable when some trials provide IPD (so that both ANCOVA and final score treatment effect estimates can be derived) and other trials without IPD only provide the final score treatment effect estimate.

Estimation of the bivariate meta-analysis

The parameters of the bivariate meta-analysis model were estimated using restricted maximum likelihood (REML). The standard errors calculated accounted for uncertainty in the estimated between-study correlation and confidence intervals were derived using the normal distribution and implemented in Stata using White (2011)'s mvmeta package. The BoS statistic was calculated as described in previous chapters (Sections 1.4.8 and 3.2.5).

7.2.7 A real dataset for the application of methods

The data used to investigate the new method in this chapter is from a previous meta-analysis by Wang et al. (2005). The dataset contains IPD from 10 RCTs that investigated whether active hypertension treatments lower systolic blood pressure (SBP) or diastolic blood pressure (DBP) compared to placebo or no treatment. This provided IPD for a total of 25,581 patients as detailed in Riley et al. (2013). Each participant in each trial recorded a baseline and final measurement for SBP and DBP. For illustration, this chapter focuses only on the SBP. The distribution of baseline values were very similar between the treatment and control groups for the SBP in each trial, and thus baseline balance was good, although not perfect (Table 7.1).

Table 7.1: The baseline means and standard deviations for the systolic blood pressure from the 10 Hypertension trials from the control and treatment groups. The treatment effect estimates and standard errors from the ANCOVA and final score models and the within-study correlations calculated using bootstrapping.

Trial	Baseline Mean (SD)		Treatment Effect Estimates (SE)		Within-study correlations
	Control	Treatment	θ_A	θ_F	$\rho_{W_{(A,F)i}}$
ANBP	153.05 (15.73)	152.28 (15.25)	-6.664 (0,850)	-6.899 (0.884)	0.962
COOP	191.55 (17.64)	189.94 (16.15)	-14.166 (2,174)	-14.834 (2.300)	0.943
EWPH	178.23 (15.06)	177.33 (15.85)	-12.881 (3.211)	-13.573 (3.682)	0.871
HDFP	151.00 (19.53)	151.68 (19.83)	-8.709 (0.543)	-8.440 (0.585)	0.933
MRC1	156.65 (15.96)	156.50 (16.09)	-8.702 (0.375)	-8.760 (0.406)	0.928
MRC2	182.13 (12.73)	182.19 (12.63)	-10.602 (0.764)	-10.587 (0.774)	0.988
SHEP	170.12 (9.24)	170.49 (9.50)	-11.357 (0.546)	-11.137 (0.569)	0.963
STOP	194.15 (11.16)	194.68 (12.21)	-17.926 (2.413)	-17.655 (2.518)	0.950
SYCH	170.25 (11.41)	170.73 (10.90)	-6.548 (0.644)	-6.357 (0.669)	0.966
SYSE	173.94 (10.07)	173.75 (9.86)	-10.256 (0.442)	-10.373 (0.476)	0.920

The treatment effect estimates for the SBP were all negative, which means that the ac-

tive hypertension treatments reduce SBP compared to control or no treatment (Table 7.1). For each trial, the treatment effect estimates from each model were very similar. In this meta-analysis example, the treatment effect estimates from the final score model were less than the treatment effect estimates from the ANCOVA model for five trials and vice versa in the remaining five trials. This observation will be investigated further in Section 7.3.2 when we consider the impact of missing treatment effect estimates.

The within-study correlations of the final score and ANCOVA treatment effect estimates were calculated using the bootstrap method. The magnitude of all the within-study correlations were all large (Table 7.1).

7.2.8 Generating missing data scenarios

Making ANCOVA results missing at random

The bivariate method was applied to the complete data from all trials, for which ANCOVA and final score are available for all. In truth, it would have been unnecessary to do a bivariate meta-analysis with complete IPD, since the ANCOVA treatment effect estimates could be derived and pooled in a univariate meta-analysis, thus negating the need for the final score treatment effect estimates. In practice it is unlikely that complete data will be available, either from IPD or aggregate data, and therefore missing data scenarios were generated for application. These contained complete data for five trials (50% of trials) and missing treatment effect estimates for the ANCOVA for the remaining 5 trials. The scenarios were generated by setting the ANCOVA treatment effect estimates and standard errors to randomly missing in five of the trials (i.e. a missing at random situation).

Making missing ANCOVA results missing not at random

Also, two other missing data scenarios were created by making the missingness in the ANCOVA treatment effect estimate dependent upon the relationship between the ANCOVA and final score treatment effect estimates. The aim was to generate a

situation where the new method would fail, by wrongly borrowing strength, due to a missing not at random (MNAR) situation. For five of the trials (ANBP, COOP, EWPH, MRC1 and SYSE), denoted as 'MNAR1', the magnitude of the treatment effect estimate from the ANCOVA was greater than the magnitude of the treatment effect estimate from the final score. Alternatively, for the remaining five trials (HDFP, MRC2, SHEP, STOP and SYCH), denoted as 'MNAR2', the magnitude of the treatment effect estimate from the ANCOVA was less than the treatment effect estimates from the final score. There are two questions of interest here. Firstly, does the bivariate meta-analysis provide misleading results compared to the complete data case? Secondly, if the missing treatment effect estimates are known to be MNAR, is a univariate meta-analysis approach more appropriate than a bivariate approach?

Finally, to avoid spurious findings based on a select few missing data scenarios, all possible permutations of missing data scenarios for the 10 Hypertension trials with five trials missing ANCOVA estimates were considered. For 10 trials with five studies missing ANCOVA estimates, there were a total of 252 ($C(10,5) = \frac{10!}{5!(10-5)!}$) unique permutations of missing data scenarios. Each meta-analysis approach was applied to each set of treatment effect estimates for each of the 252 permutations and the results were summarised across the 252 meta-analysis datasets. The mean of the summary treatment effect estimates from each meta-analysis and the mean standard error of the summary treatment effect estimate were recorded for each meta-analysis approach, alongside the range of the summary treatment effect estimates and the standard errors.

7.3 Results

The various models are now applied to the Hypertension meta-analysis dataset, and their results compared, for each complete and missing data scenarios.

7.3.1 Results when there are complete data

Consider first the complete data case where the ANCOVA and final score treatment effect estimates are available from all 10 Hypertension studies. The summary treatment effect estimates for the ANCOVA were equal, -10.167 (Table 7.2), from the univariate meta-analysis and the bivariate meta-analysis. The summary treatment effect estimates were also very similar for the final score. From the univariate meta-analysis, the final score summary treatment effect was -10.138 and from the bivariate meta-analysis it was -10.119.

Table 7.2: Results from the complete data case; all ANCOVA treatment effect estimates available for all 10 Hypertension studies

Type of treatment effect measure	Summary treatment effect estimate	Standard error	95% confidence interval	Observed BoS (%)	Predicted BoS using model 4.7 (%)
Bivariate meta-analysis					
ANCOVA	-10.167	0.972	-12.072 to -8.263	<0.001	7.7
Final score	-10.119	0.964	-12.008 to -8.229	1.4	7.7
Univariate meta-analyses					
ANCOVA	-10.167	0.967	-12.063 to -8.271		
Final score	-10.138	0.980	-12.060 to -8.217		

The standard errors of the estimates were also similar between the univariate and the bivariate meta-analysis results (Table 7.2), which was expected as there is little borrowing of strength with complete data in a multivariate meta-analysis (Riley et al. 2007*a*, Jackson et al. 2017). For the ANCOVA, the standard error was slightly lower in the bivariate meta-analysis approach, 0.967, compared to the univariate model, 0.980 and BoS was close to 0%. Conversely, the standard error was slightly higher in the bivariate meta-analysis approach, 0.972 for the final score, compared to the univariate approach, 0.964 and BoS was 1.4%. However, overall, the new bivariate model adds

very little in this situation of complete data. This is expected, as there is no missing data, and so Chapter 6 suggests BoS will be close to zero, and the predicted BoS from the model developed in Chapter 4 is 7.7%.

7.3.2 Results when some ANCOVA treatment effect estimates are missing

Missing At Random

A scenario is now considered when some ANCOVA results are missing at random (MAR) from five of the ten Hypertension trials, MRC1, MRC2, SHEP STOP and SYSE, but the final score results are always available. In this situation, unlike the complete data, there appears to be some advantage of the bivariate method. This is expected, as the previous chapter showed that the BoS could be up to 50% when 50% of studies have the outcome missing, and the predicted BoS based on prediction model 4.7 is about 30%. The summary ANCOVA treatment effect estimate for the bivariate meta-analysis (-10.080) was closer to the summary ANCOVA treatment effect estimates for the complete data (-10.167) than the separate univariate approach for ANCOVA (-9.022) (Table 7.3).

Table 7.3: Results from a missing at random example; there were no ANCOVA estimates for trials MRC1, MRC2, SHEP, STOP and SYSE

Model for treatment effect measure	Summary treatment effect estimate	Standard error	95% confidence interval	Observed BoS (%)	Predicted BoS using model 4.7 (%)
Bivariate meta-analysis					
ANCOVA	-10.080	0.948	-11.939 to -8.221	53.1	30.6
Final score	-10.108	0.972	-12.012 to -8.204	0.6	4.8
Separate univariate meta-analysis					
ANCOVA	-9.022	1.488	-11.939 to -6.105		
Final score	-10.138	0.980	-12.060 to -8.217		
Cochrane combined meta-analysis (5 trials ANCOVA and 5 trials final score)					
Combined	-10.082	0.931	-11.908 to -8.257		

The Cochrane combined summary estimate, -10.082 (Table 7.3), was contained between the summary estimates from the bivariate meta-analysis for the ANCOVA and the final score, which were -10.080 and -10.108, respectively. Additionally, the Cochrane combined summary estimate was more similar in value to the ANCOVA bivariate estimate than the final score bivariate estimate.

The standard error of the ANCOVA summary treatment effect estimate was smaller in the bivariate meta-analysis model, 0.948 (Table 7.3), compared to the univariate approach, 1.488, due to the gain in information borrowed from the five studies with final score values. However, the standard error from the bivariate meta-analysis for the ANCOVA was similar to the Cochrane combined approach, 0.931, where the five trials with ANCOVA results are combined with the five trials with only final score results.

Missing Not At Random

Consider now the results when missing data were not missing at random (MNAR). One of the MNAR scenarios investigated here had missing ANCOVA treatment effect estimates for the five trials (ANBP, COOP, EWPH, MRC1 and SYSE), where the actual ANCOVA treatment effect estimate was greater than the final score treatment effect estimate, denoted as MNAR1 (Table 7.4). The other MNAR scenario had missing ANCOVA treatment effect estimates for the five trials (HDFP, MRC2, SHEP, STOP and SYCH), where the actual ANCOVA treatment effect estimate was less than the final score treatment effect estimate, denoted as MNAR2.

Table 7.4: Results from missing not at random examples when there was no IPD for half the trials. MNAR1: missing ANCOVA estimate for ANBP, COOP, EWPH, MRC1 and SYSE. MNAR2: missing ANCOVA estimate for HDFP, MRC2, SHEP, STOP, SYCH.

Scenario	Model for treatment effect measure	Summary treatment effect estimate	Standard error	95% confidence interval	Observed BoS (%)	Predicted BoS using model 4.7 (%)
MNAR1	Bivariate meta-analysis					
	ANCOVA	-10.336	0.984	-12.264 to -8.408	46.5	31.0
	Final score	-10.207	0.999	-12.166 to -8.248	0.2	5.3
	Separate univariate meta-analysis					
	ANCOVA	-10.582	1.654	-13.824 to -7.340		
	Final score	-10.138	0.980	-12.060 to -8.217		
	Cochrane combined meta-analysis					
	Combined	-10.276	0.996	-12.228 to -8.324		

Continued on next page

Table 7.4: continued

Scenario	Type of treatment effect measure	Summary treatment effect estimate	Standard error	95% confidence interval	Observed BoS (%)	Predicted BoS using model 4.7 (%)
MNAR2	Bivariate meta-analysis					
	ANCOVA	-9.966	0.951	-11.829 to -8.102	52.5	30.7
	Final score	-10.093	0.960	-11.976 to -8.211	1.0	4.9
	Separate univariate meta-analysis					
	ANCOVA	-9.798	1.281	-12.310 to -7.287		
	Final score	-10.138	0.980	-12.060 to -8.217		
	Cochrane combined meta-analysis					
	Combined	-10.030	0.951	-11.895 to -8.166		

The summary treatment effect estimates from the MNAR1 and MNAR2 scenarios were quite different for all the methods (Table 7.4). The borrowing of strength assumes that the relationship between the ANCOVA and final score treatment effect estimates observed in those five trials that provide both, is the same as in those trials providing only final score. However, these scenarios were purposely created so that this was not correct.

Despite this concern, the bivariate method showed improvement over a univariate meta-analysis of just five trials providing ANCOVA estimates. The ANCOVA summary treatment effect estimates from the bivariate meta-analyses in the MNAR1 and MNAR2 scenarios were -10.336 and -9.966 (Table 7.4), respectively. These are closer to the ANCOVA summary treatment effect estimates of -10.167 (Table 7.2) from the complete data case (where all 10 trials gave ANCOVA results), than the ANCOVA results of -10.582 and -9.798 from a separate univariate meta-analyses of just five trials. The standard errors were also much smaller in the bivariate model.

However, the Cochrane combined approach appears to perform best, as it gives summary treatment effect estimates for MNAR1 and MNAR2 scenarios of -10.276 and -10.030, respectively (Table 7.4). These are closer in value to the ANCOVA treatment effect estimates from the complete data case, -10.167 (Table 7.2), than the ANCOVA summary treatment effect estimates from the new bivariate approach, 10.336 and -9.966 (Table 7.4). Similarly to the MAR scenario, the Cochrane combined summary treatment effect estimate was contained between the ANCOVA and final score summary treatment effect estimates from the bivariate meta-analysis. For example, in scenario MNAR1 the Cochrane combined summary treatment effect estimate was -10.276 and the ANCOVA and final score summary treatment effect estimates from the bivariate meta-analysis were -10.336 and -10.207 (Table 7.4), respectively.

The standard errors for the summary treatment effect estimates were similar

from the bivariate meta-analysis approach and the Cochrane combined approach for both the MNAR scenarios.

7.3.3 Investigating all permutations of missing data scenarios

The results from all the meta-analysis approaches for all the permutations of missing data scenarios are summarised in Table 7.5. There were 252 different permutations of missing data scenarios with five missing ANCOVA treatment effect estimates, which includes the three examples presented previously.

Bivariate meta-analysis compared to separate univariate meta-analyses

The means of the ANCOVA and final score summary treatment effect estimates from the bivariate meta-analyses were similar to the mean summary treatment effect estimates from the separate univariate meta-analyses (Table 7.5). For example, for the ANCOVA model, the mean summary treatment effect estimate was -10.150 from the bivariate meta-analyses and -10.284 from the separate univariate meta-analyses method. However, the range of the summary treatment effect estimates from the meta-analysis methods differed between the bivariate meta-analysis and the separate univariate meta-analysis. The width of the range was greater for the separate univariate meta-analyses, ranging from -12.692 to -7.934, compared to the range from the bivariate meta-analysis approach, -10.449 to -9.606. The borrowing of strength leads to more information and thus, the bivariate meta-analysis results are more precise (and thus less variable) when there are missing data, compared to the separate univariate meta-analysis results.

Table 7.5: Summary results from the meta-analyses of all 252 different permutations of missing data scenarios

Model for treatment effect measure	Mean summary treatment effect estimate	Range of summary treatment effect estimates	Mean standard error	Range of standard errors	Mean observed BoS (%)	Range of observed BoS (%)
Bivariate meta-analysis						
ANCOVA	-10.150	-10.449 to -9.606	0.940	0.002 to 1.612	49.43	40.00 to 58.48
Final score	-10.113	-10.311 to -9.668	0.935	0.002 to 1.597	0.74	0.01 to 2.90
Separate univariate meta-analysis						
ANCOVA	-10.284	-12.692 to -7.934	1.388	0.430 to 2.317		
Final score	-10.138					
Cochrane combined meta-analysis						
Combined	-10.152	-10.276 to -10.030	0.973	0.914 to 1.030		
Separate univariate of complete data						
ANCOVA	-10.167		0.967			

The mean standard error of the ANCOVA summary treatment effect estimates from the separate univariate meta-analyses, 1.388 (Table 7.5), was much larger than the mean standard error from the bivariate meta-analyses, 0.940. The range of standard errors from the separate univariate meta-analyses, 0.431 to 2.317, was greater in width and encompassed larger values than the range of standard errors from the bivariate meta-analyses, 0.002 to 1.612.

Bivariate meta-analysis compared to Cochrane combined meta-analysis

The mean summary treatment effect estimate from the Cochrane combined meta-analyses, -10.152 (Table 7.5), was similar to the means from the bivariate meta-analyses for the ANCOVA, -10.150, and the final score, -10.113. The range of the summary treatment effect estimates was narrower for the Cochrane combined meta-analyses, -10.276 and -10.030, than the range of the summary treatment effect estimates for the bivariate meta-analyses, -10.449 to -9.606.

Similarly, the mean standard errors between the Cochrane combined meta-analyses, 0.973 (Table 7.5), and the bivariate meta-analyses for the ANCOVA results were similar, 0.973 and 0.940, respectively. The range of the standard errors was very narrow for the Cochrane combined meta-analysis (range: 0.914 to 1.030), compared to the range of standard error for the bivariate meta-analyses for ANCOVA (range: 0.002 to 1.612).

7.4 Discussion

This chapter proposed a multivariate meta-analysis approach for the synthesis of treatment effect estimates from RCTs with continuous outcomes modelled using ANCOVA or final score models. This was contrasted with two univariate approaches: synthesis of the treatment effect estimates for each analytical method separately, or a single univariate meta-analysis that combines treatment effect estimates from different analytical methods together; the latter method is advocated by the Cochrane Collaboration

(Higgins & Green 2008). The current and newly proposed meta-analysis methods were compared using an existing IPD dataset of 10 Hypertension trials (Wang et al. 2005, Riley et al. 2013). With the complete IPD dataset the treatment effects from each analytical model could be estimated and hypothetical missing data scenarios explored. The key findings of the comparison are summarised in Figure 7.1.

Figure 7.1: Key findings and recommendations from the application of meta-analysis methods to an IPD dataset with continuous outcomes

Key findings and recommendations:

- The multivariate meta-analysis approach is an alternative method of synthesising treatment effect estimates from RCTs with continuous outcomes that have been derived through a combination of ANCOVA and final score models.
- For complete data (i.e. ANCOVA available for all trials), there was little difference in the summary treatment effect estimates and standard errors between the separate univariate meta-analyses and the newly proposed bivariate meta-analysis, as expected.
- For the 252 permutations of missing data scenarios (irrespective of missing data mechanism), the mean summary treatment effects from the bivariate meta-analyses and the mean summary treatment effects from the Cochrane combined meta-analyses were similar to each other, but different to those from the separate univariate meta-analyses. Furthermore, the Cochrane combined approach and bivariate meta-analysis approach gave closer results to those from the complete data case than the separate univariate meta-analysis.
- The mean standard errors from the bivariate meta-analysis are smaller than those from the separate univariate or Cochrane combined meta-analyses.

However, the range of standard errors was narrower for the Cochrane combined approach than the bivariate approach. The range of standard errors for the separate univariate approach was larger than the ranges from the bivariate or Cochrane combined approach.

- The bivariate meta-analysis approach is potentially most important if there is evidence of baseline imbalance and some studies are missing ANCOVA results. In other situations, the Cochrane combined meta-analysis is advantageous over the bivariate approach when IPD is unavailable, since the Cochrane combined approach does not require the within-study correlations.

The results of this chapter demonstrate that the separate univariate meta-analysis method should not be used for the synthesis of treatment effect estimates from trials with a continuous outcome modelled using ANCOVA and/or final score models. In the presence of missing treatment effect estimates, the summary treatment effect estimates from the separate univariate meta-analyses differed from the summary treatment effect estimates from the complete data.

In comparison to the bivariate approach for the missing data scenarios, the separate univariate meta-analysis summary treatment effect estimate was further from the summary treatment effect estimates from the complete data case. Additionally in the presence of missing treatment effect estimates, the standard errors for the summary treatment effect estimates were greater from the separate univariate meta-analyses than the bivariate meta-analyses. Furthermore, for all the permutations of missing data scenarios the ranges of summary treatment effect estimates and their standard errors were large for the separate univariate meta-analyses compared to the bivariate approach or the Cochrane combined approach.

The bivariate meta-analysis approach and the Cochrane combined meta-analysis

approach treatment effect estimates and standard errors were similar for the selected MAR and MNAR scenarios (Section 7.3.2). The mean summary treatment effect estimates and the mean standard errors were similar between two methods from the 252 missing data permutations. Therefore, for the situation where there is baseline balance and half the ANCOVA treatment effect estimates are missing the choice between method is unimportant with regard to the accuracy of results. However, there are benefits to utilising the bivariate meta-analysis approach, including borrowing strength about the relationship between the treatment effect estimates from the final score and the ANCOVA models. An additional benefit is the interpretability of the results from a bivariate meta-analysis. The bivariate meta-analysis provides a summary treatment effect estimate for each analytical model included in the analysis (Jackson et al. 2011). Whereas, the Cochrane combined meta-analysis provides a singular summary treatment effect estimate which is synthesised from treatment effect estimates from different analytical models (Higgins & Green 2008).

7.4.1 Considerations for a bivariate meta-analysis

Conversely, this benefit of utilising a bivariate approach may also lead to technical difficulties, e.g. how to draw conclusions based on two summary treatment effect estimates compared to one summary treatment effect estimate.

Another possible consideration for the utilisation of a bivariate approach is the need for IPD for the derivation of the within-study correlation matrix. IPD is an important component to the bivariate approach since, without IPD for some trials the bivariate approach is unlikely to be usable, due to unavailable within-study correlations through lack of reporting (Riley 2009). An advantage of IPD is it can be assessed whether there is baseline imbalance present.

Consequently, the decision between the bivariate meta-analysis approach and the Cochrane combined meta-analysis approach has many considerations. It is

important to consider the number of studies that provide both estimates together with the quantity of IPD that is obtainable. Furthermore, there are some examples when the use of the Cochrane combined approach should be considered over the bivariate approach including when there is a struggle to obtain IPD. Another consideration is if the ANCOVA is provided for the majority of trials and in the remaining trials there is no evidence of baseline imbalance.

7.4.2 Limitations

The methods in this chapter were compared using only one dataset where only one outcome (SBP) was studied. Therefore, it is difficult to be certain how the conclusions from this chapter transfer to other datasets.

This chapter studied different missing data scenarios and the impacts missing data had on results from different meta-analysis methods. Although this chapter studied missing data there were restrictions on the quantity of missing data. For each missing data scenario there was always missing data for half of the ANCOVA treatment effect estimates. It is unclear how different proportions of missing data for the ANCOVA treatment effect estimates might affect results from the meta-analysis methods.

To avoid estimation problems for the bivariate method, treatment effect estimates from two analytical models were chosen for the analysis. In this chapter, the ANCOVA and final score were selected, since in trials the standard derivations of changes from baseline are often not reported. Although the change score model was not selected, it might be expected that similar results would be observed had this also been analysed instead of the final score results. However, the final score results were chosen as these are more commonly reported and can be derived more easily from reported aggregate data.

7.4.3 Further research

In this chapter, the trials within the dataset all had good baseline balance across the treatment and control groups. However, the impact that baseline imbalance might have on the comparison of methods is important to address in further research. The summary treatment effect estimate may be biased if the baseline imbalance across treatment groups has not been accounted for. Consider an example where trials with baseline imbalance do not provide IPD and report treatment effect estimates from final score models and the trials with baseline balance provide IPD (i.e. treatment effect estimates from ANCOVA and final score models). It is unknown, based upon the data available to the meta-analyst, that there is baseline imbalance in the trials without IPD. In this situation which method should be recommended? The use of estimation methods for univariate meta-analysis methods for the synthesis of summary treatment effect estimates from RCTs with continuous outcomes has been studied in a simulation study (McKenzie et al. 2016), although the bivariate meta-analysis approach has not been considered.

Before recommendations on which meta-analysis method should be utilised in future, further work is required. A focus for future work should include examining the methods in further data sets to check the results from this chapter hold. This future work could include different proportions of missing ANCOVA treatment effect estimates and treatment effect estimates from change score models rather than final score models.

Next steps

The next chapter will continue this work through a simulation study with the aim of assessing each meta-analysis approach under varying percentages of missing data, as well as baseline balance and baseline imbalance conditions.

Chapter 8

A simulation study to assess the performance of a bivariate meta-analysis method for randomised control trials with continuous outcomes

8.1 Background

In the previous chapter, a new bivariate approach for synthesising treatment effect estimates from randomised control trials (RCTs) with continuous outcomes, measured at baseline and follow-up, was proposed and explored. The approach was studied using an individual participant data (IPD) meta-analysis for interventions for the treatment of hypertension. For each trial, the treatment effect estimates were derived from both ANCOVA model and final score model. To study the application of the bivariate approach, missing data scenarios were generated by forcing some of the trials to have missing treatment effect estimates from the ANCOVA model. The bivariate approach was applied and compared to two other methods: (i) a univariate approach of trials that provide ANCOVA results, and (ii) combining (in a univariate meta-analysis) ANCOVA results from some trials and final score results in other trials. The latter approach is the current approach recommended by Cochrane (Higgins & Green 2008).

The results suggest little difference in the meta-analysis approaches given com-

plete data (i.e. ANCOVA results from all trials). However, for settings with missing ANCOVA estimates, the results from the bivariate meta-analysis approach and the Cochrane approach were similar and improved upon findings from a separate univariate meta-analysis of just ANCOVA results. Thus, it was concluded that the use of separate univariate meta-analyses in this setting should be avoided, and either the Cochrane approach or the bivariate approach could be used. However, this recommendation was only based on a single application, and so in this chapter a more detailed simulation study is conducted to compare the various methods across a range of different scenarios, including missing data scenarios. The simulation study aims to make recommendations about which approach should be used and the settings that influence this decision.

8.2 Models and setting

8.2.1 Description of models for continuous outcomes

For each trial in the simulation study, the continuous outcome, systolic blood pressure (SBP), was modelled using the ANCOVA and final score models. The details of these models were provided in Chapter 7, Section 7.2.1.

8.2.2 Setting for simulation study

The setting for the simulation study was based on the Hypertension data analysed in Chapter 7, and so the outcome simulated was systolic blood pressure, which is a continuous outcome.

In this simulation study, we investigate the effects of baseline (im)balance on summary measures from different meta-analysis methods. Baseline imbalance occurs when the mean of the baseline continuous outcome in each treatment arm differs (Fu & Holmer 2015). Baseline imbalance can occur due to an inadequate randomisation method, selection bias or occur by chance (Fu & Holmer 2016). Previous studies have

concluded that final score models and change score models produce biased estimates in the presence of baseline imbalance (Fu et al. 2013, Fu & Holmer 2015, Vickers & Altman 2001). However, the ANCOVA method produces unbiased estimates (provided a linear adjustment is correct), since it adjusts for the baseline measurement and thus accounts for the baseline imbalance.

Although the ANCOVA model is advocated as the method of choice, RCTs are still analysed using final score models or change score models (Fu & Holmer 2015). This impacts a researcher doing a meta-analysis as they are likely to encounter a selection of treatment effect estimates from different models.

8.3 Simulation study methods

The simulation study is described below in steps and the simulation code can be found in Appendix F.3.

8.3.1 Simulation plan

Step 1: Generating data for one IPD meta-analysis

For the generation of IPD for the simulation study, the Hypertension data was used to provide reasonable values and distributions for parameters (Table 8.1) using the ANCOVA model (Equation 7.3). The first step for the generation of the data was to set the number of trials per meta-analysis, either 5, 10 or 20. Within each trial, the patients were generated with unique identifiers and the number of patients per trial was fixed, at either 100 or 1000. In each trial, the patients were allocated into either a treatment group or a control group using a binomial distribution with an equal probability of 0.5.

Table 8.1: Parameters used in the simulation study to generate scenarios with baseline balance

Parameter	Values	Information
k	5, 10, 20	Number of studies in each meta-analysis
m	0, 10, 20, 40, 60	Percentage of MAR ANCOVA treatment effects
n	100, 1000	Number of participants per study
θ	-10	Average treatment effect
τ^2	0, 2.25, 9	Between-study variance
σ^2	225	Variance of random error term
ϕ_i	N(80,25)	Intercept
β_i	N(0.5,0.0025)	Baseline adjustment
x_{ij}	N(165, 324)	Baseline systolic blood pressure
	N(170, 324)	Baseline systolic blood pressure for imbalance settings

Step 1a: Simulation of the baseline balance setting

Next, the baseline systolic blood pressure (SBP) for each participant in each trial was generated from a normal distribution with a mean of 165 and variance 324. For the baseline balance setting, the baseline SBP generation mechanism was the same regardless of which treatment group a participant was allocated to.

Step 1b: Simulation of baseline imbalance settings

There were three baseline imbalance settings that were considered in this chapter. The first setting was the unconditional baseline imbalance, where all the trials within the meta-analysis have baseline imbalance. This was simulated by generating baseline SBP values for the participants in the control group from a normal distribution with mean 170 and variance 324. The baseline SBP values for the treatment group were drawn from the same distribution as the baseline balance setting i.e. a normal distribution with mean 165 and variance 324. Thus, the treatment group has a 5mmHG lower SBP at baseline than the control group, on average in the trials.

The remaining two baseline imbalance settings were generated by conditioning which trials contained baseline imbalance across treatment groups. The first conditional baseline imbalance setting generated baseline imbalance across treatment groups for trials which will be simulated to have missing ANCOVA treatment effect estimates (representing trials with no IPD available). The first x trials in the simulated meta-analysis will have missing ANCOVA treatment effect estimates, where x is the value that satisfies the percentage of missing data simulated. Therefore, the trials with IPD (or complete treatment effect estimate data) will have baseline balance and the trials with only treatment effect estimates from the final score model will have baseline imbalance.

The second conditional baseline imbalance setting generated baseline imbalance for trials simulated to have both ANCOVA and final score treatment effect estimates, thereby representing trials with IPD available. Therefore, the trials with IPD will have baseline imbalance and the trials with only treatment effect estimates from the final score model will have baseline imbalance. The baseline balance and imbalance were generated in the same manner for both the conditional baseline imbalance settings. For the trials with baseline imbalance, the baseline SBP values were generated in the same way as the unconditional baseline imbalance setting, from a normal distribution with mean 170 and variance 324. For the trials with baseline balance, the baseline SBP values were generated as in Step 1a for the baseline balance setting, from a normal distribution with mean 165 and variance 324.

Step 2: Simulation of ANCOVA model

The remaining parameters for the ANCOVA model (Equation 7.3) that require generating include the error term, e_i, the treatment effect θ, the baseline adjustment, β_i and the intercept, ϕ_i. The error term was normally distributed with mean 0 and variance σ^2, which was fixed at 225 for all simulations. The intercept was drawn from

a normal distribution with mean 80 and variance 25 and the baseline adjustment was also drawn from normal distribution with mean 0.5 and variance 0.0025. The treatment effect was also drawn from a normal distribution with a mean, -10, and the variance was the between-study variance, τ^2, either 0, 2.25 or 9 depending on the chosen scenario.

Based on the parameter values chosen, the final SBP for each individual in each trial was then simulated using the ANCOVA model (Equation 7.3). This provided one simulated meta-analysis dataset of IPD.

Step 3: Simulate 1000 IPD meta-analysis datasets

To generate 1000 IPD meta-analysis datasets, steps 1 and 2 were repeated 1000 times.

Step 4: First stage of IPD meta-analysis

For each simulated trial in each IPD meta-analysis, the final score model (Equation 7.2) and the ANCOVA model (Equation 7.3) were fitted. The trial-specific treatment effect estimates and standard errors from both the final score model and the ANCOVA model were obtained. The within-study correlations between the pairs of treatment effect estimates from the ANCOVA and the final score models were obtained using seemingly unrelated regression (Section 4.2.4) (Riley et al. 2015). Bootstrapping (Section 4.2.4) was considered too computationally intensive for the requirements in this chapter.

Step 5: Generating missing data

A key aim of the simulation study was to investigate how each method performs in the presence of missing data; that is, some trials have missing ANCOVA estimates. Therefore, settings with different percentages of missing data were generated as part of the simulation study. The percentage of missing data was kept constant across each meta-analysis in the setting and ranged from 10% to 60%(e.g. 10%, 20%, 40%, 60%), as well as complete data with 0% missing data. The missing data was generated for the

ANCOVA treatment effect estimates under the missing completely at random (MCAR) assumption, following the fitting of models in Step 3. Since the data was generated at random, the first x ANCOVA treatment effect estimates (and their variances) were forced to be missing such that x satisfies the percentage of missing data required. To fully simulate the setting of unavailable IPD for the trials with missing data, the within-study correlations for these trials were also subsequently forced to be missing under the MCAR assumption.

Step 6: Second stage of IPD meta-analysis

Recall, the methods for comparison were univariate meta-analysis of ANCOVA treatment effects only, bivariate meta-analysis of ANCOVA and final score treatment effects and the Cochrane univariate meta-analysis approach (either ANCOVA or final score treatment effects from each study, with ANCOVA preferred if both are available). All the methods for the comparison used a random-effects meta-analysis and utilised the REML estimation method for the estimation of the between-study variance. The standard errors calculated did not account for the uncertainty in the estimation of between-study variance and covariance (for multivariate meta-analysis). The 95% confidence intervals were derived using the normal distribution.

An IPD meta-analysis was excluded from the simulation if a meta-analysis method failed to converge within 200 iterations or if a meta-analysis method failed to estimate the summary treatment effect or standard error. This meant that for each simulation scenario and percentage of missing data setting, the number of meta-analysis results were the same across each meta-analysis method, which allowed for fair comparisons between methods.

Step 7: Performance measures

For the comparison of meta-analysis methods, several performance measures were calculated to investigate the summary treatment effect estimate and the confidence in-

terval, for each simulation scenario. The percentage bias, empirical standard error and the mean square error were calculated to assess the performance of the summary treatment effect estimate in each meta-analysis. Additionally, for each meta-analysis, the coverage of the 95% confidence interval (proportion of meta-analyses the 95% confidence interval contained the true value for the estimate of interest, in this case the summary treatment effect estimate) (Burton et al. 2006) and the power (proportion of meta-analyses that gave a significant p-value$<$0.05) were calculated to assess the confidence intervals from each meta-analysis method. Furthermore, the mean BoS for both ANCOVA and final score outcomes from the bivariate meta-analysis were calculated.

8.3.2 Scenarios

The simulation study investigated the performance of the meta-analysis methods using nine different simulation scenarios (Table 8.2 and Table F.1 in Appendix F.1). The simulation study scenarios differed by the number of trials included in the meta-analysis and the value of the between-study variance of the treatment effect. The number of trials in the meta-analysis were decided upon to simulate scenarios with small (5), medium (10) and large (20) number of trials. Similarly for the between-study variance, the scenarios included no between-study variance (therefore a between-study correlation of zero), small (2.25) and large (9) between-study variance.

Table 8.2: Table of scenarios for the simulation study for the baseline balance and imbalance settings

Scenario	Number of studies	Between-study variance	Percentage of missing data	Number of participants
Baseline balance setting				
1	5	0	0, 20, 40, 60	100, 1000
2	10	0	0, 10, 20, 40, 60	100, 1000
3	20	0	0, 10, 20, 40, 60	100, 1000
4	5	2.25	0, 20, 40, 60	100, 1000
5	10	2.25	0, 10, 20, 40, 60	100, 1000
6	20	2.25	0, 10, 20, 40, 60	100, 1000
7	5	9	0, 20, 40, 60	100, 1000
8	10	9	0, 10, 20, 40, 60	100, 1000
9	20	9	0, 10, 20, 40, 60	100, 1000
Baseline imbalance setting				
Unconditional	10	2.25	0, 10, 20, 40, 60	1000
1st Conditional	10	2.25	0, 10, 20, 40, 60	1000
2nd Conditional	10	2.25	0, 10, 20, 40, 60	1000

Each scenario in Table 8.2 was simulated for each percentage of missing data with both 100 participants per trial and 1000 participants per trial. These sub-scenarios' characteristics are tabulated in further detail in Table F.2 in Appendix F.1.

There were three different baseline imbalance settings, an unconditional and two conditional baseline imbalance settings. The effects of baseline imbalance were investigated using Scenario BI, 10 trials of 1000 participants in the meta-analysis and a between-study variance of 2.25 (Table 8.2).

8.4 Results: Part 1 - Baseline balance scenarios

The results of the simulation study are now provided, for each of the different scenarios, considering situations with missing data and baseline balance (baseline imbalance is considered in Section 8.5). A key focus of the results section is the comparison between the bivariate meta-analysis method for the ANCOVA outcome and the Cochrane recommended meta-analysis method.

8.4.1 Magnitude of BoS from the bivariate approach

In the bivariate model, the BoS statistic for the ANCOVA treatment effect outcome increased as the percentage of missing data for the ANCOVA treatment effect estimate increased (Figures 8.1 and 8.2). This is consistent with the findings in Chapters 4 and 6, which showed that the percentage of missing data for the outcome of interest was an important predictor for the magnitude of BoS. Since BoS increases with increasing missing data, the bivariate approach will be most beneficial over the univariate approach when there are trials with missing ANCOVA estimates.

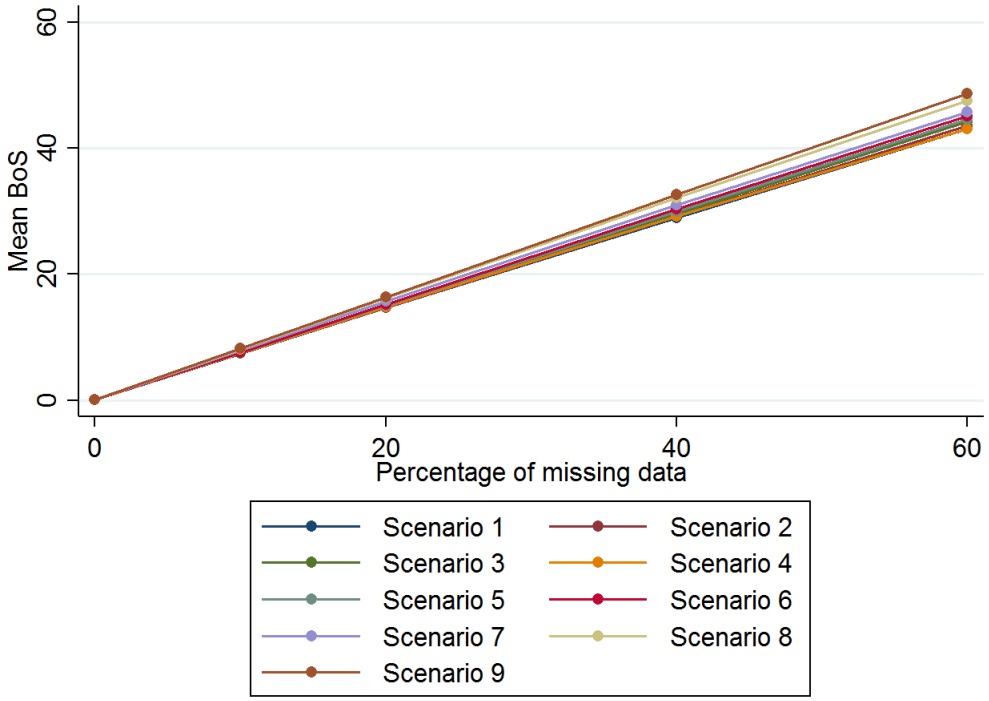


Figure 8.1: The magnitude of BoS in relation to the percentage of missing data for scenarios 1 to 9 for trials with 100 participants

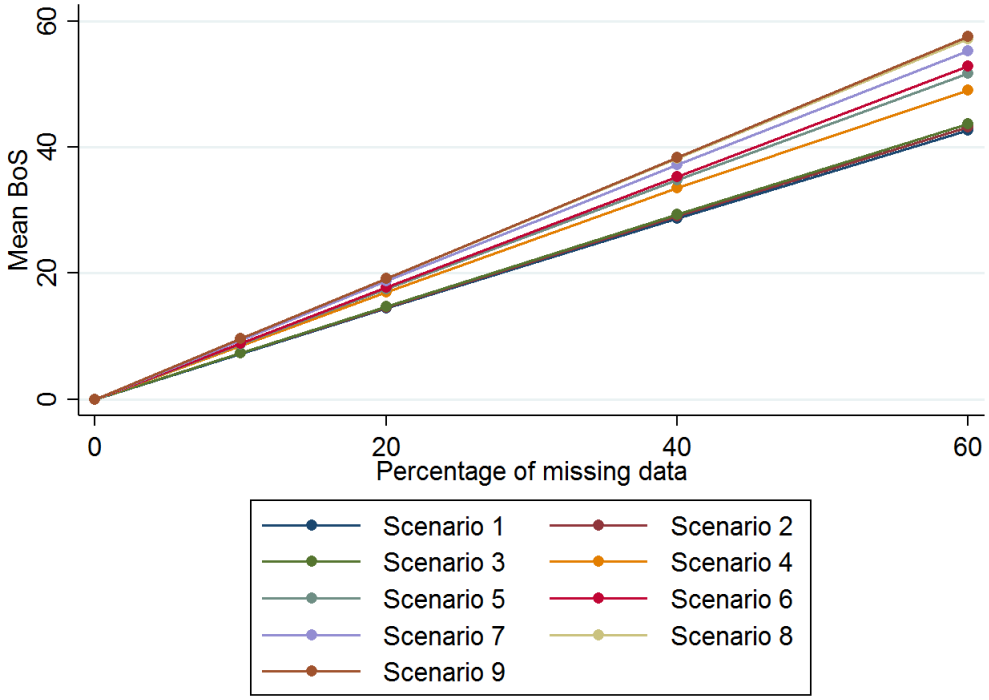


Figure 8.2: The magnitude of BoS in relation to the percentage of missing data for scenarios 1 to 9 for trials with 1000 participants

8.4.2 Comparison of percentage bias in summary treatment effect estimate

The percentage bias for summary estimates was similar for the Cochrane recommended approach and the bivariate approach (Figures 8.3 and 8.4). For example, in Figure 8.4 Scenario 8, with 60% missing data, the percentage bias was tiny; for the Cochrane approach it was -0.189% and for the bivariate approach for ANCOVA was -0.136%. In fact, the greatest difference in percentage bias between the two approaches for any of the sub-scenarios was 0.153%. The magnitude of percentage bias was also small for the univariate method for ANCOVA (Figures 8.3 and 8.4). For example, in Figure 8.4 Scenario 8 the percentage bias was -0.249% when there was no missing data and was -0.415% when the percentage of missing data increased to 60%.

The magnitude of the percentage bias increased as the between-study variance increased for all meta-analysis methods, which can be observed by looking down the columns of graphs in Figures 8.3 and 8.4. Additionally, the percentage bias was reduced when there was an increase in the number of participants in each trial and/or an increase in the number of trials in the meta-analysis (Figures 8.3 and 8.4).

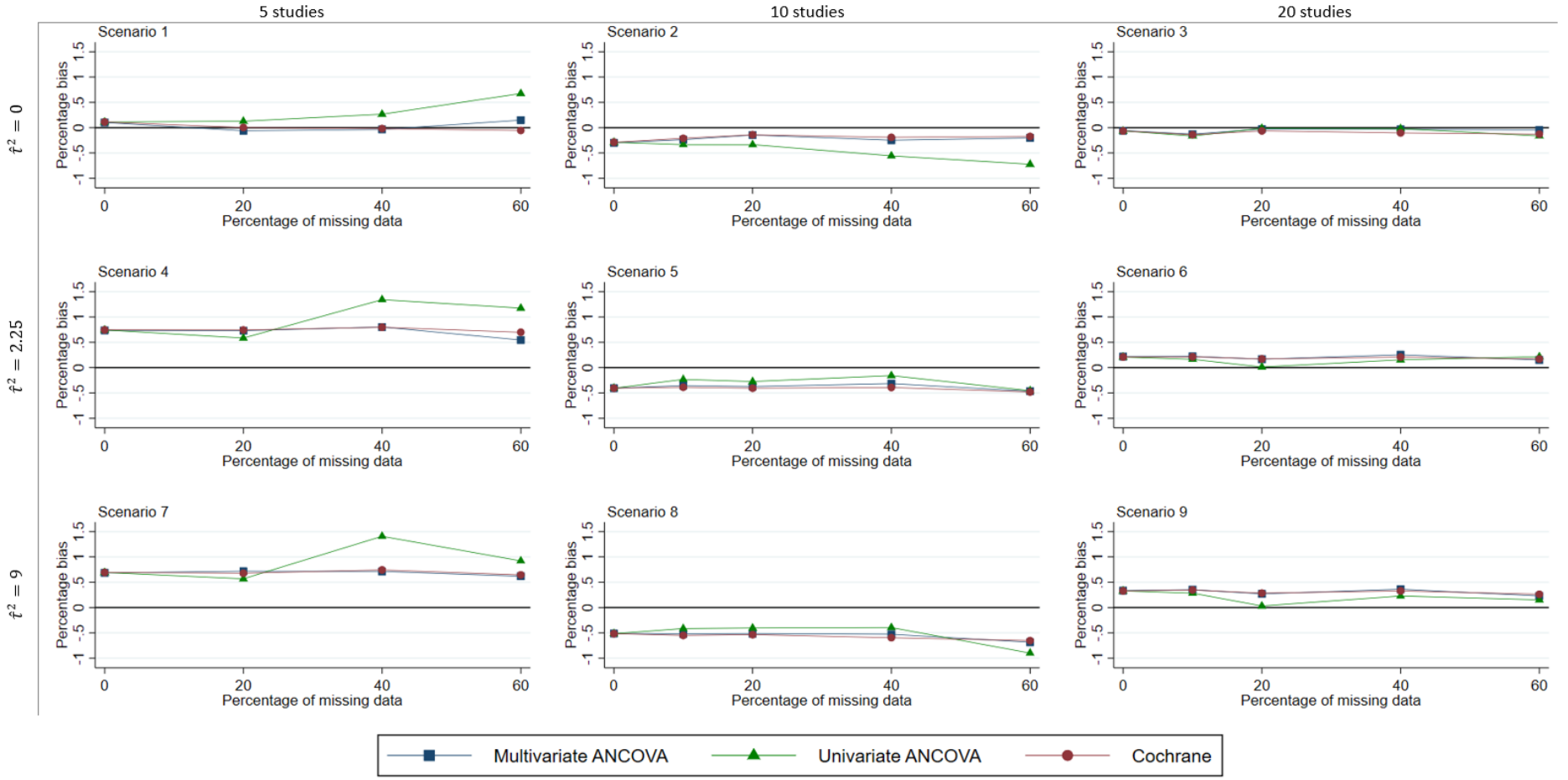


Figure 8.3: The percentage bias from the simulation of nine scenarios for trials with 100 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

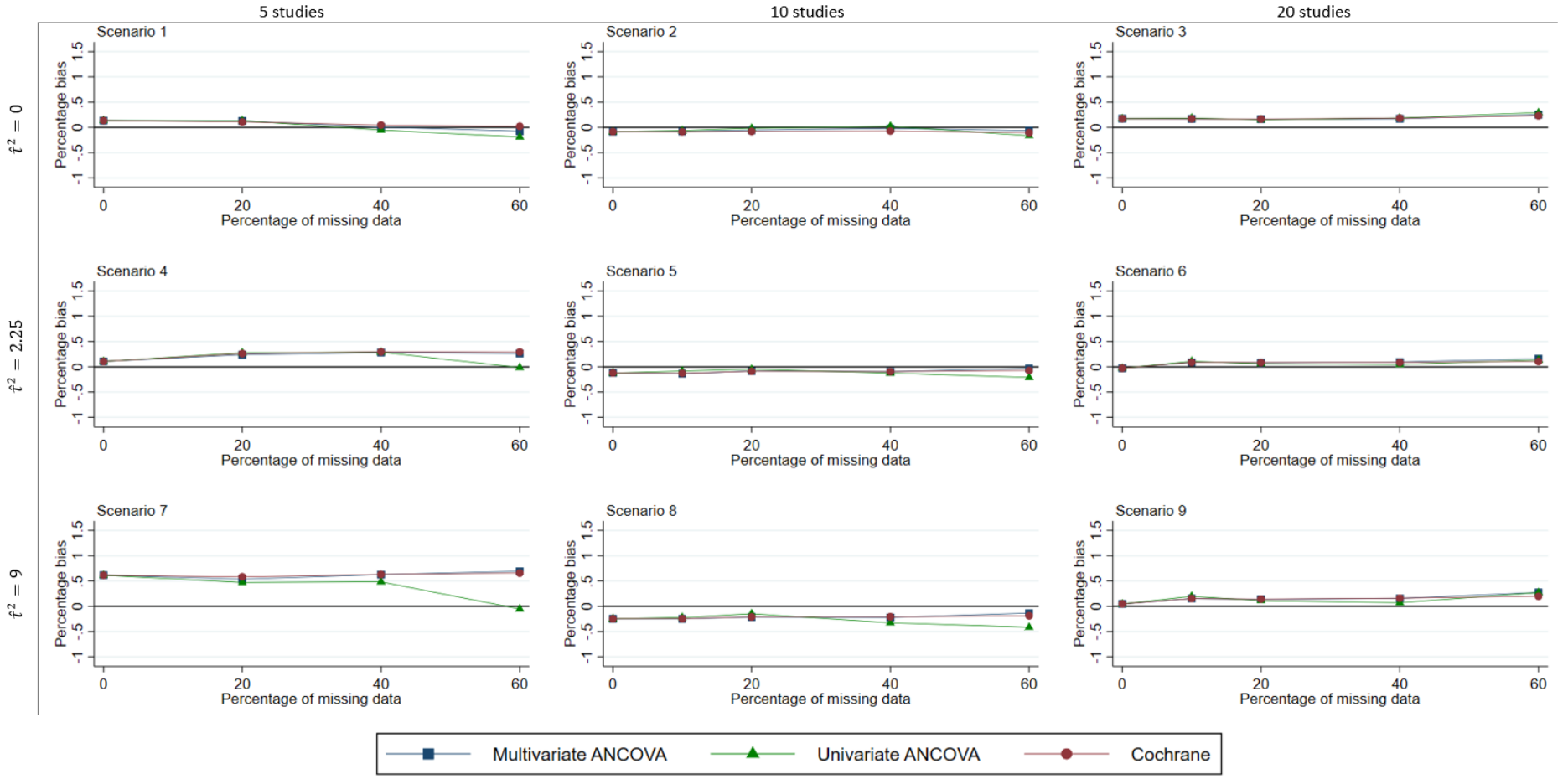


Figure 8.4: The percentage bias in nine scenarios for trials with 1000 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

8.4.3 Comparison of meta-analysis methods: mean squared error

Mean squared error was similar between Cochrane approach and bivariate approach for ANCOVA

The mean squared errors for the Cochrane recommended meta-analyses were similar to the mean squared errors for the bivariate meta-analyses for the ANCOVA outcome (Figures 8.5 and 8.6). It is observable in Figures 8.5 and 8.6, that there was little difference in the mean squared errors for the two methods. For example, in Figure 8.6 Scenario 9, the mean squared errors for the bivariate meta-analysis method for the ANCOVA outcome and the recommended by Cochrane meta-analysis method were 0.561 and 0.547, respectively, when there was 60% missing data.

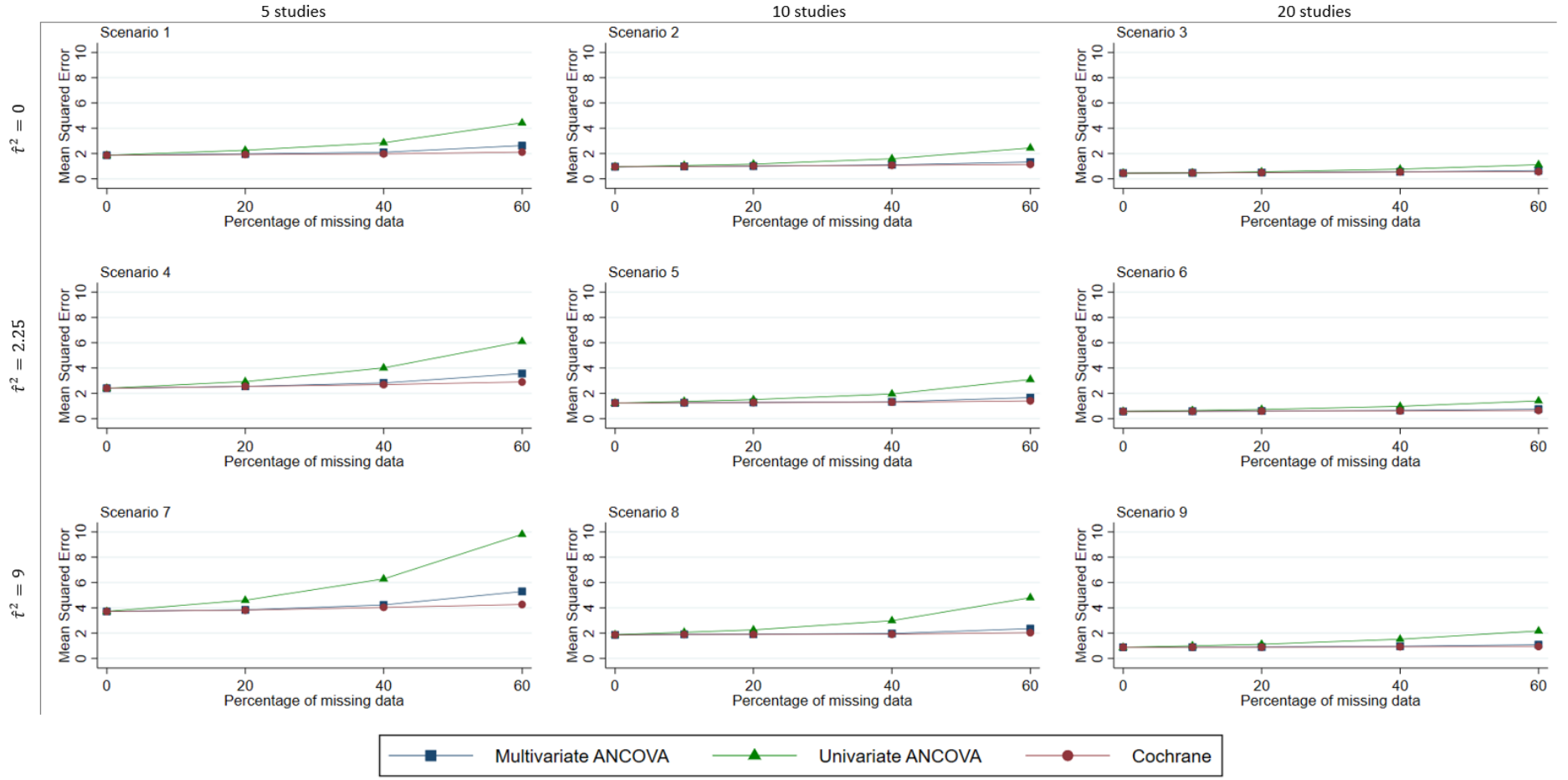


Figure 8.5: The mean squared error from the simulation of nine scenarios with baseline balance for trials with 100 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

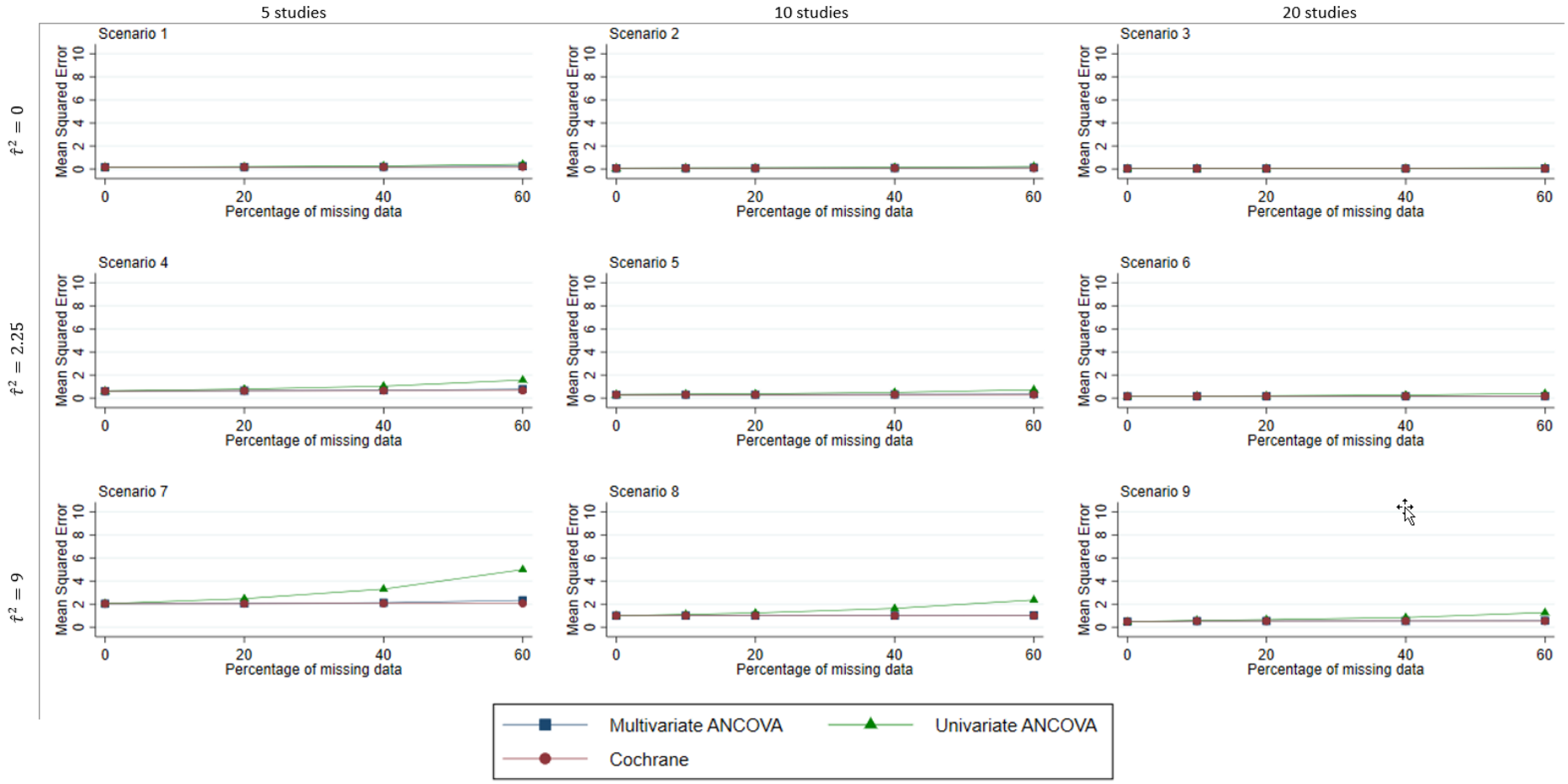


Figure 8.6: The mean squared error from the simulation of nine scenarios with baseline balance for trials with 1000 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

Mean square error increased for univariate approach as percentage of missing data increased

As the percentage of missing data increased for the ANCOVA treatment effect estimate, the mean squared error increased for the univariate meta-analysis method for the ANCOVA outcome (Figures 8.5 and 8.6). For example, in Figure 8.6 Scenario 7, the mean squared error from the univariate meta-analysis method was 2.033 for 0% missing data and was 4.991 for 60% missing data. For all other methods in Scenario 7, the mean squared error ranged from 2.033 to 2.090 for 0% missing data and ranged from 2.065 to 2.331 for 60% missing data.

8.4.4 Comparison of meta-analysis methods: mean empirical standard error

Empirical standard error smaller for Cochrane approach than bivariate approach for ANCOVA

The mean empirical standard error was often observed to be slightly smaller for the Cochrane recommended meta-analysis method than the bivariate meta-analysis method for the ANCOVA outcome (Figures 8.7 and 8.8). For example, in Figure 8.5 Scenario 8 with 60% missing data, the mean empirical standard errors for the Cochrane recommended method and the bivariate method for ANCOVA outcome were 1.412 and 1.555, respectively. This was more evident in Figure 8.7 when the number of participants in each trial was 100, particularly when the number of trials in the meta-analysis was small e.g. 5 trials.

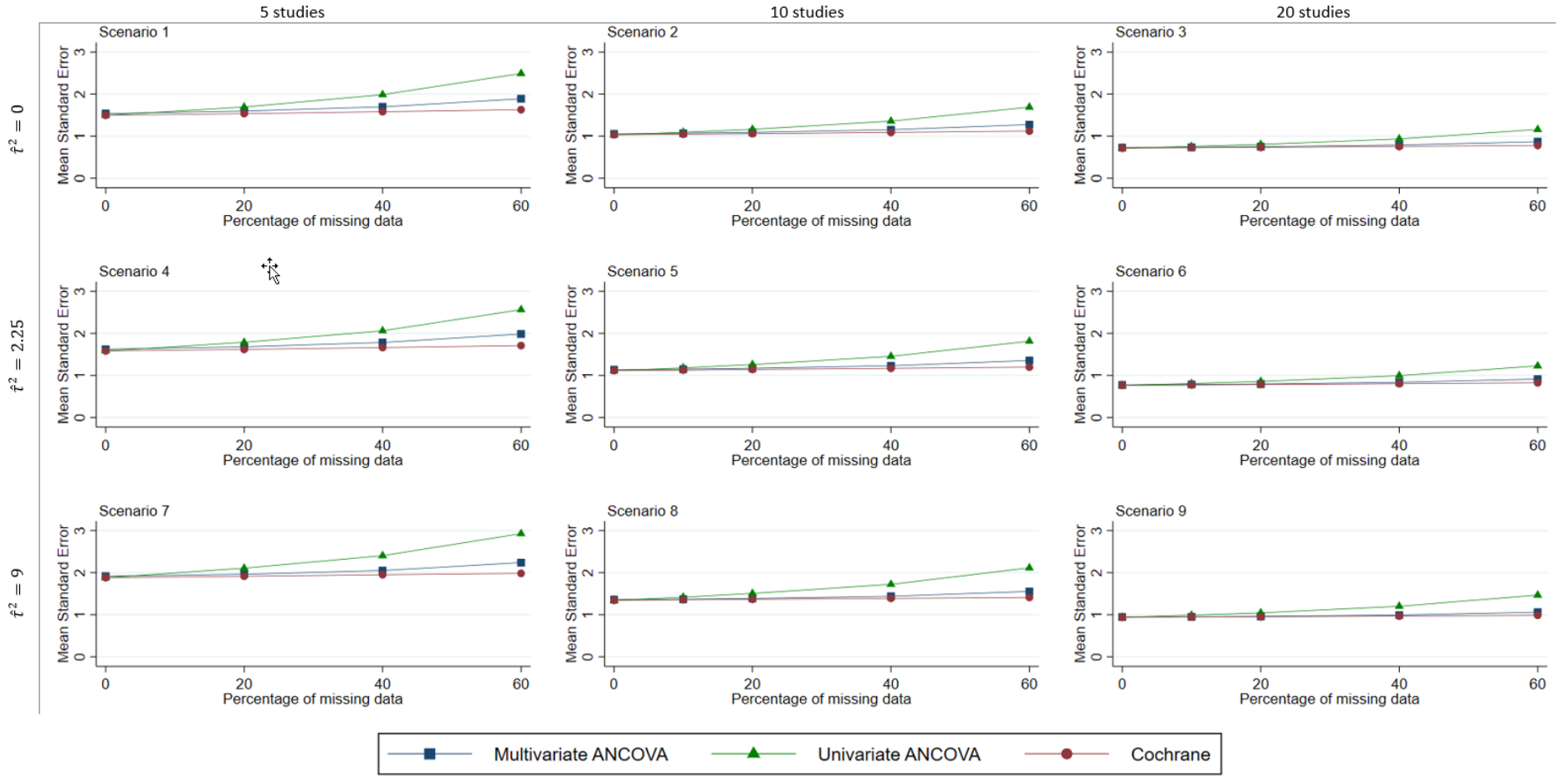


Figure 8.7: The mean empirical standard error from the simulation of nine scenarios with baseline balance for trials with 100 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

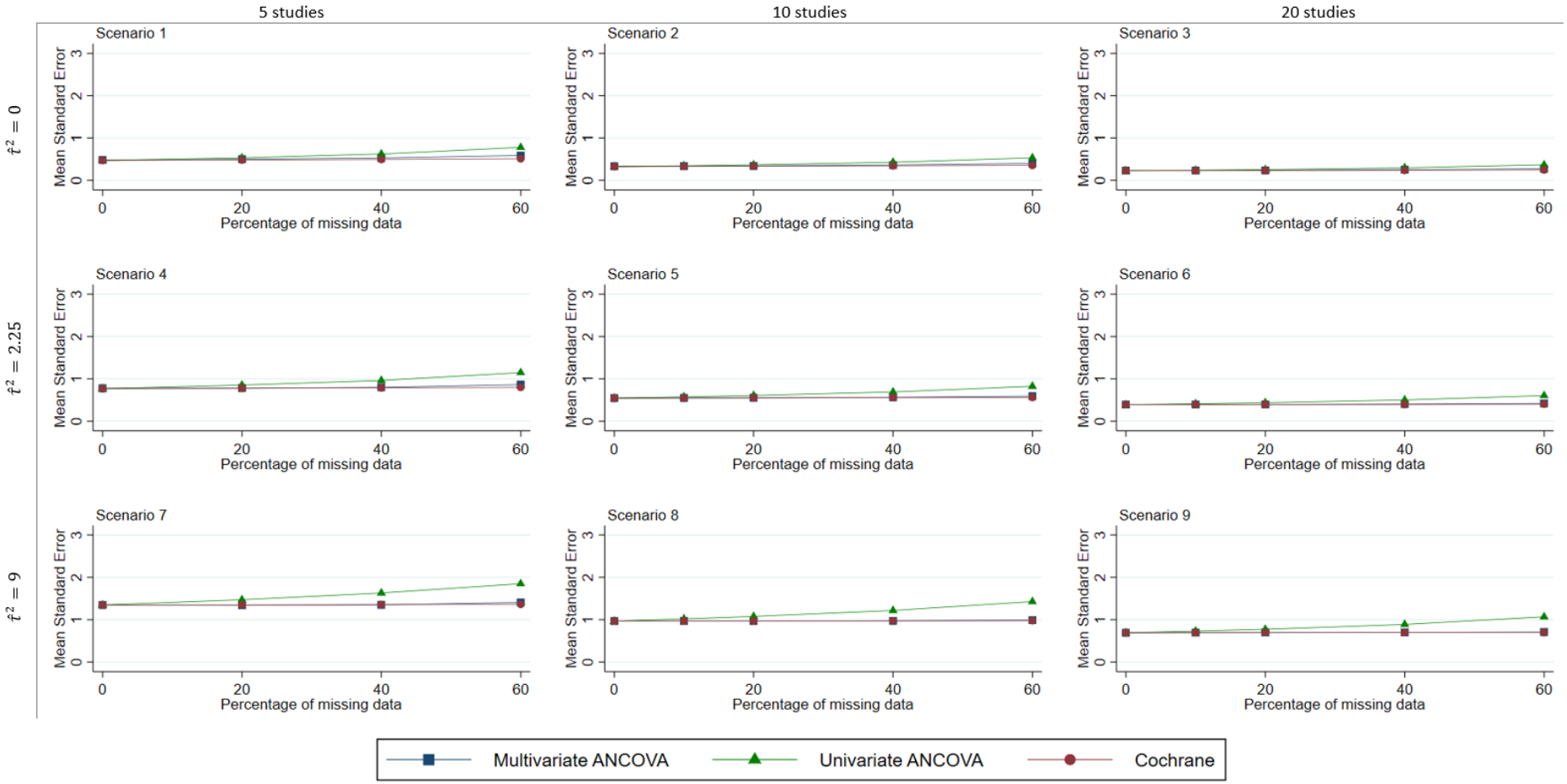


Figure 8.8: The mean empirical standard error from the simulation of nine scenarios with baseline balance for trials with 1000 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

Empirical standard error is large for univariate approach for ANCOVA when the percentage of missing data is large

The mean empirical standard error increased for the univariate approach as the percentage of missing data for the ANCOVA treatment effect estimates increased (Figures 8.5 and 8.6). This is evident in Figure 8.6 Scenario 9, since the mean empirical standard increases from 0.691 for 0% missing data to 1.069 for 60% missing data. The mean empirical standard errors remain similar across the different percentages of missing data for the other meta-analysis methods. For example, in Figure 8.6 Scenario 9, the mean empirical standard error for the Cochrane recommended method was 0.691 for 0% missing data and 0.699 for 60% missing data.

8.4.5 Comparison of meta-analysis methods: performance of 95% confidence interval

Coverage was similar between Cochrane approach and bivariate approach for ANCOVA

The coverage of the 95% confidence interval was similar between the Cochrane recommended meta-analysis method and the bivariate meta-analysis method for the ANCOVA outcome (Figures 8.9 and 8.10). For example, in Figure 8.10 Scenario 7, with 60% missing data for the ANCOVA treatment effect estimate the coverage for the Cochrane recommended meta-analysis method was 89.325%, compared to a coverage of 87.714% for the bivariate meta-analysis method. Interestingly, they were both much less than 95%, suggesting that corrections, such as Hartung-Knapp would be useful.

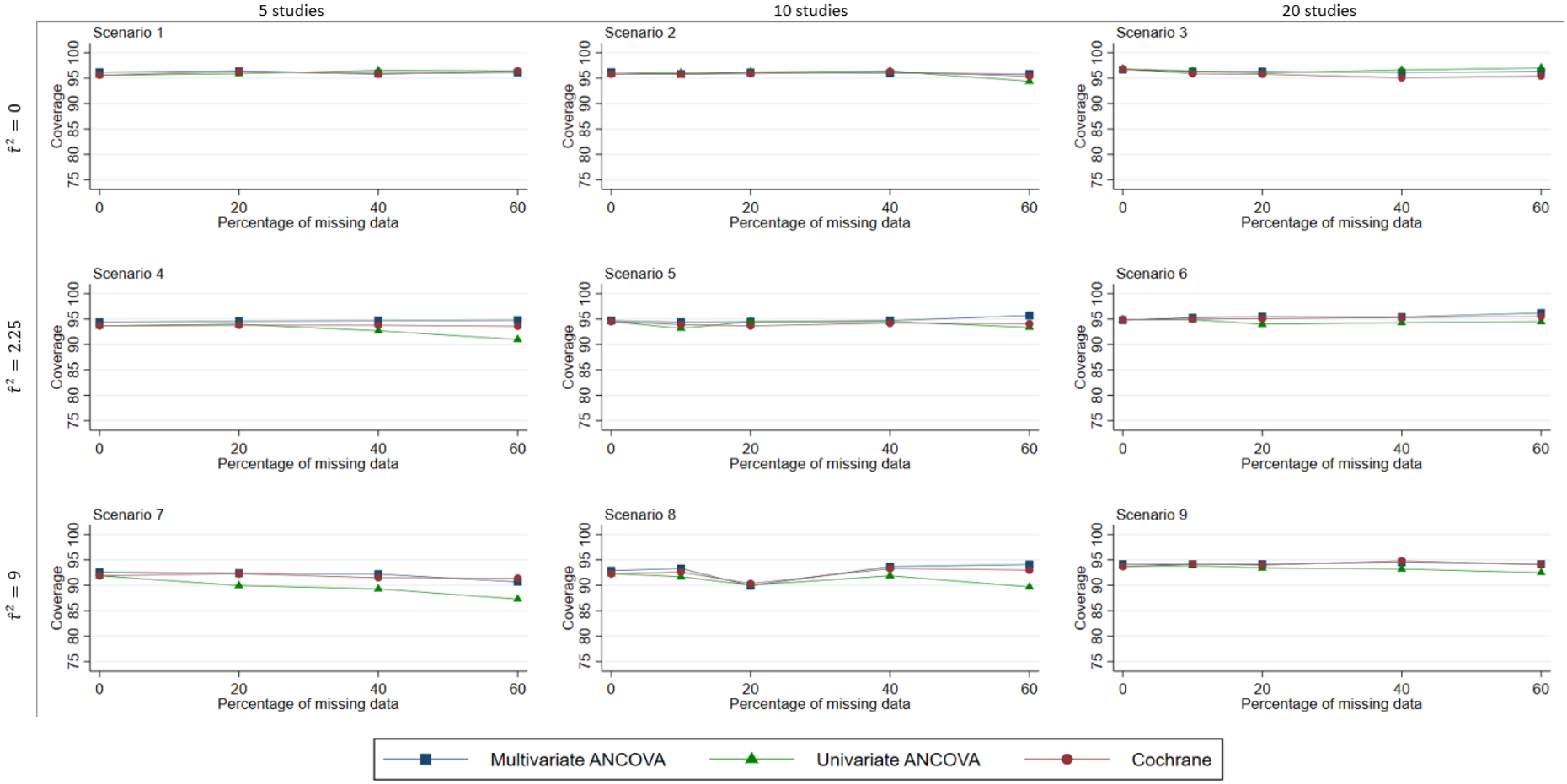


Figure 8.9: The coverage of the 95% confidence interval from the simulation of nine scenarios with baseline balance for trials with 100 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

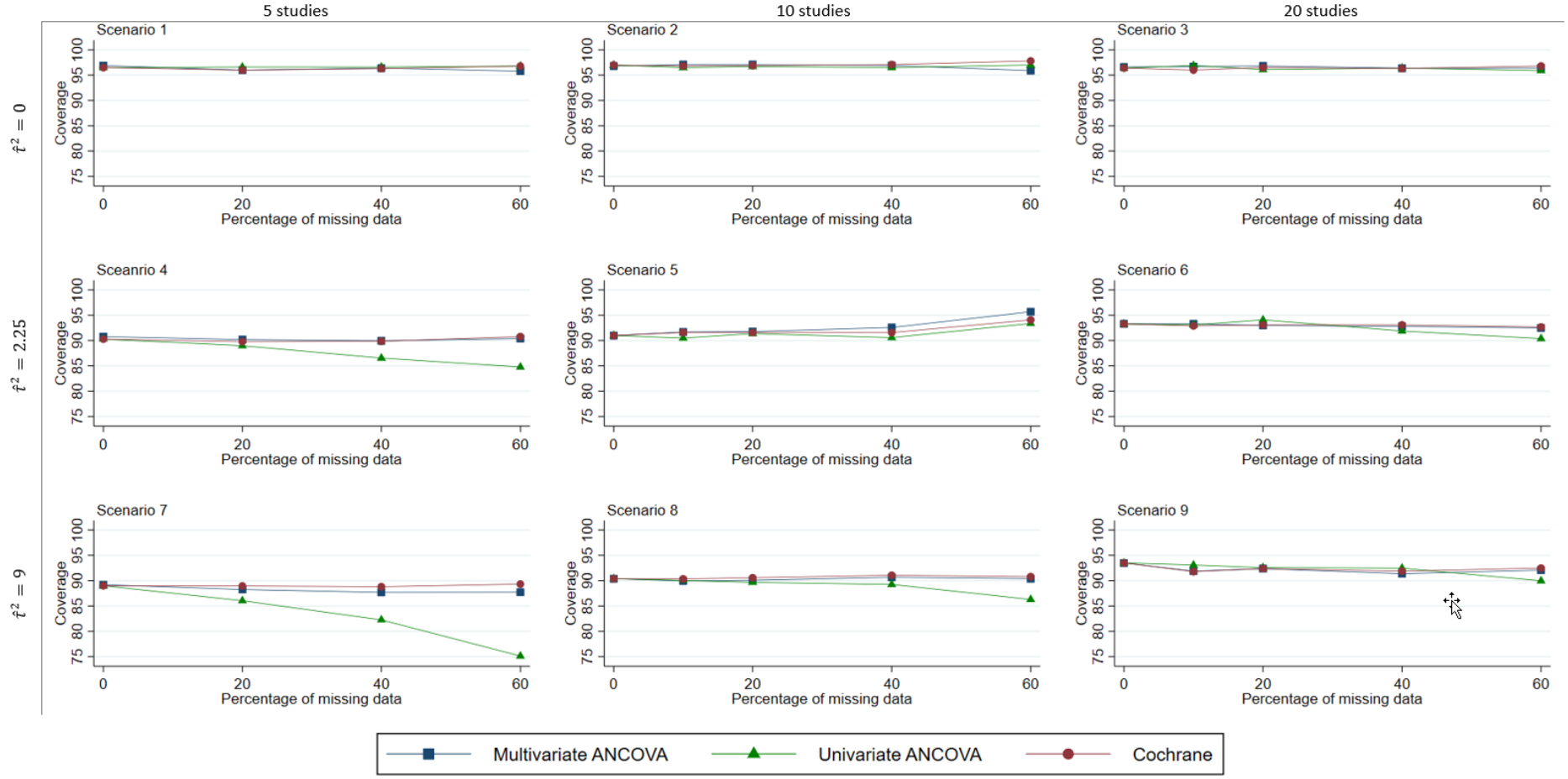


Figure 8.10: The coverage of the 95% confidence interval from the simulation of nine scenarios with baseline balance for trials with 1000 participants for the ANCOVA outcome from the multivariate and univariate meta-analyses and for the Cochrane meta-analysis method

Power was similar for the bivariate and Cochrane methods

The power was similar between all the meta-analysis methods (Table F.3 in Appendix F.2). In the majority of scenarios, the power was 100% for all meta-analysis methods. For all the meta-analysis methods excluding the univariate meta-analysis method for the ANCOVA outcome the power ranged from 97.898% to 100%. The power was less for the univariate meta-analysis method when there were large quantities of missing data, the number of participants in each trial was small and the between-study variance was large. For example, for scenario 7, with 100 participants per trial the power was 86.587% when there was 60% missing data, much larger than those for the Cochrane or bivariate methods.

Coverage from univariate approach for ANCOVA decreased as percentage of missing data increased

The coverage was worse for the univariate meta-analysis for only ANCOVA treatment effect estimates compared to the multivariate meta-analysis or the Cochrane meta-analysis approach. As the percentage of missing data increased for the ANCOVA treatment effect estimates, the coverage of the 95% confidence interval for the univariate meta-analysis method for the ANCOVA outcome decreased (Figures 8.9 and 8.10). This was particularly evident when there was a small number of studies included in the meta-analysis, as shown in Scenarios 4 and 7, in Figures 8.9 and 8.10. For example, in Figure 8.10 scenario 7, the coverage for 0% missing data was 89.0% compared to 75.126% when the missing data was 60%.

8.4.6 Comparison of meta-analysis methods: bias in heterogeneity estimates

Bias in between-study variance often less for Cochrane approach than the bivariate approach for ANCOVA

The magnitude of the bias in the estimate of the between-study variance was often greater for the bivariate meta-analysis method for the ANCOVA outcome compared to the Cochrane method (Figures 8.11 and 8.12). For example, in Figure 8.12 scenario 8 (recall that Scenario 8 had a true between-study variance of 9), when there was 60% missing data the bias for the Cochrane recommended meta-analysis method was 0.056 compared to -0.153 for the bivariate meta-analysis method for the ANCOVA outcome.

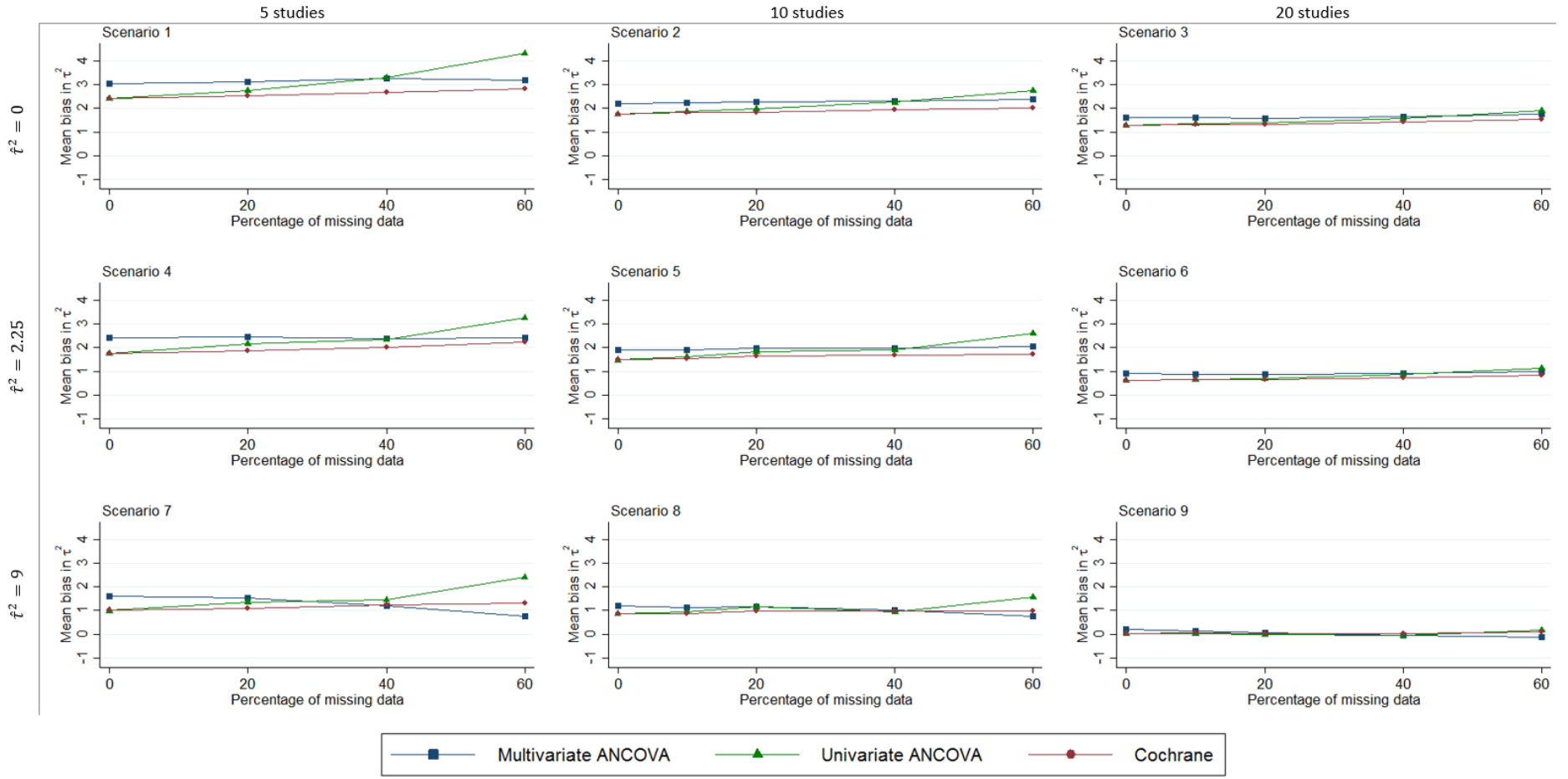


Figure 8.11: The bias in the estimate of the between-study variance from the simulation of nine scenarios with baseline balance for trials with 100 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

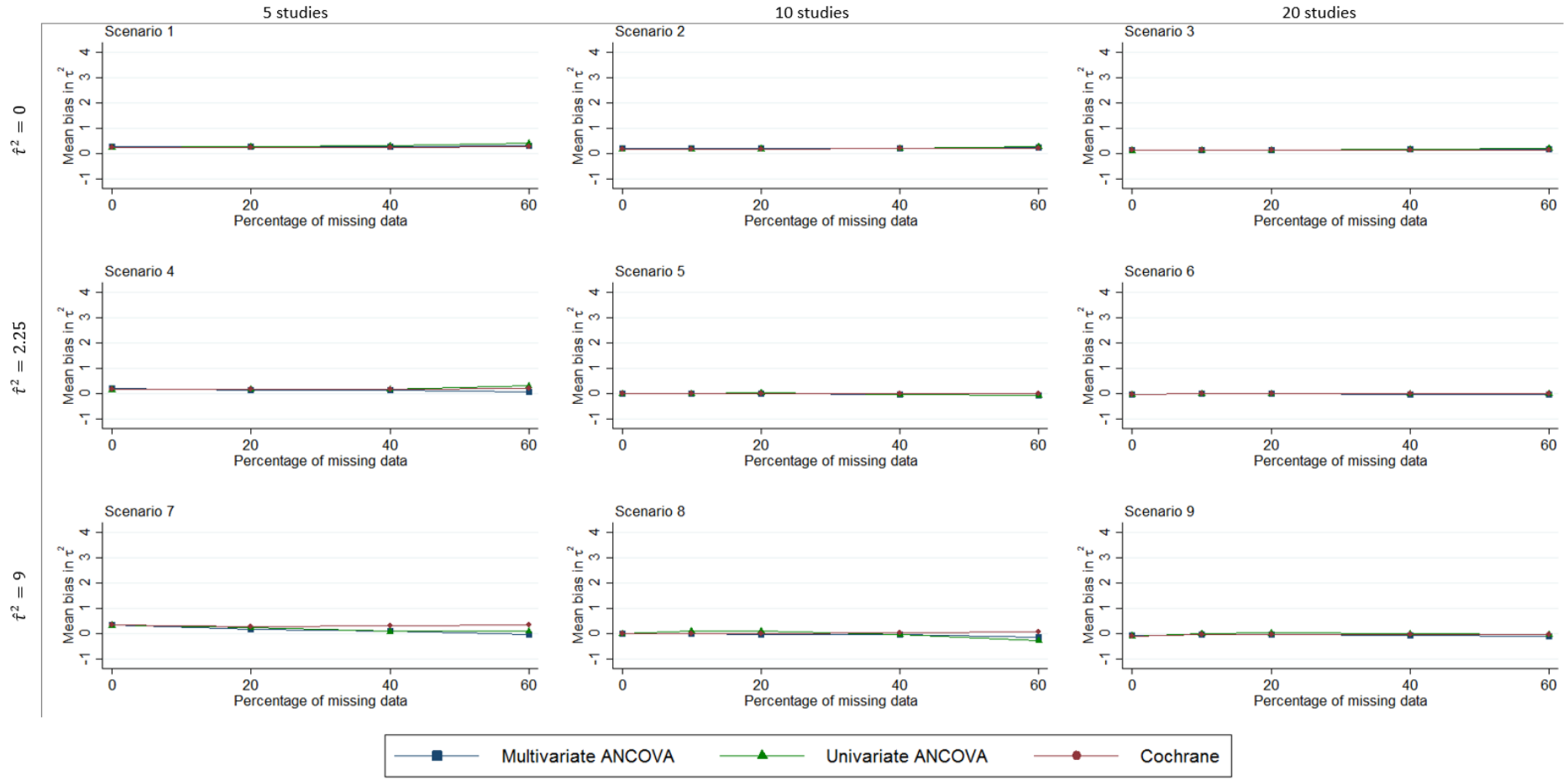


Figure 8.12: The bias in the estimate of the between-study variance from the simulation of nine scenarios with baseline balance for trials with 1000 participants for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

8.5 Results: Part 2 - Baseline imbalance

The results for the scenarios with baseline imbalance are now described.

8.5.1 Unconditional baseline imbalance across all trials

Comparison between the bivariate and Cochrane recommended methods

For the unconditional baseline imbalance setting, each trial in the meta-analysis had baseline imbalance across treatment groups. In this setting, the bivariate meta-analysis method for the ANCOVA treatment effect estimates had more desirable performance measure results when there was missing data, compared to the alternative Cochrane recommended meta-analysis method (Figure 8.13). The magnitude of the percentage bias was smaller for the bivariate approach for the ANCOVA treatment effect estimates (range: -0.113 to 0.007) compared to the Cochrane approach (range: -0.129 to 14.408); this was especially apparent when the percentage of missing data was large.

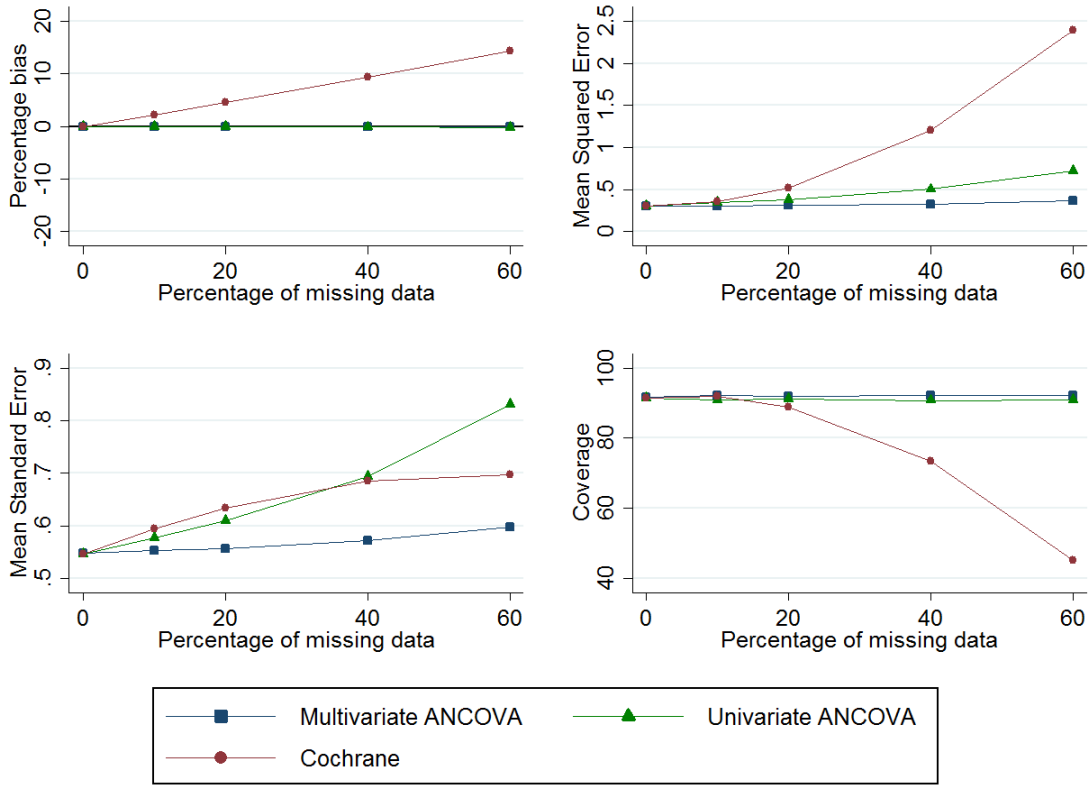


Figure 8.13: Performance measures for the analysis of the unconditional baseline imbalance for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

Additionally, the mean squared errors and the mean empirical standard errors were greater for the Cochrane recommended meta-analysis method than the bivariate meta-analysis method for the ANCOVA outcome when the percentage of missing data was large (Figure 8.13). For example, the mean squared errors ranged from 0.306 to 0.363 for the bivariate meta-analysis method for the ANCOVA compared to 0.306 to 2.396 for the Cochrane recommended meta-analysis method.

For increasing missing data, the coverage of the confidence interval decreased for the Cochrane recommended meta-analysis method. For example, when there was 60% missing data, the coverage was 45%, whereas the coverage for the bivariate method for the ANCOVA treatment effect estimates was close to 95% (range: 91.8% to 92.3%) regardless of the quantity of missing data in the meta-analysis.

Univariate meta-analysis for the ANCOVA outcome

The separate univariate meta-analysis method for the treatment effect estimates from the ANCOVA model had desirable performance measure results (Figure 8.13). The percentage bias was small in magnitude (range: -0.06% to -0.20%) and the coverage for the confidence interval was close to 95% (range: 90.8% to 91.5%) for each percentage missing data quantity. This is since the method only ever uses studies with available ANCOVA results, and thus the baseline imbalance is addressed directly. However, a downside is that this method does not use all available studies and therefore is not utilising all available information.

8.5.2 Conditional baseline imbalance

First conditional baseline imbalance setting

The first conditional baseline imbalance setting was the trials with no IPD (missing ANCOVA treatment effect estimates), which had baseline imbalance across their treatment arms (treatment or control) and the trials with IPD had baseline balance across their treatment arms. The performance measures were very similar between the Cochrane approach and the bivariate approach for the ANCOVA treatment effect estimates (Figure 8.14). For example, when there was 10% missing data, the percentage bias for the Cochrane approach and the bivariate approach were 4.567% and 4.248%, respectively. For settings with a percentage of missing data greater than 20%, the performance measure results were less desirable for both methods. For example, when there was 60% missing data, the percentage bias for the Cochrane approach and the bivariate approach were 14.431% and 12.7384%, respectively. The bivariate approach for the ANCOVA treatment effect estimates had slightly greater coverage than the Cochrane method for 60% missing data with a coverage of 56.9% compared to 44.2%. Both coverage percentages were considerably lower than the 95% that would be desired.

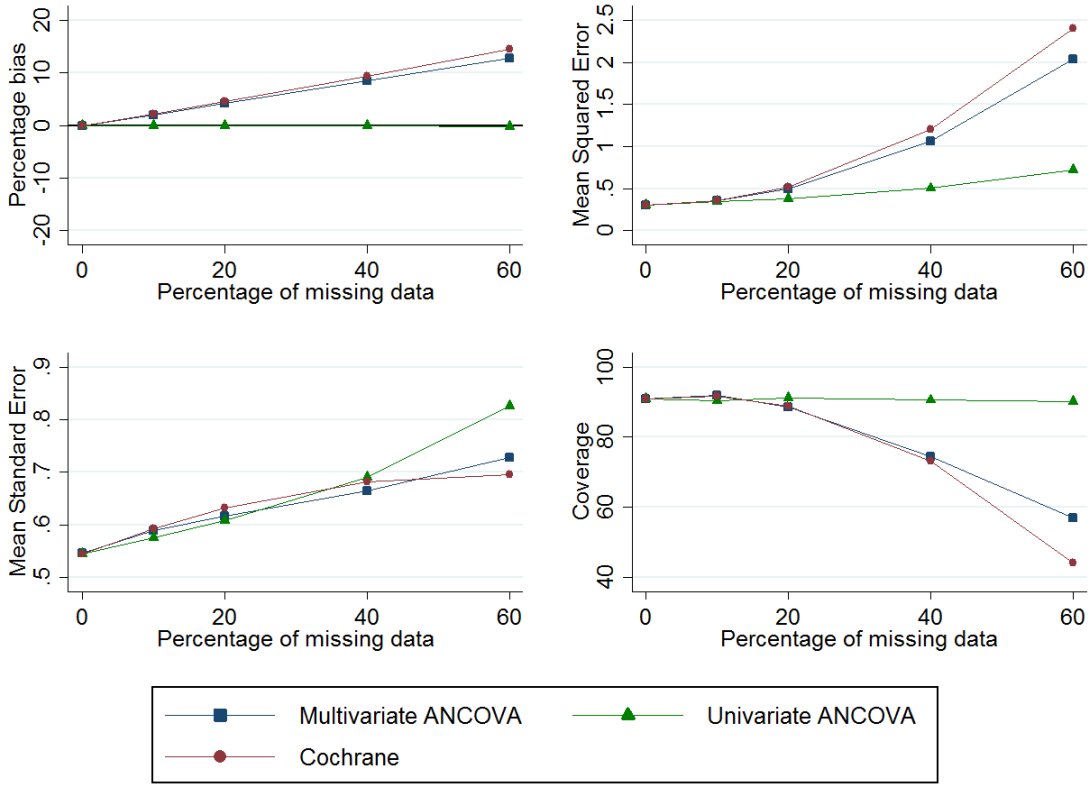


Figure 8.14: Performance measures for the analysis of the conditional baseline imbalance, the trials with no IPD have baseline imbalance across treatment groups, for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

In this baseline imbalance setting, applying a separate univariate meta-analysis for the ANCOVA treatment effect estimates had more desirable performance measures than the other methods (Figure 8.14). For example, the percentage bias from the univariate meta-analysis method for the ANCOVA outcome was -0.175% whereas for the other methods the percentage bias ranged from 12.738% to 14.802%. It was expected, in this baseline imbalance setting, that the separate univariate meta-analysis would have better performance than the other methods. This is because the separate univariate meta-analysis for the ANCOVA treatment effects is the only meta-analysis method that provides results that are not influenced by the baseline imbalance in the trials without IPD.

Second conditional baseline imbalance setting

The second conditional baseline imbalance setting that was studied in this chapter had baseline imbalance across treatment groups for trials that had available IPD. For the trials with missing ANCOVA treatment effect estimates, there was baseline balance across the treatment groups. In this baseline imbalance setting, the Cochrane method had more desirable performance measures compared to the bivariate approach for the treatment effect estimates from the ANCOVA model (Figure 8.15). The percentage bias magnitude for the ANCOVA treatment effect estimates from the bivariate approach increased as the percentage of missing data increased. For example, the percentage bias for the bivariate approach for the ANCOVA outcome was -0.074 when there was 0% missing data and was -12.664 when there was 60% missing data. However, the percentage bias magnitude did not increase for the Cochrane approach but was close to 0%, ranging from 0.029 to 0.217 across all the percentages of missing data.

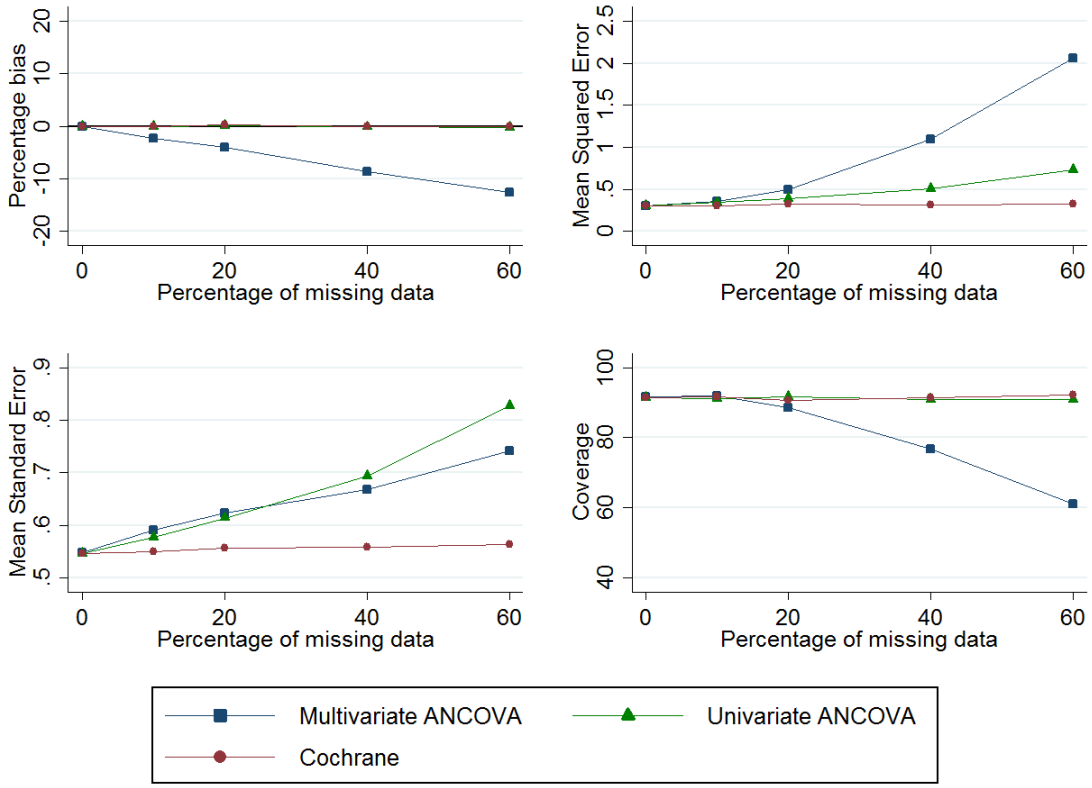


Figure 8.15: Performance measures for the analysis of the conditional baseline imbalance, the trials with IPD have baseline imbalance across treatment groups, for the methods: ANCOVA outcome form multivariate meta-analysis, the univariate meta-analysis for the ANCOVA and the Cochrane meta-analysis method

Additionally, the coverage of the 95% confidence interval for the Cochrane approach was closer to 95% (range: 90.644% to 92.285%) than the bivariate approach for the ANCOVA treatment effects (range: 61.222% to 91.968%) (Figure 8.15). The mean squared error was smaller for the Cochrane approach than the bivariate approach for the ANCOVA outcome when there was increasing missing data. For example, when there was 60% missing data, the mean squared errors for the Cochrane approach and the bivariate approach for the ANCOVA outcome were 0.321 and 2.059, respectively.

8.6 Discussion

This chapter assessed the performance of meta-analysis methods for the synthesis of treatment effect estimates from randomised control trials with continuous outcomes modelled using either the ANCOVA or final score models. The meta-analysis methods analysed were the current univariate approaches (separate and Cochrane recommended) and the newly proposed bivariate approach. These were analysed under a variety of different simulated scenarios to assess how each meta-analysis method's performance is affected by particular conditions. The key findings from the simulation study are described in Figure 8.16 and discussed in further detail below.

Figure 8.16: Key findings from the simulation study for the analysis of meta-analysis methods

Key Findings:

- The performance measures for the Cochrane recommended meta-analysis method were similar to the performance measures for the bivariate meta-analysis for scenarios with baseline balance.
- The performance measures for the univariate meta-analysis for the ANCOVA outcome were poor compared to the performance measures for the other meta-analysis methods in situations where trials have baseline balance
- When all the trials with IPD have baseline balance across treatment arms, the choice between the Cochrane recommended meta-analysis method and the bivariate meta-analysis method for the ANCOVA outcome had limited importance.
- When there is baseline imbalance in trials without ANCOVA estimates, both the bivariate method and the Cochrane method performed poorly; rather, it is better to use a separate univariate meta-analysis of just those

trials that provided ANCOVA estimates.

- When there is baseline imbalance in trials with ANCOVA estimates but not in others then the Cochrane method or bivariate method could be used.

Comparison between Cochrane recommended meta-analysis method and bivariate meta-analysis method for the ANCOVA treatment effect estimates

The performance measures were similar between the Cochrane recommended meta-analysis method and the bivariate meta-analysis method for the ANCOVA outcome when there was baseline balance, irrespective of the quantity of missing data. Both methods were affected in the same ways from increases in the between-study variance, increases in the number of trials within the meta-analysis and increases in the number of participants in each trial. Therefore, similar results were seen between these two methods from different scenarios.

The Cochrane recommended meta-analysis method and the bivariate meta-analysis method for the ANCOVA outcome performed very differently within the baseline imbalance settings. This simulation study considered three different baseline imbalance settings. The first baseline imbalance setting was the unconditional baseline imbalance setting, where all trials had baseline imbalance across the treatment arms. In this setting, the bivariate method for the ANCOVA outcome had more desirable performance measure values than the Cochrane recommended method.

The remaining two baseline imbalance settings were conditional upon whether or not a trial had available IPD (reported the treatment effect estimate from the ANCOVA model). In the first conditional baseline imbalance setting, there was baseline imbalance across treatment arms for trials without IPD (missing treatment effect estimates from the ANCOVA model). The bivariate meta-analysis for the ANCOVA outcome method's performance measure values were more similar to the Cochrane

recommended method than for the unconditional baseline imbalance setting. Although the performance measures were more similar, the performance measure values were more desirable for the bivariate method than the Cochrane recommended method.

In the second conditional baseline imbalance setting, there was baseline imbalance across treatment arms for trials with IPD. The Cochrane recommended meta-analysis method had more desirable performance measures in this setting than the bivariate meta-analysis method for the ANCOVA treatment effects. This was the only baseline imbalance setting studied in this chapter that favoured the Cochrane recommended approach.

Separate univariate meta-analysis method

The baseline balance and the unconditional baseline imbalance settings agreed with the findings from Chapter 7; the univariate meta-analysis for the ANCOVA treatment effects performs poorly in the presence of missing ANCOVA treatment effect estimates when there is baseline balance in other trials that provide other results. This is because the univariate meta-analysis is ignoring information from studies without ANCOVA treatment effect estimates. This information is informative to the true treatment effect and unbiased given the groups are balanced in those trials. Therefore, the separate univariate meta-analyses should be avoided when there are missing ANCOVA treatment effect estimates from RCTs when there is baseline balance.

However, for the conditional baseline imbalance settings, the separate univariate meta-analysis performed well compared to the other meta-analysis methods. Although it performs well, the separate univariate meta-analysis does not include information from studies that do not report the ANCOVA treatment effect estimates, causing a loss of information. The separate univariate meta-analysis is observed to perform well since the ANCOVA model accounts for the baseline imbalance, therefore all the information in the univariate meta-analysis is not influenced by the baseline imbalance

as it pools only ANCOVA results.

8.6.1 Limitations and further research

Due to estimation problems and the relationship between the analytical models, multivariate meta-analyses of all three were not used. The simulation study used a bivariate meta-analysis of ANCOVA treatment effects and final score treatment effects. The final score model was used over the change score model, since the final score treatment effects are more commonly reported and can be derived from aggregated data more easily. Although it might be expected that results would be similar.

This thesis has studied settings with multiple correlated outcomes. Further research is needed for meta-analyses of multiple correlated continuous outcomes measured at baseline and follow-up.

8.6.2 Recommendations

The Cochrane recommended meta-analysis method of combining ANCOVA and final score estimates performs well when there are no concerns regarding baseline balance across treatment groups in studies that did not provide ANCOVA estimates. The Cochrane recommended method performs as well as the bivariate meta-analysis for the ANCOVA outcome when there is baseline balance across treatment groups for all trials. The Cochrane recommended meta-analysis method is advantageous over the bivariate meta-analysis method, since it is less complicated and does not require IPD for the calculation of within-study correlations.

When there is baseline imbalance across treatment groups in some trials, it is recommended that a separate univariate meta-analysis of the available ANCOVA treatment effect estimates is used. This is irrespective of whether the baseline imbalance is conditional on available ANCOVA treatment effects or unconditional across all trials. In these settings the performance measures for the Cochrane method

and bivariate method were poor compared to had ANCOVA results been available for all trials.

Chapter 9

Discussion

This chapter provides a summary of the key findings from this thesis and discusses their implications. An aim of this chapter is to seek to answer some further questions that arise from the key findings and to make recommendations for future meta-analyses of multiple outcomes. The limitations of this work are presented and areas of further research are discussed (Section 9.5). Lastly, this thesis closes with some final conclusions (Section 9.6).

9.1 Overview of the thesis

There are different meta-analysis approaches to analysing multiple outcomes from independent studies to obtain an effect estimate for each outcome. The most common approach is separate univariate meta-analyses for each outcome; however, there exist multivariate meta-analysis methods that can jointly analyse multiple outcomes, utilising within-study and between-study correlations. This thesis studied the benefits and application of multivariate meta-analysis over the univariate meta-analysis approach.

In Chapter 2, the results from multivariate meta-analyses were compared to results from separate univariate meta-analyses. A Health Technology Assessment (HTA) report that investigated the effects of diet and exercise interventions during pregnancy for the reduction of weight gain on maternal and foetal outcomes was reanalysed (Rogozińska et al. 2017). The HTA report studied multiple outcomes for

mothers and foetuses in separate univariate meta-analyses as well as developing a maternal composite outcome and a foetal composite outcome which were analysed in univariate meta-analyses (Rogozińska et al. 2016). For all trials in the HTA report, the individual participant data (IPD) was available and therefore the within-study correlations of each pair of outcomes were available. The HTA report had a large number of trials, however there was missing data for some outcomes from some trials, thus there was an opportunity to borrow strength via the correlation of outcomes in a multivariate meta-analysis. However, the findings show the results between the univariate meta-analyses and the multivariate meta-analysis were generally similar for the maternal and foetal outcomes e.g. there were no changes in statistical or clinical significance. However, potentially clinically important differences were identified in two maternal outcomes, pre-eclampsia or pregnancy induced hypertension (PE or PIH) and gestational diabetes. The odds of gestational diabetes for the diet and exercise intervention compared to control were estimated to be lower from the multivariate meta-analysis than from univariate meta-analysis, similarly for PE or PIH.

There are many papers that have investigated the difference in multivariate and univariate results, and concluded that often there was little difference between the univariate and multivariate results (Sohn 2000, Simel & Bossuyt 2009, Trikalinos et al. 2013, 2014). Despite the fact that the benefits of multivariate meta-analysis were limited in the HTA report, the benefits of multivariate meta-analysis are well documented in the literature (Riley et al. 2007*a*,*b*, Hamza et al. 2009, Jones et al. 2009, Jackson et al. 2011, Kirkham et al. 2012, Riley et al. 2015, Frosi et al. 2015). Benefits of multivariate meta-analysis include reduction in bias due to partial reporting (Kirkham et al. 2012, Frosi et al. 2015), improved statistical properties (Riley et al. 2007*a*,*b*), and joint inferences (Riley et al. 2015) (Chapter 1). A specific benefit of multivariate meta-analysis, that was of particular interest in this thesis, is borrowing strength across outcomes (Riley et al. 2007*a*, 2015, Jackson et al. 2017, Copas et al. 2018).

The borrowing of strength (BoS) quantifies the gain in precision from utilising extra information in a multivariate meta-analysis compared to a univariate meta-analysis. The BoS statistics were larger for the outcomes, PE or PIH and gestational diabetes, where potentially important clinical differences were observed between the multivariate and univariate meta-analysis results, than the outcomes with no difference between the multivariate and univariate results.

In Chapter 3, the distribution of BoS was studied in the Trikalinos et al. (2013, 2014) review, which concluded there were generally limited differences between multivariate and univariate meta-analysis results. The aim was to investigate whether the BoS statistic identified settings where the multivariate meta-analysis was beneficial over the univariate approach. In the data from the Trikalinos et al. (2013, 2014) review very large BoS statistics were calculated e.g. 51.4% (see Section 3.3.3) in a few of the applications. It was observed empirically that the BoS statistic was greater when there were greater differences between the multivariate and univariate meta-analysis results. However, the magnitude of BoS statistic was only observed to identify differences in the widths of confidence intervals, and may not be sensitive to changes in the pooled estimate.

The BoS statistic is calculated during a multivariate meta-analysis. For the BoS statistic to be used in determining whether a univariate or multivariate meta-analysis would be beneficial, there was a need to be able to predict the magnitude of BoS in the setting before any analysis. Some meta-analysis level characteristics of univariate and multivariate meta-analyses were selected as possible candidate predictors and in Chapter 4 prediction models for the magnitude of BoS from fixed-effect and random-effects meta-analyses were developed. The common predictors in the two parsimonious final models (one for each fixed-effect and random-effects meta-analyses) were the number of studies in the meta-analysis, percentage of missing

data for the outcome of interest and the maximum within-study correlation. The relationship between BoS and the within-study correlations has been discussed previously (Jackson et al. 2017), such that no borrowing of strength will occur when there is no within-study correlation in a fixed-effect meta-analysis. The percentage of missing data was expected to be important, since there is a greater need for borrowing strength with an increase in missing data (Riley et al. 2007*a*). The number of studies is important, as the more studies available, the more outcome estimates available to borrow strength from.

The prediction models for the BoS statistic for the fixed-effect and the random-effects meta-analyses were embedded in interactive graphical tools using R shiny in Chapter 5. These tools allow for researchers to more easily predict the magnitude of the BoS statistic in their meta-analysis setting. Following the importance of the meta-analysis level characteristics as predictors for BoS in Chapter 4 and Chapter 5, the relationship between the magnitude of BoS and the meta-analysis level characteristics was explored through an interactive tool developed in Chapter 5 (available from: `https://mhattle.shinyapps.io/BoSstatistic/`). The interactive tool was developed to allow researchers to visually explore the impacts the meta-analysis characteristics have upon the magnitude of BoS, with particular interest in the percentage of missing data for the outcome of interest. Through the exploration of the interactive graph in Chapter 5, it was proposed that a bounded relationship might exist between the BoS statistic and the percentage of missing data for the outcome of interest (Chapter 6). Using mathematical assumptions in Chapter 6, the BoS statistic was mathematically proved to be bounded by the percentage of missing data for the outcome of interest for both fixed-effect and random-effects bivariate meta-analyses with complete and missing data for the alternative outcome.

Chapter 7 investigated whether the benefits of multivariate meta-analysis could be extended to a novel application; the meta-analysis of a continuous outcome from

RCTs. In RCTs, continuous outcomes are measured at baseline and follow-up, and can be modelled using three different models: Final score, Change score and ANCOVA models (Vickers & Altman 2001). The ANCOVA model is the preferred model but is often not reported (Frison & Pocock 1997, Vickers & Altman 2001, Van Breukelen 2006, Zhang et al. 2014). Therefore for a meta-analysis, researchers are restricted to the available treatment effect estimates from the models reported in the trial publications, if trial IPD is unavailable. The current methods for meta-analysis of continuous outcomes analysed using ANCOVA, final score or change score models are univariate meta-analyses; either three separate univariate meta-analyses for treatment effect estimates from each model or an alternative univariate meta-analysis method advocated by Cochrane (Higgins & Green 2008). The Cochrane method is a univariate meta-analysis that analyses one treatment effect estimate from any model per trial with priority to treatment effect estimates from the ANCOVA model. In Chapter 7, a bivariate meta-analysis approach that jointly analyses treatment effect estimates from the final score model and treatment effect estimates from the ANCOVA model was proposed. Each meta-analysis method (separate univariate method, Cochrane method and bivariate method) was applied to an IPD dataset of 10 RCTs for the treatment of Hypertension, where the continuous outcome was systolic blood pressure. For each of the trials, there was baseline balance across treatment groups. Missing data scenarios of 50% missing ANCOVA treatment effect estimates were generated to replicate when ANCOVA treatment effect estimates are unavailable. All possible permutations of missing data scenarios for 50% missing were analysed and summarised in Chapter 7. The bivariate and Cochrane univariate meta-analysis had similar results to the results from the complete data case, although the mean standard errors were smaller for the bivariate meta-analyses than the Cochrane univariate meta-analyses. To investigate the comparison of methods further and to include different settings, like baseline imbalance, a simulation study was undertaken in Chapter 8. The simulation study analysed each meta-analysis' performance measures in different settings: including baseline balance, baseline imbalance, changes in between-study variance, number of

studies and number of participants. The simulation study found that the Cochrane meta-analysis method can be recommended if non-ANCOVA effect estimates are only from trials with baseline balance across treatment arms. Furthermore, if there are a mixture of balance and imbalanced trials, then the Cochrane method can still be used as long as the non-ANCOVA effect estimates are only from balanced trials. The performance of the bivariate meta-analysis method is similar to the Cochrane method, however it is more complex to undertake. When all the trials have baseline imbalance, the separate univariate meta-analysis of the available ANCOVA effect estimates performs the best.

9.2 Key findings

This section aims to reiterate the key findings, which are summarised in Figure 9.1. Then some of the key findings are discussed in further detail.

Figure 9.1: Key findings

Key findings:

- Although there is published research that suggests that results from multivariate meta-analysis do not differ from the univariate meta-analyses, this is not true for all meta-analyses (Sohn 2000, Simel & Bossuyt 2009, Trikalinos et al. 2013, 2014). In some studies, statistical and clinical differences occur when comparing multivariate and univariate results.
- The Borrowing of strength (BoS) statistic can identify when multivariate meta-analysis results may differ from univariate and therefore when multivariate may be beneficial over a univariate meta-analysis (Jackson et al. 2017, Copas et al. 2018).
- The BoS statistic can be predicted from meta-analysis level characteristics prior to any analysis or decisions regarding analysis.

- The key characteristics of a meta-analysis that are important predictors for the magnitude of BoS are the percentage of missing data for the outcome of interest, the number of studies and the maximum within-study correlation.
- The magnitude of BoS can be explored through an interactive tool developed for the visualisation of BoS in relation to meta-analysis characteristics.
- The BoS statistic, under certain conditions, is bounded by the percentage of missing data for the outcome of interest.
- The application of the bivariate meta-analysis approach to continuous outcomes from RCTs modelled using final score and ANCOVA models, in general, performs as well as a Cochrane advocated univariate meta-analysis approach, where there is baseline balance across treatment groups in the trials available for meta-analysis (Higgins & Green 2008).
- When there is baseline imbalance across treatment groups in included trials, the Cochrane meta-analysis and the bivariate meta-analysis had poor performance measures when compared to the separate univariate meta-analysis for the available treatment effect estimates from the ANCOVA model.

9.2.1 BoS statistic identified differences between the univariate and multivariate meta-analysis results

An aim of this thesis was to explore whether, through the exploration and further understanding of borrowing strength, the application of multivariate meta-analysis could be improved. A key focus was aiding researchers to know when the application of multivariate meta-analysis would be beneficial over a univariate meta-analysis. This is particularly important, since the multivariate meta-analysis is not widely utilised, even though the benefits of multivariate meta-analysis has been widely documented

(Riley et al. 2007*a*,*b*, Hamza et al. 2009, Jones et al. 2009, Jackson et al. 2011, Kirkham et al. 2012, Riley et al. 2015, Frosi et al. 2015). For example, the Cochrane collaboration to date have not published multivariate meta-analyses of multiple outcomes and multivariate meta-analysis for multiple outcomes is not a method detailed in their handbook for systematic reviews of interventions (Higgins & Green 2008).

The differences between the univariate and multivariate meta-analysis have been reviewed in multiple published works (Sohn 2000, Simel & Bossuyt 2009, Trikalinos et al. 2013, 2014, Price et al. 2019). Sohn (2000), Simel & Bossuyt (2009) and Trikalinos et al. (2013, 2014) concluded that the results from the multivariate and the univariate meta-analyses are often similar, and therefore recommendations have been made that there is limited practical importance in the decision between the meta-analysis methods. However, through the reanalysis of the data from the Trikalinos et al. (2013, 2014) review in Chapter 3, there were results from meta-analyses with potentially important clinical differences and differences in statistical significance. Therefore, this thesis recommends that the multivariate meta-analysis method should not be disregarded in all situations, rather there is a need to determine when differences may occur and when the multivariate approach is most beneficial.

For the recommendation that the multivariate meta-analysis should not be disregarded and should be utilised in beneficial situations, there needs to be a practical way for researchers to identify when the multivariate meta-analysis should be used. The BoS statistic was developed as a measure to quantity the gain in information from utilising a multivariate meta-analysis compared to a univariate meta-analysis (Jackson et al. 2017). It was recommended by Jackson et al. (2017) that the BoS statistic should routinely accompany multivariate meta-analysis results. Through the investigation of differences between univariate and multivariate results in the Trikalinos et al. (2014) review and the corresponding magnitudes of BoS, the BoS statistic was identified as sensitive to differences in the precision between the univariate

and multivariate meta-analysis results. Through the magnitude of the BoS statistic, it is hoped that researchers may be able to determine when multivariate meta-analyses is most beneficial.

9.2.2 Meta-analysis characteristics can inform the magnitude of BoS that could be expected/predicted

The BoS statistic is easily calculated within meta-analysis computer packages or from the multivariate meta-analysis results (White 2011, Jackson et al. 2017, Copas et al. 2018). This is insufficient if the recommendation is to use the BoS statistic to inform the decision on which analysis to utilise. Therefore, the BoS statistic was further analysed to increase the understanding of what influences the magnitude of BoS.

In Chapters 3 and 4, meta-analysis level characteristics from the 43 meta-analyses in the Trikalinos et al. (2013, 2014) review were compared to the magnitude of the corresponding BoS statistics. In Chapter 4, prediction models were developed from meta-analysis level characteristics for the prediction of the magnitude of BoS. Separate prediction models were developed for fixed-effect and random-effects meta-analyses, since the BoS statistics were observed to differ between the fixed-effect and the random-effects meta-analyses (Chapter 3). The prediction models were applied to examples of meta-analyses, for which the true BoS statistics were known. Although the predictions from the prediction models were not exact, in the examples the prediction models were able to predict large BoS values for true large BoS values. Therefore, in Chapter 4, it was concluded that it is possible to broadly predict the BoS statistic from meta-analysis level characteristics. As a result, the magnitude of the predicted BoS statistic could be used to inform the decision as to whether there may be differences between the univariate and multivariate results.

The impact of individual meta-analysis level characteristics were explored further through an interactive tool (Chapter 5). Through the visual exploration of

the tool, the tool implied that specific relationships for the BoS statistic and some meta-analysis characteristics may exist. There was prior knowledge from a paper by Jackson et al. (2017) that discussed the relationship between the within-study correlation and the magnitude of BoS. Jackson et al. (2017) discussed that for studies with complete data the BoS statistic is bounded by the magnitude of the within-study correlation. However, in addition to the within-study correlation relationship, the interactive tool indicated a relationship between BoS and the percentage of missing data for the outcome of interest, which was also an important predictor for the magnitude of BoS in the prediction models (Chapter 4). The bounded relationship between the magnitude of BoS and the percentage of missing data for the outcome of interest was proved in Chapter 6. This relationship is important for researchers to quickly and easily estimate the BoS that they might expect in their meta-analysis. This relationship and the prediction models could help researchers determine whether the multivariate method is beneficial in their setting. That is, they could predict BoS from either the percentage of studies missing the outcome of interest, or the prediction models developed. If BoS is predicted to be 'large', then they may consider the multivariate approach.

9.2.3 Bivariate meta-analysis does not perform as well as univariate methods for continous outcomes analysed using final score and ANCOVA models

For RCTs with continuous outcomes, there are three analytical models researchers can use to model the data to estimate the treatment effect. The recommended method for the estimation of treatment effect estimates is the ANCOVA model, since it has greater statistical power to identify a treatment effect (Frison & Pocock 1997, Vickers & Altman 2001, Van Breukelen 2006, Zhang et al. 2014). However, the ANCOVA model is rarely utilised since it is more complex than the other two analytical models. A researcher doing a meta-analysis of RCTs with a continuous outcome of interest is limited to the treatment effect estimates reported in the trial's publication. Consequently, researchers

may have treatment effect estimates estimated using different models. A bivariate meta-analysis method was proposed for this situation and compared to two different univariate methods, Cochrane method and separate univariate method (Chapter 7).

The Cochrane univariate meta-analysis and bivariate meta-analysis' performance measures were similar when there was baseline balance

The Cochrane collaboration recommend a method which analyses one treatment effect estimate from each trial irrespective of model (with priority allocated to treatment effect estimates from the ANCOVA model) in a univariate meta-analysis (Higgins & Green 2008). A simulation study in Chapter 8 compared the Cochrane univariate meta-analysis method to the bivariate meta-analysis method for scenarios with baseline balance. The performance measures (percentage bias, mean squared error, coverage of 95% confidence interval and power) for the Cochrane method were similar to those for the bivariate method. For the meta-analysis of RCTs with continuous outcomes with baseline balance, it is recommended that the Cochrane univariate meta-analysis is used to analyse treatment effect estimates.

The performance of univariate meta-analysis methods when there was baseline imbalance

For the meta-analysis of RCTs with continuous outcomes with baseline imbalance, the separate univariate meta-analysis of treatment effect estimates from the ANCOVA model had better performance measures than the Cochrane method or the bivariate method. Therefore, when there is baseline imbalance, it is better to disregard any treatment effect estimates from the final score models than include these in a meta-analysis, due to the bias in the effect estimates. Alternatively, if there is a mixture of trials, some with baseline balance and some with baseline imbalance, then the Cochrane meta-analysis method could be used as long as the non-ANCOVA effect estimates are only from trials with baseline balance.

9.3 Borrowing of strength: is it more than a summary measure?

9.3.1 Is it realistic to expect researchers to rely on BoS for the choice of analysis method?

Prior to this thesis, the borrowing of strength was discussed in the literature as a benefit from the application of multivariate meta-analysis and as a measure for the quantity of information gained, a summary statistic provided after a multivariate meta-analysis (Riley et al. 2007*a*, Jackson et al. 2017, Copas et al. 2018). However, this thesis proposes that the BoS statistic has more potential uses than just a post-analysis measure. In this thesis, the magnitude of BoS has been shown to identify differences between results from the univariate meta-analysis and multivariate meta-analysis results. Although the question of whether it is reasonable to expect researchers to rely on BoS for decision making has not been answered.

Although the BoS statistic identified differences between the univariate and multivariate results, this is not the full story since the BoS did not identify differences in the summary effect estimate when the width of the confidence intervals were equal for the univariate and multivariate results. Additionally, it was observed that a large BoS statistic does not necessarily equate to differences in univariate and multivariate results. For example, from the Trikalinos et al. (2013, 2014) data in meta-analysis 27 for outcome two, the univariate and multivariate REML meta-analysis results were -0.94 (95% C.I.: -1.10 to -0.78) and -0.90 (95% C.I.: -1.03 to -0.78), respectively and the BoS was 29.9% (see Section 3.3.2 in Chapter 3). Conversely, there are benefits of multivariate meta-analysis that are not necessarily identified through the magnitude of BoS or differences in the results between univariate and multivariate results, for example the relationship between outcomes can be described (Jackson et al. 2011). Finally, the magnitude of BoS does not provide any indication whether

a multivariate meta-analysis is feasible. For a multivariate meta-analysis, it needs to be considered whether the within-study correlations are available either through the availability of IPD or from published material. If the within-study correlations are not available, the methods for the estimation of the overall correlations provides no information regarding the within-study and between-study correlation and, as a result, the heterogeneity also can not be examined (Riley et al. 2008). Additionally, the within-study correlations are rarely available from publications (Riley et al. 2007*a*, 2008, Riley 2009). Therefore, the collection of IPD is the preferred method for the collection of within-study correlations. However, the collection of IPD is not without its own caveats, which include time requirements, resources, statistical expertise, and selection bias may still occur (Debray et al. 2015*a*). Additionally, when IPD is available for an IPD meta-analysis, a decision needs to be made whether a one-stage or two-stage meta-analysis approach is used, this is particularly important since the two approaches can produce different effect estimates (Debray et al. 2013, Burke et al. 2017).

Therefore, it would be alarming to recommend that researchers rely solely on the expected or predicted magnitude of BoS for decision making for the choice of analysis method. On the other hand, for the planning of meta-analysis studies where the collection of IPD or estimation of within-study correlations can be included, the expected BoS will be useful for researchers to inform the choice of methods. After all, this thesis has shown the BoS is useful for identifying differences between the univariate and multivariate results.

Can the BoS inform the decision on whether to collect IPD?

A logical extension to whether the BoS statistic should be relied on for the choice of meta-analysis method, is to ask whether the magnitude of BoS should be used to determine whether researchers put time and resources into collecting IPD. The concern with proposing the BoS statistic to inform whether IPD should be collected is this

assumes that the benefits of the multivariate meta-analysis are the only advantages to obtaining IPD. However, the advantages of IPD reach beyond availability of within-study correlations required for multivariate meta-analysis. Such advantages include identification of treatment effect modifiers, reducing publication bias, standardisation of statistical analysis, and model assumptions can be checked in each study. (Riley et al. 2010, Debray et al. 2015*a*). Therefore, the BoS should not be solely used as a decision making tool for the collection of IPD. Although the magnitude of BoS may be of interest to aid this decision, the other advantages of IPD should also be considered along with the time and resource constraints.

9.3.2 How easily implemented are the methods for predicting BoS?

Prediction models and prediction interactive graphs

The prediction models in Chapter 4 are easily implemented with regards to mathematical calculation, since the prediction model has only linear terms and no interactions. Alternatively, the prediction model could be built into a calculator, for example, on a website, that would only require a user to input the requested values. For the use of the prediction models, a user would need to decide whether their analysis would use a fixed-effect or random-effects assumption, to inform which prediction model is applicable to their setting. This decision should not be seen as a problem, since the decision of fixed-effect or random-effects assumption should be made prior to analysis. The predictors in each of the prediction models related to the studies and outcomes, include the total number of studies in the meta-analysis, percentage of missing data and number of studies with a particular outcome. These predictors are easily available from IPD and aggregate data (AD). However, the within-study correlations are not necessarily known without IPD and definitely won't be known if it is planned for the overall correlation to be estimated methods (Riley et al. 2008, Riley 2009). If the expected within-study correlation is known, this could be used with caution.

Interactive tool from the BoS equation

The interactive tool is easily accessed through a website where it is embedded. However, the interactive tool requires larger quantities of information than the prediction models. Additionally, some of this information required is estimated during the multivariate meta-analysis, for example the between-study correlation and between-study variance. Furthermore, the within-study correlations are also required for the interactive graph; in the same way as this was problematic for the prediction model, this is information that is not always available if IPD is not available.

Additionally for the set up of the interactive tool, mathematical assumptions were made about the studies in the meta-analysis. These assumptions including equal within-study correlations and within-study variances across all studies are simplistic and are unlikely to occur in practice. Therefore, caution needs to be taken when using the interactive tool as the accuracy of magnitude of BoS is uncertain if violations of the mathematical assumptions occur.

Through the implementation of sliders on the interactive graph, it is very easy for a user to vary their expected characteristics of meta-analysis and visually explore the impacts these changes may have upon the magnitude of BoS. This makes the interactive graph very usable for researchers. This is particularly advantageous if there is uncertainty around the expected value of a meta-analysis characteristic. Therefore, the interactive graph is a useful tool for researchers who are considering a meta-analysis of data which contains multiple outcomes.

Bounded relationship for BoS

The expected value of BoS is bounded by the known percentage of missing data for the outcome of interest, under certain conditions. Although it is useful to know that such a relationship exists between the BoS and the percentage of missing data for the outcome of interest, the mathematical assumptions made are too simplistic. When

the mathematical assumptions are violated, the BoS is not always bounded by the percentage of missing data for the outcome of interest. On the other hand, an advantage of the bounded relationship between the BoS statistic and the percentage of missing data for the outcome of interest is it does not require lots of information that may be unknown. It is easily understood and can be quickly calculated by researchers without the need for calculations and exploration of graphs. Therefore, it provides a useful rule of thumb to be used (even if cautiously) by researchers who would like to know how much BoS to expect.

9.3.3 Which method for predicting BoS should be used?

The progression of research into the prediction of BoS raises the questions; how useful are the prediction models if, from the bounded relationship, we know how much BoS to expect? And which method should be used to inform the magnitude of BoS? The bounded relationship is simplistic due to the mathematical assumptions made, however the information required is easily accessible from aggregate data (AD). This does not negate the need for the prediction models, since the mathematical assumptions in the bounded relationship can be violated in practice, which may result in the BoS exceeding the percentage of missing data for the outcome of interest. Therefore, if all the required predictors (for the prediction model) are known or can be approximated, it is recommended that the prediction model is used to predict the magnitude of BoS. If some of the predictors are unknown, the bounded relationship provides a reasonable rule of thumb. A researcher may also wish to use the interactive graph to observe how the unknown predictors may influence the magnitude of BoS.

9.3.4 What magnitude of BoS is 'large'?

A question many researchers may ask is "what BoS is 'large'?". In this thesis, providing a categorisation of BoS into large, medium and small cut-off has been avoided. There are concerns with providing such a scale for recommendations for a categorisation of BoS, since a BoS statistic that is considered large in one setting may not be considered

large in another. Therefore, a cut-off between what is a small and large BoS will not be provided.

9.4 Recommendations

The recommendations from this thesis have been discussed in Sections 9.1 and 9.3, and are summarised in Figure 9.2.

Figure 9.2: Key recommendations

Recommendations:

- Multivariate meta-analysis is a useful tool to borrow strength across correlated outcomes to improve precision and the quantity of evidence toward summary results.
- As a multivariate approach is more complex than a univariate approach and may require the estimation of within-study correlations, it is not always required.
- When there is complete data, generally a multivariate approach is not needed.
- The magnitude of BoS might be considered when deciding whether a multivariate or univariate meta-analysis approach should be used for analysis.
- Where possible, the BoS should be predicted before analysis using the developed prediction models or the interactive graphical tools for the prediction of BoS.
- If some of the predictors for the prediction model are unavailable the boundedness of BoS (by the proportion of studies missing the outcome of interest) is an adequate rule of thumb to use and the interactive graph can be used

to identify how the characteristics of meta-analysis might affect the magnitude of BoS.

- When there is baseline balance the Cochrane recommended univariate meta-analysis is recommended for RCTs with continuous outcomes modelled using final score and/or ANCOVA models.
- When there is baseline imbalance across treatment group, a separate univariate meta-analysis for ANCOVA treatment effect estimates is recommended.

9.5 Limitations and further research

In this section, the limitations of this thesis and areas for potential future research are discussed.

9.5.1 Consideration of other distributions

In this thesis, to model the multivariate meta-analysis methods the normal distribution for the study effects was assumed. However it is known that the normality assumption is difficult to verify for small numbers of studies in a meta-analysis (Jackson et al. 2011). Although the multivariate normal meta-analysis is the most commonly used multivariate meta-analysis method, there are other distributions that could be assumed particularly for settings where it is inappropriate to assume the effect sizes are normally distributed. Copulas are a family of multivariate distributions that have a uniform[0,1] marginal distribution (Genest & Mackay 1986, Nelsen 2007, Danaher & Smith 2011, Nikoloulopoulos 2015, Dimier & Todd 2017). Copulas are starting to be applied to multivariate meta-analysis and therefore it is important to investigate whether the findings from this thesis hold for multivariate meta-analyses modelled using Copulas (Nelsen 2007, Nikoloulopoulos 2015, Dimier & Todd 2017).

9.5.2 Consideration of Bayesian methods

This thesis was undertaken using frequentist methods, however, a frequentist approach is not the only approach to multivariate meta-analysis. A limitation of this thesis is that Bayesian analysis methods were not considered for the univariate or multivariate meta-analyses. In recent years there has been research in extending methods of Bayesian univariate meta-analysis methods to Bayesian multivariate meta-analysis methods (Wei & Higgins 2013*a*, Nam et al. 2003, Bujkiewicz et al. 2013). This thesis did not assess the performance of Bayesian multivariate meta-analysis for the synthesis of effect estimates from continuous outcomes in RCTs. However, a Bayesian approach might be beneficial in these situations. The advantage of a Bayesian meta-analysis over a frequentist approach is external evidence can be included in the analysis (Bujkiewicz et al. 2013, Mavridis & Salanti 2013). Information to inform the choice of priors can arise from studies external to the meta-analysis (e.g. observational studies, clinical trials, expert's (clinicians) opinions). This is particularly important in the age of evidence-based medicine when the aim is to use all available relevant evidence to make informed decisions and Bayesian methods can provide the opportunities to include further information (Ashby & Smith 2000). Therefore, an important research question for future research to answer is whether a Bayesian multivariate meta-analysis is beneficial for the synthesis of summary treatment effect estimates for randomised control trials with continuous outcomes analyses using a mixture of ANCOVA, final score and change score models.

9.5.3 Extending to different settings

The settings in this thesis were all meta-analyses of RCTs (Chapters 2, 3, 4, 7 and 8). The HTA report reanalysed in Chapter 2 was a meta-analysis of RCTs that investigated the effect of diet and exercise interventions during pregnancy for managing weight gain on maternal and foetal outcomes. The Trikalinos et al. (2014) review data contained 43 Cochrane reviews that contained at least seven RCTs for each outcome, that were then analysed in a multivariate meta-analysis. The setting for

the new novel application of multivariate meta-analysis were RCTs with continuous outcomes measured at baseline and follow-up.

Meta-analysis is not unique to settings with RCTs and the recommendations in this thesis need to be researched for other settings. For future work, research using non-randomised settings including cohort studies, case-control studies and cross-sectional studies should be conducted to extend the recommendations from this thesis to these settings.

Furthermore, the magnitude of BoS compared to multivariate and univariate results was assessed in meta-analyses of binary outcomes (Chapter 3). The prediction models in Chapter 4 were developed using this data. However, there are other data types which were not considered in this thesis, including time-to-event and continuous, for which the multivariate meta-analysis methods might be beneficial and currently are not used, particularly since they are not advocated, for example by Cochrane (Higgins & Green 2008).

In the comparison of methods study (Chapter 7), each permutation had treatment effect estimates from the ANCOVA model for 50% of the trials and were missing for the remaining 50%. In practice, this is unlikely to always be true. It is important for different proportions of treatment effect estimates to be investigated to determine whether the recommendation made should be made for all missing data scenarios (greater than 60% missing). Therefore, it is a priority for future research to assess the effects different proportions of missing treatment effect estimates from the ANCOVA model have on the recommendations made.

The three different baseline imbalances scenarios were investigated for simulations of 10 trials with 1000 participants, between study variance of 2.25 and 50% missing data for the ANCOVA treatment effect estimate. For future research, the

meta-analysis characteristics and proportions of missing data should be varied to explore the effect these characteristics may have on recommendations for settings with baseline imbalance.

9.5.4 Extending bivariate meta-analysis

In this thesis, the majority of applications of multivariate meta-analysis were the specific case of multivariate for two outcomes, the bivariate meta-analysis. The results and implications from the thesis may be extendable to meta-analyses with more than two outcomes, however this was not investigated. In particular, the prediction models were developed from meta-analysis level characteristics from bivariate settings. Therefore, it is uncertain how accurate predictions for meta-analysis settings with more than two outcomes may be. That is not to say that the prediction models should not be used in these settings, but researchers should be cautious with interpreting the predicted BoS. To conclusively make recommendations for multivariate meta-analyses, the studies in this thesis need to be extended to settings with more than two outcomes.

It is worth noting that in Chapters 7 and 8 that studied a new application of a bivariate meta-analysis, a trivariate meta-analysis (meta-analysis of three outcomes) was considered but not investigated further since the treatment effect estimates from the three models are too closely related and therefore estimation problems were likely.

9.5.5 Prediction model validation

There are limitations to the prediction models developed in Chapter 4. The prediction models were developed based only on 86 values of BoS from 43 bivariate meta-analyses with binary outcomes. The prediction models were not developed from different types of data, for example, continuous, time-to-event or diagnostic outcomes, and the prediction model was not externally validated.

It is important to externally validate a prediction model, in addition to internal

validation. External validation is used to investigate the performance of the model in a different population, such that the model is developed to be generalisable (Altman & Royston 2000, Altman et al. 2009, Collins et al. 2014). However, the prediction model developed in Chapter 4, was not externally validated as there was insufficient data considered appropriate for the external validation. Externally validating the models is a priority for future research, especially if the model is to be used.

9.6 Final conclusions

In conclusion, the multivariate meta-analysis is beneficial for many applications of meta-analysis, but it also has limitations compared to the univariate method. The application of multivariate meta-analysis should be encouraged, particularly in situations with large quantities of missing data and large within-study correlations. However, in some settings, multivariate meta-analysis may be redundant, particularly when univariate meta-analysis results are the same. With further developments in methods such as the borrowing of strength statistic, the multivariate meta-analysis may be seen in more meta-analysis reviews. However, until such time that multivariate meta-analysis is advocated, e.g. by the Cochrane collaboration, it is likely to remain an underutilised method for meta-analysis based on published results. However, with the growing use and collection of IPD, the opportunities for multivariate meta-analysis increase.

Appendices

Appendix A

A.1 Hypertension example fixed-effect forest plots

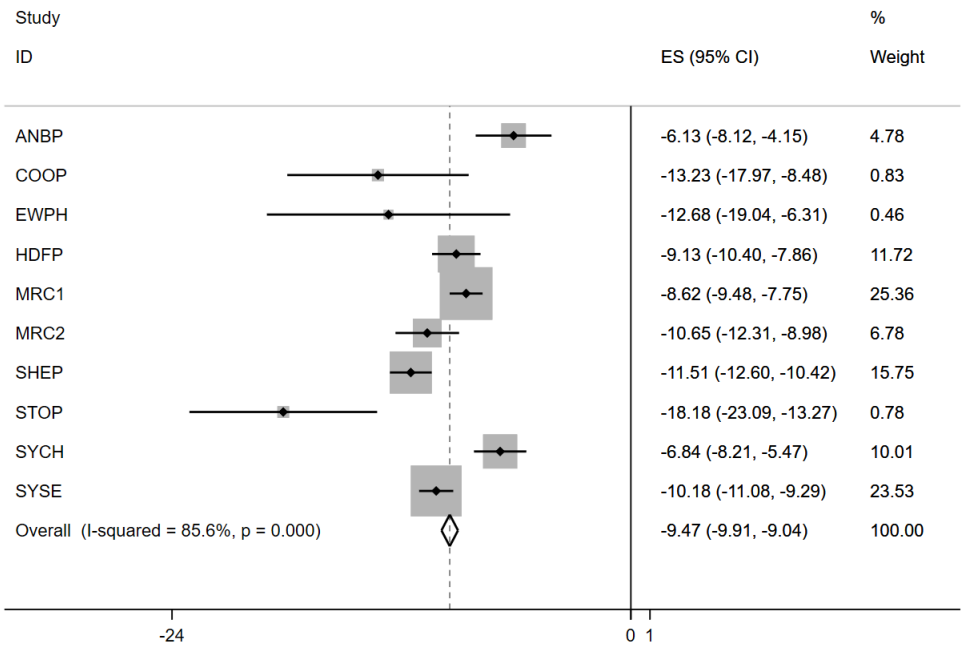


Figure A.1: Fixed-effect meta-analysis of treatment effect estimates from the change score model

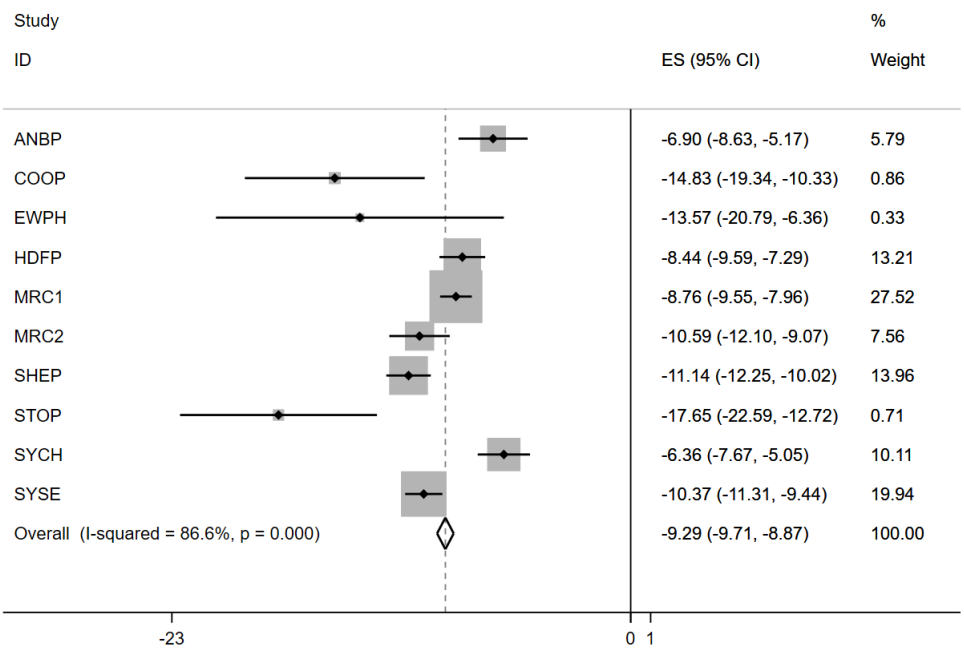


Figure A.2: Fixed-effect meta-analysis of treatment effect estimates from the final score models

A.2 Hypertension example random-effects forest plots

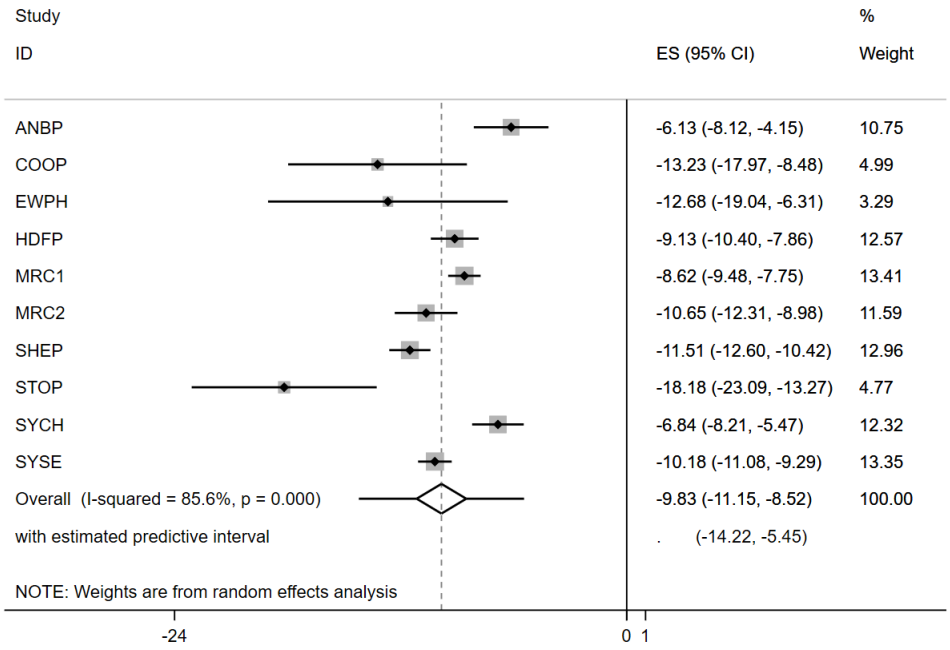


Figure A.3: Random-effects meta-analysis of treatment effect estimates from the change score models

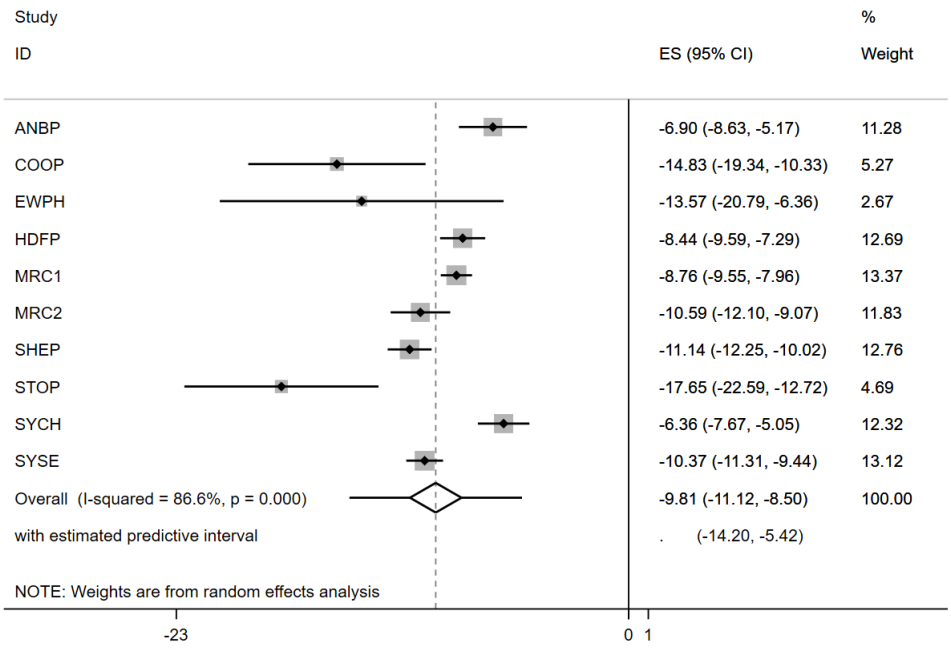


Figure A.4: Random-effects meta-analysis of treatment effect estimates from the final score models

A.3 Conference Presentations

International Society for Clinical Biostatisticians (ISCB) 2016

"When should we use multivariate meta-analysis? Predictors of Borrowing of Strength in 43 bivariate meta-analyses with Cochrane."

Young Statistician's Meeting (YSM) 2017

"Can we use predictors of Borrowing of Strength to flag when a multivariate meta-analysis is most needed?"

Royal Statistical Society (RSS) Conference 2018

"Multivariate meta-analysis of correlated outcomes: how much borrowing of strength do we expect?"

Appendix B

B.1 Sensitivity analysis

Table B.1: The results from the sensitivity analysis with different values for the imputed within-study correlations for the foetal outcomes in studies where it was not provided

Correlation	Odds Ratio	95% confidence interval	BoS
0	0.97	0.38 to 2.48	5.1
	1.06	0.93 to 1.20	3.4
	0.89	0.77 to 1.03	2.2
	0.99	0.81 to 1.20	3.6
0.55	1.48	0.46 to 4.73	48.2
	1.01	0.88 to 1.16	20.0
	0.84	0.70 to 1.01	13.9
	0.98	0.80 to 1.21	16.6
-0.39	0.81	0.31 to 2.11	5.6
	1.06	0.93 to 1.20	13.0
	0.87	0.72 to 1.05	7.4
	0.98	0.80 to 1.19	12.4

Table B.2: The results from the sensitivity analysis with different values for the imputed within-study correlations for the maternal outcomes

Correlation	Odds Ratio	95% confidence interval	BoS
0	0.85	0.67 to 1.08	17.1
	0.84	0.68 to 1.04	10.1
	0.94	0.76 to 1.15	2.6
	0.89	0.80 to 0.99	8.1
0.69	0.80	0.63 to 1.03	21.4
	0.81	0.64 to 1.01	13.4
	0.92	0.74 to 1.14	11.1
	0.88	0.79 to 0.98	16.2
-0.69	0.70	0.50 to 0.99	28.5
	0.74	0.56 to 0.98	19.4
	0.81	0.61 to 1.09	11.1
	0.88	0.78 to 0.98	19.6

Appendix C

C.1 Comparison of univariate and multivariate meta-analyses with corresponding BoS

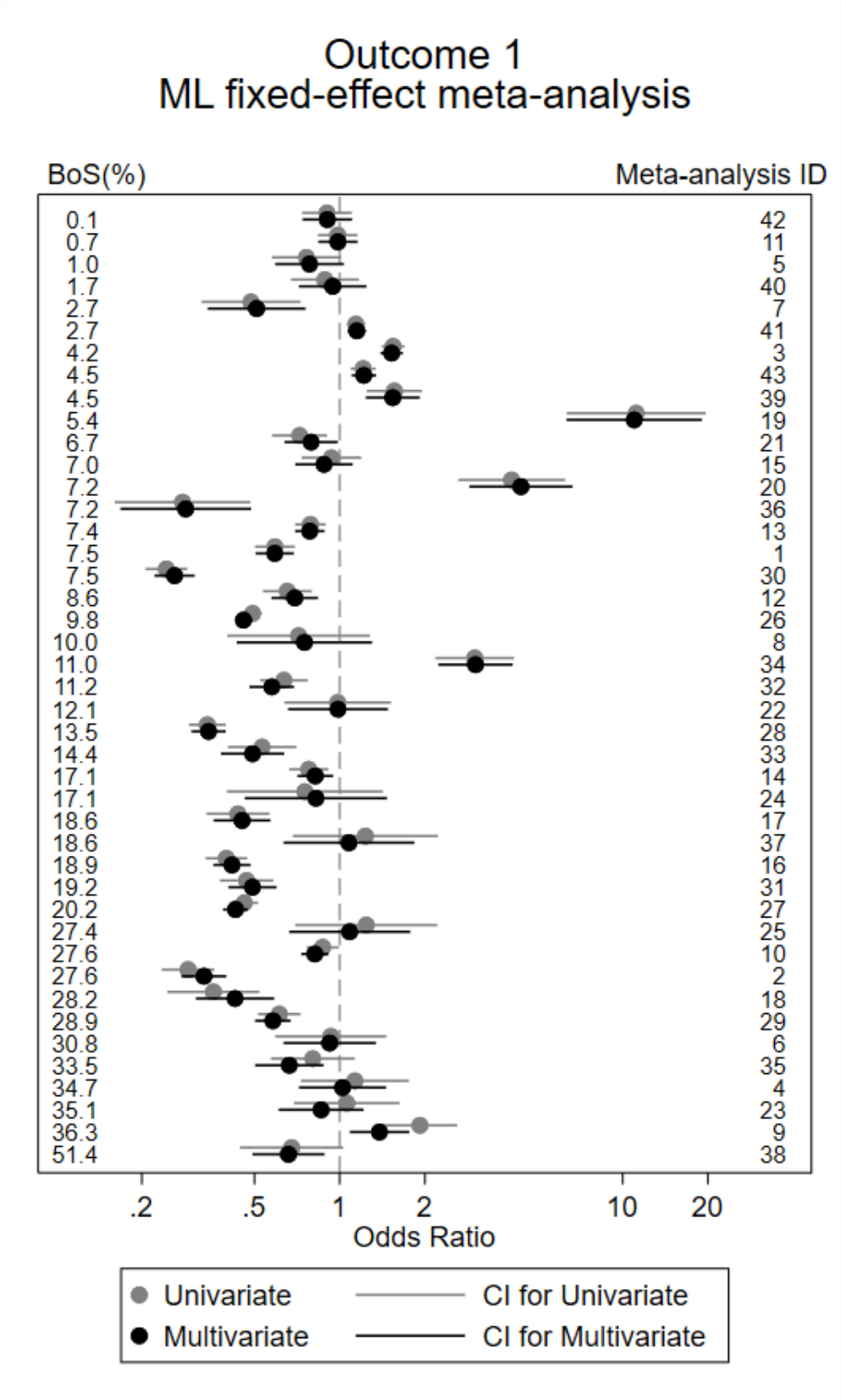


Figure C.1: Comparison of the univariate and multivariate meta-analysis results for outcome one from the fixed-effect meta-analysis ordered by the BoS statistic

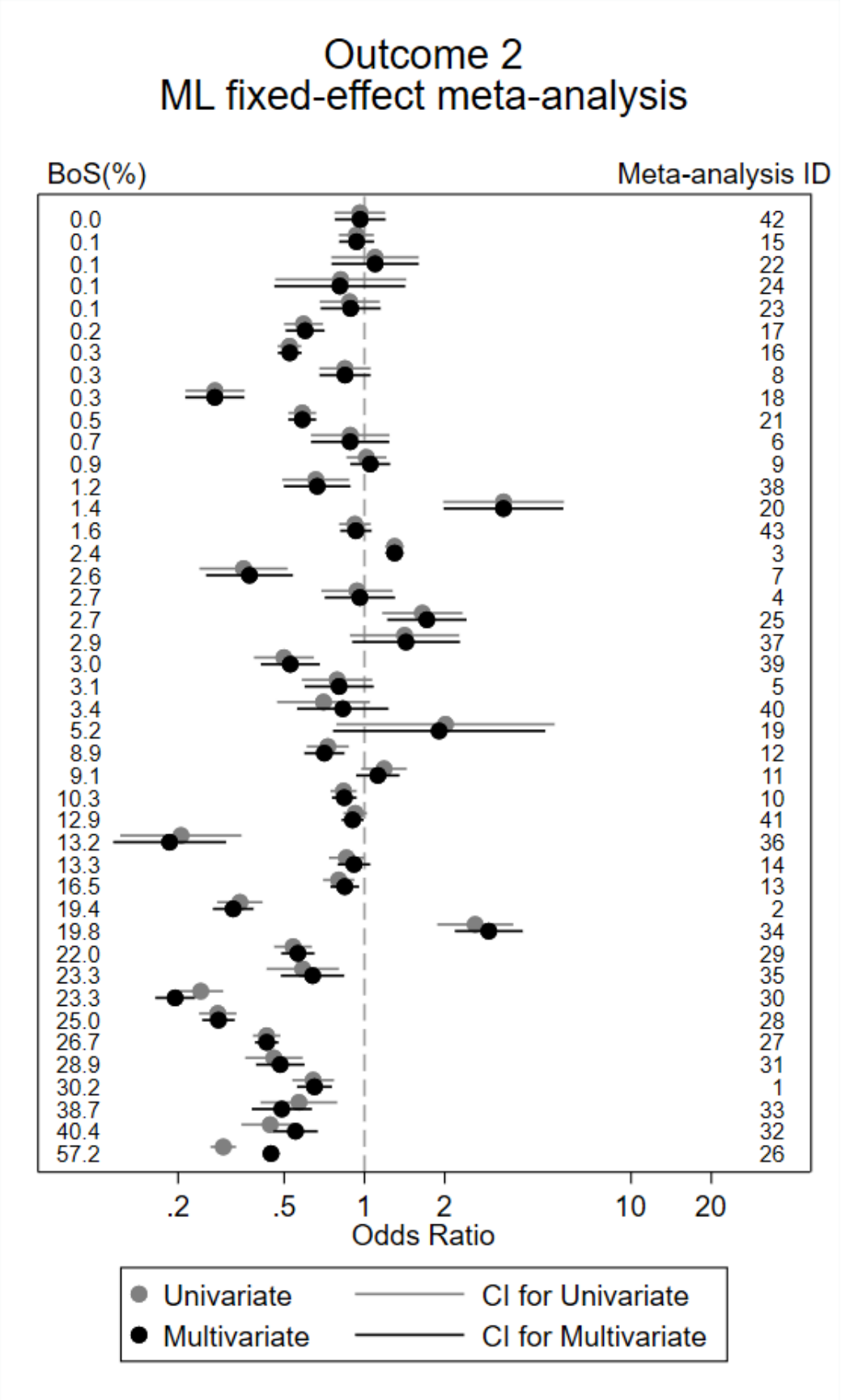


Figure C.2: Comparison of the univariate and multivariate meta-analysis results for outcome two from the fixed-effect meta-analysis ordered by the BoS statistic

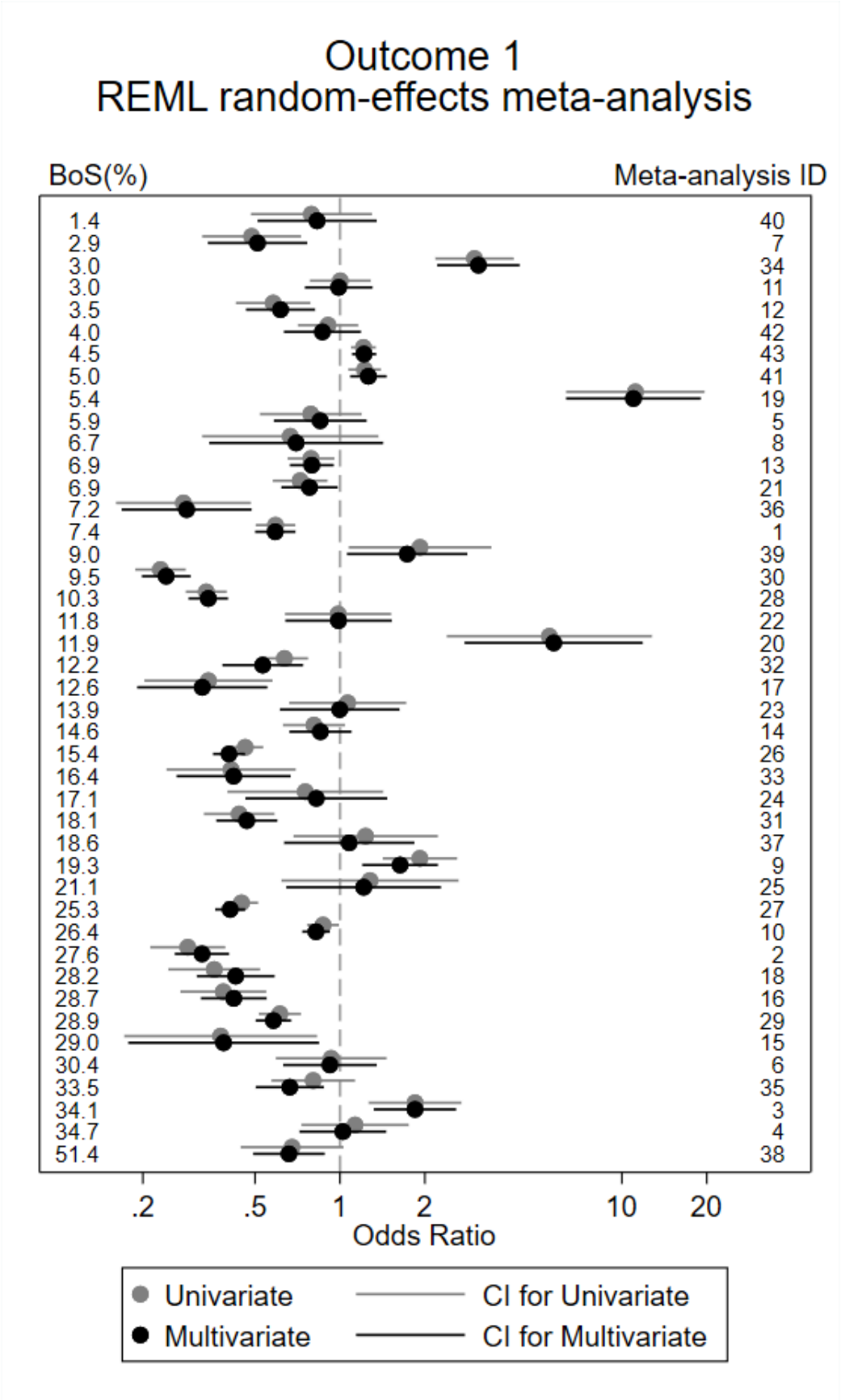


Figure C.3: Comparison of the univariate and multivariate meta-analysis results for outcome one from the REML random-effects meta-analysis ordered by the BoS statistic

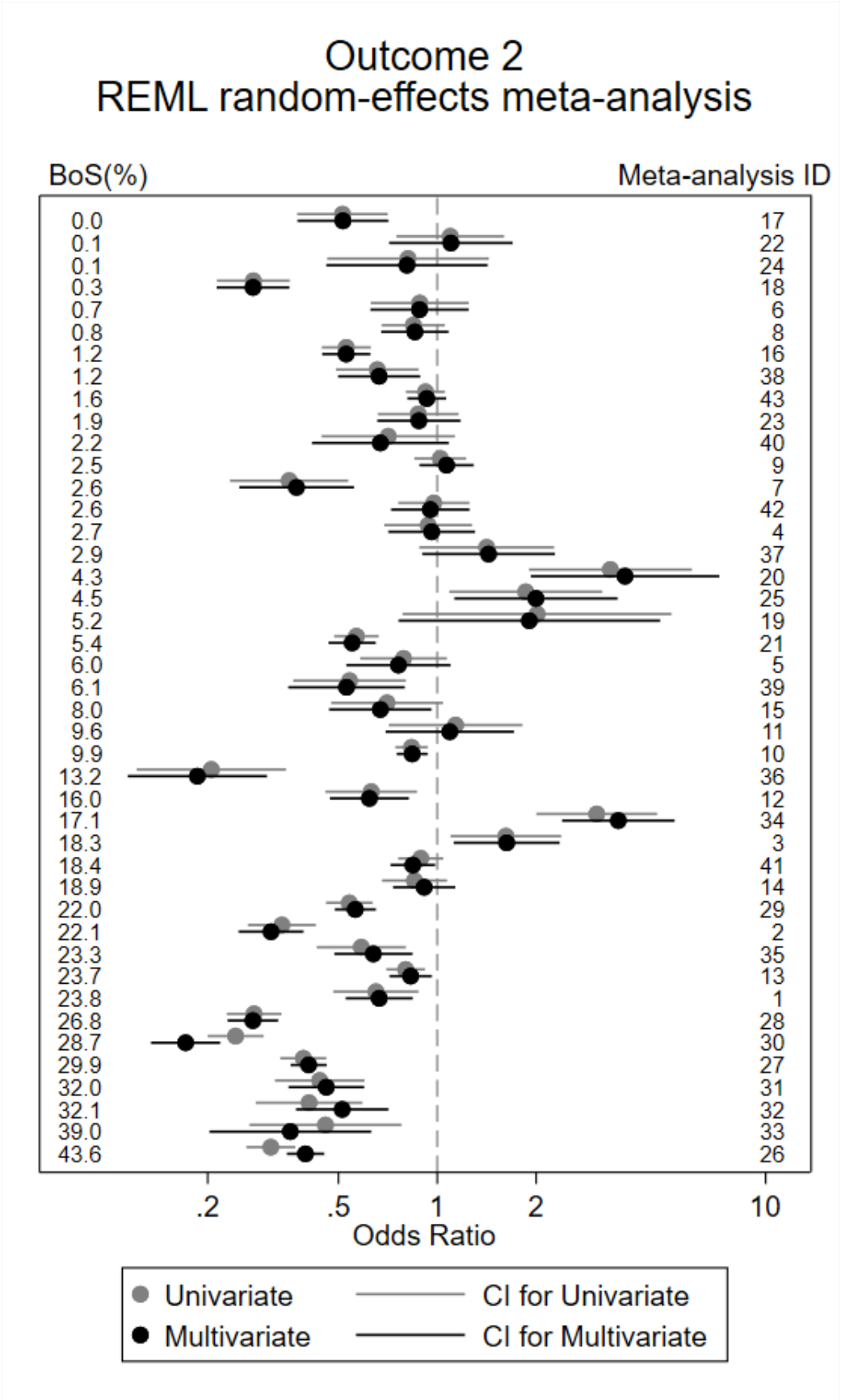


Figure C.4: Comparison of the univariate and multivariate meta-analysis results for outcome two from the REML random-effects meta-analysis ordered by the BoS statistic

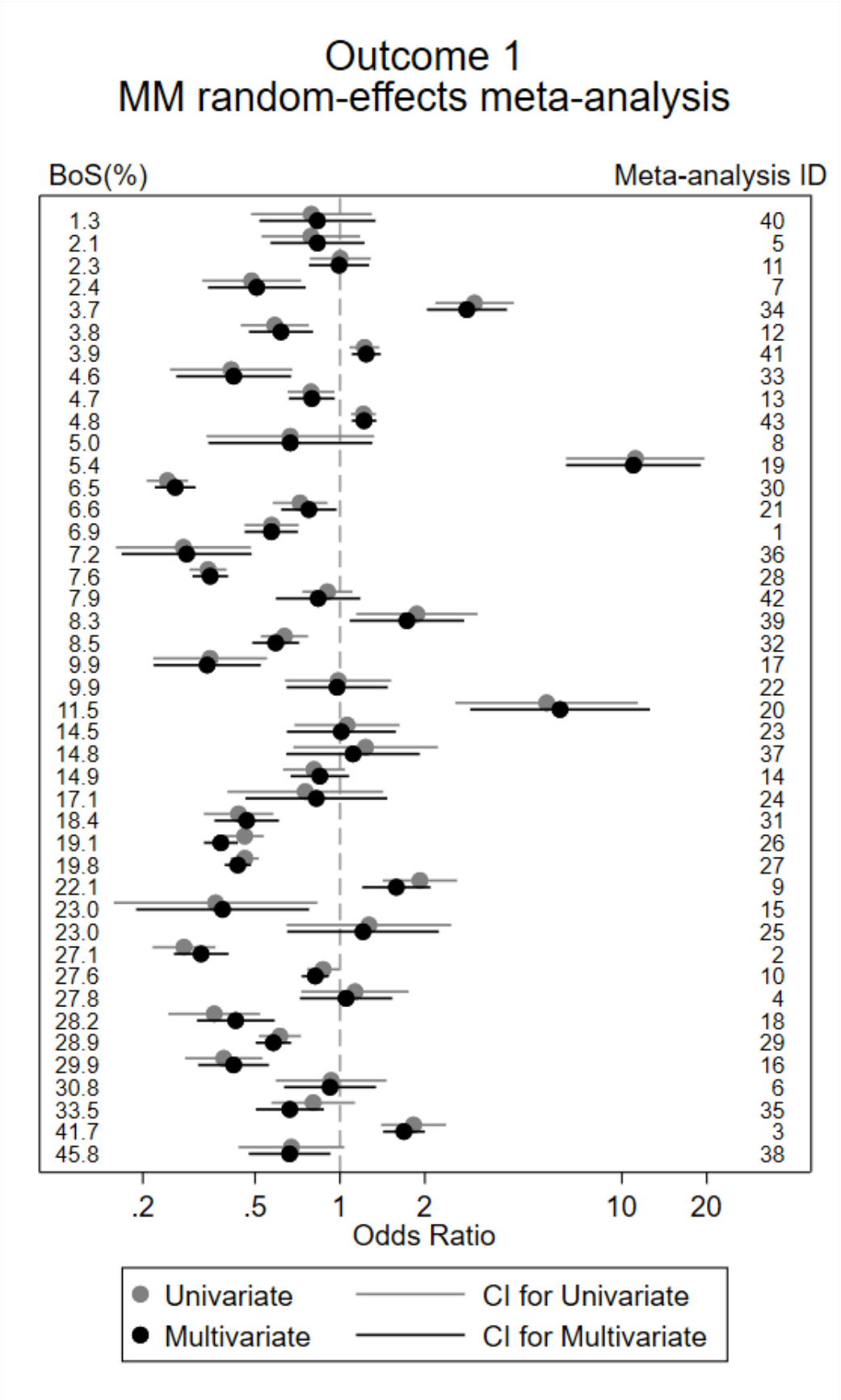


Figure C.5: Comparison of the univariate and multivariate meta-analysis results for outcome one from the MM random-effects meta-analysis ordered by the BoS statistic

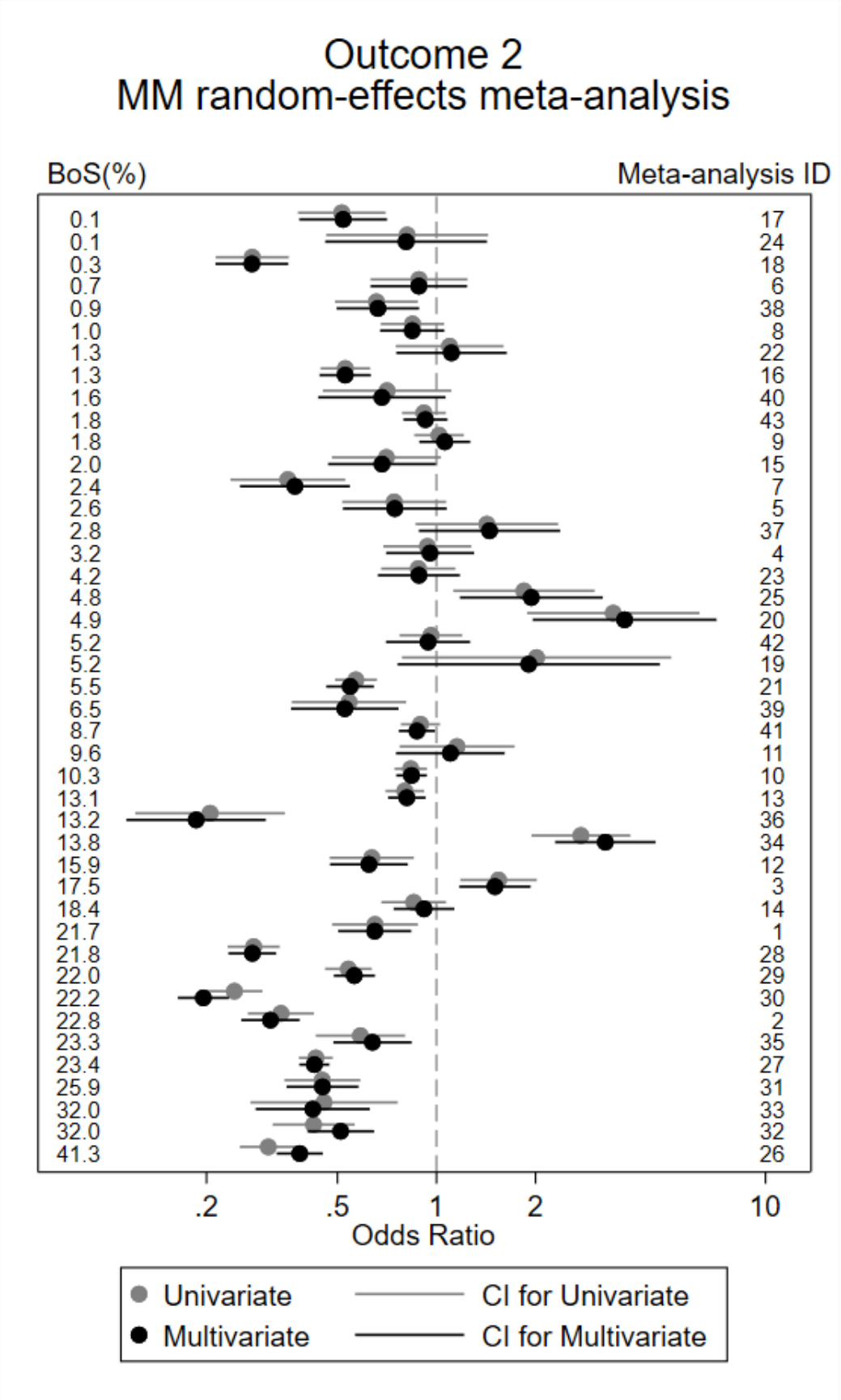


Figure C.6: Comparison of the univariate and multivariate meta-analysis results for outcome two from the MM random-effects meta-analysis ordered by the BoS statistic

C.2 Univariable regression

Table C.1: Table of results from the univariable mixed-effect multilevel models for the possible predictors of BoS statistic from the fixed-effect meta-analyses

Covariate	Intercept	Residual Variance, σ_e^2	Heterogeneity, σ_u^2	Regress R^2
Number of studies	8.486	142.402	3.11×10^{-22}	0.145
Percentage missing across both outcomes	6.700	142.758	1.55×10^{-21}	0.143
Percentage missing for outcome of interest	6.235	103.594	0.755	0.373
Number of studies with outcome of interest	9.726	155.464	4.80×10^{-22}	0.066
Number of studies with both outcomes	7.677	147.796	5.82×10^{-23}	0.113
Average within-study correlation	6.393	145.645	3.81×10^{-17}	0.125
Average absolute within-study correlation	-4.597	140.696	7.50×10^{-22}	0.155
Maximum within-study correlation	3.216	143.05	1.12×10^{-22}	0.141
Maximum absolute within-study correlation	-11.938	135.218	8.54×10^{-29}	0.188

Table C.2: Table of results from the univariable mixed-effect multilevel models for the possible predictors of BoS statistic from the REML meta-analyses

Covariate	Intercept	Residual Variance, σ_e^2	Heterogeneity, σ_u^2	Regress R^2
Number of studies	10.001	121.384	1.34×10^{-18}	0.133
Percentage missing across both outcomes	7.011	111.378	3.23×10^{-22}	0.204
Percentage missing for outcome of interest	6.851	66.662	5.389	0.485
Number of studies with outcome of interest	11.227	132.193	1.23×10^{-13}	0.055
Number of studies with both outcomes	9.745	128.1119	6.37×10^{-15}	0.085
Average within-study correlation	8.295	124.573	5.13×10^{-19}	0.110
Average absolute within-study correlation	-0.382	122.759	4.50×10^{-18}	0.123
Maximum within-study correlation	5.6295	122.908	9.91×10^{-18}	0.1217
Maximum absolute within-study correlation	-6.429	118.994	7.25×10^{-23}	0.150
Multivariate between-study variance	13.399	138.771	1.34×10^{-9}	0.008
Univariate between-study variance	13.881	139.780	6.69×10^{-15}	0.001
White's I^2	13.466	139.488	6.73×10^{-17}	0.003

Continued on next page

Table C.2: continued

Covariate	Intercept	Residual Variance, σ_e^2	Heterogeneity, σ_u^2	Regress R^2
Jackson's I^2	13.275	139.433	2.46×10^{-13}	0.004
Between-study correlation	12.314	136.257	7.04×10^{-15}	0.026
Absolute between-study correlation	5.588	138.981	7.83×10^{-18}	0.007

Table C.3: Table of results from the univariable mixed-effect multilevel models for the possible predictors of BoS statistic from the MM meta-analyses

Covariate	Intercept	Residual Variance, σ_e^2	Heterogeneity, σ_u^2	Regress R^2
Number of studies	9.035	108.814	1.86×10^{-18}	0.132
Percentage missing across both outcomes	6.281	100.398	1.13×10^{-21}	0.199
Percentage missing for outcome of interest	6.149	61.359	5.042	0.470
Number of studies with outcome of interest	10.118	118.038	3.58×10^{-22}	0.058
Number of studies with both outcomes	8.777	114.698	7.85×10^{-21}	0.085
Average within-study correlation	7.095	109.944	2.58×10^{-18}	0.123
Average absolute within-study correlation	-0.918	109.655	2.41×10^{-17}	0.125
Maximum within-study correlation	4.711	109.382	7.39×10^{-18}	0.127
Maximum absolute within-study correlation	-5.412	108.628	4.68×10^{-18}	0.133
Multivariate between-study variance	12.545	124.938	5.28×10^{-13}	0.003
Univariate between-study variance	12.653	125.065	6.04×10^{-11}	0.002
White's I^2	12.535	125.139	6.57×10^{-14}	0.001

Continued on next page

Table C.3: continued

Covariate	Intercept	Residual Variance, σ_e^2	Heterogeneity, σ_u^2	Regress R^2
Jackson's I^2	11.972	124.761	5.07×10^{-17}	0.004
Between-study correlation	12.718	119.629	8.54×10^{-12}	0.045
Absolute between-study correlation	13.955	125.011	7.87×10^{-16}	0.002

Table C.4: Table of results from the univariable mixed-effect multilevel models for the difference between the univariate and multivariate summary estimates divided by the standard error for each estimation method

Estimation Method	Intercept	Residual Variance, σ_e^2	Heterogeneity, σ_u^2
Fixed-effect	12.977	158.033	6.55×10^{-13}
REML	14.128	139.860	8.89×10^{-16}
MM	12.916	125.209	6.12×10^{-17}

Appendix D

D.1 Model assumptions checked

The assumptions made for the development of multivariable linear models were checked using the residuals in plots for each model. Figure D.1, shows the plots for each model of the fitted values against the studentised residuals. These plots are used to investigate the assumption of constant variance. The plots do not appear to indicate towards a strong violation of the assumption of constant variance. There do not appear to be any patterns that would suggest that there is a non-linear term missing.

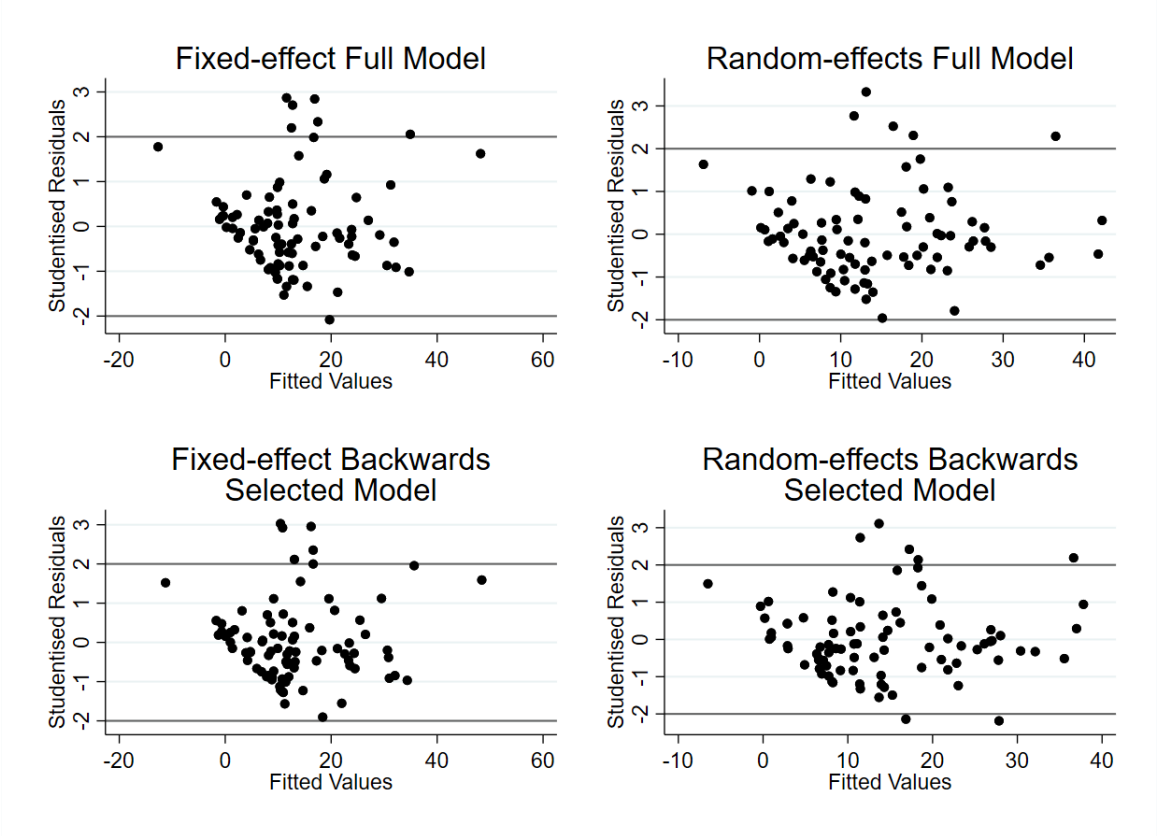


Figure D.1: Scatter plots of the studentised residuals against the fitted values from the models

The normal probability plots of the standardised residuals for each of the models are shown in Figure D.2. For a perfect normal distribution all of the plots should follow the normal line. In each plot there are deviations from this line in multiple places, but this was considered mild, as the plots approximately follow the line. Due to the BoS having a lower bound of zero it was expected that non-normality was a possibility. To avoid this we could have transformed BoS onto another scale, however this would not have fully resolved any non-normality problems and the interpretability of BoS would have been lost. Therefore, the residuals were thus deemed appropriately approximately normally distributed for our application.

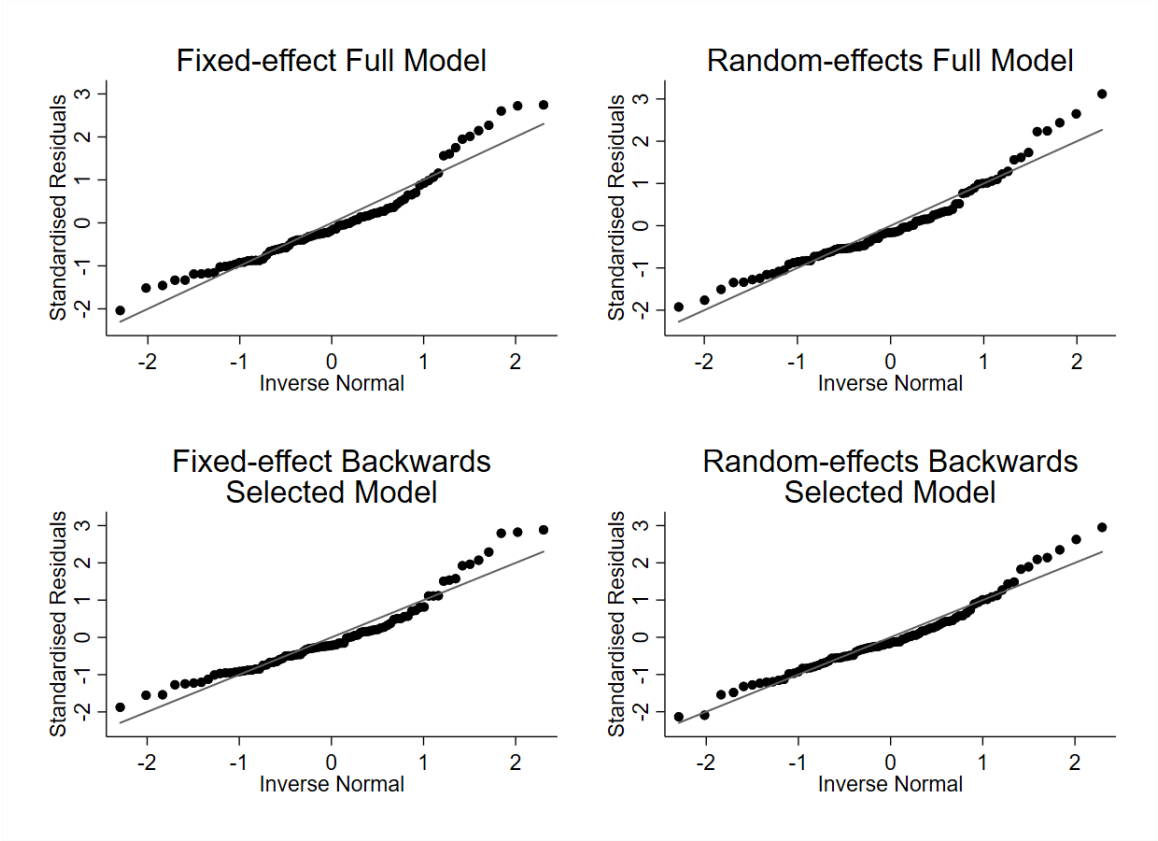


Figure D.2: Normal probability plots for the standardised residuals

D.2 Exploratory analysis: consideration of interactions

D.2.1 Method for exploratory analysis

Selective interactions were considered for exploratory purposes only. Only select variables were considered for interactions to minimise the effect of overfitting due to a large number of variables. Interactions with the absolute between-study correlation and the average absolute within-study correlations were chosen. The absolute between-study correlation and the average absolute within-study correlations were investigated (without interactions) using the full models in Section 4.3.1.

Parsimonious models were developed using backwards selection with a signifi-

cance level of 0.1 for the removal of covariates or interactions. A covariate was only removed from the model if the covariate was not contained within an interaction in the model. For the fixed-effect meta-analyses, only interactions with the average absolute within-study correlation were included. It was decided that the average would be considered rather than the maximum within-study correlation, as the average is contributed to by all the observations. For the random-effects meta-analyses, the interactions for the average absolute within-study correlations and the absolute between-study correlation were included for exploratory analysis. These models are included here for exploratory purposes only, since there were concerns around overfitting, due to the number of covariates included in the models compared to the number of observations.

D.2.2 Results for exploratory analysis

For the fixed-effect meta-analyses, the interactions that remained in the model were the percentage of missing data for all outcomes, the number of studies with the outcome of interest and the number of studies with both outcomes, which were interacted with the average absolute within-study correlation (Table D.1).

Table D.1: Table of multivariable results for the backwards selection models with interactions of the possible predictors of BoS statistic from the fixed-effect meta-analyses

Covariate	Coef.	S.E.	p-value	95% C.I.
Percentage missing across both outcomes	-0.7672	0.4561	0.097	-1.6756 to 0.1413
Average absolute within-study correlation	-43.2249	19.5695	0.030	-82.2009 to -4.2488
Percentage missing for outcome of interest	0.4939	0.1225	<0.001	0.2500 to 0.7378
Number of studies with outcome of interest	1.9776	0.9110	0.033	0.1632 to 3.7920
Number of studies with both outcomes	-3.4105	1.2254	0.007	-5.8511 to -0.9700
Maximum absolute within-study correlation	35.15478	11.7625	0.004	11.7277 to 58.5818
Interaction between percentage missing across both outcomes and average absolute within-study correlation	1.0979	0.5772	0.061	-0.0518 to 2.2476
Interaction between number of studies with outcome of interest and average absolute within-study correlation	-3.2931	1.1816	0.007	-5.6463 to -0.9398
Interaction between the number of studies with both outcomes and average absolute within-study correlation	5.7664	1.7475	0.001	2.2859 to 9.2469
Intercept	2.4412	8.6555	0.779	-14.7976 to 19.6801

The two interactions with the within-study correlation that were kept in the model for BoS from the random-effects meta-analyses were statistically significant (Table D.2). The two interactions were with the percentage of missing data for the outcome of interest and Jackson's I^2. The interactions with the absolute between-study correlations were not statistically significant. For example, the interaction between the number of studies and the absolute between-study correlation had a p-value of 0.050 in the model.

Table D.2: Table of multivariable results for the backwards selection models with interactions of the possible predictors of BoS statistic from the REML random-effects meta-analyses

Covariate	Coef.	S.E.	p-value	95% C.I.
Number of studies	59.4363	29.4621	0.047	0.7046 to 118.168
Percentage missing across both outcomes	-28.4982	14.2491	0.049	-56.9034 to -0.0931
Percentage missing for outcome of interest	0.1285	0.1548	0.409	-0.1801 to 0.4371
Average absolute within-study correlation	-10.8463	10.1920	0.291	-31.1637 to 9.4712
Number of studies with both outcomes	-96.2018	50.7979	0.062	-197.4656 to 5.0619
Maximum absolute within-study correlation	44.6690	11.3910	<0.001	21.9614 to 67.3766
Jackson's I^2	0.7555	0.2430	0.003	0.2712 to 1.2399
Absolute between-study correlation	-502.7593	303.4062	0.102	-1107.589 to 102.07
Interaction between percentage missing for outcome of interest and average absolute within-study correlation	0.6626	0.2187	0.003	0.2266 to 1.0986
Interaction between Jackson's I^2 and average absolute w-s corr	-1.1069	0.3935	0.006	-1.8913 to -0.3225

Continued on next page

Table D.2: continued

Covariate	Coef.	S.E.	p-value	95% C.I.
Interaction between number of studies and absolute between-study correlation	-58.9213	29.4976	0.050	-117.7236 to -0.1190
Interaction between percentage missing across both outcomes and absolute between-study correlation	28.2533	14.2630	0.051	-0.1794 to 56.6860
Interaction between number of studies with both outcomes and absolute between-study correlation	95.2028	50.8792	0.065	-6.2231 to 196.6287
Intercept	481.5283	303.3555	0.117	-123.1999 to 1086.256

Appendix E

E.1 Interactive tool stills

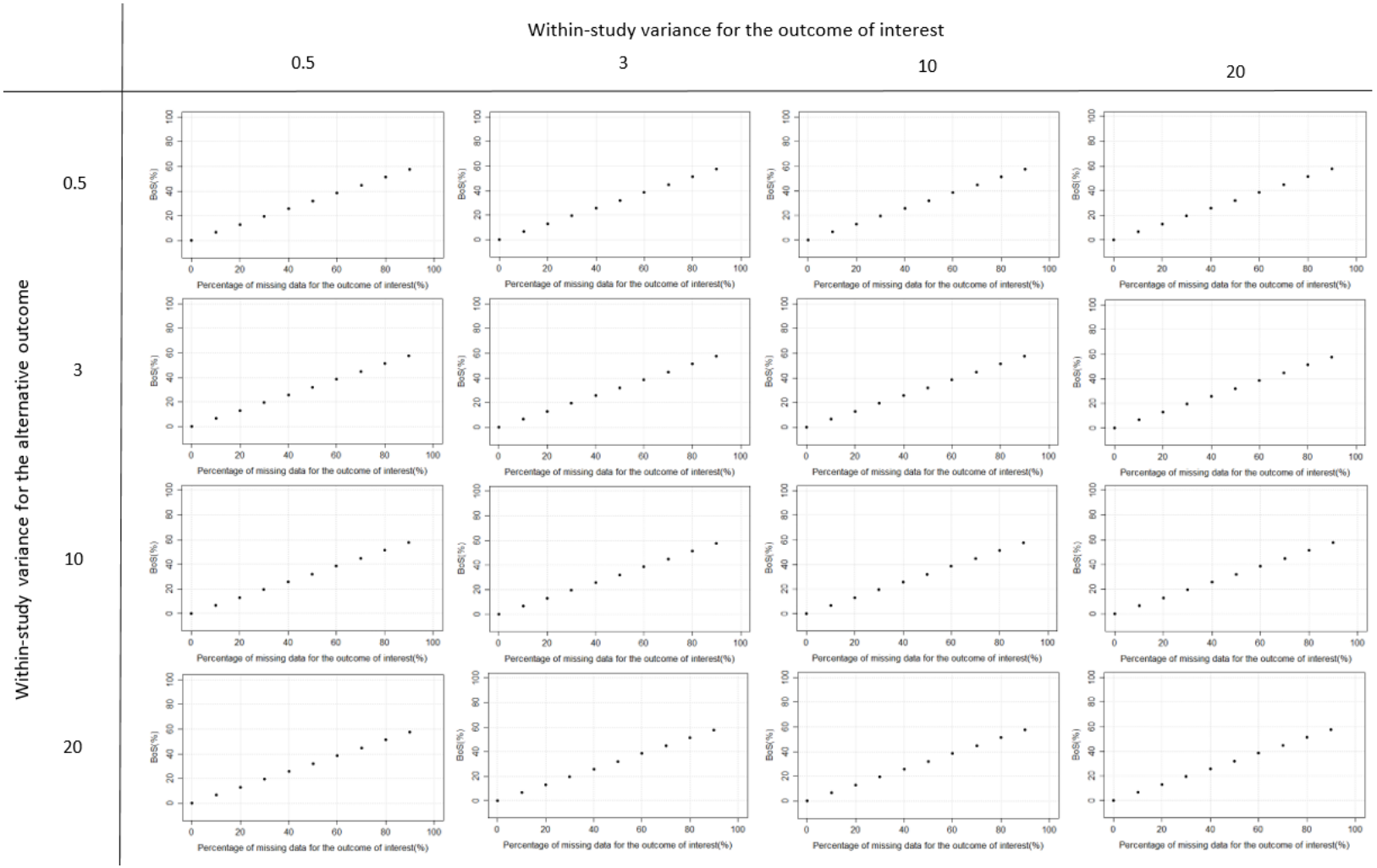


Figure E.1: The effect of the within-study variances on the magnitude of BoS from a fixed-effect meta-analyses with no missing data for the alternative outcome, within-study correlation of 0.8 and 20 studies

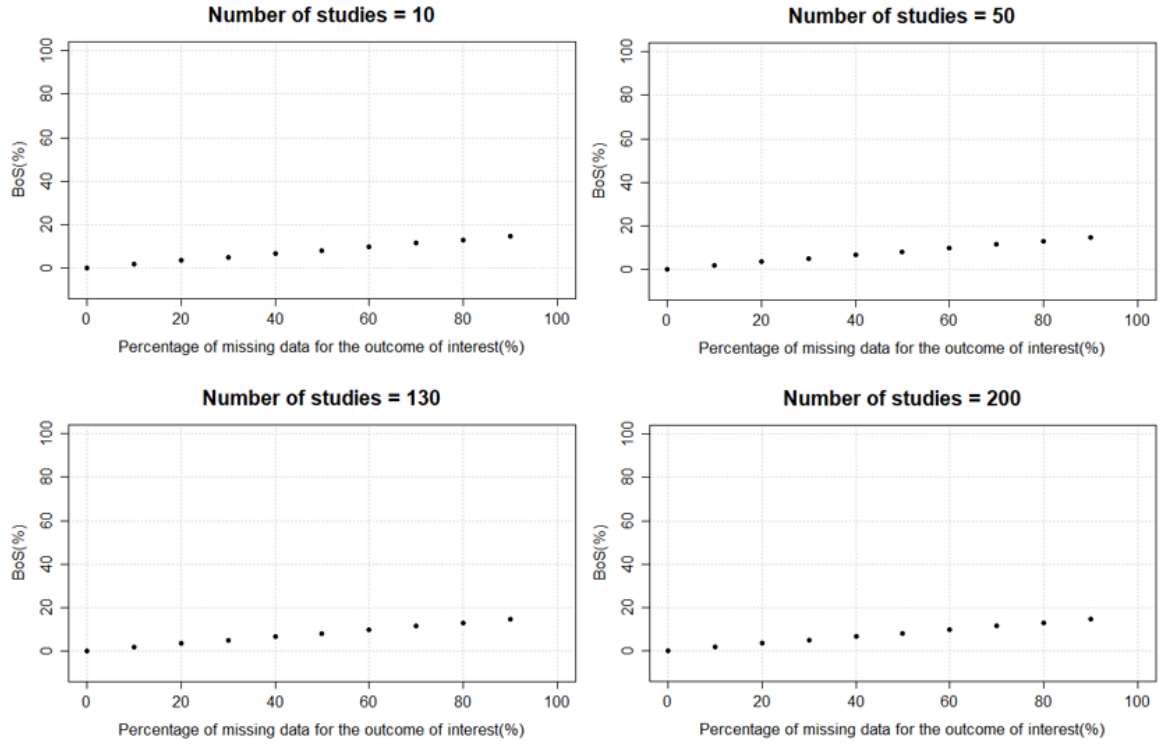


Figure E.2: The effect of the number of studies in the meta-analysis on the magnitude of BoS from a fixed-effect meta-analyses with no missing data for the alternative outcome, within-study correlation of 0.4 and within-study variances of 5

E.2 Interactive tool code for the prediction of BoS for bivariate fixed-effect meta-analysis

server code

```
library(shiny)
shinyServer(function(input, output){
  i<-reactive({
    n<-input$n
    ws<-input$ws
    -13.020+0.962*(-0.13*n+26.590*ws)
  })
  j<-reactive({
```

```
    n<-input$n
    ws<-input$ws
    -13.020 +0.962*(0.655*n+29.4+26.590*ws)
  })
  k<-reactive({
    n<-input$n
    ws<-input$ws
    ((13.020/0.962)+0.13*n-26.59*ws)/(0.294+0.00785*n)
      })
  l<-reactive({
    n<-input$n
    ws<-input$ws
    ((113.020/0.962)+0.13*n-26.59*ws)/(0.294+0.00785*n)
  })
output$thePlot<-renderPlot({
   plot(c(0,100), c(0,100), type = "n", xlab = "Percentage of missing
        data for the outcome of interest (%)", ylab = "BoS (%)",
        cex.axis=1.3, cex.lab=1.4)
   grid()
   if(i()<0 & j()>100){
     segments(x0=0,y0=0, x1=k(),y1=0)
     segments(x0=k(),y0=0,x1=l(),y1=100)
     segments(x0=l(),y0=100,x1=100,y1=100)
   }
   if(i()<0 & j()<=100){
     segments(x0=0,y0=0,x1=k(),y1=0)
     segments(x0=k(),y0=0,x1=100,y1=j())
   }
   if(i()>=0 & j()>100){
```

```
    segments(x0=0, y0=i(),x1=l(),y1=100)
    segments(x0=l(),y0=100,x1=100,y1=100)
  }
  if(i()>=0 & j()<=100){
    segments(x0=0,y0=i(),x1=100,y1=j())
  }
  })
 output$blank<-renderText({
  " "
 })
})
```

ui code

```
library(shiny)
fluidPage(
  titlePanel("Predicted BoS for bivariate fixed-effect meta-analysis",
    windowTitle = "Predicted BoS"),
  fluidRow(
    column(6,
           sliderInput("n", "Number of studies", 10, 200, 200, step=10),
           sliderInput("ws", "Maximum absolute within-study
              correlation", 0, 1, 0, step=0.05)
    )
  ),
  column(width = 12, offset = 0.5,
              plotOutput("thePlot")
  )
)
```

E.3 Interactive tool code for the prediction of BoS for bivariate random-effects meta-analysis

server code

```
library(shiny)
shinyServer(function(input, output, session){
  observe({
    a<-input$aws
    updateSliderInput(session,"ws", value=a, min=a, max=1, step=0.05)
  })
  i<-reactive({
    n<-input$n
    ws<-input$ws
    a<-input$aws
    k<-input$k
    -7.561+0.925*(0.477*n-0.274*(k/2)-0.907*n*((100-k)/100)-19.102*a
        +39.297*ws)
  })
  j<-reactive({
    n<-input$n
    ws<-input$ws
    a<-input$aws
    k<-input$k
    31.289+0.925*(0.477*n-19.102*a+39.297*ws)
  })
  h<-reactive({
    n<-input$n
    ws<-input$ws
    a<-input$aws
```

```
  k<-input$k
  ((7.561/0.925)+0.43*n+19.102*a-39.297*ws+0.137*k-0.00907*n*k)/
       (0.42+0.00907*n)
})
l<-reactive({
  n<-input$n
  ws<-input$ws
  a<-input$aws
  k<-input$k
  ((107.561/0.925)+0.43*n+19.102*a-39.297*ws+0.137*k-0.00907*n*k)/
       (0.42+0.00907*n)
})
r<-reactive({
  k<-input$k
  100-k
})
q<-reactive({
  n<-input$n
  ws<-input$ws
  a<-input$aws
  k<-input$k
  31.199+0.925*(0.477*n-0.557*k-19.102*a+39.297*ws)
})
  output$thePlot<-renderPlot({
  k<-input$k
  plot(c(0,100), c(0,100), type="n", cex.main=1.8, xlab = "Percentage
     of missing data for the outcome of interest (%)",
     ylab = "BoS (%)", cex.axis=1.3, cex.lab=1.4)
  grid()
```

```
if(k<=0){
  if(i()<0 & j()>100){
    segments(x0=0,y0=0, x1=h(),y1=0)
    segments(x0=h(),y0=0,x1=l(),y1=100)
    segments(x0=l(),y0=100,x1=100,y1=100)
  }
  if(i()<0 & j()<=100){
    segments(x0=0,y0=0,x1=h(),y1=0)
    segments(x0=h(),y0=0,x1=100,y1=j())
  }
  if(i()>=0 & j()>100){
    segments(x0=0, y0=i(),x1=l(),y1=100)
    segments(x0=l(),y0=100,x1=100,y1=100)
  }
  if(i()>=0 & j()<=100){
    segments(x0=0,y0=i(),x1=100,y1=j())
  }
}
else if(k>0){
  if(i()<0 & q()<=0){
    segments(x0=0,y0=0, x1=r(),y1=0)
  }
  if(q()<=100 & i()<0){
    segments(x0=0, y0=0, x1=h(), y1=0)
    segments(x0=h(),y0=0, x1=r(),y1=q())
  }
  if(i()<0 & q()>100){
    segments(x0=0,y0=0,x1=h(),y1=0)
    segments(x0=h(),y0=0,x1=l(),y1=100)
```

```
      }
      if(i()>=0 & q()>100){
        segments(x0=0,y0=i(),x1=l(),y1=100)
        segments(x0=l(),y0=100,x1=r(),y1=100)
      }
      if(i()>=0 & q()<=100){
        segments(x0=0,y0=i(),x1=r(),y1=q())
      }
    }
      })
})
```

ui code

```
library(shiny)
fluidPage(
  titlePanel("Predicted BoS for bivariate random-effects meta-analysis",
         windowTitle = "Predicted BoS"),
  fluidRow(
    column(6,
    sliderInput("n", "Number of studies", 10, 200,200, step=10),
    sliderInput("k", "Percentage of missing data for the alternative
          outcome", 0, 90, 0, step=10)
    ),
    column(6,
           sliderInput("aws", "Average absolute within-study
                 correlation", 0, 1, 0, step=0.05),
        sliderInput("ws", "Maximum absolute within-study
          correlation", 0, 1, 0, step=0.05)
    )
```

```
  ),
  column(width = 12, offset = 0.5,
         plotOutput("thePlot")
  )
)
```

E.4 Interactive tool code for BoS from its mathematical equation

server code

```
library(shiny)
shinyServer(function(input, output){
  ws<-reactive({
    if(input$ws==1){
      ws<-0.999
    } else if(input$ws==-1){
      ws<--0.999
    } else {
      ws<-input$ws
    }
  })
  lambda<- reactive({
    ws<-ws()
    vara<-input$vara
    varb<-input$varb
    ws*sqrt(vara)*sqrt(varb)
  })
  eta<- reactive({
    bs<-input$bs
```

```
  tausq1<-input$tausq1
  tausq2<-input$tausq2
  bs*sqrt(tausq1)*sqrt(tausq2)
})
m<-reactive({
  if(input$k==10){
    m<-c(0,10,20,30,40,50,60,70,80)
  } else if(input$k==20){
    m<-c(0,10,20,30,40,50,60,70)
  } else if(input$k==30){
    m<-c(0,10,20,30,40,50,60)
  } else if(input$k==40){
    m<-c(0,10,20,30,40,50)
  } else if(input$k==50){
    m<-c(0,10,20,30,40)
  } else if(input$k==60){
    m<-c(0,10,20,30)
  } else if(input$k==70){
    m<-c(0,10,20)
  } else if(input$k==80){
    m<-c(0,10)
  } else if(input$k==90){
    m<-c(0)
  } else {
    m<-c(0,10,20,30,40,50,60,70,80,90)
  }
})
bos<-reactive({
  vara<-input$vara
```

```
varb<-input$varb
n<-input$n
k<-input$k
mlambda<-0
mvar<-100000
tau1<-input$tausq1
tau2<-input$tausq2
(1-(((((1/(tau1+mvar))*((m()/100)*n))+((1/(tau1+vara))*((k/100)*n))
       +((1/(tau1+vara))*(n-((m()/100)*n)-((k/100)*n))))
       *((((tau1+mvar)/(((tau1+mvar)*(tau2+varb))
       -((eta()+mlambda)^2)))*((m()/100)*n))
       +(((tau1+vara)/(((tau1+vara)*(tau2+mvar))
       -((eta()+mlambda)^2)))*((k/100)*n))+(((tau1+vara)
       /(((tau1+vara)*(tau2+varb))-((eta()+lambda())^2)))
       *(n-((m()/100)*n)-((k/100)*n)))))/((((((tau1+mvar)
       /(((tau1+mvar)*(tau2+varb))-((eta()+mlambda)^2)))
       *((m()/100)*n))+(((tau1+vara)/(((tau1+vara)*(tau2+mvar))
       -((eta()+mlambda)^2)))*((k/100)*n))+(((tau1+vara)
       /(((tau1+vara)*(tau2+varb))-((eta()+lambda())^2)))
       *(n-((m()/100)*n)-((k/100)*n))))*((((tau2+varb)/(((tau1+mvar)
       *(tau2+varb))-((eta()+mlambda)^2)))*((m()/100)*n))
       +(((tau2+mvar)/(((tau1+vara)*(tau2+mvar))
       -((eta()+mlambda)^2)))*((k/100)*n))+(((tau2+varb)
       /(((tau1+vara)*(tau2+varb))-((eta()+lambda())^2)))
       *(n-((m()/100)*n)-((k/100)*n)))))-(((((eta()+mlambda)
       /(((tau1+mvar)*(tau2+varb))-((eta()+mlambda)^2)))
       *((m()/100)*n))+(((eta()+mlambda)/(((tau1+vara)*(tau2+mvar))
       -((eta()+mlambda)^2)))*((k/100)*n))+(((eta()+lambda())
       /(((tau1+vara)*(tau2+varb))-((eta()+lambda())^2)))
```

```
        *(n-((m()/100)*n)-((k/100)*n))))^2))))*100
})


output$thePlot<-renderPlot({
  m<-m()
  bos<-bos()
  plot(m, bos,pch = 19,ylim=c(-10,100),xlim = c(0,100), cex.main=1.8,
  xlab = 'Percentage of missing data for the outcome of interest',
  ylab = 'BoS(%)', cex.axis=1.3, cex.lab=1.5, grid())
  abline(0,1, col="red")
})
output$blank<-renderText({
  " "
})

output$wscorr<-renderText({
  if(input$ws==1){
    "Note: The within-study correlation can not be equal to 1,
    rather the within-study correlation is estimated as 0.999."
  } else if (input$ws==-1){
    "Note: The within-study correlation can not be equal to -1,
    rather the within-study correlation is estimated as -0.999."
  } else {
    " "
  }
})

output$BoS<-renderText({
```

```
    paste(bos())
  })
})
```

ui code

```
library(shiny)
fluidPage(
  titlePanel("BoS", windowTitle = "BoS"),
  fluidRow(
    column(4,
      sliderInput("n", "Number of studies", 10, 200,200, step=10),
      sliderInput("k", "Percentage of missing data for the alternative
      outcome", 0, 90, 0, step=10)
    ),
    column(4,
      sliderInput("ws", "Within-study correlation", -1, 1, 0.8,
      step=0.05),
      sliderInput("vara", "Within-study variance for available data for
      outcome of interest", 0.5, 20, 20, step=0.1),
      sliderInput("varb", "Within-study variance for available data for
      the alternative outcome", 0.5, 20, 20, step=0.1)
    ),
  column(4,
      sliderInput("bs", "Between-study correlation", -1, 1, 0, step=0.05),
      sliderInput("tausq1", "Between-study variance estimate for outcome of
       interest", 0, 20, 0, step=0.1),
      sliderInput("tausq2", "Between-study variance estimate for the
      alternative outcome", 0, 20, 0, step=0.1)
    ),
```

```
  fluidRow(
           column(12,style = "padding-left:30px",
                  plotOutput("thePlot"),
                  textOutput("blank"),
                  textOutput("wscorr")
                  ))
    )
)
```

Appendix F

F.1 Details for all sub-scenarios

Table F.1: Table of sub-scenarios for the simulation study for the baseline balance setting

Scenario No.	No. of Trials	No. of Participants	% missing	α_i	β_{1i}	β_{2i}	σ_i^2	Baseline	Treatment proportion
1a	5	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
1b	5	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
1c	5	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
1d	5	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
1e	5	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
1f	5	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
1g	5	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
1h	5	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2a	10	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2b	10	100	10	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2c	10	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2d	10	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2e	10	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2f	10	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2g	10	1000	10	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2h	10	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2i	10	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
2j	10	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3a	20	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3b	20	100	10	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3c	20	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3d	20	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3e	20	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3f	20	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3g	20	1000	10	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5

Continued on next page

Table F.1: continued

Scenario No.	No. of Trials	No. of Participants	% missing	α_i	β_{1i}	β_{2i}	σ_i^2	Baseline	Treatment proportion
3h	20	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3i	20	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
3j	20	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,0)	225	N(165,324)	0.5
4a	5	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
4b	5	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
4c	5	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
4d	5	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
4e	5	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
4f	5	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
4g	5	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
4h	5	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5a	10	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5b	10	100	10	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5c	10	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5d	10	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5e	10	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5f	10	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5g	10	1000	10	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5h	10	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5i	10	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
5j	10	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6a	20	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6b	20	100	10	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6c	20	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6d	20	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6e	20	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5

Continued on next page

Table F.1: continued

Scenario No.	No. of Trials	No. of Participants	% missing	α_i	β_{1i}	β_{2i}	σ_i^2	Baseline	Treatment proportion
6f	20	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6g	20	1000	10	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6h	20	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6i	20	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
6j	20	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	0.5
7a	5	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
7b	5	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
7c	5	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
7d	5	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
7e	5	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
7f	5	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
7g	5	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
7h	5	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8a	10	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8b	10	100	10	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8c	10	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8d	10	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8e	10	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8f	10	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8g	10	1000	10	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8h	10	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8i	10	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
8j	10	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9a	20	100	0	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9b	20	100	10	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9c	20	100	20	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5

Continued on next page

Table F.1: continued

Scenario No.	No. of Trials	No. of Participants	% missing	α_i	β_{1i}	β_{2i}	σ_i^2	Baseline	Treatment proportion
9d	20	100	40	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9e	20	100	60	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9f	20	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9g	20	1000	10	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9h	20	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9i	20	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5
9j	20	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,9)	225	N(165,324)	0.5

Table F.2: Table of sub-scenarios for the simulation study for the baseline imbalance settings

Scenario No.	No. of Trials	No. of Participants	% missing	α_i	β_{1i}	β_{2i}	σ_i^2	Baseline		Treatment proportion
For Unconditional, Conditional 1 and Conditional 2 scenarios										
a	10	1000	0	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	N(170, 324)	0.5
b	10	1000	10	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	N(170, 324)	0.5
c	10	1000	20	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	N(170, 324)	0.5
d	10	1000	40	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	N(170, 324)	0.5
e	10	1000	60	N(80,25)	N(0.5, 0.0025)	N(-10,2.25)	225	N(165,324)	N(170, 324)	0.5

F.2 Power of the confidence interval results

Table F.3: Table of power for each meta-analysis method for each sub-scenario for the baseline balance settings

Scenario No.	No. of Trials	No. of Participants	Between-study Variance	Power of 5% level test				
				Bivariate ANCOVA	Univariate ANCOVA	Cochrane	Bivariate Final	Univariate Final
1a	5	100	0	100	100	100	100	100
1b	5	100	0	100	100	100	100	100
1c	5	100	0	100	99.900	100	100	100
1d	5	100	0	99.7	96.497	100	100	100
1e	5	1000	0	100	100	100	100	100
1f	5	1000	0	100	100	100	100	100
1g	5	1000	0	100	100	100	100	100
1h	5	1000	0	100	100	100	100	100
2a	10	100	0	100	100	100	100	100
2b	10	100	0	100	100	100	100	100
2c	10	100	0	100	100	100	100	100
2d	10	100	0	100	100	100	100	100
2e	10	100	0	100	99.900	100	100	100
2f	10	1000	0	100	100	100	100	100
2g	10	1000	0	100	100	100	100	100
2h	10	1000	0	100	100	100	100	100
2i	10	1000	0	100	100	100	100	100
2j	10	1000	0	100	100	100	100	100
3a	20	100	0	100	100	100	100	100
3b	20	100	0	100	100	100	100	100
3c	20	100	0	100	100	100	100	100
3d	20	100	0	100	100	100	100	100

Continued on next page

Table F.3: continued

Scenario No.	No. of Trials	No. of Participants	Between-study Variance	Power of 5% level test				
				Bivariate ANCOVA	Univariate ANCOVA	Cochrane	Bivariate Final	Univariate Final
3e	20	100	0	100	100	100	100	100
3f	20	100	0	100	100	100	100	100
3g	20	1000	0	100	100	100	100	100
3h	20	1000	0	100	100	100	100	100
3i	20	1000	0	100	100	100	100	100
3j	20	1000	0	100	100	100	100	100
4a	5	100	2.25	100	100	100	99.900	100
4b	5	100	2.25	100	99.800	100	100	100
4c	5	100	2.25	99.900	99.700	100	100	100
4d	5	100	2.25	99.7	95.295	100	100	100
4e	5	100	2.25	100	100	100	100	100
4f	5	1000	2.25	100	100	100	100	100
4g	5	1000	2.25	100	100	100	100	100
4h	5	1000	2.25	100	100	100	100	100
5a	10	1000	2.25	100	100	100	100	100
5b	10	100	2.25	100	100	100	100	100
5c	10	100	2.25	100	100	100	100	100
5d	10	100	2.25	100	100	100	100	100
5e	10	100	2.25	100	100	100	100	100
5f	10	1000	2.25	100	100	100	100	100
5g	10	1000	2.25	100	100	100	100	100
5h	10	1000	2.25	100	100	100	100	100
5i	10	1000	2.25	100	100	100	100	100
5j	10	1000	2.25	100	100	100	100	100
6a	20	100	2.25	100	100	100	100	100
6b	20	100	2.25	100	100	100	100	100

Continued on next page

Table F.3: continued

Scenario No.	No. of Trials	No. of Participants	Between-study Variance	Power of 5% level test				
				Bivariate ANCOVA	Univariate ANCOVA	Cochrane	Bivariate Final	Univariate Final
6c	20	100	2.25	100	100	100	100	100
6d	20	100	2.25	100	100	100	100	100
6e	20	100	2.25	100	100	100	100	100
6f	20	1000	2.25	100	100	100	100	100
6g	20	1000	2.25	100	100	100	100	100
6h	20	1000	2.25	100	100	100	100	100
6i	20	1000	2.25	100	100	100	100	100
6j	20	1000	2.25	100	100	100	100	100
7a	5	100	9	99.900	99.900	99.900	99.400	99.400
7b	5	100	9	99.900	98.796	99.900	99.398	99.398
7c	5	100	9	99.300	96.100	99.800	99.400	99.400
7d	5	100	9	97.898	86.587	99.800	99.400	99.400
7e	5	100	9	100	100	100	100	100
7f	5	1000	9	100	100	100	100	100
7g	5	1000	9	100	99.798	100	100	100
7h	5	1000	9	99.799	96.073	100	100	100
8a	10	100	9	100	100	100	100	100
8b	10	100	9	100	100	100	100	100
8c	10	100	9	100	100	100	100	100
8d	10	100	9	100	99.900	100	100	100
8e	10	100	9	100	98.800	100	100	100
8f	10	1000	9	100	100	100	100	100
8g	10	1000	9	100	100	100	100	100
8h	10	1000	9	100	100	100	100	100
8i	10	1000	9	100	100	100	100	100
8j	10	1000	9	100	100	100	100	100

Continued on next page

Table F.3: continued

Scenario No.	No. of Trials	No. of Participants	Between-study Variance	Power of 5% level test				
				Bivariate ANCOVA	Univariate ANCOVA	Cochrane	Bivariate Final	Univariate Final
9a	20	100	9	100	100	100	100	100
9b	20	100	9	100	100	100	100	100
9c	20	100	9	100	100	100	100	100
9d	20	100	9	100	100	100	100	100
9e	20	100	9	100	100	100	100	100
9f	20	1000	9	100	100	100	100	100
9g	20	1000	9	100	100	100	100	100
9h	20	1000	9	100	100	100	100	100
9i	20	1000	9	100	100	100	100	100
9j	20	1000	9	100	100	100	100	100

F.3 Baseline balance and imbalance simulation code

```
/* Number of studies */
local nstudy 10

*/ set missing percentage and calculate number of studies missing */
local percentmiss 60
local miss=‘nstudy’*(‘percentmiss’/100)

*/ Number of participants in each study */
local parti 1000


/* number of simulations */
local nsim 1000
/* proportion of participants in each arm */
local trtproportion 0.5


/* For baseline balance settings */
     /* Mean for baseline of systolic blood pressure
     local baselinemean 165


/* For baseline imbalance settings*/
     /* mean for baseline of systolic blood pressure for control */
     local baselinemeanc 170
     /* mean for baseline of systolic blood pressure for treatment
     group */
     local baselinemeant 165


/* standard deviation of baseline sbp */
local baselinesd 18
/* error term */
local errorsig 15
/* mean treatment effect */
local trteff -10
/* standard deviation of treatment effect, set to 0 for same
   treatment effect across studies */
local trteffsd 1.5
/* baseline adjustment */
local baseadj 0.5
/* baseline adjustment standard deviation, set to 0 for same baseline
   adjustment across studies */
local baseadjsd 0.05
/* intercept */
```

```
local inter 80
/* intercept standard deviation, set to 0 for same intercept across
   studies */
local intersd 5



/* seed */
set seed 453
/* file for meta-analysis results */
postfile study metaid multancova multancovase multfinal ///
    multfinalse iterations bos1 bos2 tau2ma tau2mf bscorr i2ma ///
    i2mf uniancova uniancovase i2ua tau2ua unifinal unifinalse ///
    i2uf tau2uf Cochrane Cochranese i2c tau2c using ///
    "study.dta", replace
/* simulation */
forvalues i=1/‘nsim’ {
    clear
    /* loop to generate number of participants per trial
       trial and participant id
       stack data from each study and save data */
    forvalues j=1/‘nstudy’{
        clear
        set obs ‘parti’
        gen trialid=‘j’
        gen patientid=_n
        save data‘j’, replace
        if ‘j’>1{
        append using data1
        save data1, replace
        }
    }
    /* open data from all studies that have been stacked */
    use data1, replace
    sort trialid
    /* proportion of participants in each arm */
    gen trtgrp=rbinomial(1,‘trtproportion’)

/* Use the following for baseline balance setting */
        /* Baseline systolic blood pressure */
    gen baseline=rnormal(‘baselinemean’, ‘baselinesd’)

/* Use one of the following for desired baseline imbalance setting */
        /* For baseline imbalance setting unconditional all
           encompassing */
        /* Baseline systolic blood pressure */
        gen baseline=rnormal(‘baselinemeanc’, ‘baselinesd’) if ///
        trtgrp==0
```

```
    replace baseline=rnormal(‘baselinemeant’, ‘baselinesd’) if ///
    trtgrp==1

    /* For baseline imbalance setting conditional, baseline ///
       imbalance for missing treatment effect estimates from the ///
       ANCOVA model (i.e. for trials without IPD) */
    /* Baseline systolic blood pressure */
    gen baseline= rnormal(‘baselinemeanc’, ‘baselinesd’) if ///
    trtgrp==0 & trialid<=‘miss’
    replace baseline=rnormal(‘baselinemeant’, ‘baselinesd’) if ///
    baseline==.

    /* For baseline imbalance setting conditional, baseline ///
       imbalance for available treatment effect estimates from ///
       the ANCOVA model (i.e. for trials with IPD) */
    /* Baseline systolic blood pressure */
    gen baseline= rnormal(‘baselinemeanc’, ‘baselinesd’) if ///
    trtgrp==0 & trialid>‘miss’
    replace baseline=rnormal(‘baselinemeant’, ‘baselinesd’) if ///
    baseline==.


/* error term */
gen error=rnormal(0,‘errorsig’)
/* the beta2, beta1, alpha one for each study (repeated for each
    participant for calculation purposes for final measurement) */
    /* generate treatment effect. For same treatment effect across
       studies set standard deviation to 0. */
    gen beta2a=rnormal(‘trteff’,‘trteffsd’) if patientid==1
    by trialid: egen beta2=sum(beta2a)
    /* generate baseline adjustment. For same baseline adjustment
       across studies set standard deviation to 0. */
    gen beta1a=rnormal(‘baseadj’,‘baseadjsd’) if patientid==1
    by trialid: egen beta1=sum(beta1a)
    /* generate intercept. For same intercept across studies set
       standard deviation to 0. */
    gen alpha1=rnormal(‘inter’,‘intersd’) if patientid==1
    by trialid: egen alpha=sum(alpha)
/* calculate final scores for participants */
gen final=alpha + beta1*baseline + beta2*trtgrp + error
/* file for models fitted */
postfile model studyid b1 V11 b2 V22 corr using "model.dta", ///
replace
/* Fit each model separately for each study */
forvalues j=1/‘nstudy’{
    regress final baseline trt if trialid==‘j’
    mat b=e(b)
    mat V=e(V)
```

```
    local b1=b[1,2]
    local V11=V[2,2]
    regress final trt if trialid==‘j’
    mat b=e(b)
    mat V=e(V)
    local b2=b[1,1]
    local V22=V[1,1]
    sureg (final = baseline trt) (final = trt) if trialid==‘j’ ///
    , corr
    mat V=e(V)
    local corr=V[2,4]/(sqrt(V[2,2])*sqrt(V[4,4]))
    post model (‘j’) (‘b1’) (‘V11’) (‘b2’) (‘V22’) (‘corr’)
}
postclose model
/* open and merge data from models and correlations */
use "model.dta", clear
gen V12=sqrt(V11)*sqrt(V22)*corr
replace b1=. if studyid<=‘miss’
replace V11=. if b1==.
replace V12=. if b1==.
/* multivariate met-analysis */
mvmeta b V, reml iterate(200) wt(keepmat(wts)) nounc
mat b=e(b)
mat V=e(V)
local mb1=b[1,1]
local mb2=b[1,2]
local mb1se=sqrt(V[1,1])
local mb2se=sqrt(V[2,2])
matrix A=wtsborrowed
matrix B=J(rowsof(A),1,1)
matrix C=B’*A
local bos1=C[1,1]
local bos2=C[1,2]
mat sig=e(Sigma)
local tau2_ma=sig[1,1]
local tau2_mf=sig[2,2]
local bscorr=sig[1,2]/sqrt(sig[1,1]*sig[2,2])
mat Q=e(Q)
mat Qa=e(Qa)
local i_2ma=max(0, 100*(1-(Qa[1,1]/Q[1,1])))
local i_2mf=max(0, 100*(1-(Qa[2,2]/Q[2,2])))
local ic=e(ic)
/* univariate ma for ANCOVA */
mvmeta b V, reml iterate(200) vars(b1) nounc
mat b=e(b)
mat V=e(V)
local ub1=b[1,1]
local ub1se=sqrt(V[1,1])
```

```
    mat Q=e(Q)
    mat Qa=e(Qa)
    local i_2ua=max(0, 100*(1-(Qa[1,1]/Q[1,1])))
    mat sig=e(Sigma)
    local tau2_ua=sig[1,1]
    /* univariate meta-analysis for final */
    mvmeta b V, reml iterate(200) vars(b2) nounc
    mat b=e(b)
    mat V=e(V)
    local ub2=b[1,1]
    local ub2se=sqrt(V[1,1])
    mat Q=e(Q)
    mat Qa=e(Qa)
    local i_2uf=max(0, 100*(1-(Qa[1,1]/Q[1,1])))
    mat sig=e(Sigma)
    local tau2_uf=sig[1,1]
    /* Cochrane meta-analysis */
    gen b3=b1
    gen V33=V11
    replace b3=b2 if b3==.
    replace V33=V22 if V33==.
    mvmeta b V, reml iterate(200) vars(b3) nounc
    mat b=e(b)
    mat V=e(V)
    local ub3=b[1,1]
    local ub3se=sqrt(V[1,1])
    mat Q=e(Q)
    mat Qa=e(Qa)
    local i_2c=max(0, 100*(1-(Qa[1,1]/Q[1,1])))
    mat sig=e(Sigma)
    local tau2_c=sig[1,1]
    post study (‘i’) (‘mb1’) (‘mb1se’) (‘mb2’) (‘mb2se’) (‘ic’) ///
    (‘bos1’) (‘bos2’) (‘tau2_ma’) (‘tau2_mf’) (‘bscorr’) (‘i_2ma’) ///
    (‘i_2mf’) (‘ub1’) (‘ub1se’) (‘i_2ua’) (‘tau2_ua’) (‘ub2’) ///
    (‘ub2se’) (‘i_2uf’) (‘tau2_uf’) (‘ub3’) (‘ub3se’) (‘i_2c’) ///
    (‘tau2_c’)
}
postclose study
```

List of references

Aiello, F., Attanasio, M. & Tinè, F. (2011), 'Assessing covariate imbalance in meta-analysis studies', *Statistics in Medicine* **30**(22), 2671–2682.
URL: *http://doi.wiley.com/10.1002/sim.4311*

Akobeng, A. K. (2005), 'Understanding systematic reviews and meta-analysis', *Archives of Disease in Childhood* **90**(8), 845–848.
URL: *http://dx.doi.org/10.1136/adc.2004.058230*

Altman, D. G. & Royston, P. (2000), 'What do we mean by validating a prognostic model?', *Statistics in medicine* **19**(4), 453–473.
URL: *https://doi.org/10.1002/(SICI)1097-0258(20000229)19:4%3C453::AID-SIM350%3E3.0.CO;2-5*

Altman, D. G., Vergouwe, Y., Royston, P. & Moons, K. G. M. (2009), 'Prognosis and prognostic research: validating a prognostic model', *BMJ* **338**.

Ashby, D. & Smith, A. F. (2000), 'Evidence-based medicine as Bayesian decision-making', *Statistics in Medicine* **19**(23), 3291–3305.
URL: *https://doi.org/10.1002/1097-0258(20001215)19:23%3C3291::AID-SIM627%3E3.0.CO;2-T*

Austin, P. C. & Steyerberg, E. W. (2015), 'The number of subjects per variable required in linear regression analyses', *Journal of Clinical Epidemiology* **68**(6), 627–636.
URL: *http://dx.doi.org/10.1016/j.jclinepi.2014.12.014*

Bagos, P. G. (2012), 'On the covariance of two correlated log-odds ratios', *Statistics in medicine* **31**(14), 1418–1431.

Bär, F. W., Meyer, J., Boland, J., Betriu, A., Artmeyer, B., Charbonnier, B., Michels, H. R., Tebbe, U., Spiecker, M., Vermeer, F., Von Fisenne, M. J., Hopkins, G. R. & Barth, H. (1998), 'Bolus administration of saruplase in Europe (BASE), a pilot study in patients with acute myocardial infarction', *Journal of Thrombosis and Thrombolysis* **6**(2), 147–153.
URL: *https://doi.org/10.1023/A:1008809907268*

Becker, B. J. (2000), Multivariate meta-analysis, *in* 'Handbook of Applied Multivariate Statistics and Mathematical Modeling', Academic Press:San Diego, pp. 499–525.

Bland, J. M. & Altman, D. G. (2000), 'Statistics notes. The odds ratio.', *BMJ (Clinical research ed.)* **320**(7247), 1468.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1127651/*

Bleich, S. D., Adgey, A. A. J., McMechan, S. R. & Love, T. W. (1998), 'An angiographic assessment of alteplase: double-bolus and front-loaded infusion regimens in myocardial infarction', *American heart journal* **136**(4), 741–748.
URL: *https://doi.org/10.1016/S0002-8703(98)70024-9*

Bode, C., Smalling, R. W., Berg, G., Burnett, C., Lorch, G., Kalbfleisch, J. M., Chernoff, R., Christie, L. G., Feldman, R. L., Seals, A. A. & Weaver, W. D. (1996), 'Randomized Comparison of Coronary Thrombolysis Achieved With Double-Bolus Reteplase (Recombinant Plasminogen Activator) and Front-Loaded, Accelerated Alteplase (Recombinant Tissue Plasminogen Activator) in Patients With Acute Myocardial Infarction', *Circulation* **94**(5), 891–898.
URL: *https://doi.org/10.1161/01.CIR.94.5.891*

Bonnett, L. J., Snell, K. I., Collins, G. S. & Riley, R. D. (2019), 'Guide to presenting clinical prediction models for use in clinical settings', *BMJ* **365**, l737.
URL: *http://dx.doi.org/10.1136/bmj.l737*

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009), *Introduction to Meta-Analysis*, John Wiley & Sons, Ltd, Chichester, UK.

Bradburn, M. J., Deeks, J. J. & Altman, D. G. (1999), 'metan-an alternative meta-analysis command', *Stata Technical Bulletin* **8**(44).
URL: *http://stata-press.com/journals/stbcontents/stb44.pdf*

Brockwell, S. E. & Gordon, I. R. (2001), 'A comparison of statistical methods for meta-analysis', *Statistics in Medicine* **20**(6), 825–840.
URL: *https://doi.org/10.1002/sim.650*

Brown, H. K. & Kempton, R. A. (1994), 'The application of REML in clinical trials', *Statistics in Medicine* **13**(16), 1601–1617.
URL: *https://doi.org/10.1002/sim.4780131602*

Bujkiewicz, S., Thompson, J. R., Sutton, A. J., Cooper, N. J., Harrison, M. J., Symmons, D. P. & Abrams, K. R. (2013), 'Multivariate meta-analysis of mixed outcomes: A Bayesian approach', *Statistics in Medicine* **32**(22), 3926–3943.
URL: *https://doi.org/10.1002/sim.5831*

Burke, D. L., Ensor, J. & Riley, R. D. (2017), 'Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ', *Statistics in medicine* **36**(5), 855–875.
URL: *https://doi.org/10.1002/sim.7141*

Burton, A., Altman, D. G., Royston, P. & Holder, R. L. (2006), 'The design of simulation studies in medical statistics', *Statistics in Medicine* **25**(24), 4279–4292.
URL: *http://doi.wiley.com/10.1002/sim.2673*

Buzzetti, E., Kalafateli, M., Thorburn, D., Davidson, B. R., Thiele, M., Gluud, L. L., Del Giovane, C., Askgaard, G., Krag, A., Tsochatzis, E. & Gurusamy, K. S. (2017), 'Pharmacological interventions for alcoholic liver disease (alcohol-related liver disease)', *Cochrane Database of Systematic Reviews* **3**.
URL: *http://doi.wiley.com/10.1002/14651858.CD011646.pub2*

Cannon, C. P., Gibson, C. M., McCabe, C. H., Adgey, A. A. J., Schweiger, M. J., Sequeira, R. F., Grollier, G., Giugliano, R. P., Frey, M., Mueller, H. S., Steingart, R. M., Weaver, W. D., Van de Werf, F. & Braunwald, E. (1998), 'TNK–Tissue Plasminogen Activator Compared With Front-Loaded Alteplase in Acute Myocardial Infarction: results of the TIMI 10B trial', *Circulation* **98**(25), 2805–2814.
URL: *https://doi.org/10.1161/01.CIR.98.25.2805*

Cantwell, R., Clutton-Brock, T., Cooper, G., Dawson, A., Drife, J., Garrod, D., Harper, A., Hulbert, D., Lucas, S., McClure, J., Millward-Sadler, H., Neilson, J., Nelson-Piercy, C., Norman, J., O'Herlihy, C., Oates, M., Shakespeare, J., de Swiet, M., Williamson, C., Beale, V., Knight, M., Lennox, C., Miller, A., Parmar, D., Rogers, J. & Springett, A. (2011), 'Saving Mothers' Lives: Reviewing maternal deaths to make motherhood safer: 2006-2008. The Eighth Report of the Confidential Enquiries into Maternal Deaths in the United Kingdom.', *BJOG : an international journal of obstetrics and gynaecology* **118**, 1–203.
URL: *http://doi.wiley.com/10.1111/j.1471-0528.2010.02847.x*

Chen, H., Manning, A. K. & Dupuis, J. (2012), 'A method of moments estimator for random effect multivariate meta-analysis', *Biometrics* **68**(4), 1278–84.
URL: *http://doi.wiley.com/10.1111/j.1541-0420.2012.01761.x*

Cochran, W. G. (1954), 'The combination of estimates from different experiments', *Biometrics* **10**(1), 101.
URL: *https://www.jstor.org/stable/3001666?origin=crossref*

Cochrane Collaboration UK (no date), 'Cochrane - About Us'.
URL: *https://www.cochrane.org/about-us*

Collins, G. S., de Groot, J. A., Dutton, S., Omar, O., Shanyinde, M., Tajar, A., Voysey, M., Wharton, R., Yu, L.-M., Moons, K. G. & Altman, D. G. (2014), 'External validation of multivariable prediction models: a systematic review of methodological conduct and reporting', *BMC Medical Research Methodology* **14**(1), 40.
URL: *https://doi.org/10.1186/1471-2288-14-40*

Cook, D. J., Mulrow, C. D. & Haynes, R. B. (1997), 'Systematic reviews: synthesis of best evidence for clinical decisions.', *Annals of internal medicine* **126**(5), 376–80.
URL: *http://annals.org/article.aspx?doi=10.7326/0003-4819-126-5-199703010-00006*

Copas, J. B. (1983), 'Regression, Prediction and Shrinkage', *Journal of the Royal Statistical Society. Series B (Methodological)* **45**(3), 311–354.
URL: *https://doi.org/10.1111/j.2517-6161.1983.tb01258.x*

Copas, J. B., Jackson, D., White, I. R. & Riley, R. D. (2018), 'The role of secondary outcomes in multivariate meta-analysis', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **67**(5), 1177–1205.
URL: *http://doi.wiley.com/10.1111/rssc.12274*

Cornell, J. E., Mulrow, C. D., Localio, R., Stack, C. B., Meibohm, A. R., Guallar, E. & Goodman, S. N. (2014), 'Random-effects meta-analysis of inconsistent effects: a time for change', *Annals of Internal Medicine* **160**(4), 267–270.
URL: *http://annals.org/article.aspx?doi=10.7326/M13-2886*

Courvoisier, D. S., Combescure, C., Agoritsas, T., Gayet-Ageron, A. & Perneger, T. V. (2011), 'Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure', *Journal of clinical epidemiology* **64**(9), 993–1000.
URL: *https://doi.org/10.1016/j.jclinepi.2010.11.012*

Cox, D. R. (1958), 'The regression analysis of binary sequences', *Journal of the Royal Statistical Society: Series B (Methodological)* **20**(2), 215–232.
URL: *http://doi.wiley.com/10.1111/j.2517-6161.1958.tb00292.x*

Cox, D. R. (1972), 'Regression models and life-tables', *Journal of the Royal Statistical Society: Series B (Methodological)* **34**(2), 187–202.
URL: *https://www.jstor.org/stable/2985181*

Danaher, P. J. & Smith, M. S. (2011), 'Modeling multivariate distributions using copulas: applications in marketing', *Marketing Science* **30**(1), 4–21.
URL: *https://doi.org/10.1287/mksc.1090.0491*

Debray, T. P. A., Moons, K. G. M., Abo-Zaid, G. M. A., Koffijberg, H. & Riley, R. D. (2013), 'Individual participant data meta-analysis for a binary outcome: one-stage or two-stage?', *PloS ONE* **8**(4), e60650.
URL: *https://doi.org/10.1371/journal.pone.0060650*

Debray, T. P. A., Moons, K. G. M., van Valkenhoef, G., Efthimiou, O., Hummel, N., Groenwold, R. H. H. & Reitsma, J. B. (2015*a*), 'Get real in individual participant data (IPD) meta-analysis: a review of the methodology', *Research Synthesis Methods* **6**(4), 293–309.
URL: *http://doi.wiley.com/10.1002/jrsm.1160*

Debray, T. P., Vergouwe, Y., Koffijberg, H., Nieboer, D., Steyerberg, E. W. & Moons, K. G. (2015*b*), 'A new framework to enhance the interpretation of external validation studies of clinical prediction models', *Journal of Clinical Epidemiology* **68**(3), 279–289.
URL: *http://dx.doi.org/10.1016/j.jclinepi.2014.06.018*

den Heijer, P., Vermeer, F., Ambrosioni, E., Sadowski, Z., López-Sendón, J. L., von Essen, R., Beaufils, P., Thadani, U., Adgey, J., Pierard, L., Brinker, J., Davies, R. F., Smalling, R. W., Wallentin, L., Caspi, A., Pangerl, A., Trickett, L., Hauck, C., Henry, D. & Chew, P. (1998), 'Evaluation of a weight-adjusted single-bolus plasminogen activator in patients with myocardial infarction: a double-blind, randomized angiographic trial of lanoteplase versus alteplase', *Circulation* **98**(20), 2117–2125.
URL: *https://www.ahajournals.org/doi/10.1161/01.CIR.98.20.2117*

DerSimonian, R. & Laird, N. (1986), 'Meta-analysis in clinical trials', *Controlled Clinical Trials* **7**(3), 177–188.
URL: *https://doi.org/10.1016/0197-2456(86)90046-2*

Dimier, N. & Todd, S. (2017), 'An investigation into the two-stage meta-analytic copula modelling approach for evaluating time-to-event surrogate endpoints which comprise of one or more events of interest', *Pharmaceutical Statistics* **16**(5), 322–333.
URL: *http://doi.wiley.com/10.1002/pst.1812*

Egger, M., Davey-Smith, G. & Altman, D. G. (2008), *Systematic Reviews in Health Care : Meta-Analysis in Context.*, Vol. 2nd ed, BMJ Books.

Eikelboom, J. W., Mehta, S. R., Pogue, J. & Yusuf, S. (2001), 'Safety outcomes in meta-analyses of phase 2 vs phase 3 randomized trials: Intracranial hemorrhage in trials of bolus thrombolytic therapy', *JAMA* **285**(4), 444–450.
URL: *http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.285.4.444*

Feinberg, J., Nielsen, E. E., Korang, S. K., Halberg Engell, K., Nielsen, M. S., Zhang, K., Didriksen, M., Lund, L., Lindahl, N., Hallum, S., Liang, N., Xiong, W., Yang, X., Brunsgaard, P., Garioud, A., Safi, S., Lindschou, J., Kondrup, J., Gluud, C. & Jakobsen, J. C. (2017), 'Nutrition support in hospitalised adults at nutritional risk', *Cochrane Database of Systematic Reviews* **5**.
URL: *http://doi.wiley.com/10.1002/14651858.CD011598.pub2*

Frison, L. J. & Pocock, S. J. (1997), 'Linearly divergent treatment effects in clinical trials with repeated measures: efficient analysis using summary statistics', *Statistics in Medicine* **16**(24), 2855–2872.
URL: *http://doi.wiley.com/10.1002/%28SICI%291097-0258%2819971230%2916%3A24%3C2855%3A%3AAID-SIM749%3E3.0.CO%3B2-Y*

Frosi, G., Riley, R. D., Williamson, P. R. & Kirkham, J. J. (2015), 'Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews', *Journal of Clinical Epidemiology* **68**(5), 542–550.
URL: *https://doi.org/10.1016/j.jclinepi.2014.11.017*

Fu, R. & Holmer, H. K. (2015), 'Change score or followup score? An empirical evaluation of the impact of choice of mean difference estimates', *Agency for Healthcare Research and Quality (US)* **April**, 1–45.
URL: *https://europepmc.org/books/NBK285983;jsessionid=F28D8F0CD1EB1F2C58836E0*

Fu, R. & Holmer, H. K. (2016), 'Change score or follow-up score? Choice of mean difference estimates could impact meta-analysis conclusions', *Journal of Clinical Epidemiology* **76**, 108–117.
URL: *http://dx.doi.org/10.1016/j.jclinepi.2016.01.034*

Fu, R., Vandermeer, B. W., Shamliyan, T. A., O'Neil, M. E., Yazdi, F., Fox, S. H. & Morton, S. C. (2013), 'Handling Continuous Outcomes in Quantitative Synthesis', *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* .
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24006546*

Genest, C. & Mackay, J. (1986), 'The joy of copulas: bivariate distributions with uniform marginals', *The American Statistician* **40**(4), 280–283.
URL: *http://www.tandfonline.com/doi/abs/10.1080/00031305.1986.10475414*

Glass, G. V. (1976), 'Primary, secondary, and meta-analysis of research', *Educational Researcher* **5**(10), 3–8.
URL: *http://journals.sagepub.com/doi/10.3102/0013189X005010003*

Hamza, T. H., Arends, L. R., van Houwelingen, H. C. & Stijnen, T. (2009), 'Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds', *BMC Medical Research Methodology* **9**(1), 73.
URL: *https://doi.org/10.1186/1471-2288-9-73*

Hardy, R. J. & Thompson, S. G. (1996), 'A likelihood approach to meta-analysis with random effects', *Statistics in Medicine* **15**(6), 619–629.
URL: *https://doi.org/10.1002/(SICI)1097-0258(19960330)15:6%3C619::AID-SIM188%3E3.0.CO;2-A*

Harris, R. J., Deeks, J. J., Altman, D. G., Bradburn, M. J., Harbord, R. M. & Sterne, J. A. C. (2008), 'Metan: Fixed- and Random-Effects Meta-Analysis', *The Stata Journal: Promoting communications on statistics and Stata* **8**(1), 3–28.
URL: *https://doi.org/10.1177/1536867X0800800102*

Hartung, J. & Knapp, G. (2001*a*), 'A refined method for the meta-analysis of controlled clinical trials with binary outcome', *Statistics in Medicine* **20**(24), 3875–3889.
URL: *http://doi.wiley.com/10.1002/sim.1009*

Hartung, J. & Knapp, G. (2001*b*), 'On tests of the overall treatment effect in meta-analysis with normally distributed responses', *Statistics in Medicine* **20**(12), 1771–1782.

Higgins, J. P. & Green, S. (2008), *Cochrane Handbook for Systematic Reviews of Interventions*, John Wiley & Sons, Ltd, Chichester, UK.
URL: *http://doi.wiley.com/10.1002/9780470712184*

Higgins, J. P. T. & Thompson, S. G. (2002), 'Quantifying heterogeneity in a meta-analysis', *Statistics in Medicine* **21**(11), 1539–1558.
URL: *http://doi.wiley.com/10.1002/sim.1186*

Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. (2003), 'Measuring inconsistency in meta-analyses', *BMJ* **327**(7414), 557–560.

Higgins, J. P. T., Thompson, S. G. & Spiegelhalter, D. J. (2009), 'A re-evaluation of random-effects meta-analysis', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(1), 137–159.

Higgins, J. P. T., Whitehead, A., Turner, R. M., Omar, R. Z. & Thompson, S. G. (2001), 'Meta-analysis of continuous outcome data from individual patients', *Statistics in Medicine* **20**(15), 2219–2241.
URL: *https://doi.org/10.1002/sim.918*

Jackson, D. & Riley, R. D. (2014), 'A refined method for multivariate meta-analysis and meta-regression', *Statistics in Medicine* **33**(4), 541–554.
URL: *https://doi.org/10.1002/sim.5957*

Jackson, D., Riley, R. & White, I. R. (2011), 'Multivariate meta-analysis: Potential and promise', *Statistics in Medicine* **30**(20), 2481–2498.
URL: *http://doi.wiley.com/10.1002/sim.4172*

Jackson, D., White, I. R., Price, M., Copas, J. & Riley, R. D. (2017), 'Borrowing of strength and study weights in multivariate and network meta-analysis', *Statistical Methods in Medical Research* **26**(6), 2853–2868.
URL: *http://journals.sagepub.com/doi/10.1177/0962280215611702*

Jackson, D., White, I. R. & Riley, R. (2013), 'A matrix based method of moments for fitting the random effects model for meta-analysis and meta-regression', *Biometrical Journal* **55**(2), 231–245.

Jackson, D., White, I. R. & Riley, R. D. (2012), 'Quantifying the impact of between-study heterogeneity in multivariate meta-analyses', *Statistics in Medicine* **31**(29), 3805–3820.
URL: *https://doi.org/10.1002/sim.5453*

Jackson, D., White, I. R. & Thompson, S. G. (2010), 'Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses', *Statistics in Medicine* **29**(12), 1282–1297.
URL: *http://doi.wiley.com/10.1002/sim.3602*

Jones, A. P., Riley, R. D., Williamson, P. R. & Whitehead, A. (2009), 'Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials', *Clinical trials* **6**(1), 16–27.
URL: *https://doi.org/10.1177/1740774508100984*

Kawai, C., Yui, Y., Hosoda, S., Nobuyoshi, M., Suzuki, S., Sato, H., Takatsu, F., Motomiya, T., Kanmatsuse, K., Kodama, K., Yabe, Y., Minamino, T., Kimata, S.-i. & Nakashima, M. (1997), 'A prospective, randomized, double-blind multicenter trial of a single bolus injection of the novel modified t-PA E6010 in the treatment of acute myocardial infarction: comparison with native t-PA', *Journal of the American College of Cardiology* **29**(7), 1447–1453.
URL: *https://doi.org/10.1016/S0735-1097(97)00074-0*

Kirkham, J. J., Riley, R. D. & Williamson, P. R. (2012), 'A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews', *Statistics in Medicine* **31**(20), 2179–2195.
URL: *https://doi.org/10.1002/sim.5356*

Knapp, G. & Hartung, J. (2003), 'Improved tests for a random effects meta-regression with a single covariate', *Statistics in Medicine* **22**(17), 2693–2710.
URL: *https://doi.org/10.1002/sim.1482*

Kontopantelis, E. & Reeves, D. (2010), 'metaan: Random-effects meta-analysis', *Stata Journal* **10**(3), 395–407.
URL: *http://www.stata-journal.com/article.html?article=st0201*

Kontopantelis, E. & Reeves, D. (2012*a*), 'Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A comparison between DerSimonian–Laird and restricted maximum likelihood', *Statistical Methods in Medical Research* **21**(6), 657–659.
URL: *http://journals.sagepub.com/doi/10.1177/0962280211413451*

Kontopantelis, E. & Reeves, D. (2012*b*), 'Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study', *Statistical Methods in Medical Research* **21**(4), 409–426.
URL: *https://doi.org/10.1177/0962280210392008*

Kulinskaya, E., Morgenthaler, S. & Staudte, R. G. (2008), *Meta analysis: a guide to calibrating and combining statistical evidence*, Vol. 756, John Wiley & Sons.

Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W. & Simmonds, M. (2019), 'A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses', *Research Synthesis Methods* **10**(1), 83–98.
URL: *http://doi.wiley.com/10.1002/jrsm.1316*

Ma, Y. & Mazumdar, M. (2011), 'Multivariate meta-analysis: A robust approach based on the theory of U-statistic', *Statistics in Medicine* **30**(24), 2911–2929.
URL: *https://doi.org/10.1002/sim.4327*

Masic, I., Miokovic, M. & Muhamedagic, B. (2008), 'Evidence Based Medicine - New Approaches and Challenges', *Acta Informatica Medica* **16**(4), 219.
URL: *http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3789163*

Mavridis, D. & Salanti, G. (2013), 'A practical introduction to multivariate meta-analysis', *Statistical Methods in Medical Research* **22**(2), 133–158.
URL: *https://doi.org/10.1177/0962280211432219*

McKenzie, J. E., Herbison, G. P. & Deeks, J. J. (2016), 'Impact of analysing continuous outcomes using final values, change scores and analysis of covariance on the performance of meta-analytic methods: a simulation study', *Research Synthesis Methods* **7**(4), 371–386.
URL: *https://doi.org/10.1002/jrsm.1196*

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P. & Stewart, L. A. (2015), 'Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement', *Systematic Reviews* **4**(1), 1.
URL: *https://doi.org/10.1186/2046-4053-4-1*

Mulrow, C. D. (1994), 'Systematic Reviews: Rationale for systematic reviews', *BMJ* **309**(6954), 597–599.
URL: *http://www.bmj.com/cgi/doi/10.1136/bmj.309.6954.597*

Nam, I. S., Mengersen, K. & Garthwaite, P. (2003), 'Multivariate meta-analysis', *Statistics in Medicine* **22**(14), 2309–2333.
URL: *https://doi.org/10.1002/sim.1410*

Nelsen, R. B. (2007), *An introduction to copulas*, Springer Science & Business Media.

Nikoloulopoulos, A. K. (2015), 'A mixed effect model for bivariate meta-analysis of diagnostic test accuracy studies using a copula representation of the random effects distribution', *Statistics in Medicine* **34**(29), 3842–3865.
URL: *https://doi.org/10.1002/sim.6595*

Normand, S.-L. T. (1999), 'Meta-analysis: formulating, evaluating, combining, and reporting', *Statistics in Medicine* **18**(3), 321–359.
URL: *https://doi.org/10.1002/(SICI)1097-0258(19990215)18:3%3C321::AID-SIM28%3E3.0.CO;2-P*

Novianti, P. W., Roes, K. C. B. & van der Tweel, I. (2014), 'Estimation of between-trial variance in sequential meta-analyses: A simulation study', *Contemporary Clinical Trials* **37**(1), 129–138.
URL: *http://dx.doi.org/10.1016/j.cct.2013.11.012*

Park, S.-J. (1998), 'Comparison of double bolus urokinase versus front-loaded alteplase regimen for acute myocardial infarction', *The American Journal of Cardiology* **82**(6), 811–813.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0002914998004445*

Peduzzi, P., Concato, J., Feinstein, A. R. & Holford, T. R. (1995), 'Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates', *Journal of clinical epidemiology* **48**(12), 1503–1510.
URL: *https://doi.org/10.1016/0895-4356(95)00048-8*

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. & Feinstein, A. R. (1996), 'A simulation study of the number of events per variable in logistic regression analysis', *Journal of clinical epidemiology* **49**(12), 1373–1379.
URL: *https://doi.org/10.1016/S0895-4356(96)00236-3*

Price, M. J., Blake, H., Kenyon, S., White, I. R., Jackson, D., Kirkham, J. J., Neilson, J., Deeks, J. J. & Riley, R. D. (2019), 'Empirical comparison of univariate and multivariate meta-analysis in cochrane pregnancy and childbirth reviews with multiple binary outcomes', *Research synthesis methods* .
URL: *https://doi.org/10.1002/jrsm.1353*

Raudenbush, S. W., Becker, B. J. & Kalaian, H. (1988), 'Modeling multivariate effect sizes', *Psychological Bulletin* **103**(1), 111–120.
URL: *http://dx.doi.org/10.1037/0033-2909.103.1.111*

Riley, R. D. (2004), Evidence synthesis of prognostic marker studies, PhD thesis, University of Leicester.

Riley, R. D. (2009), 'Multivariate meta-analysis: the effect of ignoring within-study correlation', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(4), 789–811.
URL: *http://doi.wiley.com/10.1111/j.1467-985X.2008.00593.x*

Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J. & Thompson, J. R. (2007*a*), 'An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes', *Statistics in Medicine* **26**(1), 78–97.
URL: *https://doi.org/10.1002/sim.2524*

Riley, R. D., Abrams, K. R., Sutton, A. J., Lambert, P. C. & Thompson, J. R. (2007*b*), 'Bivariate random-effects meta-analysis and the estimation of between-study correlation', *BMC Medical Research Methodology* **7**(1), 3.
URL: *https://doi.org/10.1186/1471-2288-7-3*

Riley, R. D., Higgins, J. P. T. & Deeks, J. J. (2011), 'Interpretation of random effects meta-analyses', *BMJ* **342**, d549.
URL: *http://www.bmj.com/cgi/doi/10.1136/bmj.d549*

Riley, R. D., Jackson, D., Salanti, G., Burke, D. L., Price, M., Kirkham, J. & White, I. R. (2017), 'Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples', *BMJ* **358**, j3932.
URL: *https://doi.org/10.1136/bmj.j3932*

Riley, R. D., Kauser, I., Bland, M., Thijs, L., Staessen, J. A., Wang, J., Gueyffier, F. & Deeks, J. J. (2013), 'Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data', *Statistics in Medicine* **32**(16), 2747–2766.
URL: *https://doi.org/10.1002/sim.5726*

Riley, R. D., Lambert, P. C. & Abo-Zaid, G. (2010), 'Meta-analysis of individual participant data: rationale, conduct, and reporting', *BMJ* **340**, c221.
URL: *http://www.bmj.com/cgi/doi/10.1136/bmj.c221*

Riley, R. D., Price, M. J., Jackson, D., Wardle, M., Gueyffier, F., Wang, J., Staesseng, J. A. & White, I. R. (2015), 'Multivariate meta-analysis using individual participant data', *Research Synthesis Methods* **6**(2), 157–174.
URL: *https://doi.org/10.1002/jrsm.1129*

Riley, R. D., Takwoingi, Y., Trikalinos, T., Guha, A., Biswas, A., Ensor, J., Morris, R. K. & Deeks, J. J. (2014), 'Meta-analysis of test accuracy studies with multiple and missing thresholds: a multivariate-normal model', *J Biomed Biostat* **5**, 196.
URL: *https://doi.org/10.4172/2155-6180.1000196*

Riley, R. D., Thompson, J. R. & Abrams, K. R. (2008), 'An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown', *Biostatistics* **9**(1), 172–186.
URL: *https://doi.org/10.1093/biostatistics/kxm023*

Rogozińska, E., D'Amico, M., Khan, K., Cecatti, J., Teede, H., Yeo, S., Vinter, C., Rayanagoudar, G., Barakat, R., Perales, M., Dodd, J., Devlieger, R., Bogaerts, A., van Poppel, M., Haakstad, L., Shen, G., Shub, A., Luoto, R., Kinnunen, T., Phelan, S., Poston, L., Scudeller, T., El Beltagy, N., Stafne, S., Tonstad, S., Geiker, N., Ruifrok, A., Mol, B., Coomarasamy, A. & Thangaratinam, S. (2016), 'Development of composite outcomes for individual patient data (IPD) meta-analysis on the effects of diet and lifestyle in pregnancy: a Delphi survey', *BJOG: An International Journal of Obstetrics & Gynaecology* **123**(2), 190–198.
URL: *http://doi.wiley.com/10.1111/1471-0528.13764*

Rogozińska, E., Marlin, N., Jackson, L., Rayanagoudar, G., Ruifrok, A. E., Dodds, J., Molyneaux, E., van Poppel, M. N., Poston, L., Vinter, C. A., McAuliffe, F., Dodd, J. M., Owens, J., Barakat, R., Perales, M., Cecatti, J. G., Surita, F., Yeo, S., Bogaerts, A., Devlieger, R., Teede, H., Harrison, C., Haakstad, L., Shen, G. X., Shub, A., Beltagy, N. E., Motahari, N., Khoury, J., Tonstad, S., Luoto, R., Kinnunen, T. I., Guelfi, K., Facchinetti, F., Petrella, E., Phelan, S., Scudeller, T. T., Rauh, K., Hauner, H., Renault, K., de Groot, C. J., Sagedal, L. R., Vistad, I., Stafne, S. N.,

Mørkved, S., Salvesen, K. Å., Jensen, D. M., Vitolo, M., Astrup, A., Geiker, N. R., Kerry, S., Barton, P., Roberts, T., Riley, R. D., Coomarasamy, A., Mol, B. W., Khan, K. S. & Thangaratinam, S. (2017), 'Effects of antenatal diet and physical activity on maternal and fetal outcomes: individual patient data meta-analysis and health economic evaluation', *Health Technology Assessment* **21**(41), 1–158.
URL: *https://www.journalslibrary.nihr.ac.uk/hta/hta21410*

Rosenberg, W. & Donald, A. (1995), 'Evidence based medicine: an approach to clinical problem-solving', *BMJ* **310**(6987), 1122–1126.
URL: *http://www.bmj.com/cgi/doi/10.1136/bmj.310.6987.1122*

Sackett, D. L. (1997), 'Evidence-based medicine', *Seminars in Perinatology* **21**(1), 3–5.
URL: *https://doi.org/10.1016/S0146-0005(97)80013-4*

Senn, S. (1993), 'Baseline distribution and conditional size', *Journal of Biopharmaceutical Statistics* **3**(2), 265–270.
URL: *https://doi.org/10.1080/10543409308835065*

Simel, D. L. & Bossuyt, P. M. M. (2009), 'Differences between univariate and bivariate models for summarizing diagnostic accuracy may not be large', *Journal of Clinical Epidemiology* **62**(12), 1292–1300.
URL: *http://dx.doi.org/10.1016/j.jclinepi.2009.02.007*

Smalling, R. W., Bode, C., Kalbfleisch, J., Sen, S., Limbourg, P., Forycki, F., Habib, G., Feldman, R., Hohnloser, S. & Seals, A. (1995), 'More rapid, complete, and stable coronary thrombolysis with bolus administration of reteplase compared with alteplase infusion in acute myocardial infarction', *Circulation* **91**(11), 2725–2732.

Sohn, S. Y. (2000), 'Multivariate meta analysis with potentially correlated marketing study results', *Naval Research Logistics* **47**(6), 500–510.
URL: *https://doi.org/10.1002/1520-6750(200009)47:6%3C500::AID-NAV3%3E3.0.CO;2-Z*

Steyerberg, E. (2008), *Clinical prediction models: a practical approach to development, validation, and updating*, Springer Science & Business Media.

Sweeting, M. J., Sutton, A. J. & Lambert, P. C. (2004), 'What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data', *Statistics in Medicine* **23**(9), 1351–1375.
URL: *https://doi.org/10.1002/sim.1761*

Tandon, S., Tudur-Smith, C., Riley, R. D., Boyd, M. T. & Jones, T. M. (2010), 'A Systematic Review of p53 as a Prognostic Factor of Survival in Squamous Cell Carcinoma of the Four Main Anatomical Subsites of the Head and Neck', *Cancer Epidemiology Biomarkers & Prevention* **19**(2), 574–587.
URL: *http://cebp.aacrjournals.org/cgi/doi/10.1158/1055-9965.EPI-09-0981*

Thangaratinam, S. & Jolly, K. (2010), 'Obesity in pregnancy: a review of reviews on the effectiveness of interventions', *BJOG: An International Journal of Obstetrics & Gynaecology* **117**(11), 1309–1312.
URL: *http://doi.wiley.com/10.1111/j.1471-0528.2010.02670.x*

The Fibrinogen Studies Collaboration (2009), 'Systematically missing confounders in individual participant data meta-analysis of observational cohort studies', *Statistics in medicine* **28**(8), 1218–1237.

Thompson, S. G. & Sharp, S. J. (1999), 'Explaining heterogeneity in meta-analysis: A comparison of methods', *Statistics in Medicine* **18**(20), 2693–2708.
URL: *https://doi.org/10.1002/(SICI)1097-0258(19991030)18:20%3C2693::AID-SIM235%3E3.0.CO;2-V*

Trikalinos, T. A., Hoaglin, D. C. & Schmid, C. H. (2013), 'Empirical and Simulation-Based Comparison of Univariate and Multivariate Meta-Analysis for Binary Outcomes Methods Research Report Empirical and Simulation-Based Comparison of Univariate and Multivariate Meta-Analysis for Binary Outcomes', *Agency for Healthcare Research and Quality (US)* **March**.
URL: *http://www.ncbi.nlm.nih.gov/books/NBK132562/*

Trikalinos, T. A., Hoaglin, D. C. & Schmid, C. H. (2014), 'An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes', *Statistics in Medicine* **33**(9), 1441–1459.
URL: *https://doi.org/10.1002/sim.6044*

Trikalinos, T. A. & Olkin, I. (2008), 'A method for the meta-analysis of mutually exclusive binary outcomes', *Statistics in Medicine* **27**(21), 4279–4300.
URL: *https://doi.org/10.1002/sim.3299*

Trowman, R., Dumville, J. C., Torgerson, D. J. & Cranny, G. (2007), 'The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study', *Journal of Clinical Epidemiology* **60**(12), 1229–1233.
URL: *https://doi.org/10.1016/j.jclinepi.2007.03.014*

Van Breukelen, G. J. P. (2006), 'ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies', *Journal of Clinical Epidemiology* **59**(9), 920–925.
URL: *https://doi.org/10.1016/j.jclinepi.2006.02.007*

van Houwelingen, H. C., Arends, L. R. & Stijnen, T. (2002), 'Advanced methods in meta-analysis: multivariate approach and meta-regression', *Statistics in Medicine* **21**(4), 589–624.
URL: *http://doi.wiley.com/10.1002/sim.1040*

Vanderschueren, S., Dens, J., Kerdsinchai, P., Desmet, W., Vrolix, M., De Man, F., Van den Heuvel, P., Hermans, L., Collen, D. & Van de Werf, F. (1997), 'Randomized coronary patency trial of double-bolus recombinant staphylokinase versus front-loaded alteplase in acute myocardial infarction', *American Heart Journal* **134**(2), 213–219.
URL: *https://doi.org/10.1016/S0002-8703(97)70127-3*

Vickers, A. J. & Altman, D. G. (2001), 'Statistics notes: Analysing controlled trials with baseline and follow up measurements.', *BMJ* **323**(7321), 1123–1124.
URL: *https://doi.org/10.1136/bmj.323.7321.1123*

Viechtbauer, W. (2005), 'Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model', *Journal of Educational and Behavioral Statistics* **30**(3), 261–293.
URL: *http://jeb.sagepub.com/cgi/doi/10.3102/10769986030003261*

Wang, J. G., Staessen, J. A., Franklin, S. S., Fagard, R. & Gueyffier, F. (2005), 'Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome', *Hypertension* **45**(5), 907–913.
URL: *https://doi.org/10.1161/01.HYP.0000165020.14745.79*

Wei, L. & Zhang, J. (2001), 'Analysis of Data with Imbalance in the Baseline Outcome Variable for Randomized Clinical Trials', *Drug Information Journal* **35**(4), 1201–1214.
URL: *https://doi.org/10.1177/009286150103500417*

Wei, Y. & Higgins, J. P. (2013*a*), 'Bayesian multivariate meta-analysis with multiple outcomes', *Statistics in Medicine* **32**(17), 2911–2934.
URL: *https://doi.org/10.1002/sim.5745*

Wei, Y. & Higgins, J. P. (2013*b*), 'Estimating within-study covariances in multivariate meta-analysis with multiple outcomes', *Statistics in Medicine* **32**(7), 1191–1205.
URL: *https://doi.org/10.1002/sim.5679*

White, I. R. (2011), 'Multivariate random-effects meta-regression: Updates to mvmeta', *The Stata Journal* **11**(2), 255–270.
URL: *https://doi.org/10.1177/1536867X1101100206*

Zhang, S., Paul, J., Nantha-Aree, M., Buckley, N., Shahzad, U., Cheng, J., DeBeer, J., Winemaker, M., Wismer, D., Punthakee, D., Avram, V. & Thabane, L. (2014), 'Empirical comparison of four baseline covariate adjustment methods in analysis of continuous outcomes in randomized controlled trials', *Clinical Epidemiology* **6**(1), 227–235.
URL: *https://doi.org/10.2147/CLEP.S56554*