

**USE OF BAYESIAN METHODS FOR THE
DESIGN, ANALYSIS AND SYNTHESIS OF
CLINICAL TRIALS**

**By
DANIELLE LISA BURKE**

**A thesis submitted to
The University of Birmingham
for the Degree of
DOCTOR OF PHILOSOPHY**

**School of Health and Population Sciences
The University of Birmingham
December 2014**

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

This thesis explores Bayesian methods for the statistical design, analysis and synthesis of clinical trials, and compares these with a frequentist approach in several settings to determine key differences, advantages and limitations.

A review of randomised trials indicates that Bayesian methods are rarely applied, but useful for making probability statements and incorporating prior evidence, especially in trials with small sample sizes. These advantages are illustrated in a trial in congenital lower urinary tract obstruction, which has few events but elicited prior distributions about treatment effect.

Bayesian methods are then developed for meta-analysis of phase II trials and multiple outcomes, with an emphasis on informing phase III trial decisions. A Bayesian random-effects logistic regression is advocated for meta-analysis of a binary outcome to account for all parameter uncertainty, and to derive prediction intervals for the treatment effect in a new phase III trial. Bayesian multivariate meta-analysis methods are then encouraged to make joint inferences across multiple outcomes and incorporate prior distributions for missing correlations. However, a simulation study identifies that external evidence or clinical guidance is needed to ensure appropriate prior distributions for between-study variances and correlations to avoid misleading results.

Researchers should thus consider Bayesian methods for clinical trials, but recognise potential difficulties when adopting the approach in practice.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors Cindy Billingham, Alan Girling and Richard Riley for their guidance and support throughout the last three years. Most of all, I wish I knew how to express how grateful and lucky I feel to have worked with Richard. Without him, I would not be where I am today. He inspired me to begin this career from those very first undergraduate lectures, and he guided me through to completion of this thesis. It hasn't been an easy journey but he has always been so dedicated to getting the best out of me by being continually positive of everything I do.

Also, thank you to Karla Hemming and Sylwia Bujkiewicz who provided helpful direction and feedback in some important stages of this thesis.

I would like to thank the MRC Midland Hub for Trials Methodology Research for providing a brilliant environment to do my thesis, whilst providing some fantastic opportunities for career development.

And then to my family and friends...where would I be without all of their love and support? To mum, thank you for your love and always reassuring me that I can do it; to Natalia, thank you for being the most supportive, kind, and loving sister I could ever ask for; and to my brothers...thank you for always making me laugh. Next to my friends - you have all been so wonderful and I could not have completed this without you.

Lastly to Kym - the most incredible friend - I dedicate this work to you. We started our friendship on the very first day of our PhDs together, and you have supported me through every single day until the end. We have laughed, cried, and everything in between. This experience would not have been the same without you, our cups of tea, our endless laughter, and, of course, your 'vault' full of treats.

CONTENTS

CHAPTER 1: INTRODUCTION.....	1
1.1 Overview of the thesis	1
1.2 Clinical trials	2
1.2.1 Phases of trials	3
1.2.2 The importance of randomisation.....	4
1.3 Bayesian statistical methods.....	5
1.3.1 Bayesian versus frequentist paradigm	6
1.3.2 Bayesian inference derived from Bayes theorem.....	6
1.3.3 The prior distribution	8
1.3.4 Schools of Bayesians	11
1.3.5 Markov chain Monte Carlo estimation method	12
1.3.6 Bayesian methods in clinical trials in practice	13
1.3.7 Simple hypothetical example of a Bayesian analysis of a clinical trial	14
1.4 Systematic review and meta-analysis	16
1.4.1 Fixed-effect meta-analysis.....	17
1.4.2 Random-effects meta-analysis	18
1.4.3 Example of a simple Bayesian and frequentist meta-analysis.....	20
1.5 Aims and structure for the thesis.....	23
1.5.1 Aims	23
1.5.2 Structure for the thesis.....	24
CHAPTER 2: THE USE OF BAYESIAN METHODS IN PUBLISHED RANDOMISED CONTROLLED TRIALS: A REVIEW OF LEADING MEDICAL JOURNALS.....	27
2.1 Introduction and aims	27
2.2 Methods of the review.....	28
2.2.1 Aims	28
2.2.2 Identifying a set of articles	29
2.2.3 Evaluation of articles.....	30
2.3 Results	32

2.3.1	Aim 1: Statistical analyses with a Bayesian approach	33
2.3.2	Aim 2: Sample size calculation in phase III randomised trials	40
2.4	Discussion	47
2.4.1	Overview of key findings	49
2.4.2	Limitations	52
2.5	Conclusions and rationale for remainder of thesis.....	53
CHAPTER 3: APPLICATION OF BAYESIAN METHODS TO A RANDOMISED		
CONTROLLED TRIAL IN A RARE DISEASE: THE PLUTO TRIAL.....		
3.1	Aims and objectives	55
3.2	Background to the trial	56
3.2.1	Fetal lower urinary tract obstruction	56
3.2.2	Inclusion criteria.....	57
3.2.3	Randomisation.....	58
3.2.4	Vesico-amniotic shunting	58
3.2.5	Conservative care.....	59
3.2.6	Failure to recruit.....	59
3.2.7	Outcome measures	59
3.2.8	Data.....	60
3.3	Methods.....	61
3.3.1	Proportional hazards assumption.....	62
3.3.2	Bayesian survival analysis with the poisson counting process notation.....	64
3.3.3	Prior distribution for the treatment effect estimate	66
3.3.4	Posterior probability estimates	70
3.3.5	Implementation	70
3.4	Results	71
3.4.1	Summary of data	71
3.4.2	Survival analyses.....	74
3.5	Discussion	82
3.5.1	Does the Bayesian framework provide any additional benefit in the PLUTO trial?	84

3.5.2	Limitations	86
3.6	Conclusion.....	86
CHAPTER 4: UNIVARIATE META-ANALYSIS OF PHASE II TRIALS		89
4.1	Context and objective	89
4.2	Data, outcomes and clinical objectives	90
4.3	Univariate fixed-effect meta-analysis methods using a frequentist approach	95
4.3.1	A general fixed-effect meta-analysis model (inverse variance)	95
4.3.2	Mantel-Haenszel method.....	96
4.3.3	Peto method.....	98
4.3.4	Fixed-effect logistic regression	99
4.4	Univariate random-effects meta-analysis methods using a frequentist approach .	100
4.4.1	A general random-effects meta-analysis model (inverse variance)	100
4.4.2	Examining heterogeneity in a meta-analysis of phase II trials	101
4.4.3	Random-effects logistic regression	102
4.5	Univariate meta-analysis methods using a Bayesian framework.....	103
4.5.1	Bayesian approach to inverse-variance meta-analysis methods.....	103
4.5.2	Bayesian approach to logistic regression meta-analysis model	104
4.5.3	Probability statements	105
4.5.4	Estimation and reporting of Bayesian models	105
4.5.5	Prior distribution for between-study variance	106
4.6	Deriving prediction intervals following a random-effects meta-analysis	107
4.6.1	Frequentist prediction intervals	108
4.6.2	Bayesian prediction interval for the true treatment effect in a new trial.....	109
4.6.3	Bayesian prediction interval for the treatment effect estimate in a trial with a given sample size	109
4.6.4	Comparison with subsequent phase III trials.....	112
4.7	Results I: Pooled results for the treatment effect.....	112
4.7.1	Original frequentist results of Eikelboom et al.	112
4.7.2	Heterogeneity results.....	113

4.7.3	Comparison of fixed-effect and random-effects meta-analysis pooled results for frequentist and Bayesian methods	116
4.8	Results II: Prediction intervals for the treatment effect in a new trial.....	126
4.8.1	Frequentist versus Bayesian prediction intervals	126
4.8.2	Bayesian predictive probability.....	127
4.8.3	Bayesian prediction intervals for estimates of treatment effect in a new trial with a given sample size.....	132
4.8.4	Comparison with subsequent phase III results.....	134
4.9	Results III: Assessment of publication bias	141
4.10	Results IV: What if we thought the phase II results have occurred by chance?	145
4.11	Results V: choice of prior distribution for between-study variance	149
4.12	Discussion.....	156
4.12.1	Is a meta-analysis useful to the phase III trial decision?.....	156
4.12.2	Advantages of Bayesian meta-analysis compared to meta-analysis in a frequentist framework.....	158
4.12.3	Quantifying heterogeneity in phase II meta-analysis	159
4.12.4	Recommendations for undertaking phase II meta-analyses of binary outcomes	159
4.12.5	Relevance of this work to recent meta-analyses of phase II trials	160
4.12.6	Further work	161
4.13	Conclusion	162
CHAPTER 5: MULTIVARIATE META-ANALYSIS OF PHASE II TRIALS WITH MULTIPLE OUTCOMES		163
5.1	Context and objective	163
5.2	Clinical data and multiple outcomes of interest	164
5.3	Multivariate random-effects meta-analysis methods using a Bayesian approach .	168
5.3.1	Random-effects multinomial regression analysis	168
5.3.2	A general bivariate normal random-effects meta-analysis model	169
5.3.3	Within-study and between-study correlation.....	171
5.3.4	Unknown within-study correlations.....	172
5.3.5	Probability statements.....	177

5.3.6	Estimation and reporting of models.....	178
5.3.7	Prediction intervals following a Bayesian bivariate random-effects meta-analysis	178
5.4	Application of the multivariate meta-analysis methods.....	179
5.4.1	Primary analysis: mortality and stroke.....	180
5.4.2	Secondary analysis of reinfarction and mortality	186
5.5	Discussion	189
5.5.1	What are the additional benefits of multivariate meta-analysis of phase II trials in comparison to univariate methods?	190
5.5.2	Is the multivariate approach appropriate in the setting of phase II trials?	191
5.5.3	Limitations and further work.....	192
5.6	Conclusion and recommendations.....	193
CHAPTER 6: A SIMULATION STUDY TO ASSESS THE CHOICE OF PRIOR DISTRIBUTION FOR THE BETWEEN-STUDY CORRELATION IN A BAYESIAN BIVARIATE META-ANALYSIS		
197		
6.1	Introduction and aims	197
6.2	Bivariate random-effects meta-analysis model.....	199
6.2.1	Prior distributions for between-study correlation	200
6.3	Motivating examples	204
6.3.1	Motivating example A	205
6.3.2	Motivating example B	209
6.3.3	Summary of key findings from motivating examples	211
6.4	Methods of the simulation study.....	225
6.4.1	How Riley et al. generated bivariate meta-analysis data for 50 studies.....	226
6.4.2	Step-by-step guide to new simulation study	228
6.4.3	Choice of prior distribution for between-study correlation.....	230
6.4.4	Choice of prior distribution for between-study variance.....	231
6.4.5	Summary of all the settings and prior distributions to be evaluated in the simulation study of bivariate meta-analysis	232
6.5	Results of the bivariate simulation study with complete data.....	233

6.5.1	Results when using prior distribution for between-study variance of $\tau_j \sim N(0,2)/(0,.)$	233
6.5.2	Results when prior distribution for between-study variance is $1/\tau_j^2 \sim \text{gamma}(0.1,0.1)$	239
6.6	Impact of prior distribution for ρ_B when there is missing outcome data.....	245
6.6.1	Results with missing data when prior distribution for τ_j is $N(0,2)/(0,.)$	246
6.6.2	Results with missing data when prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$	250
6.7	The number of trials per meta-analysis and the influence of the prior distribution for the between-study correlation on its posterior distribution	253
6.8	What if the between-study correlation is known?	256
6.9	Discussion	259
6.9.1	Overview of key findings	259
6.9.2	Recommendations for the choice of prior distributions in future bivariate meta-analyses.....	264
6.9.3	Limitations	266
6.10	Conclusion	267
CHAPTER 7: DISCUSSION.....		269
7.1	Overview of thesis	269
7.2	Key findings and recommendations	272
7.2.1	The use of Bayesian methods for the analysis of randomised clinical trials was rare in general medical journal articles	272
7.2.2	Clinical implications were identified in some specific Bayesian analyses	273
7.2.3	Bayesian methods were identified as important in the meta-analysis of phase II trials to inform phase III trial decisions.....	274
7.3	Bayesian methods in the context of current clinical research: should they be used?..	282
7.3.1	Where is the Bayesian approach most justifiable?	282
7.3.2	Will there be acceptance of Bayesian results if they depend on the choice of prior distributions?	284
7.3.3	What if a Bayesian approach is not chosen?	285
7.4	Future research needs.....	287

7.5	Final conclusions	291
APPENDIX	293
APPENDIX A:	Total articles included in the review (314 articles)	293
APPENDIX B	323
Appendix B1:	WinBUGS code for univariate random-effects meta-analysis using a logistic regression model.....	323
Appendix B2:	History and trace plots to display convergence of parameters from a univariate random-effects logistic regression model.....	324
Appendix B3:	Publication in <i>Trials</i> arising from the univariate meta-analysis of phase II trials that was explored in Chapter 4.	326
APPENDIX C	327
Appendix C1:	WinBUGS code for bivariate random-effects normal meta-analysis model (5.1).	327
Appendix C2:	WinBUGS code for bivariate random-effects normal meta-analysis model (5.2).	329
Appendix C3:	Results for secondary bivariate meta-analysis of reinfarction and mortality.	331
APPENDIX D	333
Appendix D1:	SAS code to fit Bayesian bivariate random-effects meta-analysis on 1000 simulated datasets using MCMC.....	333
Appendix D2:	Results of simulation study with various prior distributions for between-study variance in univariate meta-analysis model	336
LIST OF REFERENCES	339

LIST OF FIGURES

Figure 1.1: Hypothetical example of a conjugate analysis.	16
Figure 1.2: Example meta-analysis using a Bayesian random-effects meta-analysis.	23
Figure 2.1: Flow diagram of journal articles for inclusion in literature review.	34
Figure 2.2: Key findings from the review of articles in medical journals.	48
Figure 3.1: Log cumulative hazard plot for intervention arm under ITT analysis.	63
Figure 3.2: Sceptical and enthusiastic prior distributions for log(HR) and HR for VAS compared to conservative care.	67
Figure 3.3: Elicited prior distribution for log(HR) for intervention compared to conservative treatment.	69
Figure 3.4: Kaplan-Meier plot of survival time from conception by intervention arm.	74
Figure 3.5: Posterior distributions for HRs using all prior distributions and assuming proportional hazard rates.	77
Figure 3.6: Prior and posterior distributions overlaid.	78
Figure 3.7: Prior and posterior distributions for HR1 (period conception to (expected time of birth) 36.5 weeks) overlaid.	81
Figure 3.8: Prior and posterior distributions for HR2 (period 36.5 weeks (expected time of birth) to the end of follow-up) overlaid.	82
Figure 3.9: Key findings.	83
Figure 4.1: Forest plot for intracranial haemorrhage using the fixed-effect Peto meta-analysis model.	121
Figure 4.2: Forest plot for stroke using the fixed-effect Peto meta-analysis model.	121
Figure 4.3: Forest plot for reinfarction using the fixed-effect Peto meta-analysis model.	122
Figure 4.4: Forest plot for mortality using the fixed-effect Peto meta-analysis model.	122
Figure 4.5: Prediction intervals for <i>estimated</i> treatment effect in a new trial (Y_{new}) with ICH for given sample size assuming option (1).	132
Figure 4.6: Probability that the upper bound of the estimated OR's 95% credible interval will be less than 1 in a new trial for ICH with a given sample size.	134
Figure 4.7: Phase II and phase III meta-analyses comparison plot by Eikelboom et al. ¹⁶⁹	135
Figure 4.8: Comparison of prediction intervals for θ_{new} from a phase II meta-analysis with the estimate and 95% CI for the ln(OR) from phase III trials for ICH.	139

Figure 4.9: Comparison of prediction intervals for θ_{inew} from a phase II meta-analysis with the estimate and 95% CI for the $\ln(\text{OR})$ from phase III trials for stroke.	139
Figure 4.10: Comparison of prediction intervals for θ_{inew} from a phase II meta-analysis with the estimate and 95% CI for $\ln(\text{OR})$ from phase III trials for reinfarction.	140
Figure 4.11: Comparison of prediction intervals for θ_{inew} from a phase II meta-analysis with the estimate and 95% CI for the $\ln(\text{OR})$ from phase III trials for mortality.	140
Figure 4.12: Funnel plot of $1/\text{SE}(\log(\text{OR}))$ versus OR for ICH for phase II and phase III trials.	142
Figure 4.13: Funnel plot of $1/\text{SE}(\log(\text{OR}))$ vs. $\log(\text{OR})$ for ICH in phase II trials.	143
Figure 4.14: Funnel plot of $1/\text{SE}(\log(\text{OR}))$ vs. $\log(\text{OR})$ for stroke in phase II trials.	143
Figure 4.15: Funnel plot of $1/\text{SE}(\log(\text{OR}))$ vs. $\log(\text{OR})$ for reinfarction in phase II trials.	144
Figure 4.16: Funnel plot of $1/\text{SE}(\log(\text{OR}))$ vs. $\log(\text{OR})$ for mortality in phase II trials.	144
Figure 4.17: Posterior distributions for the treatment effect ($\log(\text{OR})$) in a new trial for reinfarction assuming a sceptical and vague prior distribution for the pooled average treatment effect.	148
Figure 4.18: Posterior distribution for the true treatment effect in a new trial (θ_{inew}) for ICH assuming two different prior distributions for the heterogeneity parameter.	150
Figure 4.19: Key findings.	157
Figure 5.1: Scatter plot of treatment effect estimates ($\log\text{OR}$) for stroke and mortality in each trial.	174
Figure 5.2: Key findings.	189
Figure 5.3: Key recommendations.	195
Figure 6.1: Density plots for prior distributions for between-study correlation:	204
Figure 6.2: Key findings.	260
Figure 6.3: Key recommendations.	265
Figure 7.1: Key methodological recommendations.	281
Figure 7.2: Key future research needs.	288

LIST OF TABLES

Table 1.1: Bayesian conjugate models.	9
Table 1.2: Example dataset of the estimated mean difference from ten studies.	21
Table 1.3: Pooled summary results of hypothetical meta-analysis.	22
Table 2.1: Summary table for articles with Bayesian analysis (Aim 1).	39
Table 2.2: Summary table for journal articles included in Aim 2 (sample size evaluation in randomised phase III trials in the Lancet).	41
Table 3.1: Dataset variables with descriptions.	61
Table 3.2: Prior distributions for log(HR) for intervention versus conservative care.	69
Table 3.3: Summary of key baseline characteristics, outcome and length of follow-up by randomised intervention group.	72
Table 3.4: Bayesian and frequentist results from a univariable Cox model assuming proportional hazards for the entire length of follow-up.	75
Table 3.5: Posterior estimates of unadjusted HR for death given time-dependent treatment effect and Bayesian p-values.	79
Table 4.1: Phase II randomised trials of bolus vs. infusion thrombolytic therapy in AMI.	92
Table 4.2: Incidence of ICH, stroke, reinfarction, and mortality in nine phase II trials of bolus vs. infusion thrombolytic therapy in AMI. ¹⁷¹⁻¹⁷⁹	93
Table 4.3: Event in study <i>i</i>	96
Table 4.4: An illustration of a trial in which a continuity correction may be required to calculate the odds ratio.	98
Table 4.5: Sensitivity analysis prior distributions for between-study variance.	107
Table 4.6: Heterogeneity measures for ICH, stroke, reinfarction, and mortality.	114
Table 4.7: Pooled treatment effect results for bolus versus infusion therapy for intracranial haemorrhage.	117
Table 4.8: Pooled treatment effect results for bolus versus infusion therapy for stroke.	118
Table 4.9: Pooled treatment effect results for bolus versus infusion therapy for reinfarction.	119
Table 4.10: Pooled treatment effect results for bolus versus infusion therapy for mortality.	120
Table 4.11: Prediction results for ICH after the random-effects meta-analyses.	128
Table 4.12: Prediction results for stroke after the random-effects meta-analyses.	129

Table 4.13: Prediction results for reinfarction after the random-effects meta-analyses.....	130
Table 4.14: Prediction results for mortality after the random-effects meta-analyses.....	131
Table 4.15: Phase III randomised trials of bolus vs. infusion thrombolytic therapy in AMI. .	137
Table 4.16: Results from subsequent randomised phase III trials comparing bolus to infusion thrombolytic therapy for the in-hospital treatment of AMI.....	138
Table 4.17: Accounting for potential optimism in phase II trials in Bayesian meta-analysis: results for reinfarction.....	146
Table 4.18: Results of sensitivity analysis on prior distribution for between-study variance for pooled results for mortality using random-effects logistic regression model.....	152
Table 4.19: Results of sensitivity analysis on prior distribution for between-study variance for prediction intervals for true treatment effect for mortality using random-effects logistic regression model.....	153
Table 4.20: Results of sensitivity analysis on prior distribution for between-study variance for pooled results for ICH using random-effects logistic regression model.....	154
Table 4.21: Results of sensitivity analysis on prior distribution for between-study variance for prediction intervals for true treatment effect for ICH using random-effects logistic regression model.....	155
Table 4.22: Recommendations for improved meta-analysis of phase II trials of binary outcomes.	160
Table 5.1: Incidence of stroke, reinfarction, and mortality in nine phase II trials of bolus vs. infusion thrombolytic therapy in AMI. ¹⁷¹⁻¹⁷⁹	166
Table 5.2: Probability of observed event for four mutually exclusive and exhaustive responses in study, i , for treatment group, l	169
Table 5.3: Pearson correlation coefficients for the estimates of log OR for stroke, reinfarction, and mortality.	174
Table 5.4: Univariate and bivariate random-effects meta-analysis results for bolus versus infusion therapy for stroke and mortality.....	181
Table 5.5: 95% prediction intervals and posterior probability estimates for stroke and mortality from model (5.1).	184
Table 5.6: Joint posterior probability estimates for stroke and mortality.....	186
Table 5.7: Univariate and bivariate random-effects meta-analysis results for bolus versus infusion therapy for reinfarction and mortality.	188

Table 6.1: Summary results for the 10 trials in the meta-analysis by Wang et al. ²²⁴ for systolic blood pressure and diastolic blood pressure.	206
Table 6.2: Summary results for the 10 trials in the meta-analysis by Riley et al. ²²⁴ for partially-adjusted and fully-adjusted log hazard ratios (log HR).	210
Table 6.3: Motivating example A - summary results from bivariate meta-analysis for various prior distributions for ρ_B and τ_j	213
Table 6.4: Motivating example B - summary results from bivariate meta-analysis for various prior distributions for ρ_B and τ_j	218
Table 6.5: All combinations of prior distributions for between-study correlation and between-study variance.	233
Table 6.6: Simulation results for 10 studies with complete data (setting 1). The within-study correlation, ρ_{Wi} was zero and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$	234
Table 6.7: Simulation results for 10 studies with complete data (setting 2). The within-study correlation, ρ_{Wi} was zero and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$	235
Table 6.8: Simulation results for 10 studies with complete data (setting 3). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$	236
Table 6.9: Simulation results for 10 studies with complete data (setting 4). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$	237
Table 6.10: Simulation results for 10 studies with complete data (setting 1). The within-study correlation, ρ_{Wi} was zero and the same for each study. The prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$ and for β_j is $N(0,1000^2)$	241
Table 6.11: Simulation results for 10 studies with complete data (setting 2). The within-study correlation, ρ_{Wi} was zero and the same for each study. The prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$ and for β_j is $N(0,1000^2)$	242
Table 6.12: Simulation results for 10 studies with complete data (setting 3). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$ and for β_j is $N(0,1000^2)$	243
Table 6.13: Simulation results for 10 studies with complete data (setting 4). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$ and for β_j is $N(0,1000^2)$	244

Table 6.14: Simulation results for 10 studies with missing data for outcome 1 (setting 3). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$	248
Table 6.15: Simulation results for 10 studies with missing data for outcome 1 (setting 4). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$	249
Table 6.16: Simulation results for 10 studies with missing data for outcome 1 (setting 3). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$ and for β_j is $N(0,1000^2)$	251
Table 6.17: Simulation results for 10 studies with missing data for outcome 1 (setting 4). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$ and for β_j is $N(0,1000^2)$	252
Table 6.18: Simulation results for 25 studies with complete data (setting 4). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$	254
Table 6.19: Simulation results for 50 studies with complete data (setting 4). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$	255
Table 6.20: Simulation results for fixed values of parameters with 10 studies with complete data (setting 4). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. In all analyses, ρ_B is fixed equal to 0.8.	258

Abbreviations

CI	Confidence interval
CrI	Credible interval
DBP	Diastolic blood pressure
ICH	Intracranial haemorrhage
IPD	Individual patient data
HR	Hazard ratio
OR	Odds ratio
MCMC	Markov chain Monte Carlo
MI	Myocardial infarction
ML	Maximum likelihood
MM	Method-of-moments
MSE	Mean-square error
RCT	Randomised controlled trial
REML	Restricted maximum likelihood
SBP	Systolic blood pressure
S.E	Standard error

CHAPTER 1: INTRODUCTION

1.1 Overview of the thesis

A fundamental component of medical research is the evaluation of new treatments and interventions using randomised clinical trials. Such trials enrol patients with a particular disease or condition, and typically randomise each patient to receive either a new treatment or an existing treatment (a control). Outcomes are then recorded for each individual over time, and the hope is that patients receiving the new treatment have improved outcomes compared to those in the control group. Medical statistics is a key part of randomised trials, and statisticians play a crucial role in their implementation: from the design and sample size, to the collection and analysis of data, and to the interpretation of results. Traditionally, the analysis of randomised trials has been undertaken in a so-called frequentist framework, which produces results (such as p-values and confidence intervals for the treatment effect) by viewing the trial data as from an experiment that can theoretically be repeated identically many times. However, an alternative approach is a Bayesian analysis framework, which takes a more subjective view by combining the trial data with any prior information, and does not require inferences to depend on long-run, repeated sampling theory. The Bayesian approach thereby allows direct probabilistic inferences about the treatment effect (such as the probability the treatment is effective), and - by combining prior evidence with the new data - many consider it to be more relevant to clinical decision making.

The primary aim for this thesis is to apply and develop Bayesian methods for the design, analysis, and synthesis of randomised clinical trials, and to compare their potential benefit to standard frequentist methods. In particular, this thesis focuses on the *application* of Bayesian methods in trials with small numbers of events, or the synthesis of small numbers of trials. This includes trials in rare diseases, where it may be difficult to recruit large sample sizes, or

early phase trials, which typically have small numbers of patients. Furthermore, in terms of the *development* of Bayesian methods, this thesis proposes and evaluates novel Bayesian approaches for meta-analysis in the context of phase II trials and multiple outcomes, especially for obtaining results that inform subsequent phase III decisions. An underlying theme throughout all the projects is how the results of interest depend on the choice of prior distributions for unknown parameters, and an extensive simulation study is undertaken to evaluate this issue.

The remainder of this chapter introduces in detail the key topics of this thesis, and summarises the core statistical methods and concepts that form the foundation of the work undertaken. The aims of the thesis are then listed and the remaining chapters are signposted.

1.2 Clinical trials

Medical research often occurs within a clinical trial (or an intervention study), which is a study in which participants receive specific interventions according to a research plan or protocol that was created by the study investigators.¹ The intervention may be a medical product, such as a drug or device or changes to a participants' behaviour, for example their diet. The new intervention must be compared to a control group, which might be the standard intervention that is already available, or a placebo that contains no active ingredients, or a historical control. At the first stage of development of a new intervention, it is often not known whether it will be beneficial, harmful, or no different to the interventions that are already available. Therefore, a clinical trial is able to determine the safety and efficacy of the intervention by measuring certain outcomes in the participants.

1.2.1 Phases of trials

In pharmaceutical development of a new drug, it is common to divide the trials into four phases, although there is commonly an overlap of intentions in practice. The earliest stage of drug development in humans occurs in phase I (or phase 1) trials. These trials may be randomised or uncontrolled, and they aim to assess the safety and determine appropriate doses of a drug.² They are commonly conducted with few subjects and may be done on healthy individuals, or patients with the target disease. For example, Desfrere et al.³ conducted a dose-finding phase I trial to determine the dose regimen of intravenous ibuprofen for the treatment of patent ductus arteriosus in preterm infants.

Phase II (or phase 2) trials continue to assess the dose level and safety of the new drug, but these trials are usually small with the intention of obtaining some early indication of a possible treatment benefit. Phase II trials may be uncontrolled and the results are compared to historical controls. For example, Foss et al. conducted a single-arm phase II trial to determine the safety and efficacy of a combination therapy in patients with newly diagnosed peripheral T-cell lymphoma.⁴ There are however advantages of conducting a *randomised* phase II trial, rather than a single arm trial, such as a reduction in selection bias.^{5,6} Also, a primary analysis may identify a statistically significant difference and, although this result should always be interpreted with caution because the sample size is small, it may help to identify the treatments that are more likely to show a benefit in a later trial. For example, Byrd et al. conducted a randomised phase II study of fludarabine with concurrent versus sequential treatment with rituximab in symptomatic, untreated patients with B-cell chronic lymphocytic leukemia.⁷ The findings of this study were in favour of the new therapy and the authors were able to recommend further studies.

The next stage of drug development is a phase III (or phase 3) trial, which is typically a large randomised trial with either a standard therapy arm or a placebo arm that aims to obtain an estimate of treatment effect. If there is evidence of a clinically and significantly important benefit of the new therapy at this stage (and evidence of its safety and cost-effectiveness have also been obtained) then this new drug may be licensed for use. An example of a large phase III trial is the ISIS-2 trial that assessed the long-term survival of patients with suspected acute myocardial infarction, and compared intravenous streptokinase, oral aspirin, both, or neither.⁸ This trial randomised 17,187 patients in 417 hospitals in 16 countries and concluded that there are survival benefits produced by intravenous therapy for at least 10 years. The first three stages of drug development can take up to 10 years for a successful new therapy.⁹

Phase IV (or phase 4) trials are usually post-marketing surveillance studies that continue to monitor the safety of the drug, which can be removed from the market if there are adverse events. For example, Aprotinin has been shown to have anti-inflammatory responses. However, Schmartz et al. assessed the effect of Aprotinin on the inflammatory reaction to cardiopulmonary bypass in a randomised phase IV trial.¹⁰

1.2.2 The importance of randomisation

This thesis focuses primarily on phase II and III *randomised* trials, where the aim is to quantify the treatment effect for either informing subsequent phase III research, or for determining whether the treatment should be recommended for clinical practice. Randomisation is fundamental for the quantification of treatment effects because it reduces selection bias by ensuring that there are no systematic differences between the treatment groups in factors, which are known or unknown that may affect the outcome.¹¹ In other

words, randomisation theoretically removes the potential for comparisons of treatment groups to be affected by confounding factors.

1.3 Bayesian statistical methods

In all phases of clinical trials there is a need for robust statistical methods. For example, statistical methods are required to determine the appropriate dose-level of a drug, to estimate a sample size for a new study that is large enough to detect the clinically-important difference between therapies, or to estimate the treatment effects and account for all uncertainty. Another important aspect is the synthesis of clinical trial results when there are multiple trials in a similar disease area, or that compared a similar therapy, and meta-analysis is required to produce evidence-based summary results.

In this thesis, the focus of the statistical methods to analyse and synthesise clinical trials is largely within a Bayesian framework. Methods within the Bayesian framework have been extensively developed¹² in several areas including dose-finding trials with methods such as the continual reassessment method,^{13,14} interim analysis of clinical trials,^{15,16} and meta-analysis.¹⁷⁻²⁰ Chapter 2 of this thesis will explore where Bayesian statistical methods are currently being used in practice and determine where there may be some benefit of the Bayesian approach in specific settings, which then stimulates work in Chapters 3 to 6.

The following sections describe the essential elements of the Bayesian framework. These include a fundamental introduction to Bayesian inference, a discussion of the types of prior distribution, and estimation of parameters using Markov chain Monte Carlo methods. These are introduced within two simple examples: an analysis of a randomised clinical trial, and a random-effects meta-analysis.

1.3.1 Bayesian versus frequentist paradigm

When given a set of data for analysis, a key difference between Bayesian and frequentist approaches to statistical inference is the view of probability. The Bayesian approach assumes that an unknown parameter (for example, a treatment effect) of interest, θ , is fixed but our beliefs about it are expressed as a subjective probability distribution, which allows subjective, direct probabilistic inferences to be made about the magnitude of θ from the data at hand and other external information. In contrast, the frequentist approach considers θ to be a fixed, unvarying quantity, and probabilistic inferences about θ arise by observing the observed frequency of events in the data at hand, and assuming that it was obtained from a process that could be infinitely and identically repeated.²¹ Another important difference is that the Bayesian approach acknowledges prior probability information about the unknown parameter of interest, which is incorporated as a prior distribution.²² This prior distribution can be derived in numerous different ways, such as from evidence in previous clinical trials or from clinical opinion. The prior distribution is a topic of much controversy and is therefore discussed in detail in section 1.3.3.

1.3.2 Bayesian inference derived from Bayes theorem

Bayesian inference is based on a paper that was published posthumously by Thomas Bayes,²³ which introduced Bayes' theorem, and this is expressed mathematically in equation (1.1) for making inferences about some unknown parameter of interest, θ , after observing data, y :

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (1.1)$$

In this equation, $p(\theta|y)$ is called the posterior probability distribution of θ having observed the data y ; $p(y|\theta)$ is the likelihood function (which summarises the observed data, and is the

focus of the frequentist approach); and $p(\theta)$ is the prior probability distribution for θ . $P(y)$ is known as the normalising factor and its value is not usually of interest since it does not contain θ . Therefore, equation (1.1) is often written as:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad (1.2)$$

When making inferences from the posterior distribution, this usually involves estimating features such as the mean and variance. These can be expressed in terms of posterior expectations of functions of θ , $g(\theta)$, given by:²⁴

$$E[g(\theta)|y] = \int g(\theta)p(\theta|y)d\theta = \int \frac{g(\theta)p(y|\theta)p(\theta)}{p(y)}d\theta \quad (1.3)$$

The mean and variance of the posterior distribution for θ are shown in equations (1.4) and (1.5), respectively.

$$E[\theta|y] = \int \theta p(\theta|y)d\theta = \int \frac{\theta p(y|\theta)p(\theta)}{p(y)}d\theta \quad (1.4)$$

$$\begin{aligned} \text{Var}[\theta|y] &= E[\theta^2|y] - (E[\theta|y])^2 \quad (1.5) \\ &= \int \frac{\theta^2 p(y|\theta)p(\theta)}{p(y)}d\theta - \left(\int \frac{\theta p(y|\theta)p(\theta)}{p(y)}d\theta \right)^2 \end{aligned}$$

Unlike frequentist statistics, within the Bayesian framework it is possible to make direct probability statements about the posterior probability of θ , relative to some value of clinical interest, x . For example:

$$p(\theta < x|y) = \int_{-\infty}^x p(\theta|y)d\theta = \int_{-\infty}^x \frac{p(y|\theta)p(\theta)}{p(y)}d\theta \quad (1.6)$$

Similar to confidence intervals in a frequentist framework, the Bayesian posterior distribution can be used to derive a $100(1-\alpha)\%$ (often $\alpha=0.05$) credible interval (CrI), which gives an interval in which θ lies with probability $(1-\alpha)$. If the credible interval is a two-sided interval with equal probability in each tail, then the lower and upper bounds of the interval (θ_L, θ_U) can be estimated such that:

$$p(\theta < \theta_L | y) = \int_{-\infty}^{\theta_L} \frac{p(y|\theta)p(\theta)}{p(y)} d\theta = \frac{\alpha}{2} \quad (1.7)$$

$$p(\theta > \theta_U | y) = \int_{\theta_U}^{\infty} \frac{p(y|\theta)p(\theta)}{p(y)} d\theta = \frac{\alpha}{2}$$

Often, rather than reporting the equal-tail credible interval, instead the highest posterior density (HPD) interval is used. The HPD interval is the interval within which the values contained have the highest posterior densities.²⁵ In this thesis, however, only equal-tailed credible intervals will be reported.

1.3.3 The prior distribution

In a Bayesian framework, in order to obtain posterior distributions of parameters of interest, prior distributions must be specified for all unknown parameters. One advantage of this framework is that it allows for all external prior information to be incorporated into the analysis alongside the data. In terms of specifying a prior distribution, for convenience in the calculations, it is often preferable to choose a prior distribution that is 'conjugate' to the likelihood function. This means that when the prior distribution is combined with the likelihood using Bayes theorem the resulting posterior distribution will be from the same family as the prior distribution.²¹ The most common conjugate models are shown below in Table 1.1.

Table 1.1: Bayesian conjugate models.

Prior	Likelihood	Posterior
Normal	Normal	Normal
Beta	Binomial	Beta
Dirichlet	Multinomial	Dirichlet
Gamma	Poisson	Gamma

An aspect of the Bayesian framework that causes concern for frequentist statisticians is that there is no ‘correct’ prior distribution for an unknown parameter. There may be a range of sources from which the prior distribution can be derived and varying levels of uncertainty that could be considered. The sources of elicitation may include subjective belief, or evidence from previous objective studies, such as previous trials in a similar disease area, earlier phase trials, observational studies, or meta-analyses.²² Prior distributions that are derived from subjective belief are frequently gained from experienced individuals, such as specialist clinicians in the area of research.²⁶ These priors are often heavily criticised by frequentist statisticians because they may be biased by the individuals’ opinion, the choice of individual and the timing of the elicitation.²² If multiple subjective beliefs have been derived from several specialists then the prior distributions may be combined to form one subjective prior distribution. This may be fairly straightforward if the specialists have similar opinions, but perhaps it would not be sensible to combine the distributions if there are opposing views because the resulting prior distribution may not represent any one of the individual opinions.

A particular range of prior distributions are sometimes called ‘default’ or ‘off-the-shelf’ priors, which include ‘sceptical’ and ‘enthusiastic’ priors. If, for example, the analysis is of a randomised clinical trial, a sceptical prior that does not allow for large treatment differences may be used to limit the potential optimism. The derivation of this prior distribution must be guided by clinical expertise or evidence from external trials so that a plausible sceptical prior

distribution is obtained.²⁷ The external information may be a trial where a similar treatment (such as a drug from the same class) was evaluated, but perhaps in a different disease area. Often these prior distributions are centred on the value of no difference between treatments such that there is only a small probability (for example, 5%) that the treatment effect is as large as the alternative hypothesis θ_A . In practice, this means that there will need to be fairly strong evidence in the data to recommend action because the posterior distribution may be drawn closer to the value of no difference. If there is a sensitivity analysis for the choice of prior distribution, this sceptical prior may not always be necessary; it may depend on whether the researchers think that the results may be optimistic for reasons such as publication bias.²⁷ Sceptical priors are considered in real examples in Chapters 3 and 4.

An enthusiastic prior distribution can be used if it is considered likely that the trial may stop early if there are results in favour of the standard treatment or control. The prior distribution may be centred on a value that represents a difference in favour of the new treatment, for example the alternative hypothesis, and variance derived such that there is only a small chance (for example, 5%) that the true treatment effect is negative.²⁸ In large sample settings, these prior distributions are not likely to be adopted because if the results are positive in favour of the new treatment and this is only due to the enthusiastic prior (i.e. the results without the prior distribution are inconclusive) then the findings are less likely to be accepted by the clinical community. However, in settings such as rare diseases or paediatric trials, the sample sizes are small and often end with non-conclusive results. However, if existing evidence can be incorporated this can effectively increase the sample size of the trial. The use and acceptance of prior evidence in these settings depends largely on the relevance of the prior information and how likely it may be to obtain further data rather than formally using existing evidence in a prior distribution. This issue will be considered again in Chapter 3.

Due to the tension involved in using informative (and possibly subjective) prior distributions, there are ‘vague’ prior distributions, which are also referred to as ‘reference’ or ‘non-informative’ priors. The intention of using vague priors is to represent almost no information about the unknown parameters of interest in order to be the least subjective. As a result, if the data is sufficiently large, the amount of information in the prior distribution is small relative to the likelihood, and thus the posterior distribution is similar to the likelihood function. This approach is useful if there is no prior information available to derive an informative prior distribution. Furthermore, the Bayesian framework can be adopted because it can handle far more complex scenarios than a frequentist framework, but perhaps the user does not want to include informative prior evidence. However, these prior distributions must be selected with caution because no prior distribution is truly non-informative. In particular, variance parameters in multilevel models tend to be highly sensitive to the choice of ‘vague’ prior distribution.²⁹ Regardless of the prior distribution that is selected, it is important to do a sensitivity analysis for the prior distributions to examine how the conclusions depend on the choice of prior. This issue will be an underlying theme in all projects in this thesis, and motivates a simulation study in Chapter 6.

1.3.4 Schools of Bayesians

Due to the different prior distributions, there are various different Bayesian schools of thought. Three broad schools may be identified and these are the ‘reference’, the ‘proper’, and the ‘decision-theoretic’. Briefly, the ‘reference’ approach adopts reference priors (defined in section 1.3.3) and uses a Bayesian interpretation for the results using the posterior distribution. This essentially means that the results are very similar to those that arise from a frequentist analysis; however the interpretation is in Bayesian terms. The ‘proper’ Bayesian adopts informative prior distributions based on previous evidence and conclusions are based on the posterior distribution. Lastly, the ‘decision-theoretic’ school (sometimes referred to as

the 'full' Bayes approach) uses utility functions to make decisions based on maximising the expected utility. In this thesis, the reference and proper Bayes approaches are used, depending on the extent of prior information available; however sensitivity to the choice of prior distribution is explored in all analyses. The reporting of results is derived from the posterior distributions.

1.3.5 Markov chain Monte Carlo estimation method

It is often difficult to derive the posterior distribution of parameter θ , $p(\theta|y)$, if there are non-conjugate distributions or nuisance parameters.²¹ The Markov chain Monte Carlo (MCMC) method is a technique that uses simulation to evaluate integrals as opposed to exact or approximate algebraic analysis.²¹ Suppose there is a p -dimensional parameter vector, θ , and the posterior distribution $p(\theta|y)$. The Gibbs Sampler is a MCMC method that samples each value of θ^t , for iteration t , from a distribution that depends only on the previously sampled values.³⁰ The Markov chain must be started at arbitrary values $\theta^0=(\theta_1^0, \theta_2^0, \dots, \theta_p^0)$. The Gibbs sampler then generates random samples from the full conditional distributions such that, for the t th iteration, the sampled values of the parameters $\theta_p^t=(\theta_1^t, \theta_2^t, \dots, \theta_p^t)$ are obtained by the sampling method shown below:³¹

$$\theta_1^t \text{ is generated from } p(\theta_1|\theta_2^{t-1}, \theta_3^{t-1}, \theta_4^{t-1} \dots, \theta_p^{t-1}, \mathbf{y}) \quad (1.8)$$

$$\theta_2^t \text{ is generated from } p(\theta_2|\theta_1^t, \theta_3^{t-1}, \theta_4^{t-1} \dots, \theta_p^{t-1}, \mathbf{y})$$

$$\theta_3^t \text{ is generated from } p(\theta_3|\theta_1^t, \theta_2^t, \theta_4^{t-1} \dots, \theta_p^{t-1}, \mathbf{y})$$

.....

$$\theta_p^t \text{ is generated from } p(\theta_p|\theta_1^t, \theta_2^t, \theta_3^t \dots, \theta_{p-1}^t, \mathbf{y})$$

When the Markov chain has converged, θ^t are samples from the marginal posterior distributions of the parameters of interest, for example $p(\theta_1|\mathbf{y})$. In this thesis, the posterior distributions are obtained using MCMC via the Gibbs Sampler within WinBUGS 1.4.3³².

The initial period of sampling is known as the ‘burn-in’ and these values are discarded. After the burn-in period, and the stationary distribution has been reached, the iterations continue to obtain a precise sample that forms the posterior distribution. The influence of the initial values and the length of the burn-in period can be assessed by sensitivity analyses. The posterior distributions must be assessed to confirm convergence has been achieved using trace, history and autocorrelation plots. If there is evidence of autocorrelation this suggests that the sampling process is not moving around the stationary distribution adequately, and consecutively sampled values are correlated. In this case, a larger number of samples may be needed to achieve an accurate representation of the posterior distribution. Alternatively, the sample can be thinned to reduce the autocorrelation by only retaining every k th sampled value.

1.3.6 Bayesian methods in clinical trials in practice

In practice, the adoption of Bayesian methods as an alternative to existing frequentist methods is still considered relatively low.³³ This is likely to be for a number of reasons. Firstly, many people feel uneasy about the potentially subjective approach to probability

within the Bayesian framework that allows for personal belief or judgement and there is still large controversy surrounding the derivation of prior distributions. If there is no prior evidence to formulate informative priors then there are issues concerned with how to derive truly non-informative priors.²⁹ Or, if the prior distribution is derived from subjective belief then the results may not be considered as objective evidence. Secondly, Bayesian analyses can be computationally complex, and although there has been major development of software to run Bayesian analyses, such as WinBUGS, it may be considered unreasonable to write new code for each trial when simpler, frequentist routines are already in place. Thirdly, some statisticians are unfamiliar with Bayesian techniques, whereas standard methods are common and may be easier to adopt rather than learning about new techniques. Finally, it is thought that regulatory authorities and journals can be hesitant in accepting new methods, and so researchers and pharmaceutical companies may not want to jeopardise their chances of a successful publication or drug to market.³³

1.3.7 Simple hypothetical example of a Bayesian analysis of a clinical trial

This section is a simple example of a conjugate analysis of a randomised clinical trial that does not require MCMC, but highlights the difference between a frequentist and Bayesian analysis. Suppose the primary analysis of the trial is to estimate an unadjusted mean difference in blood pressure, μ , which compares a new treatment with control. Assume that the likelihood has a normal distribution given by (1.9), where y_m is the estimate of mean difference from the data and m is the number of patients. It is often mathematically convenient to assume that the prior distribution, $p(\mu)$, follows a normal distribution as shown in (1.10), because then the posterior distribution will be conjugate to the prior distribution. Here, μ_0 is the mean of the prior distribution and m_0 is the 'implicit' sample size. The posterior distribution is obtained using Bayes theorem and is given by (1.11).²¹

Likelihood:
$$p(y_m|\mu) \sim N \left[\mu, \frac{\sigma^2}{m} \right] \quad (1.9)$$

Prior distribution:
$$p(\mu) \sim N \left[\mu_0, \frac{\sigma^2}{m_0} \right] \quad (1.10)$$

Posterior distribution:
$$p(\mu|y_m) \sim N \left[\frac{m_0\mu_0 + my_m}{m_0 + m}, \frac{\sigma^2}{m_0 + m} \right] \quad (1.11)$$

The posterior mean, $(m_0\mu_0 + my_m)/(m_0 + m)$, is a weighted average of the prior mean and the data, rather than just the data within a frequentist framework. Indeed, the frequentist solution is simply the maximum likelihood estimate, $\hat{\mu} = y_m$.

Assume that from the data $\hat{\mu}$ is -0.223 (mean difference of -0.223 in favour of the new treatment) and $\hat{\sigma}^2/m$ is 0.02, where m is 400 and $\hat{\sigma}^2$ is 8. Also, assume an informative prior distribution (possibly derived from clinical opinion) that is centred on no difference with variance 0.0675. Then, using equations (1.11), the posterior distribution is $p(\mu|y_m) \sim N(-0.169, 0.015)$. The distributions are shown below in Figure 1.1. The likelihood is drawn closer to the prior distribution with increased precision compared to the likelihood or prior distribution alone to produce an updated distribution (posterior).

Of course, there may be a number of different prior distributions that could have been used and the analysis of a clinical trial is unlikely to be so straightforward, but this example shows the relationship between a prior distribution, the likelihood and the resulting posterior distribution.

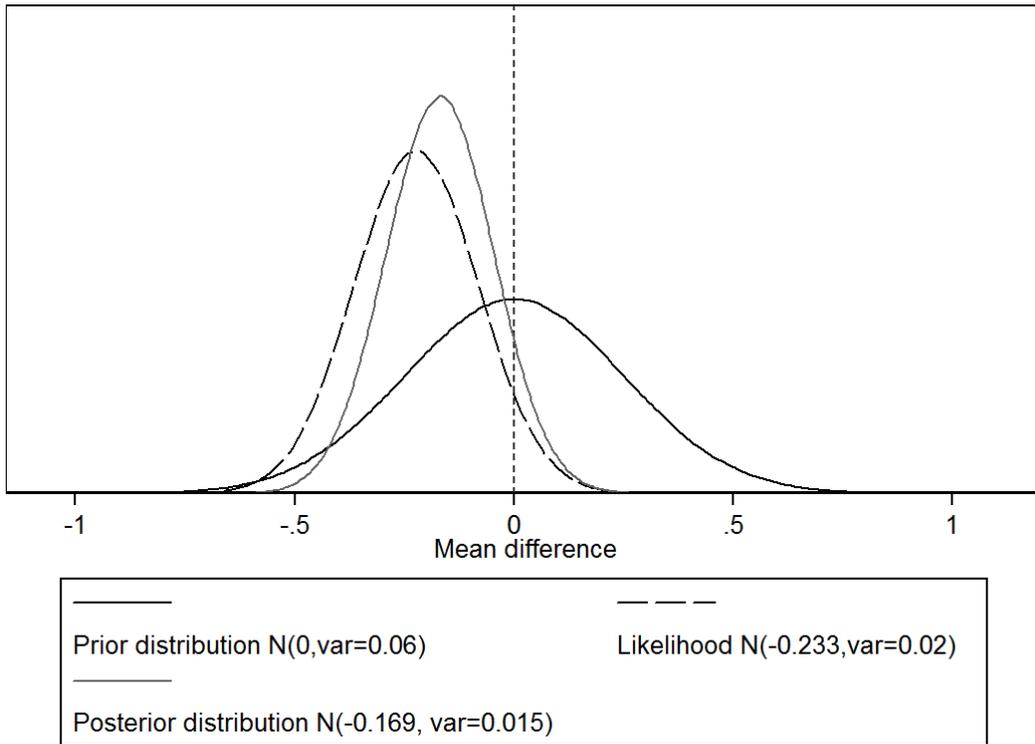


Figure 1.1: Hypothetical example of a conjugate analysis.

1.4 Systematic review and meta-analysis

Often there are several trials that have attempted to answer a similar research question. For example, there may be multiple trials that have compared similar therapies in the same population and it is important to synthesise the results from these studies. A systematic review is a formal evidence synthesis approach, which collects, critically appraises and synthesises the relevant current empirical evidence from published and unpublished studies.³⁴ Meta-analysis is the formal statistical method of a systematic review that combines quantitative data from several studies to produce pooled overall results that aid evidence-based clinical decision making.³⁵ For example, for studies that compare a type of new therapy to a standard therapy in the same disease area, the estimate of the odds ratio may be extracted from each study to produce a pooled estimate of the odds ratio. Several other

estimates can be synthesised, such as relative risk, risk difference, or survival outcomes. Whilst there has been an explosion of statistical developments in evidence synthesis, in particular since the launch of the Cochrane Collaboration in 1993, amongst these methods have also been extensive developments of Bayesian methods for meta-analysis. In particular, the use of Bayesian methods for multivariate meta-analysis with multiple outcomes^{17,18} and the derivation of prior distributions for the between-study heterogeneity to use in a future random-effects meta-analysis.³⁶⁻³⁸ The Bayesian methods tend to mirror those that have been developed within the frequentist framework; however, there may be some key advantages of the Bayesian framework. These include the ability to easily account for all parameter uncertainty in the model, the ability to include external evidence in the form of prior distributions that would otherwise be excluded, and the ability to extend the models to complex scenarios, such as combining comparative-arm and single-arm studies in a single meta-analysis.²⁰ The next two sections describe different approaches to meta-analysis with a description of how to obtain the pooled estimate across studies within Bayesian and frequentist frameworks.

1.4.1 Fixed-effect meta-analysis

Chapters 4 to 6 of this thesis use Bayesian meta-analysis to pool *treatment effect* estimates from multiple randomised trials. There are various methods that can be applied and the simplest approach is a fixed-effect analysis model, which assumes that there is a single (fixed) treatment effect across all studies, θ .³⁹ The summary (pooled) estimate, $\hat{\theta}$, from a meta-analysis gives the best estimate of this single treatment effect, using the evidence from all available studies included in the meta-analysis. The fixed-effect model can be written as:

$$Y_i \sim N(\theta, S_i^2) \tag{1.12}$$

This model assumes that the estimated treatment effect, \hat{Y}_i , for study i , follows a normal distribution with variance, S_i^2 , which is assumed known. There are various methods, both frequentist^{39,40} and Bayesian, that can be used to estimate θ , which are explored and compared in detail in Chapter 4. The most common frequentist method is the inverse-variance method, which assigns weights to each trial, which are shown in equation (1.13).³⁴ Here, S_i^2 is the variance of the treatment effect estimate from each study, i . The pooled treatment effect, θ , can be estimated using maximum likelihood estimation shown in equation (1.14) and the variance of the pooled treatment effect is calculated by equation (1.15).

$$w_i = \frac{1}{S_i^2} \quad (1.13)$$

$$\hat{\theta} = \frac{\sum_{i=1}^k Y_i w_i}{\sum_{i=1}^k w_i} \quad (1.14)$$

$$\text{var}(\hat{\theta}) = \frac{1}{\sum_{i=1}^k w_i} \quad (1.15)$$

Within a Bayesian meta-analysis, a prior distribution must be specified for the unknown pooled treatment effect, θ . For example, if there is no external evidence, a ‘vague’ normal prior distribution could be selected, such as $\theta \sim N(0, 1000^2)$.

1.4.2 Random-effects meta-analysis

In contrast to the fixed-effect approach, a random-effects approach assumes a distribution of true treatment effects across studies shown in model (1.16). Now, the estimate of treatment effect, \hat{Y}_i from each study, i , can be estimating a different true treatment effect, θ_i , and therefore the summary (pooled) estimate, $\hat{\beta}$, from the meta-analysis gives an estimate of the average treatment effect, β , across studies.³⁴ Individual studies may therefore have a true

treatment effect that varies considerably from this average value, and this between-study variance (heterogeneity) in the treatment effect is known as τ^2 .

$$Y_i \sim N(\theta_i, S_i^2) \quad (1.16)$$

$$\theta_i \sim N(\beta, \tau^2)$$

Similar to the fixed-effect model, there are numerous methods to estimate the pooled treatment effect, β , and between-study variance, τ^2 , within Bayesian and frequentist frameworks. Within a frequentist framework the DerSimonian and Laird ‘method-of-moments’ is popular to estimate β and τ^2 , which is shown in equations (1.17) to (1.20).^{34,41} In these equations, k is the number of studies in the meta-analysis, the weights, w_i , are obtained by a fixed-effect model, for example from the inverse-variance method (equation (1.13)).

$$\hat{\tau}^2 = \frac{Q - (k - 1)}{\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}} \quad \text{where } Q = \sum_{i=1}^k w_i (Y_i - \hat{\beta})^2 \quad (1.17)$$

$$w_i^* = \frac{1}{\text{var}(Y_i) + \hat{\tau}^2} \quad (1.18)$$

$$\hat{\beta} = \frac{\sum_{i=1}^k Y_i w_i^*}{\sum_{i=1}^k w_i^*} \quad (1.19)$$

$$\text{var}(\hat{\beta}) = \frac{1}{\sum_{i=1}^k w_i^*} \quad (1.20)$$

Equation (1.18) shows the weight estimation in a random-effects meta-analysis, which includes the estimate of τ^2 . Equation (1.19) is the calculation for the pooled average treatment effect estimate ($\hat{\beta}$), and equation (1.20) is the variance of the pooled average treatment effect estimate.

Within the Bayesian framework prior distributions must be specified for β and τ^2 . For example, the vague $N(0,1000^2)$ can be specified for β , however the prior distribution for τ^2 is more difficult. A previous simulation study by Lambert et al.²⁹ revealed how several ‘vague’ prior distributions for this parameter are actually fairly informative and the results can be highly sensitive to the choice of prior distribution. This is because often there are few studies within a meta-analysis and therefore it is difficult to estimate the extent of between-study heterogeneity. Thus, two previous papers modelled a large collection of meta-analysis characteristics on τ^2 to obtain empirical prior distributions for the between-study heterogeneity in a new meta-analysis of continuous outcomes³⁸ and binary outcomes.³⁶ These prior distributions have been derived for a general meta-analysis setting and various areas of healthcare, such as meta-analyses where the outcome is all-cause mortality. Consideration of prior distribution in meta-analysis is a key focus of Chapters 4 to 6; in particular, Chapter 6 examines the impact of prior distributions for between-study variance and correlation parameters when conducting a multivariate meta-analysis of randomised trials with multiple outcomes.

1.4.3 Example of a simple Bayesian and frequentist meta-analysis

This section is a brief example of a meta-analysis using a fixed-effect and random-effects approach, within both frequentist and Bayesian frameworks to highlight any simple differences between the approaches. The example is a hypothetical dataset from a paper by Riley et al.³⁹. Consider that the estimates of treatment effect are computed and synthesised from 10 studies of the same antihypertensive drug, and the mean difference is the estimate of the change in systolic blood pressure between the treatment group and the control group. Negative estimates suggest that the reduction in blood pressure was greater for those in the treatment group compared to the control group.

Table 1.2: Example dataset of the estimated mean difference from ten studies.

Study ID	Mean difference (\hat{Y}_i)	S.D(\hat{Y}_i)
1	0.0	0.42
2	0.1	0.22
3	-0.4	0.03
4	-0.8	0.20
5	-0.63	0.30
6	0.22	0.30
7	-0.34	0.07
8	-0.51	0.10
9	0.03	0.12
10	-0.81	0.30

The frequentist estimate of the pooled treatment effect within a fixed-effect model is obtained using maximum likelihood estimation, as shown in section 1.4.1 and the pooled *average* treatment effect estimate is obtained using the DerSimonian and Laird method described in section 1.4.2. The pooled treatment effect estimates in the Bayesian framework are obtained using MCMC. A vague prior distribution is specified for the pooled treatment effect in the fixed-effect model, so $\theta \sim N(0, 1000^2)$. For the random-effects model, the same vague prior distribution is specified for the pooled average treatment effect so $\beta \sim N(0, 1000^2)$, and a $N(0, 2)$ distribution, truncated at zero, is specified for the between-study standard deviation. This distribution for the between-study standard deviation is selected as a conservative vague prior distribution.²⁹ The pooled estimates of treatment effect are displayed below in Table 1.3.

Table 1.3: Pooled summary results of hypothetical meta-analysis.

Framework	Meta-analysis method	Pooled estimate (95% interval)	Between-study variance ($\hat{\tau}^2$)
Frequentist	Fixed-effect	-0.38 (-0.43 to -0.33)	-
	Random-effects	-0.33 (-0.48 to -0.18)	0.028
Bayesian	Fixed-effect	-0.38 (-0.43 to -0.34)	-
	Random-effects	-0.33 (-0.57 to -0.08)	0.084 (0.010 to 0.39)*

* 95% credible interval for $\hat{\tau}^2$

The pooled results are negative, indicating that hypertensive treatment is, on average, beneficial, although the reduction in blood pressure is clinically small. The 95% intervals are wider for the random-effects models since there is additional uncertainty by incorporating between-study heterogeneity. The frequentist and Bayesian estimates are similar, although the estimate of τ^2 is larger for the Bayesian analysis and its uncertainty is also accounted for in the Bayesian analysis. Therefore, the 95% interval for the pooled average treatment effect estimate is wider within the Bayesian framework compared to the frequentist. The results from the Bayesian random-effects meta-analysis are shown in the forest plot in Figure 1.2 below. The visual heterogeneity in results across studies, and the estimate and credible interval for τ^2 indicate that the random-effects approach is preferable and the true treatment effect varies considerably across trials. Thus, in some settings the treatment effect may be somewhat different to the average meta-analysis result. This is important if wanting to predict the potential treatment effect in a new study or to a new population, and this issue is a key focus of Chapters 4 to 6.

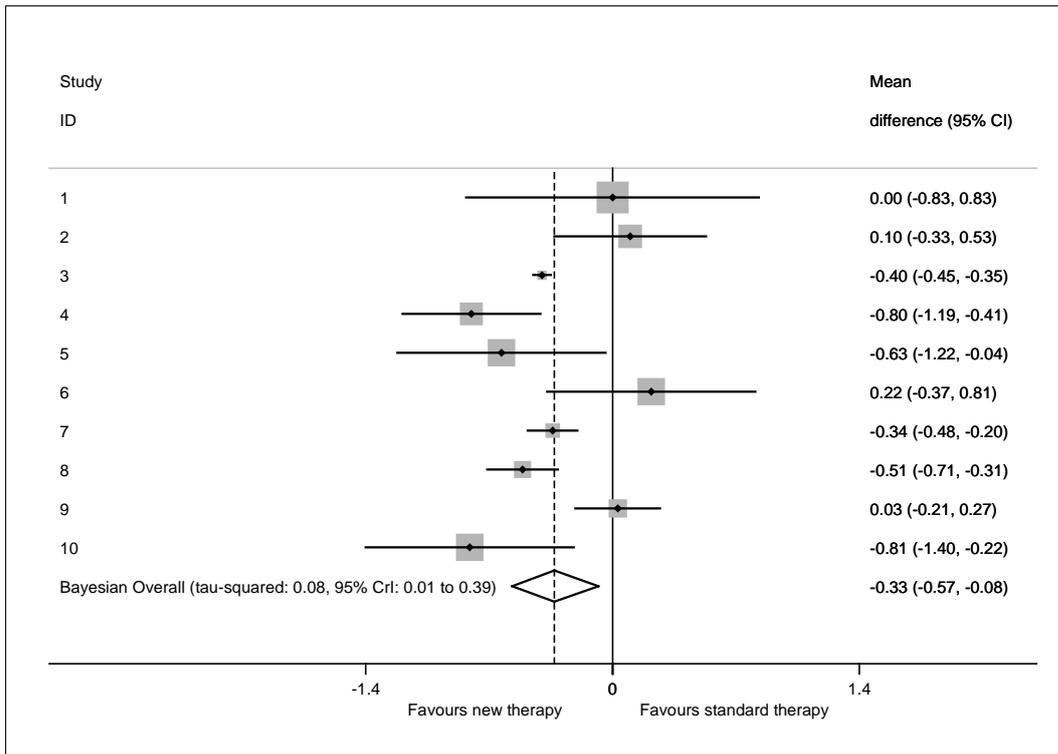


Figure 1.2: Example meta-analysis using a Bayesian random-effects meta-analysis.

1.5 Aims and structure for the thesis

1.5.1 Aims

There are several key aims of this thesis.

- To apply Bayesian methods to randomised trials with small numbers of events.
- To develop novel Bayesian approaches for meta-analysis in the context of phase II trials and multiple outcomes, with an emphasis on informing phase III decisions.
- To examine the impact of prior distributions in these contexts.

1.5.2 Structure for the thesis

Bayesian methods have been extensively developed in many areas of health research, however the prevalence of their use in applied medical research is less clear. This thesis continues in Chapter 2 by evaluating where and how often Bayesian statistical analysis methods are being adopted in applied research. A literature review is undertaken of the application of Bayesian methods in published randomised controlled trials in leading medical journals in 2012. This identifies the prevalence of Bayesian methods, the type of statistical analyses conducted using a Bayesian approach, and whether these were justified over frequentist methods. This chapter also evaluates how sample size calculations are informed with existing evidence to determine whether the design of randomised trials are informally adopting a Bayesian approach by incorporating prior information. The findings of the review inform the rationale for the remainder of the thesis.

Chapter 3 is an application of Bayesian analysis methods to a randomised controlled trial in a rare disease called the PLUTO trial (Percutaneous Shunting for Lower Urinary Tract Obstruction).⁴² This chapter investigates the use of Bayesian analysis methods in comparison to frequentist methods for a randomised trial with very small sample sizes due to failure to recruit its target sample size. Both Bayesian and frequentist survival analysis methods are explored with an emphasis on the use of various prior distributions for the estimate of treatment effect (both vague and evidence-based), and whether there are advantages and limitations of the Bayesian approach in this setting.

Chapter 4 explores the use of Bayesian and frequentist methods for the univariate meta-analysis of phase II randomised trials, which tend to be small trials by design. This chapter considers the use of meta-analysis methods in the context of phase II trials in order to inform the decision to progress to a phase III trial. An example of nine phase II trials in acute

myocardial infarction is used for illustration of fixed-effect and random-effects meta-analysis methods using frequentist and Bayesian frameworks. Several methodological issues are explored including how to assess heterogeneity, deriving prediction intervals, and the impact of prior distributions in the Bayesian approach.

Chapter 5 then extends this work to explore the use of Bayesian methods for multivariate meta-analysis of phase II randomised trials with multiple correlated outcomes. The same illustrative dataset is used to explore the methodological issues in a bivariate setting, and the results are compared to the univariate results of Chapter 4. Several different meta-analysis models are applied, and various methods that deal with unknown within-study correlations are explored.

Chapter 6 is a simulation study to assess the choice of prior distribution for the between-study correlation in a Bayesian bivariate meta-analysis of two correlated outcomes. The work extends the bivariate meta-analysis of Chapter 5 by exploring the impact of various vague prior distributions for the between-study correlation when there are few studies to estimate this parameter. The simulation study also assesses the impact of the prior distributions for the between-study correlation on the pooled treatment effect estimates and the between-study variance. Various possible settings are explored to generalise the results to difference scenarios including complete-case and missing data settings.

Finally, Chapter 7 summarises the key findings from the thesis and discusses the remaining methodological difficulties with suggestions for further research priorities.

CHAPTER 2: THE USE OF BAYESIAN METHODS IN PUBLISHED RANDOMISED CONTROLLED TRIALS: A REVIEW OF LEADING MEDICAL JOURNALS

2.1 Introduction and aims

An overarching aim of this thesis is to develop Bayesian methods for medical research. Bayesian methods for the design and analysis of clinical trials have been promoted for general use⁴³ and specifically in relation to data monitoring⁴⁴ and meta-analysis.⁴⁵ However, promotion and methodological development does not necessarily mean Bayesian methods are used in applied research. Therefore, the first aim of this chapter is to review the application of Bayesian methods in the design and analysis of randomised clinical trials in the published literature, in the leading general medical journals, in 2012. This is to identify: (i) the prevalence of Bayesian methods in the published literature compared to frequentist methods, and (ii) the type of statistical analyses conducted using a Bayesian approach, and whether or not these were justified over frequentist methods.

Another area of interest for this thesis is how randomised phase III trials are designed in terms of accounting for prior evidence of the treatment effect of interest. This evidence most likely comes from earlier exploratory phase II trials, or from existing randomised phase III trials that were conducted in a different setting, for example a trial in a different population.⁴⁶ If there is existing evidence, this could be considered in the design of the phase III trial. Later, Chapters 4 and 5 of this thesis will investigate how the results of existing randomised phase II trials can inform the design, and potential results, of a randomised phase III trial within a Bayesian framework. To facilitate this, the second objective of this review is to evaluate how sample size calculations for randomised phase III trials are informed with

existing evidence. This will identify: (i) whether the articles describe the sample size calculation; (ii) if the description contains justification for the choice of the estimates in the sample size calculation by referencing previous evidence or clinical judgement; and (iii) the types of research studies from which the supporting material was obtained, in particular whether phase II trials were referenced.

The review will therefore provide a snapshot of the current Bayesian analysis methods of randomised controlled trials in the literature and identify appropriate areas for the consideration of Bayesian methods in later chapters of the thesis.

2.2 Methods of the review

In this section, the aims and methods of the review are described in detail.

2.2.1 Aims

This review has the following two key aims.

Aim 1: Statistical analyses with a Bayesian approach

- To review the application of Bayesian methods in the design and analysis of randomised clinical trials in the published literature, in the leading general medical journals, in 2012. This will identify:
 - The prevalence of Bayesian methods in the published literature compared to frequentist methods.
 - The statistical analyses that have been conducted using a Bayesian approach, and whether or not these analyses have been justified in regard to their use over frequentist methods.

Aim 2: Sample size in randomised phase III trials

- To evaluate how sample size calculations for randomised phase III trials are informed with existing evidence. This will identify:
 - Whether the articles describe the sample size calculation.
 - If the description contains justification for the choice of the effect size and standard deviation estimates in the sample size calculation by referencing previous evidence or clinical judgement.
 - The types of research studies from which the supporting material was obtained, in particular whether phase II trials were referenced.

2.2.2 Identifying a set of articles

A wide search was undertaken to identify articles that reported any type of randomised trial, for example phase II, or phase III trials. The initial search of articles was conducted by sourcing articles using Web of Science⁴⁷ that were published in 2012 from seven of the leading general medical journals. These journals were:

- The Lancet
- PLOS Medicine
- British Medical Journal (BMJ)
- New England Journal of Medicine (NEJM)
- The Journal of American Medical Association (JAMA)
- Nature
- Annals of Internal Medicine

For example, the search strategy within Web of Science for articles published in the Lancet in 2012 was the following:

TOPIC: (randomi* trial) AND PUBLICATION NAME: (Lancet) AND YEAR
PUBLISHED: (2012).

Once the initial search had been completed, the resulting articles were reduced by reading the titles and abstracts to remove articles that were not randomised trials, such as meta-analyses and epidemiological studies. This produced a final set of articles to be evaluated in full for the review.

2.2.3 Evaluation of articles

Review of articles for Aim 1: Statistical analyses with a Bayesian approach

It was considered that the use of the Bayesian approach is unlikely to be reported in the abstract unless it is the primary analysis of the trial, and therefore the full text of *all* articles in the final set were obtained so that all Bayesian analyses could be identified. Text searches were then performed electronically within the full articles obtained, and the text searched was any of the following words:

“Bayes”, “posterior”, “credible”, “MCMC”, “Markov chain”, and “Gibbs”.

The prevalence of Bayesian methods were summarised by reporting the number of articles that contain Bayesian methods out of the total set of articles in the review. The statistical analyses that were conducted using a Bayesian approach, and whether these methods were justified over frequentist methods, were identified by the following questions:

- What was the statistical analysis conducted using a Bayesian approach?
- Did the authors describe the reason for using the Bayesian approach instead of frequentist methods?
- Was the Bayesian approach used as the primary or secondary analysis of the trial?
- Was there consideration of the impact of the prior distributions selected in the analysis?

Additional information that was of general interest was also collected. This included an overview of the trial context, which detailed the interventions that were compared, the disease setting, and the outcome measures being compared. Details of the type of randomised trial were summarised, for example, whether the trial was a cluster trial or a parallel group trial. An overview of the results presented was given, for example, whether the authors presented the means, medians, standard deviations, credible intervals and posterior probability estimates. A summary was also given for the software that each article used for the Bayesian analyses.

Review of articles for Aim 2: Sample size in randomised phase III trials

The articles that were selected for this part of the review were a subset of the full set of research articles, with randomised phase III trials being the sole trials of interest for this section. It was assumed that all phase III trials would report the sample size calculation in the statistical analysis methods; therefore only a subset of trials was deemed sufficient to give an overview of the methodology that would be applied in published phase III trials in general. Therefore, just the randomised phase III trials from the Lancet were selected to answer the aims relating to this part of the review.

Information was extracted to answer the following questions:

- Did the articles detail the sample size calculation? If so:
 - Did the authors justify their choice of the estimates used in the calculation, for example by referring to existing data or clinical judgement?
 - What types of research articles or evidence were the estimates in the sample size calculation obtained from?

For the latter question, the types of research articles or evidence referenced were summarised into: phase II trials (which may be randomised or non-randomised), randomised phase III trials (which may be in various different settings), systematic reviews and meta-analyses, observational studies (including cohort and case-control studies), clinical judgement or other source.

Also, for all articles detailing a sample size calculation, the elements of the calculation were extracted including: the estimate of the primary outcome measure in the control group (i.e. the placebo or standard therapy group); the expected value of the primary outcome measure in the experimental treatment group; the hypothetical value of treatment effect (for example, a relative risk or a hazard ratio estimate); the margins for non-inferiority or equivalence; and the intra-class correlation in cluster trials.

2.3 Results

In this section the review findings for Aims 1 and 2 are described in detail. A second reviewer independently checked a random sample of 10% of the articles to validate the data extraction results, and there were no discrepancies.

2.3.1 Aim 1: Statistical analyses with a Bayesian approach

After duplicate records were removed, the literature search in Web of Science identified 696 potential articles from 1st January 2012 to 31st December 2012. The abstracts were read and 382 articles were excluded because they were not randomised trials (e.g. meta-analyses, letters to editors, observational studies); see Figure 2.1. A final total of 314 articles were included in the review (see Appendix A). This included 23 articles in *Annals of Internal Medicine*, 14 articles in *PLOS Medicine*, one article in *Nature*, 74 articles in *NEJM*, 47 articles in the *BMJ*, 69 articles in *JAMA*, and 86 articles in the *Lancet*.

After text search of the full 314 articles, and in-depth reading of these articles, only three (1%) were identified that applied a Bayesian approach for one or more of their statistical analyses. Of these three, two were in *JAMA*^{48,49}, and one was in *Annals of Internal Medicine*⁵⁰. The three articles are summarised in Table 2.1 and are reviewed individually below.

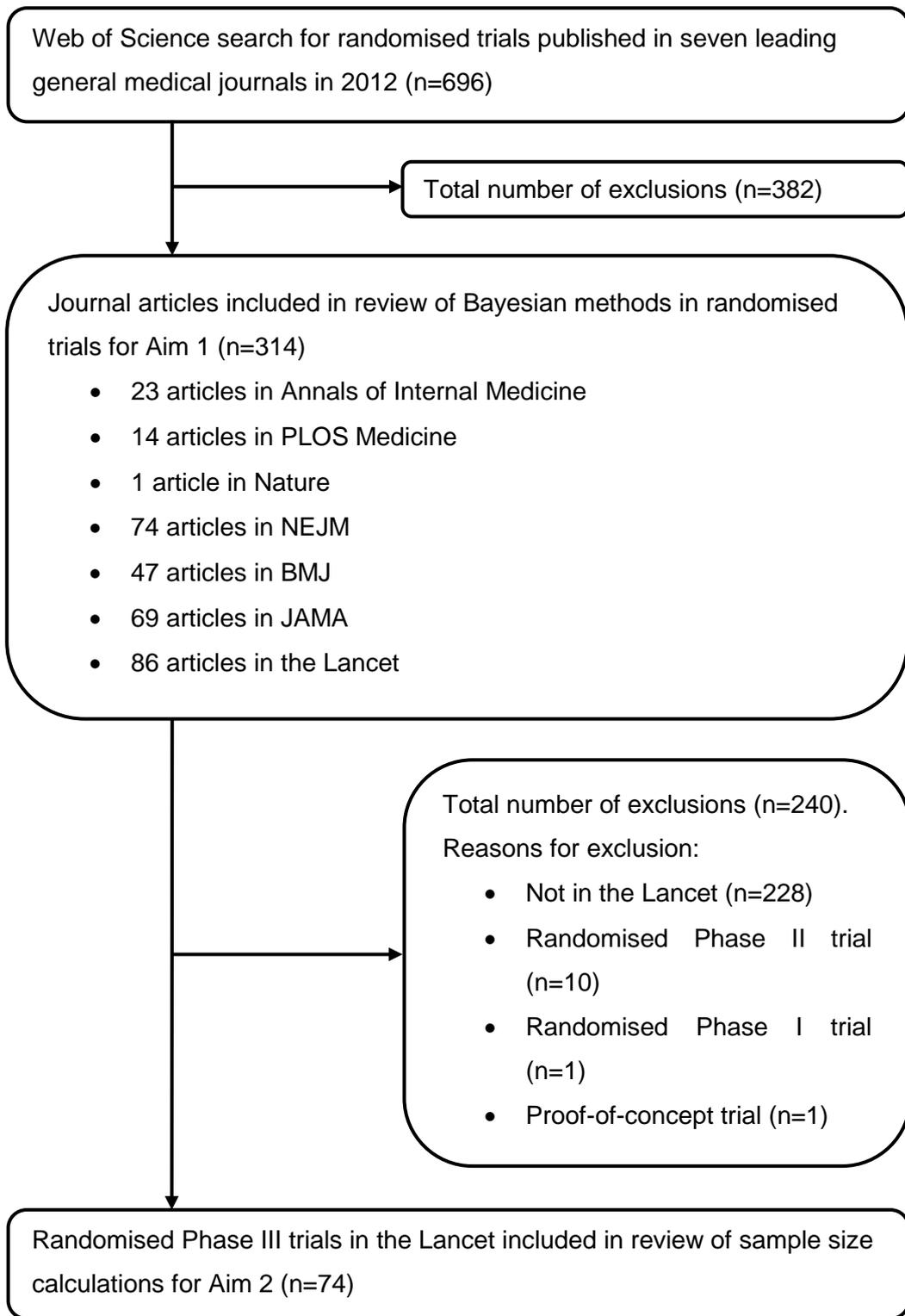


Figure 2.1: Flow diagram of journal articles for inclusion in literature

Article 1 – Lucas et al.⁴⁸ (JAMA)

The article by Lucas et al. in JAMA used a Bayesian analysis and was a cluster randomised crossover non-inferiority trial.⁴⁸ The trial compared the effect of 2- versus 4-week attending-physician rotations on several outcomes, which include the primary outcome of unplanned patient revisit to the same hospital within 30 days, which is a binary response variable. Secondary outcomes were length of hospital stay, trainee evaluations of attending-physician performance, and attending-physician burnout.

The primary analysis involved the derivation of a mixed-effects regression model with crossed random-effects⁵¹ to estimate odds ratios for unplanned patient revisits. The authors do not clarify which regression model they used, although the primary outcome is binary and the effect measure is an odds ratio, therefore it is likely to be a logistic regression model. This analysis adopted a Bayes approach with non-informative priors for model estimation. The authors justified the use of the Bayesian estimation by referring to it as equivalent to maximum likelihood methods through the use of non-informative prior distributions. Markov chain Monte Carlo (MCMC) procedures of estimation were adopted, which the authors suggested are superior to maximum likelihood because they are able to handle both crossed random-effects and discrete outcomes.

The model was estimated using MLwiN software, version 2.25, which is run from within Stata, version 12, using a command called *runmlwin*.⁵² The authors do not state the specific choice of prior distributions; however it is likely they were not chosen by the user as MLwiN uses in-built 'non-informative' prior distributions. Therefore there was no further consideration of other prior distributions.

Although the estimation method is Bayesian, the authors presented the results by 95% confidence intervals and p-values. It is unclear why the results were not presented as

posterior distributions or credible intervals, and the article did not state how the p-values were obtained. However, the authors argued that the use of non-informative prior distributions means that the method is “numerically equivalent to...maximum likelihood⁵³” and so the authors appear to have presented the Bayesian results as if they are frequentist.

Article 2 – Newman et al.⁴⁹ (JAMA)

The other article in JAMA using a Bayesian analysis reported the results of a randomised, double-blind, placebo-controlled, parallel-group trial in patients who have had coronary artery bypass graft surgery. The trial investigated the efficacy and safety of acadesine treatment that was administered in the peri-operative period and the outcomes were all-cause mortality, non-fatal stroke, and severe left ventricular dysfunction through 28 days.⁴⁹

The authors used a Bayesian approach to assess futility in two interim analyses. Although justification for this Bayesian analysis was not given, it is likely to be for the ease of obtaining direct probability estimates. The authors stipulated that the estimated probabilities of rejecting the null hypothesis at the end of the study, at 30% and 40% follow-up, must be at least 20% and 65%, respectively, to continue with the trial. The method for the interim analysis was not detailed in the article, and further details were not given in the supplementary material. Also, the authors did not reference any relevant Bayesian methodological articles, which may have informed this analysis. Therefore, it is not possible to know what analysis was conducted and whether the authors considered sensitivity to the prior distribution choices.

The results revealed that the trial stopped recruiting at the 30% futility analysis “based on the recommendation of the DSMB”,⁴⁹ however the reason for this decision was not documented. Thus, it is not possible to know whether this was due to the failure of the criteria for

continuation that would have been based on a posterior probability estimate, or another reason such as safety.

Although the interim assessment was Bayesian, the statistical analysis at the end of the trial was frequentist. The primary analysis with complete follow-up was a classical Cochran-Mantel-Haenszel hypothesis test that adjusts for a stratification factor (sex). All statistical analyses were performed in SAS version 9.

Article 3 – Hashkes et al.⁵⁰ (Annals of Internal Medicine)

The third article was in Annals of Internal Medicine and was a crossover trial that investigated the use of Riloncept for patients with Familial Mediterranean Fever that are resistant to or intolerant of colchicine.⁵⁰ This trial was a randomised, double-blind, single-participant alternating treatment study.

The primary analysis was conducted with a Bayesian approach that compared Riloncept to placebo in terms of the primary outcome of the number of attacks in each treatment group. The authors summarised this comparison with a risk ratio and modelled the number of attacks in each treatment course given the duration in a treatment group with a Poisson distribution. The authors argued for their use of a Bayesian approach by saying that it is, “recommended for studies with this design and sample size^{21”}.⁵⁰ This is because Familial Mediterranean Fever is a rare disease resulting in a small trial sample size (n=14). The authors showed consideration of the choice of prior distribution for the risk ratio parameter in the model by conducting a sensitivity analysis that applied three ‘off-the-shelf’ prior distributions for this parameter; a non-informative, an enthusiastic, and a sceptical prior distribution.²¹ The results for the primary analysis were reported for all three prior distributions. The impact of the choice of prior distribution was not discussed; however the results are consistent for the three prior distributions.

A secondary analysis of the odds ratio of “escape” from treatment was also conducted with a Bayesian approach. The posterior estimate and 95% credible interval were reported in the results.

The mean, SD and 95% credible interval for the risk ratio were reported, with a risk ratio less than one indicating a significant result in favour of Riloncept. No posterior probability estimates were given in the results for the Bayesian analyses. The Bayesian analyses were conducted with MCMC methods in WinBUGS Version 1.4, and the WinBUGS code for the primary and the secondary analyses were given in an appendix with some details and the models.

Although the authors considered the limitations of statistical analyses with small sample sizes they still conducted many secondary analyses using a frequentist approach for these outcomes.

Table 2.1: Summary table for articles with Bayesian analysis (Aim 1).

Journal article	Trial type	Bayesian analysis objective	Justification for using a Bayesian approach	Is the Bayesian analysis the primary or secondary analysis?
Article 1: Lucas, B ⁴⁸	Cluster randomised crossover non-inferiority trial.	Reference Bayes approach for a mixed-effects regression model with crossed random-effects to estimate OR for ordinal outcomes.	Authors argue Bayesian approach is equivalent to ML using non-informative priors. MCMC can easily handle crossed random-effects and discrete outcomes.	Primary analysis.
Article 2: Newman, M ⁴⁹	Randomised, double-blind, placebo-controlled, parallel-group trial.	Two interim futility analyses, which stipulated that the probability of rejecting the null hypothesis at the end of study, at 30% and 40% follow-up, must be at least 20% and 65%, respectively, for continuation.	No justification reported.	Neither. The Bayesian approach is used for interim analyses.
Article 3: Hashkes, P ⁵⁰	Randomised, double-blind, crossover trial.	A Poisson regression model to compare the number of attacks in each treatment group with a risk ratio. A logistic regression model to compare the odds of “escape” from treatment.	Bayesian approach is recommended for studies with their design and small sample size.	Primary analysis.

Table 2.1: Summary table for articles with Bayesian analysis (Aim 1).

Journal article	Are the prior distributions reported?	Is the impact of prior distributions considered?	What statistical software is used?	What results are presented?
Article 1: Lucas, B ⁴⁸	Non informative but specific type not reported.	No, used default non-informative prior distributions within MLwiN.	MLwiN version 2.25, run within Stata version 12.	Although the approach is Bayesian, results presented as 95% confidence intervals and p-values.
Article 2: Newman, M ⁴⁹	No.	No consideration of choice of prior distributions was reported.	SAS version 9.	Stopped recruitment at first interim futility analysis. Posterior probability of rejecting the null hypothesis not reported.
Article 3: Hashkes, P ⁵⁰	Yes. Non-informative RR~lognormal (0, sd=10); enthusiastic RR~lognormal (0.5, sd=0.125); sceptical RR~lognormal (0, sd=0.125)	Yes, 3 prior distributions for the risk ratio parameter: non-informative, sceptical and enthusiastic prior. The impact of the prior is not discussed but results are consistent.	WinBUGS version 1.4.	Means, standard deviations and 95% credible intervals.

2.3.2 Aim 2: Sample size calculation in phase III randomised trials

The papers evaluated for their sample size calculation were the 86 articles in the Lancet (Figure 2.1). However, twelve of these⁵⁴⁻⁶⁵ were not included because ten were randomised phase II trials,^{54-62,64} one was a randomised phase I trial,⁶⁵ and one was a proof-of-concept trial.⁶³ Thus, 74 relevant articles remained and the evaluation of sample size is summarised in Table 2.2 and below.

Table 2.2: Summary table for journal articles included in Aim 2 (sample size evaluation in randomised phase III trials in the Lancet).

Journal article	Is the sample size calculation detailed in the paper?	Are the choices of values in the sample size calculation given justification?	What type of research article or evidence supports the sample size calculation?
Budde ⁶⁶	Yes	No	-
Bolla ⁶⁷	Yes	No	-
Laharie ⁶⁸	Yes	No	-
Pickard ⁶⁹	Yes	Yes	Clinical judgement
Rini ⁷⁰	Yes	Yes	RCT, non-randomised phase II trial
Cohen ⁷¹	Yes	Yes	RCT, randomised phase II trial
Coles ⁷²	Yes	Yes	RCT, randomised phase II trial
Simmons ⁷³	No, in trial protocol	Yes	RCT, observational study
Lopes ⁷⁴	No, in design and rationale paper and an original trial analysis	Yes	Systematic review/meta-analysis
Ryan ⁷⁵	Yes	Yes	Observational study
Sabate ⁷⁶	Yes	Yes	Systematic review/meta-analysis
Camenzind ⁷⁷	No, in design and rationale paper	Yes	RCT, systematic review/meta-analysis, observational study
Wandt ⁷⁸	Yes	Yes	Clinical judgement
Nogueira ⁷⁹	Yes	Yes	Clinical judgement, non-randomised phase II trial
Scirica ⁸⁰	No, in original trial analysis	No	-
Saver ⁸¹	Yes	Yes	Clinical judgement
Rickard ⁸²	Yes	Yes	RCT
Crawley ⁸³	Yes	No	-
Fairall ⁸⁴	Yes	Yes	RCT
Zinkstok ⁸⁵	Yes	Yes	Systematic review/meta-analysis, clinical judgement
Wong ⁸⁶	Yes	No	-
Ridker ⁸⁷	No, in original trial	Yes	RCT

Table 2.2: continued.

Journal article	Is the sample size calculation detailed in the paper?	Are the choices of values in the sample size calculation given justification?	What type of research article or evidence supports the sample size calculation?
	analysis		
Gallwitz ⁸⁸	Yes	No	-
Sandberg ⁸⁹	Yes	Yes	Clinical judgement
Davalos ⁹⁰	Yes	Yes	Systematic review/meta-analysis
Hauschild ⁹¹	Yes	Yes	Clinical judgement
Kirchhof ⁹²	No, in design and rationale paper	No	-
Mees ⁹³	Yes	Yes	Randomised phase II trial
DeJesus ⁹⁴	Yes	Yes	RCT
Sax ⁹⁵	Yes	Yes	RCT
Jamieson ⁹⁶	No, in trial protocol	Yes	Systematic review/meta-analysis
Sandercock ⁹⁷	Yes	No	-
Aschner ⁹⁸	Yes	No	-
Gallwitz ⁹⁹	Yes	No	-
Bhatnagar ¹⁰⁰	Yes	Yes	Survey or audit
Cima ¹⁰¹	Yes	Yes	Quality of life questionnaire
Van der Graaf ¹⁰²	Yes	Yes	Systematic review/meta-analysis, non-randomised phase II trial
Biere ¹⁰³	Yes	Yes	RCT, systematic review/meta-analysis, observational study
Engert ¹⁰⁴	Yes	Yes	RCT
Goya ¹⁰⁵	Yes	No	-
Meirik ¹⁰⁶	Yes	No	-
Van Vollenhoven ¹⁰⁷	Yes	No	-
Guelmezoglu ¹⁰⁸	Yes	Yes	RCT, systematic review/meta-analysis
Kobayashi ¹⁰⁹	Yes	No	-
Heller ¹¹⁰	Yes	Yes	Regulatory guidelines
Garber ¹¹¹	Yes	Yes	Regulatory guidelines
Castaigne ¹¹²	Yes	No	-
Nijhof ¹¹³	Yes	No	-
Manaseki-	Yes	Yes	Observational study

Table 2.2: continued.

Journal article	Is the sample size calculation detailed in the paper?	Are the choices of values in the sample size calculation given justification?	What type of research article or evidence supports the sample size calculation?
Holland ¹¹⁴			
Corbacioglu ¹¹⁵	Yes	Yes	Observational study
Avery ¹¹⁶	No, in trial protocol	Yes	Systematic review/meta-analysis
Baird ¹¹⁷	Yes	Yes	Survey or audit
Fleshner ¹¹⁸	Yes	Yes	Observational study
Garcia ¹¹⁹	Yes	No	-
El Arifeen ¹²⁰	Yes	Yes	RCT
Soofi ¹²¹	Yes	Yes	Survey or audit
Thiele ¹²²	No, in design and rationale paper	Yes	RCT, pilot trial
Griffin ¹²³	Yes	Yes	RCT
Soofi ¹²⁴	Yes	No	-
Santolaya ¹²⁵	Yes	Yes	Phase I study
Baselga ¹²⁶	Yes	No	-
Johnston ¹²⁷	Yes	Yes	RCT
Hanney ¹²⁸	Yes	No	-
Abraham ¹²⁹	Yes	No	-
Glazener ¹³⁰	Yes	Yes	Systematic review/meta-analysis, clinical judgement
Bang ¹³¹	Yes	No	-
Mitja ¹³²	Yes	Yes	Observational study
Coovadia ¹³³	Yes	Yes	RCT
Smith ¹³⁴	Yes	Yes	RCT, survey or audit
Mehilli ¹³⁵	Yes	Yes	Observational study
Buller ¹³⁶	Yes	Yes	RCT, regulatory guidelines
Gebre ¹³⁷	Yes	No	-
Enden ¹³⁸	No, a design and rationale paper	Yes	Systematic review/meta-analysis
Smith ¹³⁹	Yes	No	-

Did the articles report the sample size calculation?

Of the 74 articles, all mentioned a sample size calculation (Table 2.2). However, ten of these articles (14%) did not discuss the calculation in detail, but instead referenced one of the following: a published trial protocol (three articles^{73,96,116}), a published design and rationale paper (five articles^{74,77,92,118,138}), or a published original trial analysis (three articles^{74,80,87} - this includes an article that also referenced a rationale paper⁷⁴). Therefore, the sample size calculation for these 10 articles was reviewed from the additional referenced source rather than from the article itself.

Is justification reported for the estimates in the sample size calculation?

In 25 articles (34% of the total 74 articles, see Table 2.2), the sample size calculation was detailed but the authors did not refer to any previous evidence or clinical judgement that specifically informed the elements of the sample size calculation.^{66-68,83,86,88,92,97-99,105-107,109,112,113,119,124,126,128,129,131,137,139} For example, in the article by Laharie et al., they say: “we estimated that randomisation of 100 patients would provide an 80% power to detect a 30% difference in failure rate between the ciclosporin and infliximab groups in a two-sided test with type I error of 5%. We worked with the initial assumption that the rate of failure would be 60% in the infliximab group and 30% in the ciclosporin group, corresponding to an OR of 3.5, and assumed a loss of 4% due to misdiagnosis.”⁶⁸ Thus, the authors made assumptions about failure rates but do not state prior sources to justify them.

The remaining 49 articles (66% of the total 74 articles) reported the details of the sample size calculation, and justified the choice of some or all of the estimates in the calculation. These are now summarised further, but note that the numbers of articles mentioned in the following sections do not sum to 49 articles as many of the papers reference more than one source of supporting evidence, and so each article may contribute to two or more groups below.

What types of research articles are referenced as supporting material?

Previous phase II trials

In six out of the 49 articles^{70-72,79,93,102} (8% of 74 articles, see Table 2.2), earlier phase II trials are referenced, in which three were randomised controlled trials⁷⁰⁻⁷² and the remaining three phase II trials were non-randomised single arm trials.^{79,93,102} For example, in the article by Cohen et al.,⁷¹ the authors reference a randomised phase II trial for the estimate of primary outcome in the control group: “On the basis of...the phase 2 study, we expected at least 20% of patients in the interferon beta 1a group to meet the disability endpoint by 24 months.”

Systematic review and meta-analysis

Twelve articles^{74,76,77,85,90,96,102,103,108,116,130,138} (16% of 74 articles, see Table 2.2) reference a systematic review or meta-analysis in the sample size calculation. For example, in the article by Sabate et al.,⁷⁶ the authors state: “To estimate the rate of events in the BMS group, we used...meta-analysis of randomised controlled trials that included patients with STEMI.”

Randomised controlled trials

Twenty one papers^{70-73,77,82,84,85,87,94,95,103,104,108,122,123,127,133,136,140,141} (28% of 74 papers, see Table 2.2) reference previous randomised controlled trials in similar settings to the current trial, which may be phase II (included in *Previous phase II trials* above) or phase III trials. This includes a setting such as a randomised controlled trial in which one of the drugs of interest in the article has previously been compared to placebo; for example the article by Rini et al.⁷⁰ who refer to two previous randomised phase III trials and say, “This study was designed to test the hypothesis that treatment would result in an improvement in median PFS from 5 months with sorafenib, based on previous clinical data, to 7 months with axitinib.” Randomised phase III trials have also been referenced where the same drugs are compared in different patient populations; for example, the article by El Arifeen et al.¹⁴¹ describes: “With data from the previous study, our estimate of overall neonatal mortality rate in the study

population after implementation of the basic neonatal care package was 36 per 1000 livebirths.”

Observational studies

Nine articles^{73,75,77,103,114,115,118,132,135} (12% of 74 articles, see Table 2.2) reference observational studies that contain evidence for the estimates of the sample size calculation. These references include different types of observational studies, for example cohort studies and case-control studies. In an article by Mitjà et al.¹³² the authors state that they expect an efficacy of 95% on standard therapy based on evidence from two observational studies.

Clinically relevant differences

In eight of the articles^{69,78,79,81,85,89,91,130} (11% of 74 articles, see Table 2.2), some of the estimates in the sample size calculation are justified by deeming them to be clinically relevant. For example, in the article by Glazener et al.¹³⁰ the authors justify the assumed true difference in primary outcome rates: “We set a difference of 15% in incontinence rates as our threshold for clinical importance - derived from discussion by clinicians and the project management group”.

Other sources

Other sources of evidence are referenced as supporting evidence for elements of the sample size in 10 articles^{85,100,101,110,111,117,121,122,125,136,140} (14% of 74 articles, see Table 2.2). These include information from a survey or an audit that has been conducted before the start of the trial in four articles.^{100,117,121,140} For example, the article by Bhatnagar et al.,¹⁰⁰ states: “On the basis of an audit of the participating hospitals, we estimated that the risk of treatment failure in the placebo group would be 25%.” Two articles^{122,125} reference the results of an earlier phase I study or a pilot trial; for example, the authors in the article by Elena et al.¹²⁵ report that: “A planned sample size of 1625 was calculated, which was guided by previous results

from a phase 1 study.” The results of a quality of life study questionnaire were referenced in one article:¹⁰¹ “We identified only one study about quality of life of tinnitus patients attending a specialised tinnitus centre. We used the reported mean change over time of 0.065...to calculate our sample size.” Regulatory guidelines were referenced in three articles,^{110,111,136} for example the article by Heller et al.¹¹⁰ states that: “Non-inferiority was confirmed if the upper limit of the 95% CI of the treatment difference was less than or equal to 0.4% points, as recommended by regulatory guidelines.”

2.4 Discussion

This review had two aims: first to identify the prevalence and use of Bayesian methods in trials within leading medical journals, and second to consider what prior evidence is cited when undertaking sample size calculations in phase III trials. The key findings and limitations are displayed in Figure 2.2 now discussed.

Figure 2.2: Key findings from the review of articles in medical journals.

Aim 1: Statistical analyses with a Bayesian approach

- In general medical journals in 2012, very few published randomised trial articles (1%: 3/314) applied a Bayesian approach to their analysis.
- Bayesian methods may be important for trials with small sample sizes (e.g. in rare diseases), as this was the justification in one of the trials that applied a Bayesian approach.
- Bayesian methods are potentially advantageous in interim monitoring of phase III trials, as this was highlighted in one of the trials that applied a Bayesian approach.
- A Bayesian approach with 'non-informative' priors is potentially advantageous compared to the frequentist approach, especially when the Bayesian approach can more easily handle a complex analysis, as was highlighted in one of the trials that applied a Bayesian approach.
- A Bayesian approach is important to make direct probability statements, as apparent in some of the articles using a Bayesian approach.

Aim 2: Sample size calculation

- All published randomised trials in the Lancet in 2012 mentioned the sample size calculation, either in the published paper or in a referenced source.
- In a third of the randomised trials in the Lancet in 2012, the paper made no reference to previous evidence or clinical judgement that supported the calculation of the sample size estimate.
- In 21 of the articles (28%) in the Lancet in 2012, values in the sample size calculations were supported by previous randomised controlled trials, which include phase II and phase III trials.
- In six articles (8%) in the Lancet, the sample size calculations were supported by evidence from previous phase II trials. Of these, three were randomised phase II trials.
- Other sources of evidence for the sample size estimates include systematic reviews or meta-analyses (16%), observational studies (12%), clinically relevant values (11%), and other sources, such as audit or surveys (14%).

2.4.1 Overview of key findings

Analysis methods with a Bayesian approach

The review of 314 articles revealed that there are very few published randomised trials that adopted a Bayesian analysis approach within leading general medical journals in 2012 (1%). Thus, even though Bayesian methodology has been well-developed in the literature, these methods are not being commonly adopted for the analysis of randomised trials in these journals. This result has also been observed in previous reviews. Chevret conducted a review of Bayesian clinical trials between 1982 and 2010 and found only 44 articles that were clinical trial reports with a Bayesian analysis out of a total of 331 articles with Bayesian methodology.¹⁴² The remaining 287 articles were either reviews (67 articles) or biostatistical papers (220 articles). Similarly, a review published by Lee and Chu in 2012 found there were only 121 papers reporting the Bayesian design and analysis of clinical trials between 1947 and 2011, and of these 60% (73 articles) were randomised trials.¹⁴³ Another review in 2009 looked at published protocols for trials between 2000 and 2005 at the MD Anderson Cancer Center (MDACC) in the United States, which is well-known for its adaptive trial designs and analyses.¹⁴⁴ Of 964 published protocols (not all randomised trials), 195 articles (20%) used a Bayesian design or analysis or both, and of these 169 trials were trials run within the MDACC only, and 52 trials were Bayesian randomised trials. The most frequent setting that used a Bayesian approach was phase II trials (131 articles).

Thus, in the literature of randomised trial reports, frequentist methods are clearly the standard approach and this is likely to be for a number of reasons that have already been discussed in Chapter 1.³³ Many people may feel uneasy regarding the potential for the Bayesian framework to include a subjective approach to probability that is based on personal belief or judgement, especially when there is no prior evidence to formulate informative priors. Bayesian analyses can be computationally complex, and although there has been

recent development of software to run Bayesian analyses, such as WinBUGS, it may be considered unreasonable to write new code for each trial, when simpler existing frequentist methods already exist. Some statisticians are unfamiliar with Bayesian techniques, whereas standard methods are common and easier to adopt rather than learning about new techniques. Finally, it is thought that regulatory authorities and journals can be hesitant in accepting new methods, and so researchers and pharmaceutical companies may not want to jeopardise their chances of a successful publication or drug to market.

However, by conducting the review in this chapter, some important insight has been obtained from those three articles that did apply Bayesian analysis methods. Bayesian methods have been promoted for trials with small sample sizes^{21,145-148} and the identified article in *Annals of Internal Medicine* used the Bayesian approach specifically for this circumstance as the data was from a rare disease.⁵⁰ Therefore, the use of Bayesian methods in trials with small sample sizes may be important and this will be considered in detail in Chapter 3. Another identified trial, in the *Journal of American Medical Association*, highlighted that Bayesian methods are advantageous in interim monitoring⁴⁴ of randomised phase III trials since the Bayesian approach can produce probability estimates of interest, whereas this is problematic if a frequentist approach was adopted. The advantage and derivation of the probability estimates is considered throughout the chapters in this thesis in particular within Chapters 4 to 6 where direct probability statements about treatment effectiveness are of interest. The third identified trial, also in the *Journal of American Medical Association*, showed that the Bayesian framework is adopted if the required analysis is particularly complex as the Bayesian framework can be more easily adopted compared to the frequentist approach. This trial used non-informative prior distributions so that the results would mimic the results from a frequentist equivalent method. The use of non-informative or vague priors is thus explored in several chapters in the remainder of this thesis; in particular, Chapter 3 considers a sensitivity analysis to the choice of prior distribution in a trial investigating the use of a

surgical intervention, and Chapter 6 investigates the choice of several prior distributions for the parameters of the between-study covariance matrix in a bivariate meta-analysis model.

Sample size calculations

The review of randomised phase III trials in the Lancet showed that a sample size description was reported in all trial reports, or in the trial protocol or published rationale and design papers. Previous reviews have studied the reporting standards of randomised trials and one review, published in 2010, compared the reporting standards of randomised trials in December 2000 with the reporting standards of randomised trials in December 2006, which was after the publication of the CONSORT statement in 2001.¹⁴⁹ There was an increase in the number of reports that stated the sample size from 27% (142/519 trials) in December 2000 to 45% (279/616 trials) in December 2006.¹⁴⁹ However, the percentage in 2006 is still lower compared to the articles in this review. This may be due to the update to the CONSORT statement published in 2010,¹⁵⁰ and it may additionally be because not all journals require publications to adhere to the CONSORT statement, whereas the Lancet does make this requirement. A review of all randomised trials in five leading medical journals that endorse the CONSORT statement, between July 2002 and June 2003 show similar results to the findings in this review.¹⁵¹ In the Lancet, 95.1% of the trials reported sample size justification, which was a higher percentage compared to the other journals such as the BMJ with 86.3%, Annals of Internal Medicine with 66.7%, and JAMA with 76.9%.¹⁵¹

Although the sample size is stated in all articles reviewed in this chapter, the sample size calculation was not explicitly supported by any reference to existing evidence in 34% of the articles. This observation may be for a number of reasons:

- i. There may genuinely be no existing prior evidence to inform and support the sample size calculation.
- ii. The authors did use prior evidence but did not explicitly state that they had.
- iii. The authors reported references to prior evidence but did not explain them fully in this context, and thus the reference was not captured by this review as supporting the use of prior evidence (in other words, incomplete reporting may have caused the use of prior evidence to be missed).

The sources of information that were referenced in the other 66% of articles were varied, with references from randomised phase III trials, phase II trials, systematic reviews and meta-analyses, observational studies, clinical judgement, and other sources, such as regulatory guidelines. There were six trials that informed the sample size calculation with phase II trials, and of those, three were *randomised*. There were 12 trials that used meta-analyses to inform the sample size calculation. The use of phase II trials and meta-analyses to inform sample size calculations will be considered in detail in Chapters 4 and 5, in particular, in order to better inform the design of phase III trials with the results of multiple earlier trials.

2.4.2 Limitations

There are several limitations of this review in terms of reviewing the prevalence of Bayesian methodology. These include the restriction of the search criteria to publications in one year (2012). If the search was extended to a wider range of publication dates, it is possible that more articles with Bayesian methods may be obtained. The review also only focussed on seven leading general medical journals, and perhaps these general journals do not tend to favour trials with alternative statistical analyses compared to well-known frequentist methods, perhaps because the audience may not be familiar with advanced concepts such as Bayesian methods. In further research, the review could be extended to include other non-

general medical journals, such as disease-specific journals (for example, British Journal of Cancer), which may more routinely publish trials that use less standard methods.

A similar limitation is apparent in the review of sample size calculations in randomised controlled phase III trials, as this was only conducted for the articles in the Lancet. It is possible that the results would differ if the review was conducted in an alternative journal due to the varying reporting specifications by different journals. However, this part of the review was intended to give an overview of the sample size calculations being used in phase III trials and it is unlikely that the key findings would differ in a different journal, especially as the Lancet is a leading journal for RCT publication and adheres to CONSORT, which states that the sample size and all the quantities in the calculation must be reported.

Also, this review is limited by reporting standards as it is only possible to review what has been reported. For example, authors may have used external evidence toward their sample size calculation but not reported this, as mentioned in point iii above. It may also be that the sample size is well-reported in the protocols of randomised trials but that this is not replicated in the trial reports, possibly due to word count limitations in papers. Further evaluation could also investigate the evidence reported in protocols, as in this review the protocols were only considered if it was referenced in the article as supporting information.

2.5 Conclusions and rationale for remainder of thesis

This review has highlighted that Bayesian methods are not commonly applied in clinical trials, which is disappointing. However, the articles that have adopted a Bayesian analysis have revealed the settings in which a Bayesian approach may be advantageous and thus motivate some of these areas to be explored in the remainder of this thesis. In particular, Bayesian methods are promoted in trials of small sample sizes and therefore the next chapter will apply Bayesian methods to a trial that failed to recruit its target sample size.

Furthermore, the review of sample size calculations in randomised phase III trials has shown that previous evidence, such as earlier trials and meta-analyses, is not always incorporated into the design of these trials. To address this, Bayesian meta-analysis methods of phase II trials will be explored in Chapters 4 and 5 to investigate how they may inform phase III trial design and decisions.

CHAPTER 3: APPLICATION OF BAYESIAN METHODS TO A RANDOMISED CONTROLLED TRIAL IN A RARE DISEASE: THE PLUTO TRIAL

3.1 Aims and objectives

The literature review in the previous chapter highlighted that Bayesian methods may be advantageous in trials with small sample sizes because they can formally incorporate previous evidence in the analysis. Therefore, this chapter demonstrates the use of existing Bayesian analysis methods in a randomised trial with a very small sample size. The analysis also incorporates a frequentist approach to illustrate the additional benefits and differences of the Bayesian approach, and to highlight any limitations. The analysis has been published as part of the trial publication.⁴²

The PLUTO trial (Percutaneous Shunting for Lower Urinary Tract Obstruction) was a randomised controlled trial of a surgical intervention that failed to recruit its target sample size. The trial failed to recruit for reasons such as the rare nature of the disease and the ethical obligation of the clinician not to randomise patients that they deemed unlikely to benefit from the intervention. It is important that the data collected in the PLUTO trial is used appropriately to further the understanding of the effectiveness of the intervention since it is unlikely that another RCT, in the same population with the same intervention, would be more successful in recruiting patients compared to the attempts of the PLUTO trial.

There were many clinical objectives of the trial but this chapter focuses on determining whether the intervention increases the probability of survival up to one year compared to those that received standard care. The data also provides a real example of a trial that can

be used to explore some statistical questions: what are the differences in the results between a Bayesian and frequentist approach if a vague prior distribution is used? Can previous evidence be incorporated into the Bayesian analysis appropriately? And, is it possible to make a clinical decision for future practice if there are several different prior distributions considered to be plausible?

The chapter is structured as follows: the next section introduces the background to the trial, including a description of the condition and the two arms of the trial; section 3.3 outlines the statistical methodology that is applied to the trial within a frequentist and Bayesian framework and provides a description of the various prior distributions that are applied within the Bayesian analysis; section 3.4 provides the results, and compares the two analysis approaches; and section 3.5 concludes with some discussion of the findings.

3.2 Background to the trial

The PLUTO trial was a multi-centre randomised control trial (RCT) in the UK, Ireland, and the Netherlands, that investigated the use of a vesico-amniotic shunt (VAS) compared to conservative care in male foetuses for the treatment of congenital lower urinary tract obstruction (LUTO).^{42,152,153} The primary clinical objective was to assess if there is any difference in the mortality rates, up to one year, in babies with congenital LUTO between those that received the vesico-amniotic shunt compared to those that received conservative care.

3.2.1 Fetal lower urinary tract obstruction

LUTO is a rare birth defect that occurs in approximately 2.2 out of 10000 births and is most commonly developed in male foetuses.^{153,154} The cause is a partial or complete obstruction of the urethra, which is the tube that connects the bladder to the amniotic fluid space around

the foetus. The flow of urine is prevented from passing from the bladder into the amniotic fluid space, which negatively impacts on the development of the foetus. Urine is an important component of amniotic fluid during gestation and contributes to fetal lung development.¹⁵⁵ If the urethra is completely blocked the urine cannot be released into the amniotic fluid space which causes the bladder to enlarge. If the amniotic fluid level is decreased the lungs may not develop properly (pulmonary hypoplasia), which is potentially life-threatening. The amniotic fluid also provides a cushioning for the foetus in the uterus, therefore low levels of fluid can lead to deformations of the face and extremities due to the pressure from the walls of the uterus. The kidneys can also become damaged due to back pressure caused by the obstruction.¹⁵⁵ Diagnosis of congenital LUTO is usually made by the routine ultrasound in the mid-second trimester (20 week scan). Further evaluation determines whether prenatal intervention may be beneficial to the foetus. If cysts are present on the kidneys this is an indication that there is likely irreversible kidney damage in which case fetal intervention will not be performed.¹⁵⁵

3.2.2 Inclusion criteria

The inclusion criteria were consenting pregnant women who has a singleton male foetus with a diagnosis of LUTO (diagnosed on the basis of the results in the ultrasound scans). The obstruction of the lower urinary tract had to be complete and isolated, and the foetus was presenting without any other abnormalities. Inclusion was also for those about whom the clinician was uncertain as to the optimum management.^{42,153} Female foetus' were excluded because they are more likely to have a complex cause of the condition and a very poor prognosis. All participants provided written informed consent.

3.2.3 Randomisation

Twenty-one fetal medicine centres in England, Scotland, Ireland, and the Netherlands recruited women to the trial. Pregnancies were allocated to the intervention or conservative management by a telephone and web-based randomisation service by the University of Birmingham Clinical Trials Unit. Randomisation by minimisation was used to achieve balance in the baseline characteristics between the treatment groups for gestational age at randomisation (<24 or ≥24 weeks), amniotic fluid volume (≤5th centile or >5th centile), and mother's age (<20, 20 to 35, or >35 years). Allocation could not be masked from the participant or the clinician because of the surgical nature of the trial. All vesico-amniotic shunts were inserted within seven days of randomisation.

3.2.4 Vesico-amniotic shunting

The VAS is a hollow tube that temporarily bypasses the obstruction in the lower urinary tract and provides an alternative path for the urine to pass from the bladder to the amniotic fluid space around the foetus.¹⁵⁵ This allows the amniotic fluid to build up, which is vital for the development of the lungs, and also helps prevent any further damage to the bladder and kidneys, which has been caused by the blockage. The procedure is performed by inserting a hollow needle through the mother's abdomen and uterus and into the fetal bladder. The VAS is passed through the needle and placed in the bladder.¹⁵⁵ After insertion of the VAS, follow-up scans were arranged no less frequent than every four weeks. Before the PLUTO trial, there have been no large-scale prospective RCTs assessing the risks and benefits of the VAS in this setting.

3.2.5 Conservative care

Conservative care in the PLUTO trial was the alternative treatment option that was compared to the intervention arm (VAS). This was a non-interventional care that included regular scans of the mother and foetus, no less frequently than every four weeks, which monitored the development of the foetus.

3.2.6 Failure to recruit

The original sample size calculation for the PLUTO trial was based on a meta-analysis of observational studies that compared the shunt with conservative care.¹⁵⁶ The target sample size was 150 pregnancies however, after the first phase of recruitment, only 31 pregnancies had been enrolled in the RCT and so accrual was halted. The trial failed to recruit for many reasons including the rare nature of the condition, the large number of women that opted for a termination of pregnancy, and the ethical obligation for clinicians to provide the best standard of care, which meant that many patients could not be randomised.

3.2.7 Outcome measures

Mothers entered the study on the date of randomisation and the endpoint is one of the following: death of the foetus, termination of the pregnancy, or the end of follow-up. Patients were censored due to loss of follow-up, if they had not died before the end of the first follow-up phase (all babies had at least one year follow-up by 1st January 2012), or if they were a termination of pregnancy that was not considered to be due to failure of the intervention. The survival time is measured as the length of time in the study, in weeks, from the estimated date of conception. The presumed conception date is calculated from the estimated date of delivery minus 267 days (date of the last menstrual period) plus 14 days. The survival time is measured from the date of conception (rather than date of randomisation) because the

gestational age of the foetus' varied when they were diagnosed (and thus randomised), but the detection of congenital LUTO at an earlier gestational age could be associated with a higher risk of mortality. It was possible to have a common start of follow-up for all foetus' by estimating the presumed date of conception.

3.2.8 Data

The analysis of the PLUTO trial was conducted assuming a censor date of 1st January 2012. All babies had been followed up for at least one year from conception but follow-up is still ongoing beyond this date. Information was collected on a number of variables at baseline, which are displayed in Table 3.1.

Table 3.1: Dataset variables with descriptions.

Variable	Description
Trial number	Patient identifier
Mother's age	Mother's age at conception of baby Continuous; years
Gestational age	Gestational age at randomisation Continuous; weeks
Liquor volume	Volume of amniotic fluid at randomisation by maximum pool depth (MPD) Binary; \leq 5 th centile ¹⁵⁷
Renal pelvis dilation	Transverse length of renal pelvis Binary; \leq 90 th centile
Kidney length	Length of left and right kidney Binary; \leq 90 th centile
Renal cysts	Renal cysts present Categorical; unilateral, bilateral or none
Bladder wall thick	Thick bladder wall Binary, \leq 3mm
Echogenicity	Increased echogenicity of the kidney Categorical; unilateral, bilateral or no
Renal pelvis hydronephrosis	Kidney swollen or stretched as a result of a build-up of urine inside the kidneys Categorical; unilateral, bilateral or no
Survival time	Time from conception to death of foetus, termination of pregnancy or end of follow-up Continuous; weeks

3.3 Methods

The analysis of the trial was conducted using both frequentist and Bayesian analysis methods. A survival analysis was performed to determine whether there is a difference in mortality rates up to one year from conception between those that received the vesico-amniotic shunt and those that received conservative care. An intention-to-treat and an as-

treated analysis were considered in the publication of the results,^{42,153} however only the results from an intention-to treat assumption are reported in this chapter.

Initially, a Kaplan-Meier plot for treatment arm was produced to show unadjusted differences in the probability of survival over time. The semi-parametric Cox proportional hazards (PH) model¹⁵⁸ was then fitted using classical and Bayesian methods¹⁵⁹ to obtain estimates of the ratio of mortality rates to compare the two randomisation arms. The Cox PH regression model is shown in equation (3.1), where $h(t;\mathbf{X})$ is the hazard rate for an individual, i , at time t , with covariate vector $\mathbf{X}=(X_1, X_2, \dots, X_p)$.

$$h(t;\mathbf{X})=h_0(t)\exp\left(\sum_{i=1}^p \beta_i X_i\right), \quad t \geq 0 \quad (3.1)$$

Here, $h_0(t)$ denotes the baseline hazard, which is unknown, and β_i denotes the regression coefficients (log hazard ratios) of interest in relation to the covariates X_i . Before the analysis it was decided that, although there are possible covariates of interest, the sample size was too small to adjust for them.¹⁶⁰ Thus, only the intervention variable was fitted in the model, where $X_i=0$ for those subjects that were randomised to conservative care, and $X_i=1$ for those that were randomised to VAS. For all results in this chapter, a hazards ratio (HR) that is less than one indicates that those in the VAS group have a lower mortality rate compared to those in the conservative care group.

3.3.1 Proportional hazards assumption

The Cox proportional hazards model in (3.1) assumes that the hazard ratio ($\exp(\beta)$) for a variable of interest, X , remains constant over time,¹⁵⁸ i.e. the hazard rates for those in the VAS group are proportional to those in the conservative care group.¹⁶¹ To check whether the

hazard rates are proportional, the log cumulative hazard ($\log(-\log(S(t)))$) was plotted against time (Figure 3.1), where $S(t)$ is the probability of surviving until at least time t .

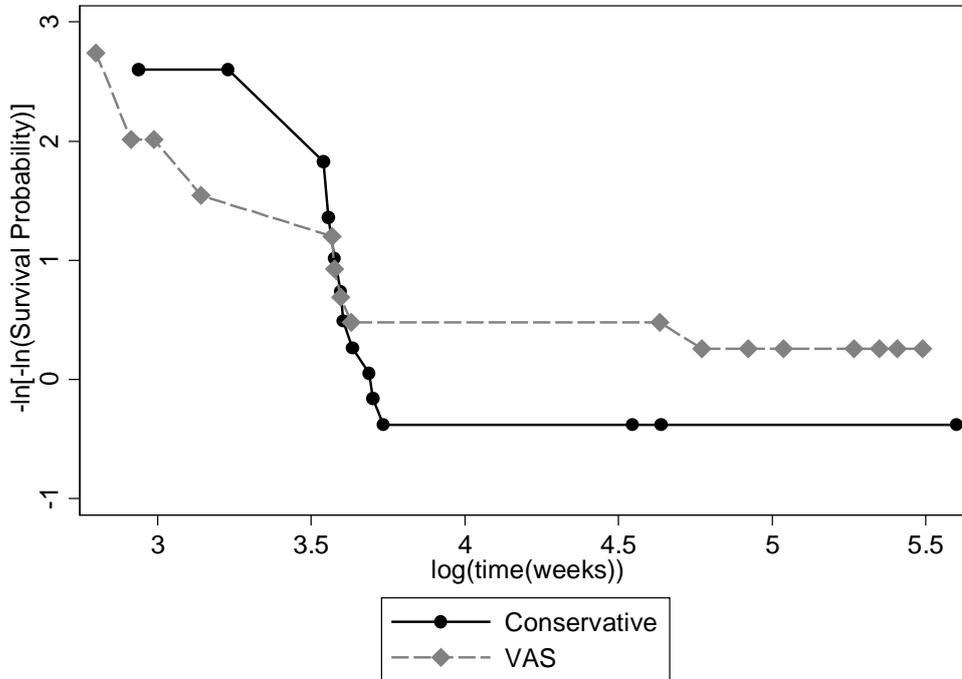


Figure 3.1: Log cumulative hazard plot for intervention arm under ITT analysis.

Figure 3.1 shows that the hazard rates for the treatment groups may not be proportional because the curves cross. However, the dataset is very small and so observed differences may simply be due to chance. Hence, the analysis was conducted under two assumptions. The first analysis assumes that the hazards are proportional over the entire length of follow-up, therefore the hazard ratio estimate is an average of the treatment effect over time. The second assumption is that the relationship that is displayed in Figure 3.1 is the true relationship and therefore the survival data was split at the time point when the lines cross (at 36.5 weeks, which is approximately the time of birth) and a Cox proportional hazards model was fitted with two coefficients for treatment effect (one for before 36.5 weeks [before birth], and one for after 36.5 weeks [after birth]). In each of the two time periods the hazard rates

are assumed proportional. In order to implement this model, a new binary variable, X_2 , was created to allow for the interaction between the time period and the intervention variable. For the first time interval $X_2=0$, and for the second time interval $X_2=1$. Thus, the Cox PH regression model can now be written as in equation (3.2):

$$h(t;\mathbf{X})=h_0(t)\exp(\beta_1X_1+\beta_2X_1X_2) \quad (3.2)$$

Here, X_1 denotes the intervention variable and $\exp(\widehat{\beta}_1)$ is the estimated hazard ratio that compares the intervention groups when $X_2=0$, i.e. the hazard ratio from conception to 36.5 weeks (expected time of birth). Therefore, $\exp(\widehat{\beta}_1+\widehat{\beta}_2)$ is the estimate of the hazard ratio for the intervention versus conservative care when $X_2=1$, i.e. the HR from 36.5 weeks (after expected time of birth) to the end of follow up.

3.3.2 Bayesian survival analysis with the poisson counting process notation

The Cox regression model using a Bayesian approach can be specified in numerous ways, such as the piecewise constant model, gamma process models or beta process models.¹⁶² In this chapter, the Cox proportional hazards model is formulated using the poisson counting process notation, which is a piecewise constant model.¹⁵⁹ With this notation, for subjects $i=1,\dots,n$, $N_i(t)$ denotes the processes that count the observed events that have occurred up to time t . $I_i(t)$ is the corresponding intensity process given by (3.3)¹⁶³

$$I_i(t)dt=E(dN_i(t)|F_t) \quad (3.3)$$

Where $dN_i(t)$ is the change in N_i in the time interval $[t, t+dt]$ and F_t denotes the data that is available just before time t . Therefore, $I_i(t)dt$ is the probability that subject i has the event of interest in this interval, since $dN_i(t)$ equals one if subject i experiences the event of interest,

and otherwise $dN_i(t)$ equals zero. As dt tends to zero the probability becomes the instantaneous hazard for subject i at time t . The intensity process is assumed to have proportional hazards form:¹⁶³

$$I_i(t) = Y_i(t)\lambda_0(t)\exp\left(\sum_{i=1}^p \beta_i X_i\right) \quad (3.4)$$

So, $Y_i(t)$ is one if subject i is in the risk set at time t and zero otherwise, and thus accounts for those subjects that have been censored. The familiar Cox model is $\lambda_0(t)\exp(\sum_{i=1}^p \beta_i X_i)$. The joint posterior distribution for the regression coefficients, β , and the underlying cumulative hazard function, $\Lambda_0(t)$, is derived using Bayes theorem and is shown in equation (3.5).¹⁶³

$$P(\beta, \Lambda_0()|D) \propto P(D|\beta, \Lambda_0())P(\beta)P(\Lambda_0()) \quad (3.5)$$

D denotes the data which includes the processes, $N_i(t)$, the covariate values and $Y_i(t)$. $P(\beta)$ is the prior distribution for the regression coefficients, $P(\Lambda_0())$ is the prior distribution for the underlying hazard function, and the likelihood is denoted by $P(D|\beta, \Lambda_0())$. Based on the assumption that censoring is non-informative, the likelihood is the same as if the counting process increments $dN_i(t)$ for the interval $[t, t+dt]$ are independent Poisson random variables with means $I_i(t)dt$, so $dN_i(t) \sim \text{poisson}(I_i(t)dt)$. The intensity process can be written as:¹⁶³

$$I_i(t)dt = Y_i(t)\exp(\beta X_i)d\Lambda_0(t) \quad (3.6)$$

In this equation, $d\Lambda_0(t)$ is the change in the cumulative baseline hazard function during the interval $[t, t+dt]$. It is assumed that $d\Lambda_0(t)$ are distributed by a gamma distribution, $d\Lambda_0(t) \sim \text{gamma}(d\Lambda^*_0(t), c)$, where $d\Lambda^*_0(t)$ is a prior belief about the unknown hazard function and c denotes the uncertainty surrounding this prior belief.¹⁶³ This assumption is convenient since the conjugate prior for the poisson likelihood is the gamma distribution. For the analysis

of the PLUTO trial, $\Lambda_0^*(t)=rdt$ where r is a guess at the failure rate per unit time and dt is the size of the time interval between failure times and the prior assumption of r was 0.1.¹⁵⁶ The value c represents the degree of confidence in this assumption (small values of c represent weak prior belief) thus $c=0.001$ since there was no prior information to inform this value.

3.3.3 Prior distribution for the treatment effect estimate

For the Bayesian analyses, prior distributions need to be specified for the estimate of treatment effect, β_1 (log(HR)). The incorporation of prior evidence may be useful to support the findings in the trial, which is largely underpowered due to its small sample size. However, it is also important to determine whether the prior distributions may be *too* informative in this dataset when there is so little data. Three off-the-shelf prior distributions (vague, enthusiastic and sceptical) were considered that used evidence from previous observational studies. Four subjective prior distributions using elicited prior beliefs from specialist clinicians were also derived prior to the trial.^{153,164}

Derivation of off-the-shelf prior distributions

To construct the off-the-shelf priors we followed methods that have been outlined by others and the method is now described.¹⁶⁵ All prior distributions for the log hazard ratio were assumed to be normally distributed. The enthusiastic prior was centred at the alternative hypothesis that PLUTO was designed to test, which was based on the results from a previous meta-analysis of observational studies.¹⁵⁶ The original meta-analysis pooled odds ratios across four studies, however the prior distribution for this trial is for a log relative risk (log(RR)). Of course, a prior distribution for the log(RR) is not a perfect match for a log(HR), however, it was considered a suitable prior distribution for the log HR by assuming that the HR is constant and instantaneous for the entire period of follow-up. Therefore, the pooled RR was derived from the summary data from the four studies using the frequentist fixed-effect

meta-analysis model described in section 1.4.1. The mean of the prior distribution is a HR of 0.646 ($\log(\text{HR})=-0.437$). The standard deviation for this prior distribution was derived by computing the value for which the enthusiastic prior has a 5% probability that the log hazard ratio favours conservative care.²² Thus, the enthusiastic prior was $\log(\text{HR})\sim N(-0.437, \text{sd}=0.266)$. The sceptical prior distribution was constructed similarly, but centred at the value of no difference (i.e. $\log(\text{HR})=0$, and so harmful effects possible) and with a 5% probability that the treatment effect was greater than a $\log(\text{HR})$ of -0.437. This prior distribution for the $\log(\text{HR})$ was therefore $N(0, \text{SD}=0.266)$. The sceptical and enthusiastic prior distributions for the $\log(\text{HR})$ and HR scales are displayed in Figure 3.2. The vague prior distribution for the $\log(\text{HR})$ was centred at the value of no effect (i.e. $\log(\text{HR})=0$) with a very large variance so $\log(\text{HR})\sim N(0, \text{sd}=1000^2)$.¹⁶⁶

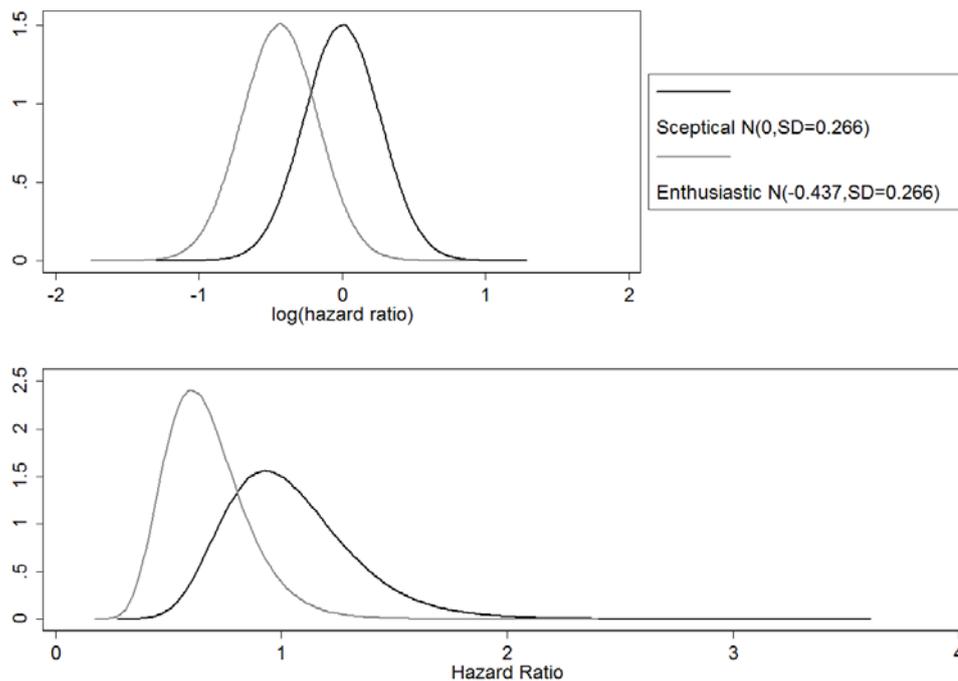


Figure 3.2: Sceptical and enthusiastic prior distributions for $\log(\text{HR})$ and HR for VAS compared to conservative care.

Elicited prior distributions

The elicited prior distributions were obtained prior to the start of the trial from consultant members of either of the British Maternal and Fetal Medicine Society, the British Association of Paediatric Nephrologists, or the British Association of Paediatric Urologists.¹⁶⁴ These three groups are referred to as fetal medicine (FM) experts (35 specialists), paediatric urologists (PU) (5 specialists) and paediatric nephrologists (PN) (10 specialists). Details of how the prior distributions were derived from each clinician, and how the prior distributions were combined, have been described extensively elsewhere.¹⁵³ Figure 3.3 displays the prior distributions by speciality and for all specialists (50 specialists) combined. All elicited prior distributions are centred on a treatment effect that favours the intervention but include the harmful effects of the intervention. The fetal nephrologists are the least enthusiastic regarding the effect ($\log(\text{HR}) \sim N(-0.03, \text{sd}=0.15)$). The fetal medicine experts and the all experts prior distributions are very similar because 35 out of 50 (70%) of the experts were fetal medicine experts. All seven prior distributions are summarized in Table 3.2.

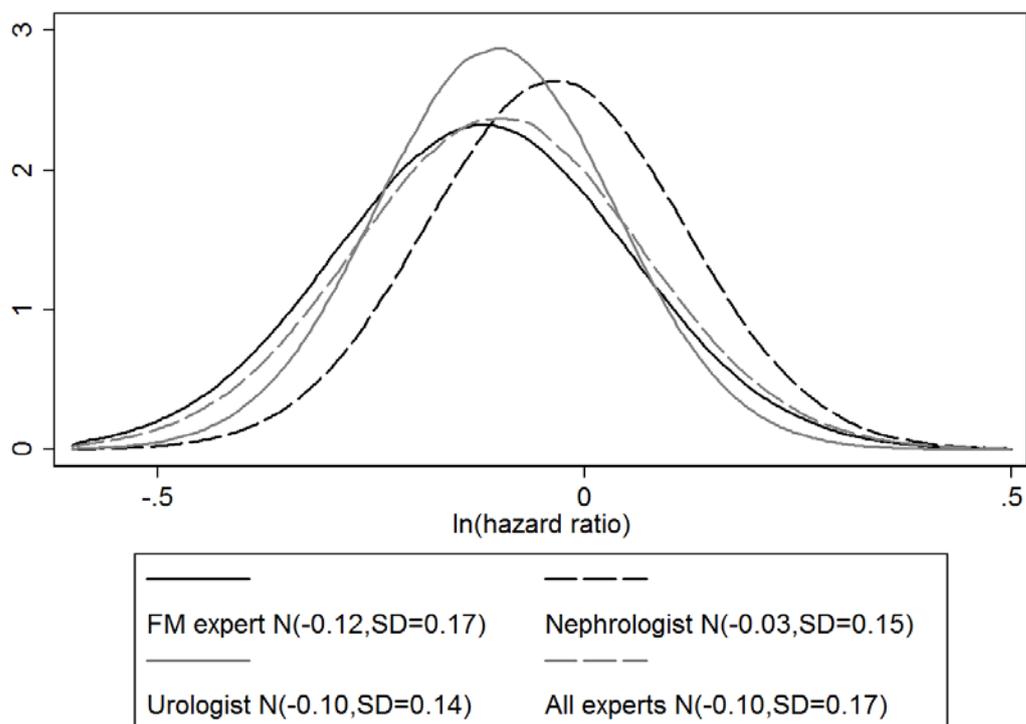


Figure 3.3: Elicited prior distribution for log(HR) for intervention compared to conservative treatment.

Table 3.2: Prior distributions for log(HR) for intervention versus conservative care.

	Prior distribution for log(HR)
Vague	$N(0, sd=1000)$
Sceptical	$N(0, sd=0.27)$
Enthusiastic	$N(-0.44, sd=0.27)$
Fetal medicine experts	$N(-0.12, sd=0.17)$
Paediatric nephrologists	$N(-0.03, sd=0.15)$
Paediatric urologists	$N(-0.10, sd=0.14)$
All experts	$N(-0.10, 0.17)$

3.3.4 Posterior probability estimates

The Bayesian framework can be used to additionally estimate posterior probabilities of effect sizes greater (or less than) some clinically meaningful value, which may aid clinical decision-making. The PLUTO trial was designed to detect a hazard ratio of 0.646 ($\log(\text{HR})=-0.437$) from the previous meta-analysis,¹⁵⁶ which is described in section 3.3.3. A discussion was held with clinicians to decide what difference they would consider as a clinically meaningful difference between the mortality rates in the VAS and conservative care treatment groups. They decided that the estimate obtained in the meta-analysis was an appropriate difference to observe. Therefore, the posterior probability that the $\text{HR}<0.646$ was considered the clinically meaningful difference. However, the clinically meaningful difference of 0.646 is a large difference and therefore posterior probability estimates for the $\text{HR}<1$ were also calculated to show the probability that the mortality rates for those on the VAS arm are lower compared to the mortality rate for those that received conservative care.

3.3.5 Implementation

The frequentist survival analyses and the exploration of the data, including the Kaplan-Meier plots and log cumulative hazard plots, were performed using Stata version 12.1.¹⁶⁷ The frequentist estimation of the hazard ratios used maximum likelihood estimation and there were no tied events. The Bayesian survival analyses were performed using MCMC methods using Gibbs sampling within WinBUGS version 1.4.³² The posterior distributions were assessed to check that they had converged to a stationary distribution by inspecting the trace and history plots. Auto-correlation was also checked by looking at the auto-correlation plots. The results were obtained with a burn in length of 20,000 iterations and a sampling length of 100,000 iterations. Bayesian summary estimates provided are the median and 95% credible intervals (CrI) from the posterior distributions.

3.4 Results

3.4.1 Summary of data

Table 3.3 is a table of baseline characteristics by intervention group. Fifteen subjects were randomised to the conservative care arm and 16 to the VAS arm. Of the 15 babies that were randomised to conservative care, 3 (20%) were alive at the end of follow-up, 2 (13.33%) were termination of pregnancies, and 10 (66.67%) died. Of the 16 babies randomised to the VAS arm, 7 (43.75%) were alive at the end of the study, 3 (18.75%) were termination of pregnancies and 6 (37.50%) died.

Despite the small sample size, the differences between the intervention groups were no more than expected by chance. However, for some characteristics the magnitude of the difference was quite large. For example, the range for gestational age for those randomised to VAS was 14 to 32 weeks, whereas the range was much narrower for those randomised to conservative care (16 to 24 weeks). Also, the mother's age is higher for those randomised to conservative care (29.1 years) compared to those on the VAS (27.7 years). If these characteristics are truly associated with a higher risk of mortality there could be some bias in the estimate of treatment effect that is not adjusted for these variables.

Out of the five mothers that had a termination of pregnancy, two were caused by a spontaneous rupture of membranes, believed to be a result of the shunt. Therefore, these two termination of pregnancies were analysed as deaths for the primary analysis and the remaining three termination of pregnancies were analysed as censored observations. There was little sensitivity to the classification of the termination of pregnancies, which has been investigated and reported elsewhere.¹⁵³

Table 3.3: Summary of key baseline characteristics, outcome and length of follow-up by randomised intervention group.

		VAS (n=16)	Conservative care (n=15)	Overall (n=31)
Baseline characteristics				
Mother's age, years	Mean (SD)	27.68 (7.16)	29.05 (6.12)	28.35 (6.60)
	Range	15.7 to 42.3	20.1 to 42.4	15.7 to 42.4
Gestational age, weeks	Mean (SD)	20.30 (4.74)	20.80 (2.30)	20.54 (3.71)
	Range	14 to 32	16 to 24	14 to 32
Liquor volume, n (%)	<5 th centile	10 (62.50)	9 (60.00)	19 (61.29)
	>5 th centile	6 (37.50)	6 (40.00)	12 (38.71)
Left kidney length, n (%)	<90 th centile	4 (26.67)	4 (26.67)	8 (26.67)
	>90 th centile	11 (73.33)	11 (73.33)	22 (73.33)
Right kidney length, n (%)	<90 th centile	1 (6.67)	5 (33.33)	6 (20.00)
	>90 th centile	14 (93.33)	10 (66.67)	24 (80.00)
Renal cysts present, n(%)	No	14 (87.50)	10 (66.67)	24 (77.42)
	Unilateral	2 (12.50)	4 (26.67)	6 (19.35)
	Bilateral	0 (0.0)	1 (6.67)	1 (3.23)
Bladder wall thick, n (%)	Yes	6 (42.86)	9 (64.29)	15 (53.57)
	No	8 (57.14)	5 (35.71)	13 (46.43)
Echogenicity, n (%)	No	7 (58.33)	5 (41.67)	12 (50.00)
	Unilateral	3 (25.00)	3 (25.00)	6 (25.00)
	Bilateral	2 (16.67)	4 (33.33)	6 (25.00)
Hydronephrosis, n (%)	No	13 (86.67)	13 (92.86)	26 (89.66)
	Unilateral	1 (6.67)	1 (7.14)	2 (6.90)
	Bilateral	1 (6.67)	0 (0.0)	1 (3.45)

Table 3.3: continued.

		VAS (n=16)	Conservative care (n=15)	Overall (n=31)
Follow-up				
Outcome, n(%)	Death	6 (37.50)	10 (66.67)	16 (51.61)
	Termination of pregnancy*	3* (18.75)	2 (13.33)	5 (16.13)
	Censored	7 (43.75)	3 (20.00)	10 (32.26)
Length of follow-up, person-weeks	Total	1568.71	868.29	2436.99

* Two of these terminations were analysed as deaths in the primary analysis of the trial.

3.4.2 Survival analyses

The Kaplan-Meier unadjusted survival curves by intervention arm are shown in Figure 3.4. This plot suggests that, for approximately the first 36.5 weeks, those who received the VAS have a lower probability of survival compared to those that received conservative care. After this time, the probability of survival is higher for those that received the VAS. The probability of survival to one year from conception is approximately 23% for those that receive conservative care compared to approximately 54% for those that receive the VAS.

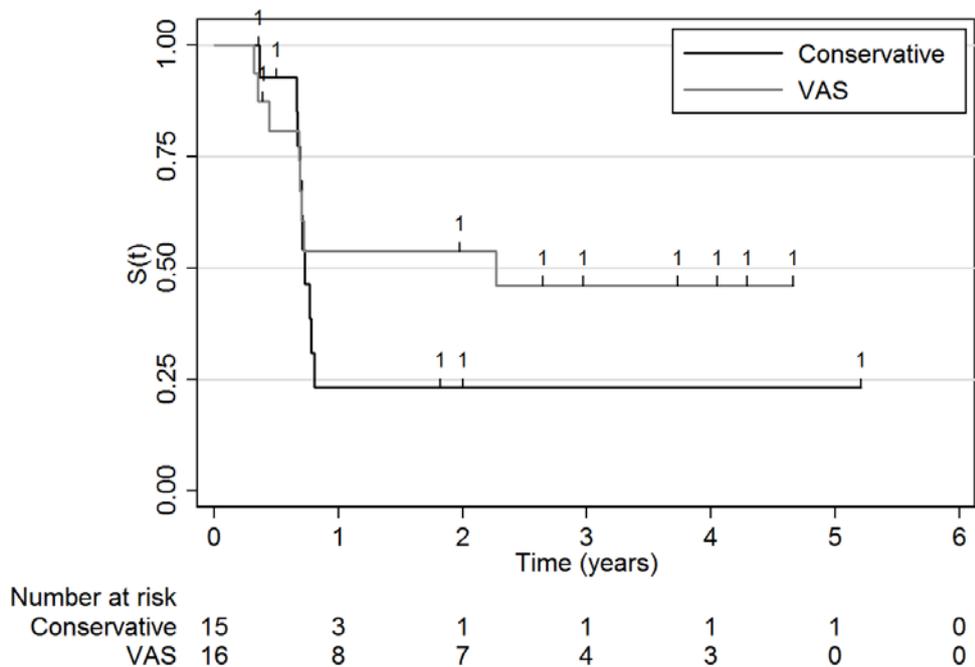


Figure 3.4: Kaplan-Meier plot of survival time from conception by intervention arm.

Results assuming the hazard rates are proportional from conception to end of follow-up

Table 3.4 presents the results under the first assumption that the hazard rates are proportional from conception to the end of follow-up, for both the frequentist and Bayesian approaches. This table includes the estimate of the HR with its 95% CrI for the Bayesian analyses, and the estimated HR with its 95% CI for the frequentist analysis. The Bayesian analyses also include two posterior probabilities: the estimated HR is less than one, and the estimated HR is less than 0.646 for all prior distributions for the log(HR).

Table 3.4: Bayesian and frequentist results from a univariable Cox model assuming proportional hazards for the entire length of follow-up.

Analysis	HR (95% CrI)	Prob (HR<1)	Prob (HR<0.646)
<i>Frequentist</i>	0.622 (0.241 to 1.605)*	-	-
<i>Bayesian</i>			
Vague: log(HR)~N(0,1000 ²)	0.616 (0.226 to 1.598)	0.840	0.542
Enthusiastic: log(HR)~N(-0.44,0.27)	0.640 (0.406 to 1.009)	0.973	0.517
Sceptical: log(HR)~N(0,0.27)	0.897 (0.570 to 1.411)	0.682	0.079
FM experts: log(HR)~N(-0.12,0.17)	0.850 (0.621 to 1.170)	0.843	0.044
PN: log(HR)~N(-0.03,0.15)	0.927 (0.700 to 1.230)	0.701	0.006
PU: log(HR)~N(-0.10,0.14)	0.874 (0.674 to 1.138)	0.842	0.011
All experts: log(HR)~N(-0.10,0.17)	0.865 (0.634 to 1.183)	0.818	0.032

* 95% CI with corresponding p-value=0.301; FM is fetal medicine experts; PN is paediatric nephrologists; PU is paediatric urologists; 0.646 was considered clinically meaningful from the original meta-analysis.¹⁵⁶

Assuming a vague prior distribution for the log(HR) within a Bayesian survival model, the posterior estimate of the HR is 0.62 (95% CrI: 0.23 to 1.60) suggesting an average reduction in mortality rate of 38.4% for those that receive the shunt compared to those that received conservative care. The frequentist estimate of the hazard ratio is 0.62 (95% CI: 0.24 to 1.61),

which corresponds to a mortality rate that is lower for those that received the shunt compared to the conservative care by 37.8%. As expected, these results are very similar to the Bayesian results with a vague prior distribution.

Perhaps surprisingly, the posterior estimate of the HR under the enthusiastic prior is a little closer to the null compared to the vague prior analysis (HR=0.64, 95% CrI: 0.41 to 1.01), however the 95% credible interval is narrower compared to the 95% CrI for the vague prior distribution analysis. As expected, the sceptical prior distribution draws the posterior distribution for the HR even closer towards the null compared to the enthusiastic prior distribution (estimated HR=0.90, 95% CrI: 0.57 to 1.41).

The posterior results with the elicited prior distributions are all similar to each other. The median of the posterior distribution suggests that the intervention reduces mortality rates compared to conservative care. For example, using the FM prior the estimated HR is 0.85 (95% CrI: 0.62 to 1.17). They all suggest a treatment effect that is closer to the value of no difference compared to the frequentist analysis and the Bayesian analyses with a vague or enthusiastic prior.

All Bayesian analyses indicate a strong probability that the VAS is effective (i.e. HR<1) with estimates of this probability greater than 0.7 for all prior distributions. For example, this probability is as high as 0.97 with the enthusiastic prior distribution. However, there is no strong evidence that the VAS is clinically relevant (i.e. HR<0.646) since even the enthusiastic prior analysis estimates this probability to be 0.52, and this probability is as low as 0.01 with the paediatric nephrologist prior distribution.

A plot of the posterior distributions using the three off-the-shelf and the four elicited prior distributions is shown in Figure 3.5. Figure 3.6 shows the prior distributions overlaid with the

updated posterior distributions. The posterior distributions are very close to their prior distributions, which is expected because there is very little data to update either the location or variance of the prior distribution.

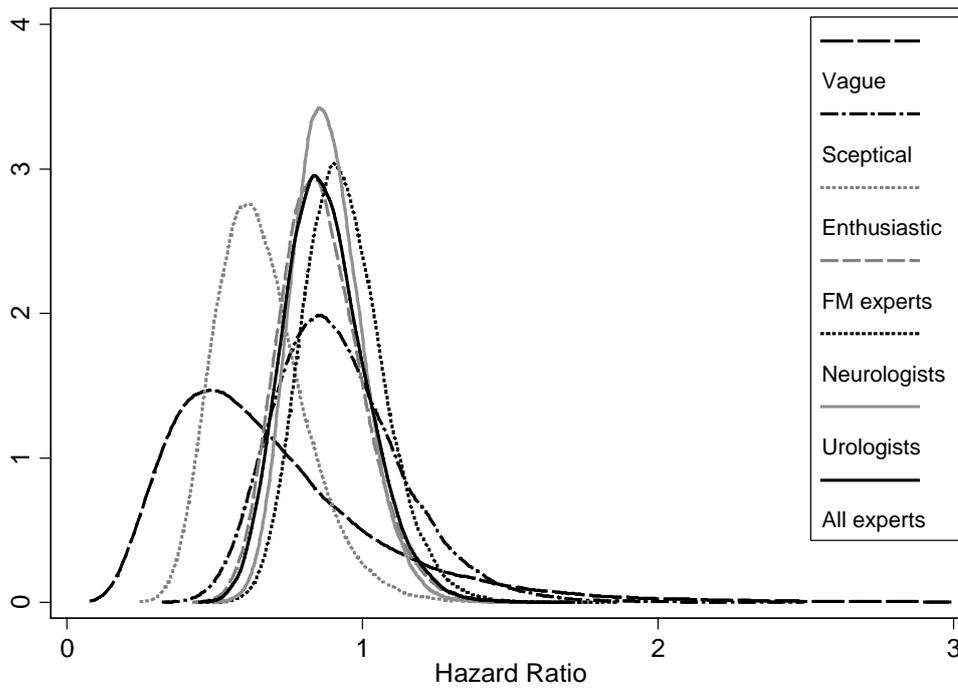


Figure 3.5: Posterior distributions for HRs using all prior distributions and assuming proportional hazard rates.

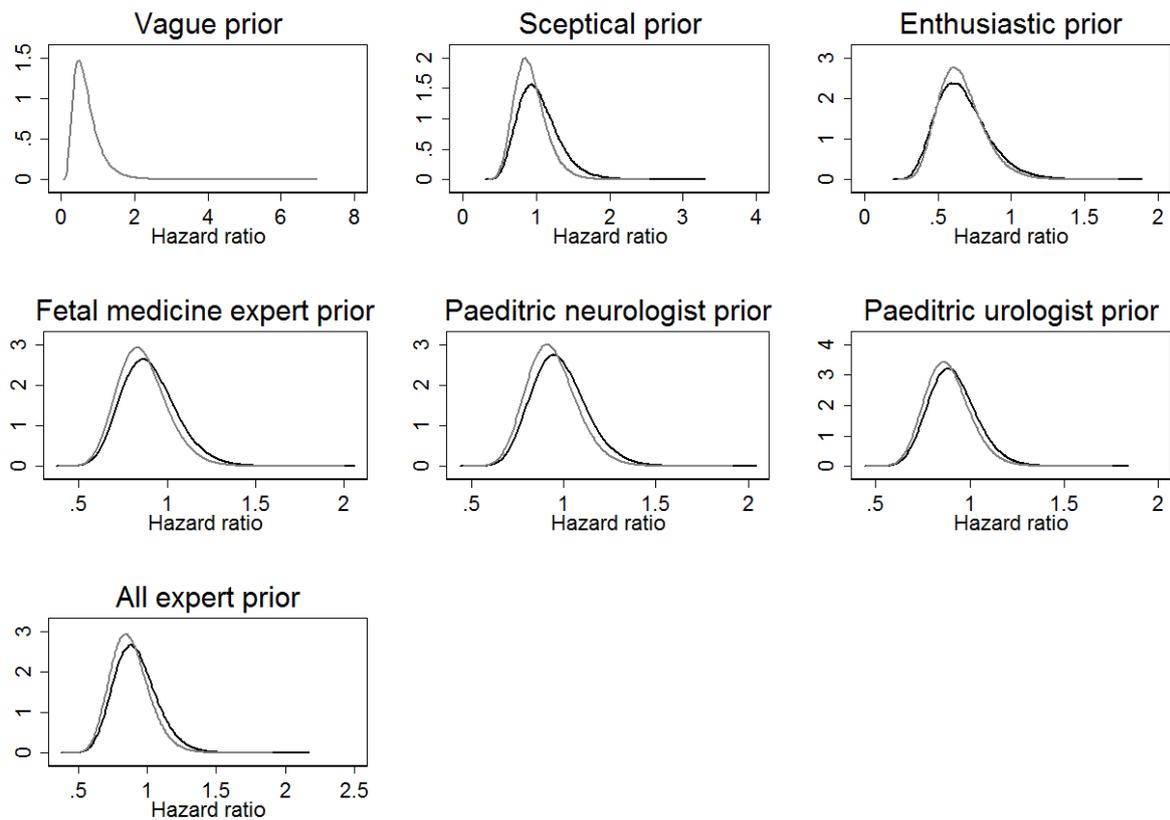


Figure 3.6: Prior and posterior distributions overlaid.

(HR scale; black: prior; grey: posterior) assuming proportional hazard rates

Results with a time-dependent treatment effect

The Bayesian and frequentist results assuming the time-dependent treatment effect are shown in Table 3.5. HR1 is the estimate of the treatment effect between conception and 36.5 weeks (expected time of birth), and HR2 is the estimate of the treatment effect between 36.5 weeks and the end of follow-up. Figure 3.7 displays the prior distributions overlaid with their updated posterior distributions for the estimate of HR1. Figure 3.8 is the corresponding figure for HR2.

Table 3.5: Posterior estimates of unadjusted HR for death given time-dependent treatment effect and Bayesian p-values.

Framework	HR1 (95% CrI)	Prob(HR1 <1)	Prob(HR1 <0.646)	HR2 (95% CrI)	Prob(HR2 <1)	Prob(HR2 <0.646)
<i>Frequentist*</i>	1.116 (0.340 to 3.657)†	-	-	0.217 (0.039 to 1.207)\$	-	-
<i>Bayesian</i>						
Vague: $\log(\text{HR}) \sim N(0, 1000^2)$	1.111 (0.328 to 3.956)	0.435	0.186	0.215 (0.027 to 1.101)	0.970	0.904
Enthusiastic: $\log(\text{HR}) \sim N(-0.44, 0.27)$	0.705 (0.438 to 1.137)	0.927	0.358	0.583 (0.360 to 0.948)	0.984	0.659
Sceptical: $\log(\text{HR}) \sim N(0, 0.27)$	1.018 (0.635 to 1.642)	0.478	0.030	0.866 (0.527 to 1.411)	0.722	0.124
FM experts: $\log(\text{HR}) \sim N(-0.12, 0.17)$	0.903 (0.653 to 1.248)	0.732	0.022	0.836 (0.601 to 1.161)	0.862	0.063
PN: $\log(\text{HR}) \sim N(-0.03, 0.15)$	0.974 (0.731 to 1.300)	0.572	0.002	0.919 (0.687 to 1.230)	0.718	0.009
PU: $\log(\text{HR}) \sim N(-0.10, 0.14)$	0.910 (0.699 to 1.186)	0.756	0.005	0.864 (0.661 to 1.131)	0.857	0.017
All experts: $\log(\text{HR}) \sim N(-0.10, 0.17)$	0.918 (0.668 to 1.262)	0.699	0.015	0.850 (0.615 to 1.173)	0.839	0.049

* 95% CI for the HR estimates are presented for frequentist analysis; † the p-value corresponding to this estimate is 0.857; \$ the p-value corresponding to this estimate is 0.081.

The frequentist results suggest that there could be a harmful or beneficial effect of the VAS in the first time period (up to the expected time of birth [36.5 weeks]) with an estimated HR1 of 1.12 (95% CI: 0.34 to 3.66), but that there is perhaps a beneficial effect of the intervention after this time period with an estimated HR2 of 0.22 (95% CI: 0.04 to 1.21). However, these results are not statistically significant. Similar to the proportional hazards assumption over the whole year, the Bayesian results are very similar to the frequentist results for the vague prior distribution analysis. The posterior estimate of HR1 is 1.11 (95% CrI: 0.33 to 3.96), and the estimate of HR2 is 0.22 (95% CrI: 0.03 to 1.10). The Bayesian posterior probability estimate that $HR1 < 1$ is only 0.44, however the probability that $HR2 < 1$ is much higher with an estimate of 0.97.

Only the results with an enthusiastic prior distribution suggests that the intervention improves survival rates from time of birth (36.5 weeks) onwards with an estimated HR2 of 0.58 (95% CrI: 0.36 to 0.95). However, the posterior probability estimates that the $HR2 < 1$ are above 0.7 for all analyses, which may aid the update of the guidelines for clinical practice.

The posterior distributions vary largely depending on the choice of prior distribution when there is such little data to update the prior distribution. For example, with a vague prior distribution, the estimated HR2 is 0.22 (95% CrI: 0.04 to 1.21) and the posterior probability that $HR2 < 1$ is 0.97. However, the estimated HR2 with the paediatric nephrologist prior distribution is 0.92 (95% CrI: 0.69 to 1.23) with an estimate of the probability $HR2 < 1$ equal to 0.72. If the decision regarding clinical practice is just based on the evidence about HR2 in the data (i.e. the vague prior or the frequentist analysis), then they may decide to use the VAS. However, the evidence based on HR2 from the data combined with the paediatric nephrologist prior evidence suggests that the VAS may not provide a benefit above the conservative care, or it may be harmful.

The posterior distributions for the elicited prior analyses are particularly similar to their prior distributions for HR1 (Figure 3.7). This is because there is very little data to update the prior distribution compared to the relatively small standard deviations in the prior distributions. For example, the fetal medicine experts' prior distribution is $\log(\text{HR}) \sim N(-0.12, \text{sd}=0.17)$. The posterior distribution for HR1 is $\log \text{HR1} \sim N(-0.10, \text{sd}=0.17)$.

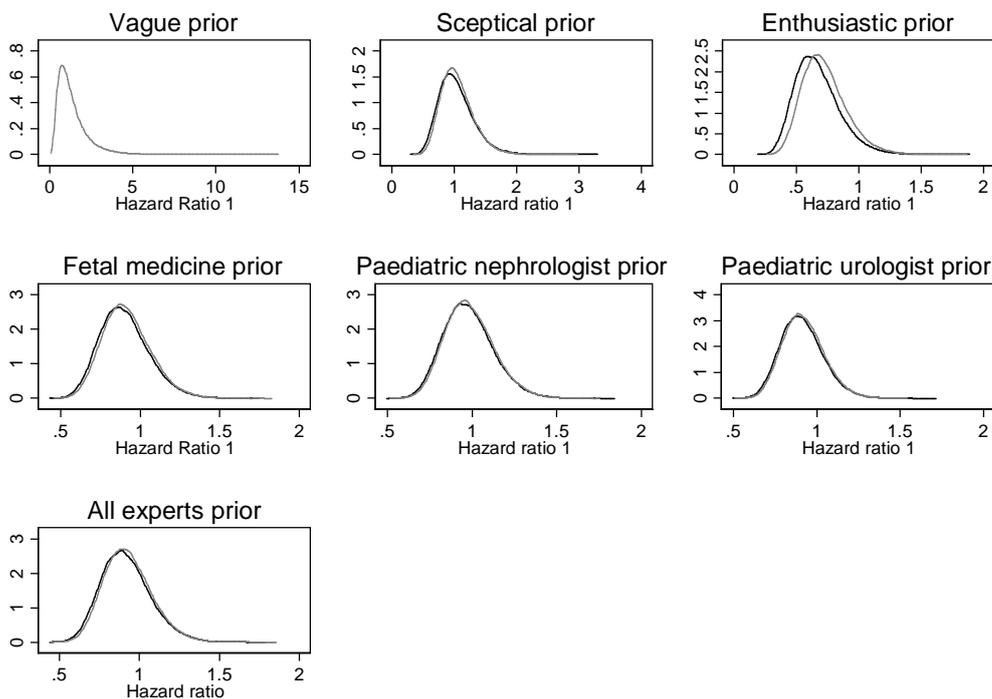


Figure 3.7: Prior and posterior distributions for HR1 (period conception to (expected time of birth) 36.5 weeks) overlaid.

(HR scale; black: prior; grey: posterior)

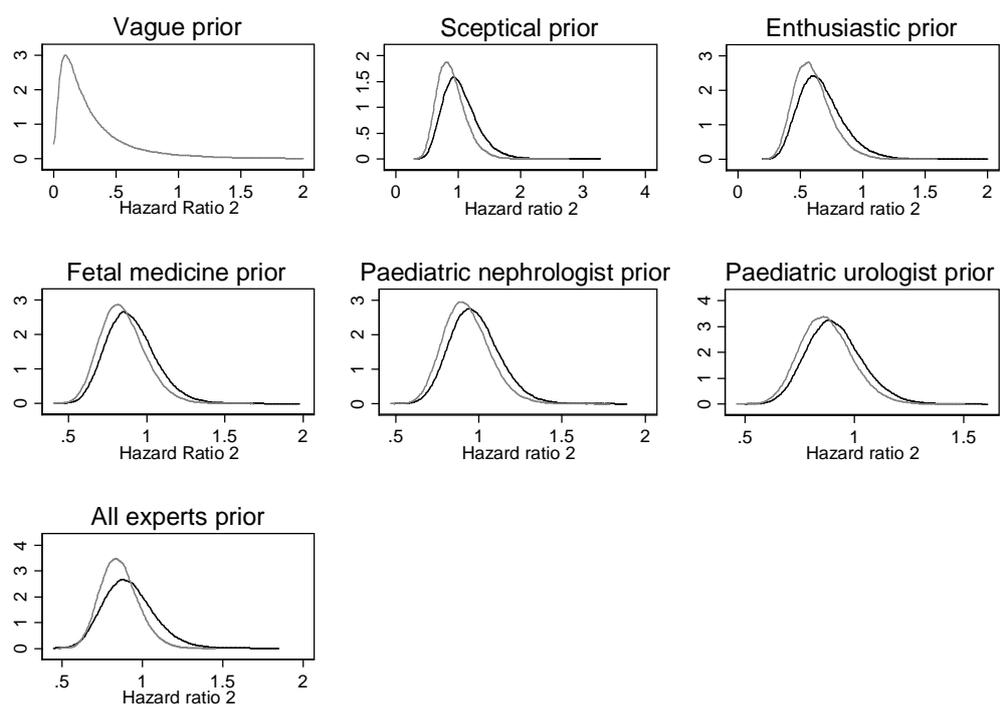


Figure 3.8: Prior and posterior distributions for HR2 (period 36.5 weeks (expected time of birth) to the end of follow-up) overlaid.

(HR scale; black: prior; grey: posterior)

3.5 Discussion

The PLUTO trial was a randomised surgical intervention trial that compared the insertion of a VAS to conservative care in foetus' with blocked lower urinary tracts. The trial failed to meet recruitment targets and closed after recruiting just 31 pregnancies. For several reasons, including the rare nature of the condition, another trial is unlikely to be more successful in terms of recruiting more patients. For this reason, and because the intervention was being used in practice under certain conditions prior to the trial, it is imperative to make the best use of the data from the PLUTO trial. Therefore, both frequentist and Bayesian survival methods have been adopted and compared. Within the Bayesian framework, several prior distributions for the treatment effect were incorporated to include previous evidence

regarding the effectiveness of the intervention. The key findings are shown in Figure 3.9 and are discussed in more detail below.

Figure 3.9: Key findings.

- There is strong evidence in all Bayesian analyses (assuming proportional hazards for the entire follow-up period) that the VAS is effective (i.e. $\text{prob}(\text{HR}<1)$ is high). But this effect is unlikely to be clinically relevant since the probability that $\text{HR}<0.646$ is low for all analyses. Further, it may be that VAS is harmful for the first 36.5 weeks (before expected time of birth) and the survival benefit of VAS is only after this point.
- A Bayesian framework can be useful to analyse a clinical trial with a small sample size to incorporate additional data in a prior distribution.
- It is difficult to assess proportional hazards in a dataset with very small numbers. Thus, several assumptions about the relationship between the hazard rates may be necessary.
- If prior distributions have small variances they may have a large influence on the posterior distributions when there are few events.
- When several, diverse prior distributions are explored it may be more difficult to make a clinical decision if the consequence is diversity in the posterior distributions. Thus, realistic prior distributions are preferred.

Under the proportional hazards assumption for the entire follow-up, the only analysis that resulted in convincing evidence that the shunting is beneficial was the Bayesian analysis with the enthusiastic prior (HR 0.640 95% CrI 0.406, 1.009), and the probability that the hazard ratio is less than one is 0.97. However, the probability that this difference is a clinically important difference (i.e. a posterior probability that the hazard ratio is less than 0.646) is just 0.52. Furthermore, under the sceptical prior, the posterior probability that the treatment effect is clinically important is just 0.08.

As expected, the results from the frequentist analysis under both assumptions regarding the proportional hazards closely mirrored the results from the corresponding Bayesian approach.

For example, under the proportional hazard assumption for the entire follow-up within a frequentist framework, the estimated HR is 0.62 (95% CI: 0.24 to 1.61). The corresponding estimate within the Bayesian framework is 0.62 (95% CrI: 0.23 to 1.60). Additionally, the Bayesian framework provided posterior probability estimates. For this analysis, the probability that the $HR < 1$ is 0.84, which suggested that there is a large probability that the VAS provides an improvement in mortality rate compared to conservative care.

Under the assumption that there is a time-dependent treatment effect, the Bayesian and frequentist results all included harmful effects of VAS from conception to 36.5 weeks (expected time of birth). The probability that the estimate of $HR_1 < 0.646$ is very low for all prior distributions with a posterior probability of just 0.03 for the sceptical prior and only 0.36 for the enthusiastic prior distribution. From 36.5 weeks onwards, only the enthusiastic prior distribution gave convincing evidence that the intervention is beneficial (HR_2 0.58, 95% CrI: 0.36 to 0.95). For all other priors the results suggested that there may be harmful effects of the intervention. However, the posterior probability that $HR_2 < 1$ is fairly high for all prior distributions (> 0.7).

3.5.1 Does the Bayesian framework provide any additional benefit in the PLUTO trial?

Classical hypothesis tests depend on the data having high power (usually 0.8 or 0.9) and a low significance level, or type one error rate (usually 0.05). However, with a small dataset like the PLUTO trial, the power is low and the data will often yield non-significant results even when in truth the intervention provides a clinically important benefit. The PLUTO trial deals with a life-threatening condition found in unborn babies and procedures need to be developed to improve the chances of survival. Therefore, even with a small amount of data, the information from the PLUTO trial is important and needs to be appropriately used to

inform clinical decisions. A Bayesian analysis provided an alternative method to deal with the data that could include additional data in the form of external evidence. The Bayesian analysis allows direct probability statements from posterior distributions for the treatment effect. Probabilities can be calculated such as the posterior probability that the treatment effect is less than some clinically important value. These probabilities are a more intuitive way of making decisions that cannot be calculated within a frequentist framework. However, the clinically meaningful difference of a HR less than 0.646 is a large difference to detect, which explains why the probabilities regarding this value are low for all analyses. It is important to ensure that any pre-specified clinically meaningful difference is also a realistic difference that the clinicians believe could be observed.

A number of prior distributions have been derived to determine whether the shunt is effective compared to conservative care, including off-the-shelf and elicited prior distributions. The vague prior allows for the results to be almost entirely based on the data and the results closely mirrored the results from the frequentist analysis. The results based on the elicited priors may represent a reflection of the updated beliefs that a clinician may have based on the additional results from the PLUTO trial. However, it is not easy to arrive at a conclusion when there are several different results as a consequence of the different prior distributions; in particular the decision for future practice is difficult when the evidence from the trial is not convincing. For example, in the results in Table 3.5, the posterior HR2 estimate (estimate of HR for expected time of birth until the end of follow-up) for the vague prior analysis is 0.212 (95% CrI: 0.027 to 1.101), whereas the posterior HR2 estimate from the paediatric nephrologist prior analysis is just 0.919 (95% CrI: 0.687 to 1.230). But, the conclusions are clearer across the various prior distribution assumptions with the proportional hazards assumption over the entire follow-up.

3.5.2 Limitations

The dataset is particularly small (31 subjects), therefore, although it is possible to conduct a Bayesian analysis as an alternative to a frequentist analysis, there must be caution interpreting any of the results without consideration of the other choices of prior distribution when making changes within clinical practice.

Due to the small sample size some important assumptions were made, including whether to assume proportional hazard rates in the Cox regression analysis. Since the assumption of proportionality could have a profound impact on the results of the trial, a number of analyses were conducted by assuming proportional hazards, and alternatively including a time-dependent treatment effect.

3.6 Conclusion

In conclusion, the PLUTO trial failed to recruit its target sample size and therefore, any frequentist hypothesis tests would be severely underpowered. A Bayesian statistical analysis offers an alternative approach alongside the classical framework to deal with a small dataset. The Bayesian approach can additionally incorporate external “data” in the form of prior distributions and calculate direct posterior probabilities that the treatment effect is less than some pre-specified clinically important difference. The posterior distributions suggest that the insertion of the VAS may increase survival rates in comparison to conservative care. However, almost all of the results indicate that there may also be harmful effects of the intervention in the weeks before birth (i.e. before 36.5 weeks). The results from this survival analysis were combined with several other analyses, including a cost-effectiveness analysis to recommend an update of the National Institute for Health and Care Excellence interventional procedures guidelines that will reflect the new evidence, whilst continuing to follow-up the babies in the PLUTO trial for long term outcomes.¹⁵³ Currently, the NICE

guidelines state that the evidence regarding the safety and efficacy of VAS for LUTO “does not appear adequate for this procedure to be used without special arrangements for consent and for audit or research.”¹⁶⁸ The implications for practice from the PLUTO trial is that the results of the RCT and observational evidence are consistent and suggest that VAS improves perinatal survival compared with conservative care, but that the long-term prognosis for these babies into infant life is poor. The health economic analysis suggests that the costs associated with this small gain in disability-free survival at the end of one year are high and therefore VAS is unlikely to be deemed cost-effective.¹⁵³

In the next chapter, Bayesian analyses of randomised trials with small sample sizes will continue but in the context of randomised phase II trials, which tend to be small by design compared to large phase III trials.

CHAPTER 4: UNIVARIATE META-ANALYSIS OF PHASE II TRIALS

4.1 Context and objective

The previous chapter explored the use of Bayesian methods to analyse a single randomised trial with a small sample size. Phase II trials also tend to have small sample sizes by design, however sometimes there may be several phase II trials that aim to answer the same question. Multiple phase II trials could thus be considered collectively before initiation of a phase III trial, for example within a meta-analysis. Meta-analysis methods were introduced in Chapter 1, and they combine quantitative data from several studies to produce pooled overall results that aid evidence-based clinical decision making.³⁵ A meta-analysis is able to quantify effect sizes and their uncertainty, and to increase sample size and thus power. Where there are multiple phase II studies, a meta-analysis could therefore be beneficial to synthesise all the evidence and inform the decision of whether or not to proceed to phase III.

The objective in this chapter is to consider the use of meta-analysis methods in the context of phase II trials. In particular, to apply Bayesian meta-analysis approaches that summarise the evidence and account for all uncertainty in order to estimate the probability that the treatment is effective on average for each outcome of interest. Furthermore, the Bayesian framework could allow one to predict whether the treatment will be effective if a new (phase III) trial of that treatment was undertaken. This information can be used to inform the decision about proceeding to phase III. However, many issues may arise for a meta-analysis in this context. Phase II trials are often small, low powered and conducted in different ways, which may cause heterogeneity in treatment effects between trials and large uncertainty in the summary results. Therefore, this chapter evaluates the benefits and limitations of Bayesian meta-

analysis of phase II trials, and exemplifies how they can be useful toward phase III decisions. The focus here is on univariate meta-analysis approaches, which is the standard method that considers a separate meta-analysis for each outcome of interest. In Chapter 5, extension is made to this work to multivariate meta-analysis, which synthesises all outcomes jointly.

4.2 Data, outcomes and clinical objectives

Eikelboom et al.¹⁶⁹ conducted a meta-analysis to examine serious adverse safety outcomes within phase II and phase III randomised controlled trials that evaluated the efficacy of bolus thrombolytic therapy compared with standard infusion therapy for the in-hospital treatment of acute myocardial infarction (AMI). Effective treatment of AMI is crucial. The incidence of hospitalised AMI in England in 2010 was 122 per 100,000 for men and 53 per 100,000 for women.¹⁷⁰ Thrombolytic therapy is an accepted form of treatment for AMI, which aims to reduce mortality and restore normal blood flow by dissolving clots in blood vessels.¹⁷¹ However, there are potential serious adverse events, such as intracranial haemorrhage (ICH) and mortality.¹⁷¹

Eikelboom et al. focussed on the disagreement between phase II and phase III results in this field. This chapter reveals new insight into this comparison. Nine randomised, controlled phase II trials were utilised by Eikelboom and in this chapter they are first used to methodologically examine univariate meta-analysis methods for combining small phase II trials.¹⁷¹⁻¹⁷⁹ They are then used to predict phase III trial results. The outcomes of interest are the four serious adverse events reported: ICH, stroke, reinfarction and mortality. The data were extracted from the original nine published papers of the phase II randomised trials and the data is shown in Table 4.1 and Table 4.2. This includes the agents compared, the total number of patients, the length of follow-up, and the percentage of patients with each adverse event. Patients entered the trials from the date of hospital admission with AMI. Similar length

of follow-up was reported for most of the trials. The trial by Vanderschueren et al.¹⁷⁴ and the BASE¹⁷⁵ study did not report the actual length of follow-up; however, the adverse event outcomes in these studies were monitored during the length of patient stay in hospital. All trials recorded the four outcomes except one by Kawai et al.,¹⁷³ which did not report the number of patients with stroke.

The clinical objective of the meta-analysis of the trials is to assess if there is any difference in the risk of adverse event outcomes for patients who receive bolus therapy compared to those patients that receive infusion therapy. The hope is that bolus therapy reduces the risk of adverse outcomes, and if so, further phase III research might be funded. For all results in this chapter, an odds ratio (OR) that is less than one indicates that those patients who receive bolus therapy have lower odds of the adverse event compared to those who receive infusion therapy.

Table 4.1: Phase II randomised trials of bolus vs. infusion thrombolytic therapy in AMI.

Trial	Bolus (No. of boluses)	Infusion[†]	N (Bolus)	N (Infusion)	Length of follow-up
RAPID, ¹⁷¹ 1995	Retepase, 5-15 MU (1 or 2)	Alteplase, 100 mg	452	154	30 days
RAPID-II, ¹⁷² 1996	Retepase, 5-10 MU (2)	Alteplase	169	155	35 days
Kawai et al., ¹⁷³ 1997	E6010, 0.22 mg/kg	Tisokinase, 14.4 MU	97	102	7 days
Vanderschueren et al., ¹⁷⁴ 1997	Staphylokinase, 15 mg doses (2)	Alteplase	50	52	Hospital stay
BASE, ¹⁷⁵ 1998	Saruplase, 40-80 mg (1 or 2)	Saruplase, 80 mg	139	53	Hospital stay
DOUBLE, ¹⁷⁶ 1998	Alteplase, 50 mg (2)	Alteplase	224	237	30 days
InTIME, ¹⁷⁷ 1998	Lanoteplase, 15-120 kU/kg (1)	Alteplase	478	124	30 days
TIMI-10B, ¹⁷⁸ 1998	Tenecteplase, 30-50 mg (1)	Alteplase	540	316	30 days
TIMIKO, ¹⁷⁹ 1998	Urokinase, up to 1.5 MU (2)	Alteplase	350	268	30 days

† Alteplase infusion was given according to the accelerated and weight-adjusted GUSTO regimen¹⁸⁰ in all trials except in RAPID in which a 3-hour infusion was used.

Table 4.2: Incidence of ICH, stroke, reinfarction, and mortality in nine phase II trials of bolus vs. infusion thrombolytic therapy in AMI.¹⁷¹⁻¹⁷⁹

Trial	Outcome	Sample size, N		No. of events (%)		Log(OR)	Var(log(OR))	OR	95% CI OR
		Bolus	Infusion	Bolus	Infusion				
RAPID, ¹⁷¹ 1995	ICH			1 (0.2)	4 (2.6)	-2.49	1.26	0.08	0.01 to 0.75
	Stroke	452	154	1 (0.7)	6 (3.9)	-1.80	0.51	0.17	0.04 to 0.67
	Reinfarction			20 (4.4)	7 (4.5)	-0.03	0.20	0.97	0.40 to 2.35
	Mortality			20 (4.4)	6 (3.9)	0.13	0.23	1.14	0.45 to 2.90
ICH					2 (1.2)	3 (1.9)	-0.50	0.85	0.61
RAPID-II, ¹⁷² 1996	Stroke	169	155	3 (1.8)	4 (2.6)	-0.38	0.60	0.68	0.15 to 3.10
	Reinfarction			8 (4.7)	7 (4.5)	0.05	0.28	1.05	0.37 to 2.97
	Mortality			7 (4.1)	13 (8.4)	-0.75	0.23	0.47	0.18 to 1.22
	ICH					0 (0)	1 (1.0)	-1.06	2.69
Kawai et al., ¹⁷³ 1997	Stroke	97	102	-	-	-	-	-	-
	Reinfarction			4 (4.1)	7 (6.9)	-0.54	0.41	0.58	0.17 to 2.06
	Mortality			4 (4.1)	1 (1.0)	1.470	1.27	4.34	0.48 to 39.57
	ICH					0 (0)	0 (0)	-0.04	3.90
Vanderschueren et al., ¹⁷⁴ 1997	Stroke	50	52	0 (0)	1 (1.9)	-1.08	2.71	0.34	0.01 to 8.54
	Reinfarction			5 (10.0)	7 (13.4)	-0.34	0.39	0.71	0.21 to 2.42
	Mortality			0 (0)	1 (1.9)	-1.08	2.71	0.34	0.01 to 8.54
	ICH					2 (1.4)	0 (0)	0.67	2.43
BASE, ¹⁷⁵ 1998	Stroke	139	53	2 (1.4)	0 (0)	0.67	2.43	1.95	0.09 to 41.19
	Reinfarction			9 (6.5)	1 (1.9)	1.28	1.14	3.60	0.44 to 29.13
	Mortality			10 (7.2)	2 (3.8)	0.68	0.63	1.98	0.42 to 9.34

Table 4.2: continued.

Trial	Outcome	Sample size, N		No. of events (%)		Log(OR)	Var(log(OR))	OR	95% CI OR
		Bolus	Infusion	Bolus	Infusion				
DOUBLE, ¹⁷⁶ 1998	ICH			2 (0.9)	1 (0.4)	0.75	1.51	2.13	0.19 to 23.61
	Stroke	224	237	6 (2.7)	2 (0.8)	1.17	0.68	3.23	0.65 to 16.19
	Reinfarction			5 (2.2)	12 (5.1)	-0.85	0.29	0.43	0.15 to 1.24
	Mortality			10 (4.5)	3 (1.3)	1.29	0.44	3.65	0.99 to 13.42
InTIME, ¹⁷⁷ 1998	ICH			0 (0)	1 (0.8)	-2.45	2.68	0.09	0.00 to 2.12
	Stroke	478	124	0 (0)	1 (0.8)	-2.45	2.68	0.09	0.00 to 2.12
	Reinfarction			9 (1.9)	8 (6.5)	-1.28	0.25	0.28	0.11 to 0.74
	Mortality			15 (3.1)	8 (6.5)	-0.76	0.20	0.47	0.19 to 1.13
TIMI-10B, ¹⁷⁸ 1998	ICH			9 (1.7)	6 (1.9)	-0.13	0.28	0.88	0.31 to 2.48
	Stroke	540	316	14 (2.6)	9 (2.8)	-0.10	0.19	0.91	0.39 to 2.12
	Reinfarction			28 (5.2)	18 (5.7)	-0.10	0.10	0.91	0.49 to 1.66
	Mortality			24 (4.4)	18 (5.7)	-0.26	0.10	0.77	0.41 to 1.44
TIMIKO, ¹⁷⁹ 1998	ICH			1 (0.3)	3 (1.1)	-1.37	1.34	0.25	0.03 to 2.45
	Stroke	350	268	1 (0.3)	3 (1.1)	-1.37	1.34	0.25	0.03 to 2.45
	Reinfarction			11 (3.1)	9 (3.4)	-0.07	0.21	0.93	0.38 to 2.29
	Mortality			16 (4.6)	12 (4.5)	0.02	0.15	1.02	0.48 to 2.20

An OR<1 favours bolus therapy

4.3 Univariate fixed-effect meta-analysis methods using a frequentist approach

There are various estimation methods that can be adopted to carry out a meta-analysis. This section will explore the available methods, and aims to determine which approaches may be more appropriate for the meta-analysis of phase II trials. The methods will be illustrated using the AMI dataset with the four outcomes, which will be analysed separately in univariate meta-analyses using both a frequentist and Bayesian framework in order to compare the results. The frequentist analyses are implemented within Stata version 12.1¹⁶⁷ and the Bayesian analyses are implemented within WinBUGS version 1.4.³² In this section frequentist fixed-effect models are introduced.

4.3.1 A general fixed-effect meta-analysis model (inverse variance)

A meta-analysis can be performed assuming a fixed-effect approach, which assumes that there is a single (fixed) treatment effect across all studies, θ .³⁹ The summary (pooled) estimate ($\hat{\theta}$) from a meta-analysis gives the best estimate of this single treatment effect, using the evidence from all available studies included in the meta-analysis. The general ('inverse variance') fixed-effect model has already been described in Chapter 1 but for completeness this model is repeated briefly here. This assumes the observed treatment effect estimate, Y_i , in study i , follows a normal distribution with known variance, S_i^2 , as shown in equation (4.1). The fixed-effect assumption implies that the observed differences between the Y_i are due to chance alone. Y_i and S_i^2 in this chapter will relate to the log odds ratio and its variance for a particular outcome.

$$Y_i \sim N(\theta, S_i^2) \tag{4.1}$$

The fixed-effect inverse variance method assigns weights to each trial,³⁴ which are shown in equation (4.2), where S_i^2 is the variance of the log odds ratio (log OR) from each study, i . The pooled treatment effect, $\hat{\theta}$ (log OR), can be estimated using maximum likelihood estimation shown in equation (4.3) and the variance of the pooled log OR is calculated by equation (4.4).

$$w_i = \frac{1}{S_i^2} \quad (4.2)$$

$$\hat{\theta} = \frac{\sum_{i=1}^k Y_i w_i}{\sum_{i=1}^k w_i} \quad (4.3)$$

$$\text{var}(\hat{\theta}) = \frac{1}{\sum_{i=1}^k w_i} \quad (4.4)$$

4.3.2 Mantel-Haenszel method

In the article by Eikelboom et al.,¹⁶⁹ the authors used a Mantel-Haenszel approach for meta-analysis,³⁴ which assumes a fixed-effect model. They also performed a separate meta-analysis for each outcome, i.e. univariate meta-analysis. The Mantel-Haenszel method is often used for pooling ORs and it has been shown to have better statistical properties when there are few events, compared to alternative methods such as the inverse variance method of equation (4.1).³⁴ Table 4.3 shows the 2x2 table for an outcome of interest, by treatment arm in study i , where $a_i + b_i + c_i + d_i = n_i$.

Table 4.3: Event in study i .

Study i	Event	Non-event
Treatment	a_i	b_i
Control	c_i	d_i

The sample estimate of the OR in study i , OR_i , is shown in equation (4.5). The Mantel-Haenszel method assigns a weight (equation (4.6)) to each trial so that the results of some studies make a greater contribution than others to the pooled result. The estimated pooled OR is obtained by equation (4.7), where k is the number of studies.

$$OR_i = \frac{a_i d_i}{b_i c_i} \quad (4.5)$$

$$w_i = \frac{b_i c_i}{n_i} \quad (4.6)$$

$$\widehat{OR} = \frac{\sum_{i=1}^k OR_i w_i}{\sum_{i=1}^k w_i} = \frac{\sum_{i=1}^k \frac{a_i d_i}{n_i}}{\sum_{i=1}^k \frac{b_i c_i}{n_i}} \quad (4.7)$$

Continuity correction in phase II meta-analysis

The individual OR estimates by the Mantel-Haenszel and inverse variance methods require a continuity correction if there are zero cells in the 2x2 table (Table 4.3). The standard approach is to add a constant correction factor of 0.5 to all cells.¹⁸¹ However, a correction factor of 0.5 may not perform well when the groups are severely unbalanced with respect to the number of subjects in each treatment arm, which is the case in this example in AMI (Table 4.2). The continuity correction factors influence the estimates of the variance, which affects the weighting given to each study, as well as the estimate of the treatment effect. Given the imbalance between treatment groups, an alternative proposed method is referred to as the ‘opposite treatment arm’ continuity correction.¹⁸¹ This method adds the reciprocal of the opposite treatment arm total sample size to the number of events and number of non-events, for each treatment arm.

Table 4.4: An illustration of a trial in which a continuity correction may be required to calculate the odds ratio.

Vanderschueren et al. ¹⁷⁴	ICH	No ICH	Total
Bolus therapy	0	50	50
Infusion therapy	0	52	52
Total	0	102	102

To illustrate this, Table 4.4 shows the trial by Vanderschueren et al., where there are no events of ICH in either treatment arm. If a correction factor of 0.5 is added to each cell, the Mantel-Haenszel method estimates the OR for this study to be 1.04 with a weighting of 1.88%. This is compared to an OR of 1 if the ‘opposite treatment arm’ continuity correction is applied with the study weighting of 0.08%, which is clearly a more appropriate estimate of the OR, and the weighting is reduced as there is very little information in this study to contribute to the pooled estimate. Due to the large imbalance in sample size in many studies, and the multiple zero event rates, the meta-analysis in this chapter will adopt the ‘opposite treatment arm’ continuity correction method, for those methods that require a continuity correction.

4.3.3 Peto method

The Peto method¹⁸² is a fixed-effect method that can only be used to pool ORs (compared to approaches that also pool alternative summary measures such as relative risks). This method estimates the OR in study i , OR_i , by equation (4.8), where a_i , b_i , c_i , d_i , and n_i are the elements of the 2 x 2 table in Table 4.3. The pooled OR is shown in equation (4.9), and the CI for the pooled OR is shown in equation (4.10), where α is the significance level that is usually chosen to be 0.05 in order to obtain a 95% CI. The Peto method does not require continuity correction. The Peto method works well when the intervention effects are small (ORs that are close to one), and when the events are rare. When this is not the case the

Peto method is not recommended since the method has been shown to give biased results.¹⁸³

$$OR_i = \exp\left(\frac{O_i - E_i}{V_i}\right); \quad (4.8)$$

$$O_i = a_i, \quad E_i = \frac{(a_i + b_i)(a_i + c_i)}{n_i},$$

$$V_i = \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{n_i^2(n_i - 1)}$$

$$\widehat{OR} = \exp\left(\frac{\sum_{i=1}^k (O_i - E_i)}{\sum_{i=1}^k V_i}\right) \quad (4.9)$$

$$CI = \exp\left(\frac{\sum_{i=1}^k (O_i - E_i) \mp z_{\alpha/2} \sqrt{\sum_{i=1}^k V_i}}{\sum_{i=1}^k V_i}\right) \quad (4.10)$$

4.3.4 Fixed-effect logistic regression

A univariate meta-analysis can also be implemented using a logistic regression model, which models the exact binomial distribution of the data. This method does not require continuity correction.¹⁸⁴ Let r_{ij} denote the binary variable for the adverse event of interest, which is 1 if subject j has the event in trial i , and zero otherwise. The event is binomially distributed with probability, π_{ij} , i.e., $r_{ij} \sim \text{Bin}(1, \pi_{ij})$, which is equivalent to a Bernoulli distribution with event probability of π_{ij} .

Model (4.11) shows the logistic regression fixed-effect meta-analysis model with a binary covariate for the treatment, $Drug_{ij}$, which represents the drug for patient j in study i ($i=1, \dots, 9$). This covariate is zero for those that receive infusion therapy, and one for those that receive bolus therapy. The model assumes that the treatment effect (θ) is fixed at the trial level, where $\text{logit}(\pi_{ij})$ denotes $\log(\pi_{ij}/(1-\pi_{ij}))$; the log odds of the event for patient j in trial i . The

model accounts for clustering of patients within trials by a separate intercept term, α_i , which denotes the baseline (control group) risk for each trial, i .¹⁸⁵ This is important because it is inappropriate to analyse the data as if it is coming from a single study.¹⁸⁵

$$\text{logit}(\pi_{ij}) = \alpha_i + \theta \cdot \text{Drug}_{ij} \quad (4.11)$$

Maximum likelihood (ML) estimation can be used to estimate the parameters of interest in this fixed-effect logistic regression meta-analysis model.

4.4 Univariate random-effects meta-analysis methods using a frequentist approach

4.4.1 A general random-effects meta-analysis model (inverse variance)

In contrast to the fixed-effect approach, a random-effects approach assumes a distribution of treatment effects across studies as shown in model (4.12), which has already been described in Chapter 1. Each study can now be estimating a different treatment effect, θ_i , and therefore the summary (pooled) estimate, $\hat{\beta}$, from the meta-analysis gives an estimate of the average treatment effect, β , across studies. Individual studies may therefore have a treatment effect that varies considerably from this average value, and this between-study variance in the treatment effect is known as tau-squared (τ^2).

$$\begin{aligned} Y_i &\sim N(\theta_i, S_i^2) \\ \theta_i &\sim N(\beta, \tau^2) \end{aligned} \quad (4.12)$$

An estimate of the between-study variance, τ^2 , must be obtained to calculate a pooled treatment effect estimate from a random-effects meta-analysis model. Restricted maximum likelihood estimation (REML) is typically preferred for the random-effects model in order to

estimate both τ^2 and β appropriately. Maximum likelihood (ML) can also be used but REML is usually preferred as ML is likely to underestimate τ^2 . The DerSimonian and Laird “method of moments” method is also popular, which is shown in equations (1.17) to (1.20) in section 1.4.2 of Chapter 1. However, REML, ML and MM estimation does not account for the fact that the estimated τ^2 has uncertainty. Some authors and packages try to account for the uncertainty post-estimation, for example by inflating the estimate of the standard error of the pooled log OR, or by using a t -distribution to derive CIs, in order to increase the width of the confidence interval (CI) for the pooled result.¹⁸⁶

4.4.2 Examining heterogeneity in a meta-analysis of phase II trials

Eikelboom et al. chose a fixed-effect approach to meta-analysis, rather than a random-effects approach, but did not explain or justify this decision. Various methods are commonly implemented to determine the level of heterogeneity in a meta-analysis¹⁸⁷; one method is the Q-statistic. The Q-statistic is a chi-squared test that tests the null hypothesis that the treatment effect is equal (fixed) in the k studies. This statistic is compared to the chi-squared distribution with $k-1$ degrees of freedom to obtain a p-value. A statistically significant p-value suggests there is more variation between the trials than expected by chance. The Q-statistic is calculated by equation (4.13), where w_i is the weight given to trial i by the fixed-effect method of choice, Y_i is the estimated treatment effect (log OR) from study i , and $\hat{\theta}$ is the pooled estimate of the treatment effect from the fixed-effect meta-analysis.

$$Q = \sum_{i=1}^k w_i (Y_i - \hat{\theta})^2 \quad (4.13)$$

This test has low power when there are a small number of studies and so a significance level of 10% is often used. The Q-statistic is a test for heterogeneity but it does not measure it.

The estimate of τ^2 after fitting equation (4.15) directly quantifies the between-study heterogeneity. Also, the I^2 statistic is an estimate of the percentage of the total variation that is due to between-trial variation. I^2 is calculated by equation (4.14), where Q is the Q statistic estimated by equation (4.13).¹⁸⁷ If the value of I^2 is large this indicates that there may be a large amount of between-study variation, and therefore a random-effects meta-analysis may be more appropriate than a fixed-effect analysis. A tentative guide to the interpretation of I^2 is: 0% represents no heterogeneity; 25% represents low heterogeneity; 50% represents moderate heterogeneity; and 75% represents high heterogeneity.¹⁸⁸

$$I^2 = 100 * \frac{(Q - (k - 1))}{Q} \quad (4.14)$$

In this chapter the p-values from the Q-statistic test will be presented and discussed alongside the I^2 estimates to determine whether these measures are useful within a meta-analysis of phase II trials.

4.4.3 Random-effects logistic regression

Model (4.15) is the random-effects logistic regression model assuming that the log odds ratio for the treatment effect is random at the trial level. The average treatment effect is denoted by β , and the study specific effect, u_i , has been included to account for the difference between trials in the true treatment effect; the variance for this is denoted by τ^2 .

$$\text{logit}(\pi_{ij}) = \alpha_i + \theta_i \cdot \text{Drug}_{ij}; \quad (4.15)$$

$$\theta_i = \beta + u_i; u_i \sim N(0, \tau^2)$$

Maximum likelihood (ML) estimation is used to estimate the parameters of interest. The log-likelihood for the random-effects model has no closed form; therefore it is approximated by

adaptive Gaussian quadrature¹⁶⁷ with five integration points used in this chapter. The estimation again does not account for the uncertainty in $\hat{\tau}^2$ although this only affects the error in the parameter estimates and not their point estimates.

4.5 Univariate meta-analysis methods using a Bayesian framework

A univariate meta-analysis can also be implemented within a Bayesian framework. Potential advantages of this approach include the ability to easily estimate the uncertainty of the between-study variance, the ability to make direct probability statements regarding quantities of interest, and to make predictive statements, conditional on the current state of knowledge.¹⁹

4.5.1 Bayesian approach to inverse-variance meta-analysis methods

Fixed-effect and random-effects meta-analysis models using a Bayesian approach to the inverse-variance method will be implemented in this chapter. The fixed-effect meta-analysis model assumes the log OR from each study, Y_i , is normally distributed about a mean θ and known variance, S_i^2 , as shown by model (4.16) The Bayesian approach additionally requires prior distributions to be specified for unknown parameters. For the fixed-effect model, in this chapter a vague normal prior distribution is assumed for the true underlying fixed treatment effect.

$$Y_i \sim N(\theta, S_i^2) \tag{4.16}$$

$$\theta \sim N(0, 1000^2)$$

The random-effects model additionally allows for heterogeneity (τ^2) about an average effect, β , as described by model (4.17). In this model, a vague normal prior distribution is assumed

for the average treatment effect and a vague normal prior distribution, truncated at zero ($N(0,1)/(0,)$), is assumed for the between-trial standard deviation. However, the correct choice of distribution for τ is often disputed and thus a sensitivity analysis for its prior distribution is described in Section 4.5.5. In contrast to the frequentist approach, the posterior distribution for β in the Bayesian models accounts for the uncertainty in τ .³⁶

$$\begin{aligned}
 Y_i &\sim N(\theta_i, S_i^2) & (4.17) \\
 \theta_i &\sim N(\beta, \tau^2) \\
 \beta &\sim N(0, 1000^2) \\
 \tau &\sim N(0, 1)/(0,)
 \end{aligned}$$

4.5.2 Bayesian approach to logistic regression meta-analysis model

A Bayesian approach to the logistic regression models in section 4.3.4 is also possible. The fixed-effect meta-analysis model is shown in model (4.18) where r_{ij} denotes the binary variable for the adverse event of interest, which is 1 if subject j has the event in trial i , and zero otherwise. The estimate of the fixed pooled logOR for bolus therapy compared to infusion therapy is denoted by θ . For this model a vague normal prior distribution is assumed for α_i , ($i=1, \dots, 9$) and θ with a mean zero and a variance 1000^2 .

$$\begin{aligned}
 r_{ij} &\sim \text{Bin}(1, \pi_{ij}) & (4.18) \\
 \text{logit}(\pi_{ij}) &= \alpha_i + \theta \cdot \text{Drug}_{ij} \\
 \alpha_i &\sim N(0, 1000^2); i=1, \dots, 9; \theta \sim N(0, 1000^2)
 \end{aligned}$$

The random-effects logistic regression model is shown in model (4.19), with the same vague normal distribution for α_i and β , and a vague normal distribution assumed for the between-

trial standard deviation. The choice of prior distribution for the between-study standard deviation will be investigated in a sensitivity analysis in section 4.5.5.

$$\text{logit}(\pi_{ij}) = \alpha_i + \theta_i \cdot \text{Drug}_{ij}; \quad (4.19)$$

$$\theta_i = \beta + u_i; \quad u_i \sim N(0, \tau^2)$$

$$\alpha_i \sim N(0, 1000^2); \quad i=1, \dots, 9; \quad \beta \sim N(0, 1000^2)$$

$$\tau \sim N(0, 1) | (0,)$$

4.5.3 Probability statements

Following a Bayesian analysis, the posterior distributions can be used to make probability statements about the probability that the treatment will be effective. These probabilities may be especially useful in the decision to proceed to phase III. For example, it is possible to calculate the probability that the treatment will reduce the odds of intracranial haemorrhage compared to the standard treatment by at least 10%, i.e. the probability that the summary OR is less than 0.90. Similarly, for each outcome, the estimate of the probability that the posterior OR is less than one can be obtained, i.e. the probability that bolus therapy is beneficial at all compared to infusion therapy.

4.5.4 Estimation and reporting of Bayesian models

Following a Bayesian analysis, the summary estimates of interest in this chapter are the median of the posterior distribution for the summary log OR, the median of the posterior distribution for the summary OR, the posterior for the standard deviation of the log OR, the 95% posterior credible interval (CrI) for the summary OR, the posterior probability that the OR is less than one, and the median of the posterior distribution of the between-trial variance (if applicable). Posterior estimates of the model parameters are obtained using the Gibbs Sampler Markov chain Monte Carlo (MCMC) method,¹⁸⁹ which is implemented in WinBUGS

version 1.4.³² The code for the random-effects logistic regression model is shown in Appendix B1. The analyses were performed with 100,000 iterations after allowing for a 100,000 iteration burn in and checking for convergence using several common measures, including the history and trace plots. An example of the history and trace plots is given for the univariate random-effects logistic regression model in Appendix B2.

4.5.5 Prior distribution for between-study variance

As discussed briefly in Chapter 1, the appropriate prior distribution for τ is often debated.²⁹ In this chapter, a sensitivity analysis is therefore performed for the prior distribution for the between-study variation within the random-effects logistic regression method as an example. This is to ensure that the selected prior distribution does not largely influence the results. The prior distributions considered are shown in Table 4.5.²⁹ Since standard deviation= $\sqrt{\text{variance}}$ = $\sqrt{1/\text{precision}}$, the prior distribution can be put on the variance, standard deviation or precision. Priors 1a to 6b are non-informative prior distributions. Prior 6a is the distribution that has been applied for all meta-analyses in this chapter. Prior 1a is a common choice of prior distribution for the between-study standard deviation. Priors 1b and 1c are the same distributional form of Prior 1a, but assess the sensitivity to the choice of parameter values. Prior 2 is a commonly used prior distribution for variance parameters with a 'spike' of probability mass close to zero. Prior 3a is a prior distribution on the log variance scale. Prior 3b is a weak informative version of Prior 3a but goes to a maximum of $\log(4.0)=1.386$, which assumes that the variance is unlikely to be greater than 4.0. Prior 4a is a prior on the variance scale and Prior 4b is a weak informative version of Prior 4a with a maximum of $\tau^2=4.0$. Prior 5a is on the precision scale with Prior 5b as a weak informative version of Prior 5a. Prior 5a is equivalent to Prior 1c. Prior 6b is a weak informative version of Prior 6a. Priors 7a and 7b are empirical prior distributions that were derived by Turner et al. for a future meta-analysis with a binary outcome where there may be few studies to estimate this

parameter.³⁶ Prior 7a is the prior distribution that was derived for a pharmacological intervention where the outcome is all-cause mortality and this prior distribution is used for the analysis of death. Prior 7b is the prior distribution for a semi-objective outcome with a pharmacological intervention and this prior distribution is used for the analysis of ICH.

Table 4.5: Sensitivity analysis prior distributions for between-study variance.

Prior	Target parameter	Distribution	Form of probability density function	Prior parameters
1a	τ	Uniform		a=0, b=10
1b	τ	Uniform		a=0, b=100
1c	τ	Uniform		a=0, b=2
2	$1/\tau^2$	Gamma	$\frac{\mu^r x^{r-1} e^{-\mu x}}{\Gamma(r)}, x>0$	$\mu=0.1, r=0.1$
3a	$\text{Log}(\tau^2)$	Uniform		a=-10, b=10
3b	$\text{Log}(\tau^2)$	Uniform		a=-10, b=1.386
4a	τ^2	Uniform		a=0.001, b=1000
4b	τ^2	Uniform		a=0.001, b=4
5a	$1/\tau^2$	Pareto	$\alpha c^\alpha x^{-(\alpha+1)}, x>c$	$\alpha=1, c=0.001$
5b	$1/\tau^2$	Pareto		$\alpha=1, c=0.25$
6a	τ	Half-Normal	$\sqrt{\frac{1}{2\pi\lambda^2}} e^{-\frac{(x-\mu)^2}{2\lambda^2}}, 0<x<\infty$	$\lambda=1, \mu=0$
6b	τ	Half-Normal		$\lambda=100, \mu=0$
7a	τ^2	lognormal	$\frac{1}{x\sqrt{2\pi\lambda}} e^{-\frac{(\ln x-\mu)^2}{2\lambda^2}}$	$\lambda=1.45^2, \mu=-4.06$
7b	τ^2	lognormal		$\lambda=1.85^2, \mu=-3.02$

4.6 Deriving prediction intervals following a random-effects meta-analysis

The random-effects meta-analyses present estimates of the average treatment effect across studies. When considering whether to conduct a phase III trial, focusing on the posterior

distribution for β may be misleading when heterogeneity is present. The effect in a new trial (θ_{new}) may be very different to the average effect (β), due to the causes of heterogeneity from trial to trial (or setting to setting)³⁹. Ideally, the factors causing the heterogeneity would be known, so that new trials could focus on implementation strategies (e.g. doses) and populations most likely to show benefit. However, identifying causes of heterogeneity is problematic given the typical (<10) studies in a meta-analysis and the potential for trial-level confounding. Therefore, the focus here is on situations where the phase II trials in the meta-analysis all include pertinent places, populations and strategies (e.g. doses, timing, length of treatment) for which the intervention effect is of interest. After a random-effects meta-analysis, a prediction interval can be calculated to give a range of values for the predicted treatment effect in a single new study, e.g. a new phase II trial or a phase III trial.³⁹ These prediction intervals can be calculated and compared in both the frequentist and Bayesian frameworks.

4.6.1 Frequentist prediction intervals

Frequentist prediction intervals are approximated based on the estimated amount of between-trial variation, the uncertainty in the pooled estimate, and a t -distribution to account for uncertainty in the estimate of τ . Equation (4.20) displays the approximate 95% prediction interval,^{39,190} where $\hat{\beta}$ represents the (pooled) average log OR; t_{k-2} is the 100(1- α /2) percentile of the t -distribution with $k-2$ degrees of freedom (k =number of studies) and α is 0.05; $\hat{\tau}$ is the estimated between-trial standard deviation; $SE(\hat{\beta})$ is the standard error of $\hat{\beta}$.

$$\hat{\beta} \pm t_{k-2} * \sqrt{\hat{\tau}^2 + SE(\hat{\beta})^2} \quad (4.20)$$

4.6.2 Bayesian prediction interval for the true treatment effect in a new trial

Bayesian prediction intervals for the treatment effect in a future study are easily derived following MCMC estimation of the model. As specified in model (4.17), $\theta_i \sim N(\beta, \tau^2)$ and so the 95% predictive distribution for $\theta_{i_{new}}$ (equation (4.21)) is the true intervention effect (log odds ratio) in a new trial, where:

$$\theta_{i_{new}} \sim N(\beta, \tau^2) \quad (4.21)$$

If the true value of β and τ^2 are known, a 95% probability interval for a new study, $\theta_{i_{new}}$, can be obtained by taking the 2.5% and 97.5% values of the posterior distribution of $\theta_{i_{new}}$. However, when β and τ^2 are unknown, this 95% interval can be obtained immediately after fitting model (4.17) and it will account for the uncertainty in β and τ^2 through samples from their posterior distributions. Therefore, the Bayesian prediction intervals will incorporate more uncertainty compared to the frequentist prediction intervals, though the t -distribution (rather than the normal distribution) tries to address this in the frequentist equation. This predictive distribution can also be used to calculate the probability that the intervention will be truly effective in a new trial,¹⁹¹ either at all (i.e. probability(true OR in a new trial < 1)) or by some clinically relevant amount, such as the odds being reduced by at least 10% (i.e. probability(true OR in a new trial < 0.9)).

4.6.3 Bayesian prediction interval for the treatment effect estimate in a trial with a given sample size

Though the *true* intervention effect ($\theta_{i_{new}}$) is of fundamental interest, a more pertinent question facing phase III funders is: what is the probability that the intervention will be

identified as beneficial in a new trial *with a given sample size*? To help answer this, during estimation of Model (4.17), one can also derive an approximate predictive distribution for the intervention effect estimate, $Y_{i_{new}}$, in a new trial of particular sample size, $N_{i_{new}}$:

$$Y_{i_{new}} \sim N(\theta_{i_{new}}, \text{var}(Y_{i_{new}})) \quad (4.22)$$

where $\theta_{i_{new}}$ is the true intervention effect in a new trial. A standard approximation of a variance of a logOR ($Y_{i_{new}}$) can be approximated by $\frac{1}{a_{i_{new}}} + \frac{1}{b_{i_{new}}} + \frac{1}{c_{i_{new}}} + \frac{1}{d_{i_{new}}}$ ¹⁹² where $a_{i_{new}}$ and $c_{i_{new}}$ are the number of events in the new trial's experimental groups and control groups, respectively; $b_{i_{new}}$ and $d_{i_{new}}$ are the number of non-events in the new trial's experimental and control groups, respectively and the total sample size $N_{i_{new}} = a_{i_{new}} + b_{i_{new}} + c_{i_{new}} + d_{i_{new}}$. Clearly, $a_{i_{new}}$, $b_{i_{new}}$, $c_{i_{new}}$ and $d_{i_{new}}$ are not known and so some assumptions are needed in order to derive them for each $Y_{i_{new}}$ sampled during the estimation process. The following two options are considered in this chapter.

Option 1: Estimate the variance of $Y_{i_{new}}$ in each drawn sample assuming a fixed baseline risk and sample size

First, assume a fixed proportion of events (baseline risk) in the control group in the new trial; this could be chosen to reflect the baseline risk in the intended population. Thus, $\frac{c_{i_{new}}}{c_{i_{new}} + d_{i_{new}}}$ is specified. Then, also specify the sample size of the control group ($nc_{i_{new}} = c_{i_{new}} + d_{i_{new}}$). The chosen baseline risk and sample size thereby fix $c_{i_{new}}$ and $d_{i_{new}}$. Also, specify the treatment group sample size ($nt_{i_{new}} = a_{i_{new}} + b_{i_{new}}$). Then, assuming that each $Y_{i_{new}}$ that is sampled during the MCMC estimation process is the same as the mean $\theta_{i_{new}}$ (which on average will be true), $a_{i_{new}}$ and $b_{i_{new}}$ can be derived as follows:

$$\theta_{i_{new}} = \log\left(\frac{a_{i_{new}} d_{i_{new}}}{b_{i_{new}} c_{i_{new}}}\right) \quad (4.23)$$

$$\Rightarrow a_{i_{new}} = \frac{\exp(\theta_{i_{new}}) n t_{i_{new}} c_{i_{new}}}{d_{i_{new}} + \exp(\theta_{i_{new}}) c_{i_{new}}}$$

$$\Rightarrow b_{i_{new}} = n t_{i_{new}} - a_{i_{new}}$$

In equation (4.23), $n t_{i_{new}}$, $c_{i_{new}}$ and $d_{i_{new}}$ are fixed (specified in advance), and $\theta_{i_{new}}$ is the current sampled true intervention effect for the new trial. In this way, each $Y_{i_{new}}$ that is sampled has a variance approximated by $\frac{1}{a_{i_{new}}} + \frac{1}{b_{i_{new}}} + \frac{1}{c_{i_{new}}} + \frac{1}{d_{i_{new}}}$, where $c_{i_{new}}$ and $d_{i_{new}}$ are chosen and $a_{i_{new}}$ and $b_{i_{new}}$ are obtained from equation (4.23) for that sample.

Option 2: Assume a fixed variance of $Y_{i_{new}}$

A simpler, but potentially less accurate approach than option (1), is to specify a fixed variance of $Y_{i_{new}}$, regardless of the $\theta_{i_{new}}$ sampled. To do this, in addition to specifying $c_{i_{new}}$ and $d_{i_{new}}$ based on the assumed baseline risk and control group sample size (see option (1)), one needs to also fix $a_{i_{new}}$ and $b_{i_{new}}$. One therefore needs to assume $\theta_{i_{new}}$ is fixed at some value ($=\log\left(\frac{a_{i_{new}} d_{i_{new}}}{b_{i_{new}} c_{i_{new}}}\right)$). For example, one might fix $\theta_{i_{new}}$ to be the mean of the posterior distribution for $\theta_{i_{new}}$ (i.e. fix it to be β). Then $a_{i_{new}}$ and $b_{i_{new}}$ can be obtained.

Options (1) or (2) allow an approximate 95% probability interval for $\theta_{i_{new}}$ to be calculated every time $Y_{i_{new}}$ is sampled:

$$Y_{i_{new}} \pm \left(1.96 * \sqrt{\frac{1}{a_{i_{new}}} + \frac{1}{b_{i_{new}}} + \frac{1}{c_{i_{new}}} + \frac{1}{d_{i_{new}}}}\right) \quad (4.24)$$

Therefore, across all samples during the estimation process, one can also derive predictive distributions for the lower and upper bounds of the 95% interval for $Y_{i_{new}}$. One can then calculate probabilities to inform phase III decisions. In particular the probability that, in a new trial with a sample size of $N_{i_{new}}$ and a control group risk of $\frac{c_{i_{new}}}{c_{i_{new}}+d_{i_{new}}}$, the upper bound of the 95% interval for $Y_{i_{new}}$ will be lower than 0 (i.e. that the lower bound of the CrI for the OR will be <1); in other words, the probability that the new trial will identify the intervention as effective by the entire 95% interval for the OR being in favour of the intervention.

4.6.4 Comparison with subsequent phase III trials

The prediction intervals from the meta-analysis of phase II trials predict treatment effects in new trials. Therefore, it is of interest to compare the results of subsequent phase III trials and see if they are consistent with the prediction intervals. Data are available from six subsequent phase III trials. The primary endpoint in these phase III trials was death from any cause.¹⁹³⁻¹⁹⁸ Secondary endpoints included ICH, other stroke, and reinfarction. When comparing meta-analysis results for phase II and III trials, Eikelboom et al. did not consider prediction intervals and rather focussed only on the compared the pooled results for the log odds ratios. The limitation of their approach is shown in section 4.8.3.

4.7 Results I: Pooled results for the treatment effect

4.7.1 Original frequentist results of Eikelboom et al.

The results for each outcome when applying the Mantel-Haenszel fixed-effect estimation method are shown in Table 4.7 to Table 4.10. This is the method used by Eikelboom et al. The pooled estimates of the ORs are in favour of the bolus therapy compared to the infusion therapy for all outcomes. For example, the estimated pooled OR for ICH is 0.552 with 95% CI (0.287, 1.063). However, the 95% CIs suggest that there is large uncertainty in the pooled

estimates and statistical significance is not obtained at the 5% significance level for any of the outcomes. This is perhaps not surprising given how small the phase II trials are.

4.7.2 Heterogeneity results

The original method of Eikelboom et al.¹⁶⁹ did not consider heterogeneity when they applied a fixed-effect meta-analysis method. To examine whether they should have accounted for heterogeneity, Table 4.6 shows the I^2 statistic, the Q-statistic chi-squared test (χ^2), and between-study variance, τ^2 , frequentist estimates for each outcome. The chi-squared test reports a p-value and τ^2 is estimated by the DerSimonian and Laird method-of-moments. The results demonstrate the difficulty in using these statistics to identify whether a random-effects meta-analysis of phase II trials is more sensible than a fixed-effect model.

Q-statistic

Most of the p-values are statistically non-significant, which suggests there is no strong evidence of heterogeneity. However, there are only nine trials and therefore there is low power to detect heterogeneity if it truly existed. The Cochrane Handbook states that it is inappropriate to decide to choose between a fixed-effect and a random-effects model based solely on a test for heterogeneity.³⁴ As a meta-analysis of phase II trials will usually contain fewer than 10 trials, the Q-statistic is thus unlikely to be useful on its own.

τ^2

Of more interest is the estimate of τ^2 , which is non-zero for stroke, reinfarction and mortality. These estimates suggest heterogeneity is present for these outcomes. Further, they seem especially large for stroke and mortality (>0.1).

Table 4.6: Heterogeneity measures for ICH, stroke, reinfarction, and mortality.

Outcome	I ² (%)	χ ² p-value	τ ²
ICH			
Fixed MH	0.0	0.661	-
Fixed IV	0.0	0.665	-
Peto	44.0	0.085	-
Random MM	0.0	0.665	<0.001
Stroke			
Fixed MH	25.6	0.225	-
Fixed IV	25.6	0.225	-
Peto	56.6	0.024	-
Random MM	25.6	0.225	0.306
Reinfarction			
Fixed MH	8.6	0.364	-
Fixed IV	8.2	0.367	-
Peto	23.1	0.238	-
Random MM	8.2	0.367	0.023
Mortality			
Fixed MH	34.3	0.144	-
Fixed IV	33.9	0.146	-
Peto	43.0	0.081	-
Random MM	33.9	0.146	0.144

MH is the Mantel-Haenszel method, IV is the inverse variance method, MM is the method-of-moments estimation by DerSimonian and Laird.

I²

The estimate of I² is less than 40%, in general, for all outcomes, even when τ² is large. In particular, given that τ² is non-zero and large for stroke and mortality it is surprising that I² is not higher than observed for these outcomes. Further, given that the value of τ² is far larger for stroke, it is also surprising that I² is similar for both mortality and stroke. The I² statistic may thus also not be the best indicator of heterogeneity in meta-analysis of phase II trials. I² is the percentage of between-trial variation out of the total variation, where the total variation is the sum of the within-study variation plus the between-study variation. I² may be lower than

expected here because the phase II trials are small and the event rate is low as the outcomes in AMI are rare. Thus, regardless of the size of τ , the total variation will always be dominated by the within-study variation, and so I^2 will usually be small, even when heterogeneity exists.¹⁹⁹

Differences between methods

The results for I^2 and the Q-statistic for the Peto method are different compared to the other methods. The Q-statistic p-values are smaller when based on the Peto method, and, for ICH, stroke and mortality, there is evidence of statistical significance at the 10% significance level in the Peto method but not the other methods. Thus, the estimated values of I^2 are larger for all outcomes. However, in this dataset of phase II trials, there are imbalances in the sample sizes for the treatment arms, and the event rate is also low for the outcomes of interest (Table 4.2). These two factors cause differences between the methods in the trial ORs and their weights since these estimates are calculated differently in the Peto method (equation (4.8)) compared to other methods, such as the Mantel-Haenszel method (equations (4.5) and (4.6)). This leads to different estimates of Q and I^2 , and thus confusion if trying to use these measures to decide between a fixed-effect and a random-effects model.

Summary

In summary, the heterogeneity statistics of Q and I^2 appear limited for identifying heterogeneity in a meta-analysis of phase II trials due to the small number of studies and the small number of patients within studies. This suggests that clinical judgement and prior beliefs about heterogeneity should also be accounted for. Given that these phase II trials were conducted separately, it seems more plausible to assume, a priori, that heterogeneity exists. This is also a more conservative assumption. Further, the estimates of τ^2 are non-zero for most outcomes, and thus a random-effects model seems more sensible.

4.7.3 Comparison of fixed-effect and random-effects meta-analysis pooled results for frequentist and Bayesian methods

Table 4.7 to Table 4.10 display the pooled results from the univariate meta-analysis models for ICH, stroke, reinfarction, and mortality for all the fixed-effect and random-effects models introduced earlier, for both frequentist and Bayesian approaches. The key findings are now discussed.

Pooled results are uncertain

The point estimates for the pooled ORs are in favour of the bolus therapy compared to infusion therapy for all outcomes (i.e. ORs are less than one), and under all meta-analysis models. However, there is large uncertainty in the estimates, resulting in wide CIs (and credible intervals) and thus the results are inconclusive. This is perhaps unsurprising given these are small phase II trials being synthesised. The forest plots from the fixed-effect Peto model are shown in Figure 4.1 to Figure 4.4. These show that the pooled results have narrower CIs than the CIs for each of the individual trials. This demonstrates an increase in the certainty of the treatment effect once the study results have been pooled, and hence why a meta-analysis of phase II trials is appealing.

Table 4.7: Pooled treatment effect results for bolus versus infusion therapy for intracranial haemorrhage.

Assumption	Frequentist Method	Log(OR)	SE(log(OR))	OR	95% CI OR	$\hat{\tau}^2$	
Fixed-effect	Mantel-Haenszel	-0.594	0.334	0.552	0.287 to 1.063	-	
Fixed-effect	Inverse variance	-0.521	0.378	0.594	0.283 to 1.249	-	
Fixed-effect	Peto	-0.644	0.351	0.525	0.264 to 1.043	-	
Fixed-effect	Logistic regression	-0.621	0.342	0.537	0.275 to 1.050	-	
Random-effects	REML (nounc)	-0.577	0.429	0.561	0.242 to 1.301	0.136	
Random-effects	REML	-0.577	0.492	0.561	0.214 to 1.473	0.136	
Random-effects	ML (nounc)	-0.521	0.379	0.594	0.283 to 1.249	<0.001	
Random-effects	ML	-0.521	0.379	0.594	0.283 to 1.249	<0.001	
Random-effects	MM (nounc)	-0.521	0.379	0.594	0.283 to 1.249	<0.001	
Random-effects	MM	-0.521	0.379	0.594	0.283 to 1.249	<0.001	
Random-effects	Logistic regression	-0.621	0.342	0.537	0.275 to 1.050	<0.001	
Assumption	Bayesian Method	Log(OR)	SE(log(OR))	OR	95% CrI OR	$\hat{\tau}^2$ (95% CrI)	Prob OR<1
Fixed-effect	Inverse variance*	-0.522	0.379	0.593	0.282 to 1.248	-	0.916
Fixed-effect	Logistic regression [†]	-0.623	0.347	0.536	0.271 to 1.057	-	0.964
Random-effects	Inverse variance [§]	-0.615	0.549	0.541	0.169 to 1.492	0.316 (0.001 to 3.095)	0.889
Random-effects	Logistic regression [#]	-0.724	0.524	0.485	0.155 to 1.266	0.435 (0.002 to 3.662)	0.937

OR<1 indicates the treatment is beneficial; REML is restricted maximum likelihood method; ML is maximum likelihood method; MM is method-of-moments method; nounc indicates assumption of no uncertainty in between-study variance estimate. *Prior distribution for $\theta \sim N(0, 1000^2)$. [†]Prior distribution for $\theta \sim N(0, 1000^2)$ and $\alpha_j \sim N(0, 1000^2)$. [§]Prior distribution for $\beta \sim N(0, 1000^2)$ and $\tau \sim N(0, 1)/(0, .)$. [#]Prior distribution for $\beta \sim N(0, 1000^2)$, $\alpha_j \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$.

Table 4.8: Pooled treatment effect results for bolus versus infusion therapy for stroke.

Assumption	Frequentist Method	Log(OR)	SE(log(OR))	OR	95% CI OR	$\hat{\tau}^2$	
Fixed-effect	Mantel-Haenszel	-0.354	0.271	0.702	0.413 to 1.192	-	
Fixed-effect	Inverse variance	-0.362	0.298	0.696	0.388 to 1.248	-	
Fixed-effect	Peto	-0.384	0.284	0.681	0.390 to 1.188	-	
Fixed-effect	Logistic regression	-0.375	0.278	0.688	0.399 to 1.186	-	
Random-effects	REML (nounc)	-0.451	0.490	0.637	0.244 to 1.666	0.648	
Random-effects	REML	-0.451	0.493	0.637	0.242 to 1.674	0.648	
Random-effects	ML (nounc)	-0.430	0.421	0.650	0.285 to 1.484	0.355	
Random-effects	ML	-0.430	0.425	0.650	0.283 to 1.496	0.355	
Random-effects	MM (nounc)	-0.425	0.408	0.654	0.294 to 1.453	0.306	
Random-effects	MM	-0.425	0.408	0.654	0.294 to 1.453	0.306	
Random-effects	Logistic regression	-0.375	0.278	0.688	0.399 to 1.186	<0.001	
Assumption	Bayesian Method	Log(OR)	SE(log(OR))	OR	95% CrI OR	$\hat{\tau}^2$ (95% CrI)	Prob OR<1
Fixed-effect	Inverse variance*	-0.364	0.299	0.695	0.387 to 1.247	-	0.889
Fixed-effect	Logistic regression [†]	-0.373	0.281	0.689	0.399 to 1.205	-	0.906
Random-effects	Inverse variance [§]	-0.432	0.515	0.649	0.219 to 1.747	0.488 (0.002 to 3.236)	0.829
Random-effects	Logistic regression [#]	-0.486	0.512	0.615	0.201 to 1.578	0.652 (0.004 to 3.896)	0.862

OR<1 indicates the treatment is beneficial; REML is restricted maximum likelihood method; ML is maximum likelihood method; MM is method-of-moments method; nounc indicates assumption of no uncertainty in between-study variance estimate. *Prior distribution for $\theta \sim N(0, 1000^2)$. [†]Prior distribution for $\theta \sim N(0, 1000^2)$ and $\alpha_j \sim N(0, 1000^2)$. [§]Prior distribution for $\beta \sim N(0, 1000^2)$ and $\tau \sim N(0, 1)/(0, .)$. [#]Prior distribution for $\beta \sim N(0, 1000^2)$, $\alpha_j \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$.

Table 4.9: Pooled treatment effect results for bolus versus infusion therapy for reinfarction.

Assumption	Frequentist Method	Log(OR)	SE(log(OR))	OR	95% CI OR	$\hat{\tau}^2$	
Fixed-effect	Mantel-Haenszel	-0.250	0.161	0.779	0.568 to 1.066	-	
Fixed-effect	Inverse variance	-0.282	0.164	0.754	0.546 to 1.041	-	
Fixed-effect	Peto	-0.255	0.164	0.775	0.562 to 1.067	-	
Fixed-effect	Logistic regression	-0.251	0.161	0.778	0.568 to 1.066	-	
Random-effects	REML (nounc)	-0.282	0.164	0.754	0.546 to 1.041	<0.001	
Random-effects	REML	-0.282	0.164	0.754	0.546 to 1.041	<0.001	
Random-effects	ML (nounc)	-0.282	0.164	0.754	0.546 to 1.041	<0.001	
Random-effects	ML	-0.282	0.164	0.754	0.546 to 1.041	<0.001	
Random-effects	MM (nounc)	-0.288	0.174	0.750	0.533 to 1.055	0.023	
Random-effects	MM	-0.288	0.174	0.750	0.533 to 1.055	0.023	
Random-effects	Logistic regression	-0.251	0.161	0.778	0.568 to 1.066	<0.001	
Assumption	Bayesian Method	Log(OR)	SE(log(OR))	OR	95% CrI OR	$\hat{\tau}^2$ (95% CrI)	Prob OR<1
Fixed-effect	Inverse variance*	-0.284	0.165	0.753	0.544 to 1.040	-	0.958
Fixed-effect	Logistic regression [†]	-0.250	0.160	0.779	0.571 to 1.066	-	0.941
Random-effects	Inverse variance [§]	-0.284	0.213	0.753	0.491 to 1.139	0.066 (0.000 to 0.758)	0.922
Random-effects	Logistic regression [#]	-0.258	0.214	0.773	0.502 to 1.179	0.076 (0.000 to 0.864)	0.901

OR<1 indicates the treatment is beneficial; REML is restricted maximum likelihood method; ML is maximum likelihood method; MM is method-of-moments method; nounc indicates assumption of no uncertainty in between-study variance estimate. *Prior distribution for $\theta \sim N(0, 1000^2)$. [†]Prior distribution for $\theta \sim N(0, 1000^2)$ and $\alpha_j \sim N(0, 1000^2)$. [§]Prior distribution for $\beta \sim N(0, 1000^2)$ and $\tau \sim N(0, 1)/(0, .)$. [#]Prior distribution for $\beta \sim N(0, 1000^2)$, $\alpha_j \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$.

Table 4.10: Pooled treatment effect results for bolus versus infusion therapy for mortality.

Assumption	Frequentist Method	Log(OR)	SE(log(OR))	OR	95% CI OR	$\hat{\tau}^2$	
Fixed-effect	Mantel-Haenszel	-0.064	0.165	0.938	0.679 to 1.294	-	
Fixed-effect	Inverse variance	-0.106	0.170	0.899	0.644 to 1.255	-	
Fixed-effect	Peto	-0.066	0.167	0.936	0.675 to 1.299	-	
Fixed-effect	Logistic regression	-0.066	0.166	0.936	0.676 to 1.296	-	
Random-effects	REML (nounc)	-0.045	0.221	0.956	0.620 to 1.475	0.1262	
Random-effects	REML	-0.045	0.244	0.956	0.592 to 1.543	0.1262	
Random-effects	ML (nounc)	-0.097	0.176	0.906	0.642 to 1.280	0.0124	
Random-effects	ML	-0.099	0.219	0.906	0.590 to 1.391	0.0124	
Random-effects	MM (nounc)	-0.039	0.227	0.962	0.617 to 1.501	0.1437	
Random-effects	MM	-0.039	0.227	0.962	0.617 to 1.501	0.1437	
Random-effects	Logistic regression	-0.066	0.166	0.936	0.676 to 1.296	<0.0001	
Assumption	Bayesian Method	Log(OR)	SE(log(OR))	OR	95% CrI OR	$\hat{\tau}^2$ (95% CrI)	Prob OR<1
Fixed-effect	Inverse variance*	-0.107	0.170	0.899	0.644 to 1.254	-	0.736
Fixed-effect	Logistic regression [†]	-0.066	0.166	0.936	0.681 to 1.303	-	0.656
Random-effects	Inverse variance [§]	-0.038	0.276	0.962	0.594 to 1.780	0.164 (0.000 to 1.443)	0.565
Random-effects	Logistic regression [#]	-0.008	0.283	0.992	0.599 to 1.843	0.213 (0.000 to 1.588)	0.513

OR<1 indicates the treatment is beneficial; REML is restricted maximum likelihood method; ML is maximum likelihood method; MM is method-of-moments method; nounc indicates assumption of no uncertainty in between-study variance estimate. *Prior distribution for $\theta \sim N(0, 1000^2)$. [†]Prior distribution for $\theta \sim N(0, 1000^2)$ and $\alpha_j \sim N(0, 1000^2)$. [§]Prior distribution for $\beta \sim N(0, 1000^2)$ and $\tau \sim N(0, 1)/(0, .)$. [#]Prior distribution for $\beta \sim N(0, 1000^2)$, $\alpha_j \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$.

Intracranial Haemorrhage

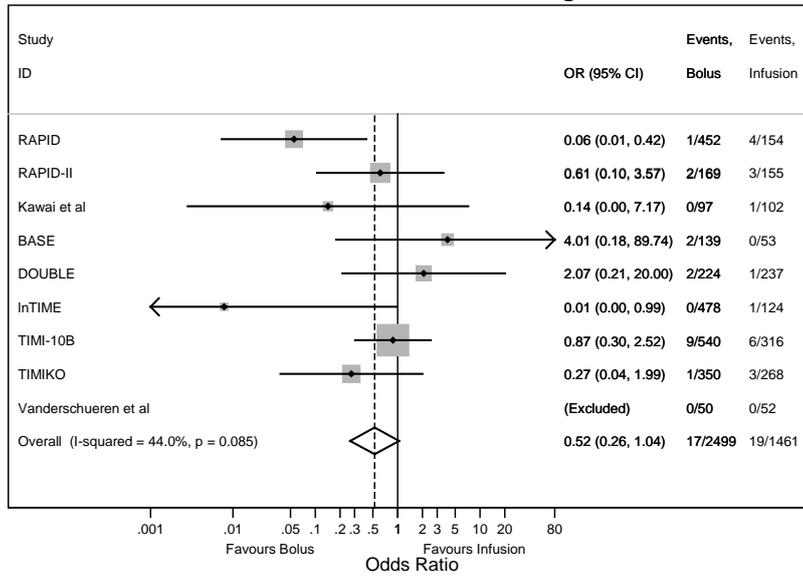


Figure 4.1: Forest plot for intracranial haemorrhage using the fixed-effect Peto meta-analysis model.

Stroke

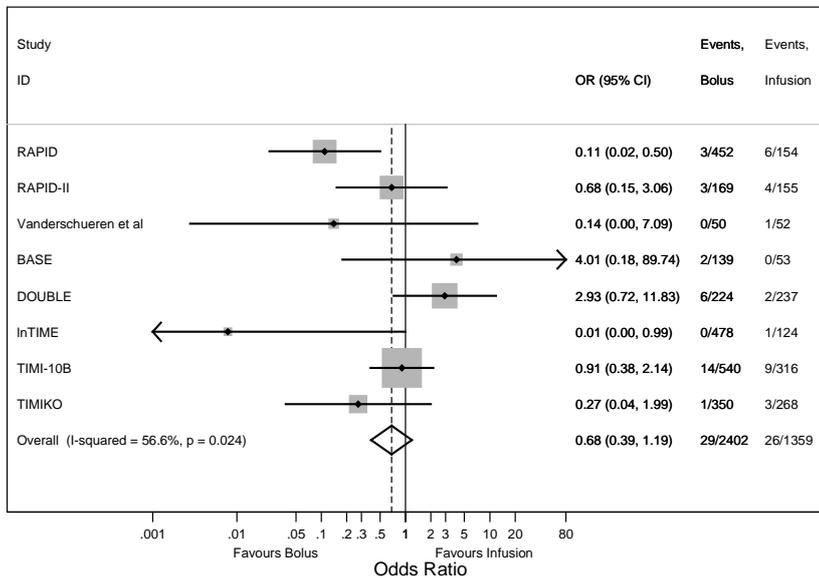


Figure 4.2: Forest plot for stroke using the fixed-effect Peto meta-analysis model.

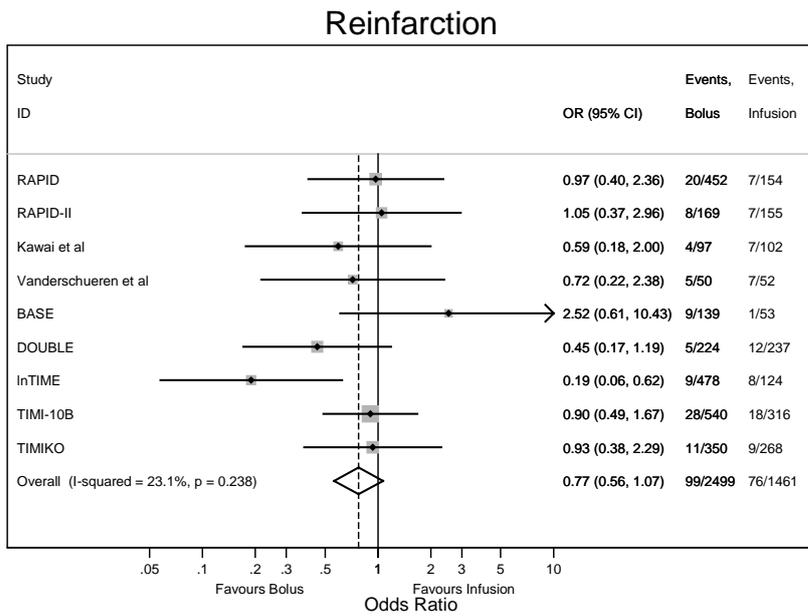


Figure 4.3: Forest plot for reinfarction using the fixed-effect Peto meta-analysis model.

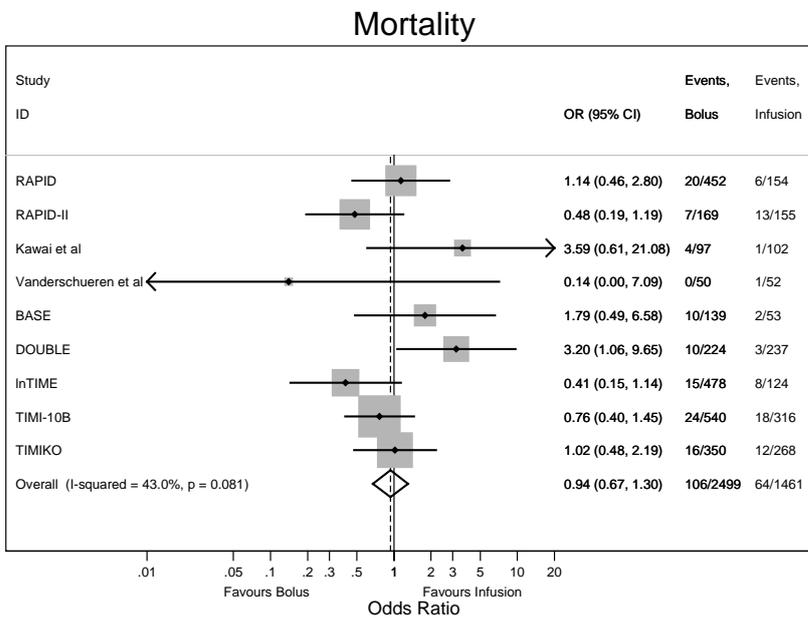


Figure 4.4: Forest plot for mortality using the fixed-effect Peto meta-analysis model.

Fixed-effect versus random-effects meta-analysis models

Under the frequentist and Bayesian frameworks, the random-effects meta-analysis models estimate a standard error of the log OR that is either similar to, or larger than, the standard error estimated by the fixed-effect methods, which results in wider CIs and CrIs. This is because the estimate of τ^2 is additionally included in the variance of the pooled log OR from the random-effects model (equation (1.18) and (1.20)). For example, for the frequentist fixed-effect Mantel-Haenszel method for mortality, the standard error of the log OR is 0.165 (95% CI OR: 0.679 to 1.294). Using REML (not accounting for uncertainty in $\hat{\tau}^2$), the standard error of the log OR is 0.221 (95% CI OR: 0.620 to 1.475).

The standard errors are only similar for the fixed-effect and random-effects models if the latter estimates τ^2 to be close to zero, because there is then only small additional variation to account for. For example, for reinfarction, $\hat{\tau}^2$ is <0.0001 using REML and the standard error of the log OR is 0.164, which is equal to the standard error that is estimated by the fixed-effect inverse variance method. This leads to the same confidence intervals under both the random-effects and fixed-effect models.

Frequentist results

Within the frequentist framework, there are differences in meta-analysis results between the estimation methods, although no estimation method results in a different overall clinical or statistical conclusion compared to the other approaches.

For the random-effects models, the standard error of the pooled log OR and $\hat{\tau}^2$ are lower for the meta-analysis method that uses ML compared to REML. For example, for mortality, the standard error is estimated to be 0.176 using ML (no uncertainty in $\hat{\tau}^2$), compared to 0.221 using REML, which results in a wider confidence interval for REML. The logistic regression

model also estimates τ^2 to be very small with $\hat{\tau}^2$ less than 0.0001, for all four outcomes. This is because the logistic regression also uses ML to estimate τ^2 . However, ML gives a downwardly biased estimate of τ , whereas REML and MM try to address this.²⁰⁰ Therefore, ML may not be the best method of estimation for meta-analysis of phase II trials where τ^2 is potentially large. For example, for stroke, τ^2 is estimated as 0.648 using REML but 0.355 using ML. Interestingly, MM and REML can also disagree. For example, for stroke the value of $\hat{\tau}^2$ is 0.306 using MM and 0.648 using REML.

As expected, for REML, ML, and MM the 95% CIs are wider when the uncertainty in the estimated between-trial variance is accounted for in the frequentist analysis by inflating the standard error of the pooled log OR. For example, for mortality, using REML the pooled OR has a 95% CI of 0.620 to 1.475 assuming no uncertainty in $\hat{\tau}^2$, but is 0.592 to 1.543 when assuming uncertainty in $\hat{\tau}^2$.

Frequentist versus Bayesian results

The pooled results within the Bayesian framework give similar conclusions to those within the frequentist framework for all four outcomes. However, the standard errors are larger for the Bayesian models, which results in wider 95% CrIs when compared to the 95% CIs in the equivalent frequentist models. For example, for reinfarction, the Bayesian logistic regression model with random-effects obtains the pooled OR of 0.773 with 95% CrI 0.502 to 1.179. The frequentist logistic regression with random-effects estimates the pooled OR to be 0.778 with 95% CI 0.568 to 1.066. The wider CrI is due to the uncertainty in $\hat{\tau}^2$ in the Bayesian model, which the frequentist model does not automatically account for. Also, the values of $\hat{\tau}^2$ are larger for the Bayesian meta-analysis models compared to the equivalent frequentist models. For the Bayesian inverse variance model, for stroke, $\hat{\tau}^2$ is 0.488 (95% CrI 0.002 to 3.236) compared to 0.306 for the frequentist inverse variance model (using MM). For the Bayesian

logistic regression model, $\hat{\tau}^2$ is 0.652 (95% CrI 0.004 to 3.896) compared to $\hat{\tau}^2 < 0.0001$ for the frequentist logistic regression model; a dramatic difference potentially due to the downward bias in the frequentist ML estimation of τ , but also may be due to the influence of the prior distribution for τ (see section 4.11).

Bayesian results

The median estimates and 95% CrI for $\hat{\tau}^2$ are larger for the logistic regression model compared to the inverse variance model. For example, for ICH and using the inverse variance method, the median value of $\hat{\tau}^2$ is 0.316 (95% CrI 0.001 to 3.095). Whereas, for the logistic regression model for ICH, the median value of $\hat{\tau}^2$ is 0.435 (95% CrI 0.002 to 3.662).

The Bayesian meta-analysis models also provide estimates of the probability that the summary OR is less than one (i.e. the probability that the effect of bolus therapy will be beneficial, on average across all populations). These estimates are shown in Table 4.7 to Table 4.10. For all estimation methods, the probability that the pooled OR is in favour of the bolus therapy is high for ICH, stroke, and reinfarction, with values greater than 0.80. For mortality, the probability is lower compared to the other outcomes with values between 0.50 and 0.75 depending on the model used.

The probability that the summary OR is less than one is lower for the random-effects analyses compared to the fixed-effect analyses as a result of the additional variation in the random-effects model. For example, for reinfarction, the fixed-effect logistic regression analysis estimates the probability to be 0.941, compared to an estimate of 0.901 that is estimated by the random-effects logistic regression analysis.

4.8 Results II: Prediction intervals for the treatment effect in a new trial

The prediction intervals for the true treatment effect in a new trial for each of the four outcomes, within the frequentist and Bayesian frameworks, are shown in Table 4.11 to Table 4.14. The Bayesian approach also allows us to obtain the predictive probability that the true OR in a new trial is less than one.

There is greater uncertainty in the predicted true ORs in a new trial compared to the average effect estimate. The 95% prediction intervals are wider compared to the corresponding 95% CIs and CrIs intervals of the pooled estimates within the frequentist and Bayesian frameworks, respectively. This is because the 95% CIs (and CrIs) relate to the uncertainty of the average effect, whereas the prediction intervals relate to the whole distribution of all effects. For example, for ICH, for the frequentist logistic regression method, the 95% CI for the summary OR is 0.275 to 1.050 and the 95% prediction interval for the true OR in a new trial is 0.239 to 1.207.

Where the estimate of $\hat{\tau}^2$ has large uncertainty in the Bayesian analyses, the estimated prediction interval is very wide. For example, for ICH, $\hat{\tau}^2$ is 0.435 (95% CrI 0.002 to 3.662) for the logistic regression analysis. The corresponding 95% CrI for the summary OR is 0.155 to 1.266 and the resulting 95% prediction interval for the true OR in a new trial is 0.049 to 3.809. The wide interval reflects the large estimated τ^2 combined with large uncertainty in β and τ .

4.8.1 Frequentist versus Bayesian prediction intervals

The Bayesian prediction intervals for the true ORs are considerably wider compared to the equivalent frequentist prediction intervals. This is because the Bayesian estimation accounts

for uncertainty in $\hat{\tau}^2$, whereas the frequentist methods do not (equation (4.20)). For example, using the Bayesian logistic regression method and for mortality, the 95% prediction interval for the true OR in a new trial is 0.261 to 4.293, whereas the 95% prediction interval for the OR in the frequentist logistic regression analysis is 0.632 to 1.386.

4.8.2 Bayesian predictive probability

The estimated probability that the true OR in a new trial will be less than one provides the probability that the bolus therapy will show a beneficial effect compared to the infusion therapy in a new trial. The predictive probability remains high for ICH, stroke, and reinfarction, with probability estimates of 0.7 or greater. These probabilities suggest that there is evidence of a potentially beneficial true treatment effect that may therefore motivate further investigation in a large phase III trial. The predictive probability for mortality is lower compared to the other outcomes, but the probability estimates are still greater than 0.50.

Table 4.11: Prediction results for ICH after the random-effects meta-analyses.

Frequentist method	Pooled log(OR)	Pooled OR	95% prediction interval for true OR in a new trial	$\hat{\tau}^2$	
REML (nounc)	-0.577	0.561	0.147 to 2.139	0.136	
REML	-0.577	0.561	0.131 to 2.405	0.136	
ML (nounc)	-0.521	0.594	0.242 to 1.456	<0.0001	
ML	-0.521	0.594	0.242 to 1.456	<0.0001	
MM (nounc)	-0.521	0.594	0.242 to 1.456	<0.0001	
MM	-0.521	0.594	0.242 to 1.456	<0.0001	
Logistic regression	-0.621	0.537	0.239 to 1.207	<0.0001	
Bayesian method	Pooled log(OR)	Pooled OR	95% prediction interval for true OR in a new trial	$\hat{\tau}^2$ (95% CrI)	Prob (new trial true OR <1)
Inverse variance*	-0.615	0.541	0.065 to 3.678	0.316 (0.001 to 3.095)	0.797
Logistic regression [†]	-0.724	0.485	0.049 to 3.809	0.435 (0.002 to 3.662)	0.824

nounc denotes assumption of no uncertainty in between-study variance estimate. *Prior distribution for $\beta \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$. [†]Prior distribution for $\beta \sim N(0, 1000^2)$, $\alpha_j \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$.

Table 4.12: Prediction results for stroke after the random-effects meta-analyses.

Frequentist method	Pooled log(OR)	Pooled OR	95% prediction interval for true OR in a new trial	$\hat{\tau}^2$	
REML (nounc)	-0.451	0.637	0.064 to 6.389	0.648	
REML	-0.451	0.637	0.063 to 6.414	0.648	
ML (nounc)	-0.430	0.650	0.109 to 3.875	0.355	
ML	-0.430	0.650	0.109 to 3.898	0.355	
MM (nounc)	-0.425	0.654	0.122 to 3.511	0.306	
MM	-0.425	0.654	0.122 to 3.511	0.306	
Logistic regression	-0.375	0.688	0.348 to 1.357	<0.0001	
Bayesian method	Pooled log(OR)	Pooled OR	95% prediction interval for true OR in a new trial	$\hat{\tau}^2$ (95% CrI)	Prob (new trial true OR <1)
Inverse variance*	-0.432	0.649	0.070 to 5.367	0.488 (0.002 to 3.236)	0.721
Logistic regression [†]	-0.486	0.615	0.052 to 5.659	0.652 (0.004 to 3.896)	0.729

nounc denotes assumption of no uncertainty in between-study variance estimate. *Prior distribution for $\beta \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$. [†]Prior distribution for $\beta \sim N(0, 1000^2)$, $\alpha_j \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$.

Table 4.13: Prediction results for reinfarction after the random-effects meta-analyses.

Frequentist method	Pooled log(OR)	Pooled OR	95% prediction interval for true OR in a new trial	$\hat{\tau}^2$	
REML (nounc)	-0.282	0.754	0.512 to 1.112	<0.0001	
REML	-0.282	0.754	0.512 to 1.112	<0.0001	
ML (nounc)	-0.282	0.754	0.512 to 1.112	<0.0001	
ML	-0.282	0.754	0.512 to 1.112	<0.0001	
MM (nounc)	-0.288	0.750	0.463 to 1.214	0.0227	
MM	-0.288	0.750	0.463 to 1.214	0.0227	
Logistic regression	-0.251	0.778	0.532 to 1.139	<0.0001	
Bayesian method	Pooled log(OR)	Pooled OR	95% prediction interval for true OR in a new trial	$\hat{\tau}^2$ (95% CrI)	Prob (new trial true OR <1)
Inverse variance*	-0.284	0.753	0.300 to 1.879	0.066 (0.000 to 0.758)	0.814
Logistic regression [†]	-0.258	0.773	0.295 to 2.054	0.076 (0.000 to 0.864)	0.787

nounc denotes assumption of no uncertainty in between-study variance estimate. *Prior distribution for $\beta \sim N(0, 1000^2)$, and $\tau \sim N(0, 1) / (0, .)$. [†]Prior distribution for $\beta \sim N(0, 1000^2)$, $\alpha_j \sim N(0, 1000^2)$, and $\tau \sim N(0, 1) / (0, .)$.

Table 4.14: Prediction results for mortality after the random-effects meta-analyses.

Frequentist method	Pooled log(OR)	Pooled OR	95% prediction interval for true OR in a new trial	$\hat{\tau}^2$	
REML (nounc)	-0.045	0.956	0.355 to 2.572	0.1262	
REML	-0.045	0.956	0.345 to 2.651	0.1262	
ML (nounc)	-0.097	0.906	0.554 to 1.483	0.0124	
ML	-0.099	0.906	0.507 to 1.618	0.0124	
MM (nounc)	-0.039	0.962	0.338 to 2.736	0.1437	
MM	-0.039	0.962	0.338 to 2.736	0.1437	
Logistic regression	-0.066	0.936	0.632 to 1.386	<0.0001	
Bayesian method	Pooled log(OR)	Pooled OR	95% prediction interval for true OR in a new trial	$\hat{\tau}^2$ (95% CrI)	Prob (new trial true OR <1)
Inverse variance*	-0.038	0.962	0.290 to 3.889	0.164 (0.000 to 1.443)	0.554
Logistic regression [†]	-0.008	0.992	0.261 to 4.293	0.213 (0.000 to 1.588)	0.520

nounc denotes assumption of no uncertainty in between-study variance estimate. *Prior distribution for $\beta \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$. [†]Prior distribution for $\beta \sim N(0, 1000^2)$, $\alpha_j \sim N(0, 1000^2)$, and $\tau \sim N(0, 1)/(0, .)$.

4.8.3 Bayesian prediction intervals for estimates of treatment effect in a new trial with a given sample size

The previous section focussed on the predicted *true* treatment effect in a new trial, $\theta_{i_{new}}$. Consider now predictions for the *estimated* treatment effect in a new trial, $Y_{i_{new}}$. For simplicity, focus solely on the outcome of ICH. Under the assumption of a baseline risk of 0.01 for ICH,²⁰¹ the 95% prediction intervals for the estimated treatment effect ($Y_{i_{new}}$) for ICH, for various given sample sizes, are shown graphically in Figure 4.5. These results were obtained using the logistic regression model (equation (4.19)) and option (1) for the variance of the treatment effect in a new trial with a given sample size (section 4.6.3). As the sample size per treatment group increases, the prediction interval narrows until approximately 2000 patients per arm where the increase in sample size reduces the width of the prediction interval by a negligible amount. Therefore, if a future trial is planned using the information in Figure 4.5, the recommendation would likely be to recruit no more than 2000 patients per treatment group.

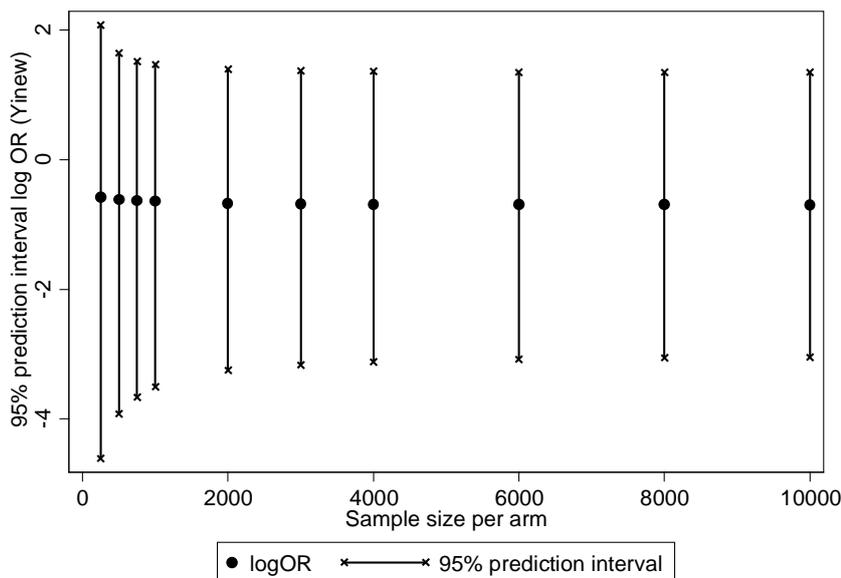


Figure 4.5: Prediction intervals for *estimated* treatment effect in a new trial ($Y_{i_{new}}$) with ICH for given sample size assuming option (1).

The probability that bolus therapy will be shown to be effective in the new trial is illustrated in Figure 4.6, for varying chosen sample sizes and for each of options (1) and (2). This refers to the proportion of samples in the MCMC estimation where the CrI around the estimated $Y_{i_{new}}$ does not contain an OR of 1. As the sample size increases, the probability of success in a new trial also increases, which reflects the narrower credible intervals that arise from larger patient numbers. Options (1) and (2) give reasonably similar results.

When the sample size is unrealistically large (10,000,000 patients per arm), such that the trial is tending toward an infinite sample size, the probability of success tends to the probability that $\exp(\theta_{i_{new}})$ is less than one, which equals 0.824 as noted in Table 4.11. For more realistic sample sizes, the probability of success is much lower. For example, with 2000 patients in each arm of the trial the probability of success is only about 0.4. However, increasing to 4000 patients per arm increases the success probability to about 0.6. Therefore, if this figure is used to design a future trial, the recommended sample size would probably not exceed 4000 patients per treatment group. This information would need to be combined with several other factors, such as costs and prevalence of the condition of interest, to determine what sample size is finally required.

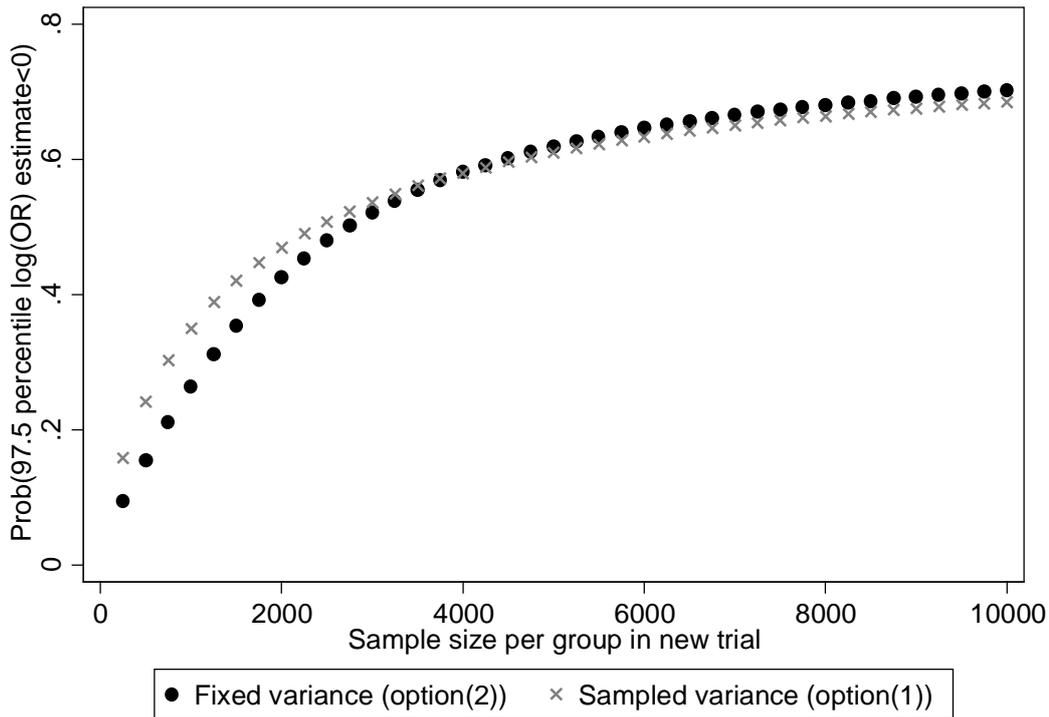


Figure 4.6: Probability that the upper bound of the estimated OR's 95% credible interval will be less than 1 in a new trial for ICH with a given sample size.

Plots such as Figure 4.5 and Figure 4.6 could be useful for the design of a new trial to determine what sample size is necessary to achieve a high probability of obtaining a result that favours the new treatment, whilst also enabling the investigator to determine where an increased sample size may not provide a reduction in the variance, or an increase in the probability of a positive result that are large enough to justify the additional cost and time of recruiting the extra patients.

4.8.4 Comparison with subsequent phase III results

Comparing summary effects

The comparison of phase II and phase III trial meta-analyses, by Eikelboom et al., for the four outcomes is displayed graphically in Figure 4.7. They only compared the summary ORs

and their CIs. The results for ICH appear to show dramatically different conclusions between phase II and phase III results, whereas the results for stroke, reinfarction and mortality appear similar in the phase II and phase III trials. For ICH, the pooled summary OR is 0.53 (95% CI OR: 0.27 to 1.01) for the phase II meta-analysis and 1.25 (95% CI OR: 1.06 to 1.49) for the phase III meta-analysis. This led Eikelboom et al. to state that the phase III trial results contradicted the phase II trial results, i.e. the phase III results were unexpected.

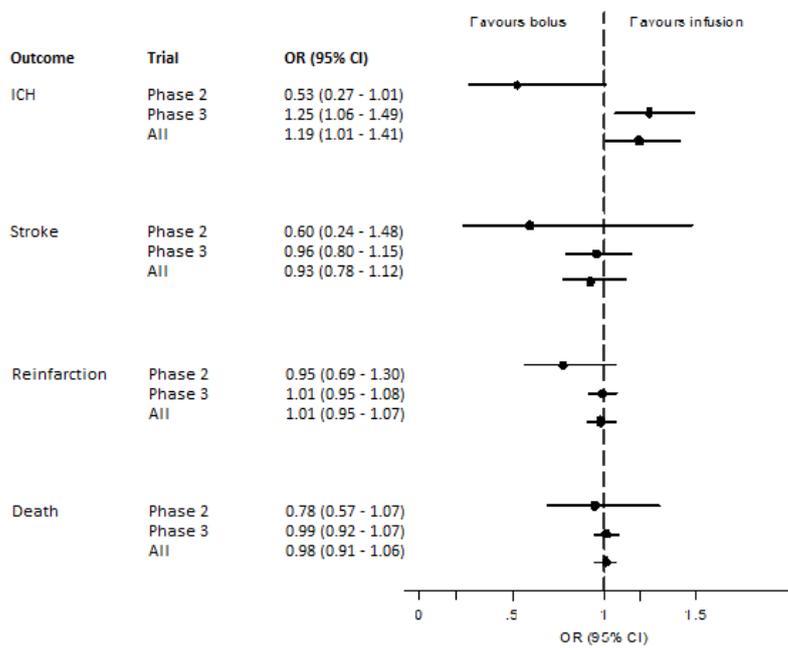


Figure 4.7: Phase II and phase III meta-analyses comparison plot by Eikelboom et al.¹⁶⁹

Comparing individual phase III trial results with phase II prediction intervals

The derived prediction intervals can also be compared to the results of the individual phase III trials themselves, not just their average effect. This comparison is actually more appropriate since the intention of the prediction intervals is to predict the true treatment effect ($\theta_{i_{new}}$) in a new study. The trial estimates from four of the subsequent phase III trials in AMI¹⁹³⁻¹⁹⁶ are displayed in and Table 4.16 (results from the other two trials from the original

paper were not accessible for use in this chapter). In contrast to the phase II trials, only some point estimates are in favour of bolus therapy compared to infusion therapy, whereas others are in favour of infusion therapy. Figure 4.8 shows the log OR estimates and their 95% CIs for ICH in the four phase III trials and compares them to the 95% prediction interval for θ_{new} from the phase II meta-analysis using the Bayesian inverse variance method and frequentist maximum likelihood. Figure 4.9 to Figure 4.11 are similar plots for stroke, reinfarction and mortality.

The estimated ORs for ICH, stroke, reinfarction and mortality in the subsequent phase III trials are supported by Bayesian and frequentist prediction intervals from the phase II meta-analysis (Figure 4.8 to Figure 4.11, Table 4.11 to Table 4.14). Therefore, if one had used the prediction intervals using the phase II trials, then all the subsequent phase III trial results were plausible. This contradicts Eikelboom et al. who considered them to be unexpected, but they ignored heterogeneity and uncertainty, and this has now been accounted for, particularly for the Bayesian models.

Summary

In summary, this work enhances the work of Eikelboom et al. and shows that after a phase II meta-analysis, the 95% prediction interval for the true treatment effect in a new trial (or the estimated treatment effect in a new trial) should be used to help decide whether to proceed to phase III, rather than just the 95% CI of the summary pooled estimate. The pooled estimate relates to the average effect across all trials, but a single trial's treatment effect may substantially differ due to heterogeneity.

Table 4.15: Phase III randomised trials of bolus vs. infusion thrombolytic therapy in AMI.

Trial	Bolus (No. of boluses)	Infusion	N (Bolus)	N (Infusion)	Length of follow-up
INJECT, ¹⁹³ 1995	Retepase, 10 MU (2)	Streptokinase, 1.5 MU	2992	2994	35 days
COBALT, ¹⁹⁴ 1997	Alteplase, 40-50 mg, (2)	Alteplase, 100 mg	3584	3585	30 days
GUSTO III, ¹⁹⁵ 1997	Retepase, 10 MU (2)	Alteplase, up to 100 mg	4921	10138	30 days
ASSENT-2, ¹⁹⁶ 1999	Tenecteplase 30-50 mg (1)	Alteplase, 100 mg	8488	8461	30 days

Table 4.16: Results from subsequent randomised phase III trials comparing bolus to infusion thrombolytic therapy for the in-hospital treatment of AMI.

Outcome	Phase III trial	No. of events on bolus/N (%)	No. of events on infusion/N (%)	OR	95% CI OR
ICH	INJECT ¹⁹³	23 / 3004 (0.8)	11 / 3006 (0.4)	2.092	1.018 - 4.300
	COBALT ¹⁹⁴	40 / 3585 (1.1)	29 / 3584 (0.8)	1.379	0.853 to 2.229
	GUSTO III ¹⁹⁵	92 / 10138 (0.91)	43 / 4921 (0.87)	1.039	0.722 to 1.494
	ASSENT-2 ¹⁹⁶	79 / 8461 (0.9)	80 / 8488 (0.9)	0.991	0.725 to 1.354
Stroke	INJECT	14 / 3004 (0.5)	19 / 3006 (0.6)	0.737	0.369 to 1.474
	COBALT	31 / 3585 (0.86)	25 / 3584 (0.7)	1.240	0.730 to 2.104
	GUSTO III	61 / 10138 (0.6)	37 / 4921 (0.7)	0.800	0.531 to 1.206
	ASSENT-2	72 / 8461 (0.85)	61 / 8488 (0.72)	1.184	0.841 to 1.668
Reinfarction	INJECT	150 / 3004 (5.0)	162 / 3006 (5.4)	0.927	0.738 to 1.164
	COBALT	140 / 3585 (3.9)	147 / 3584 (4.1)	0.952	0.752 to 1.206
	GUSTO III	426 / 10138 (4.2)	207 / 4921 (4.2)	0.999	0.843 to 1.183
	ASSENT-2	347 / 8461 (4.1)	323 / 8488 (3.8)	1.078	0.923 to 1.258
Mortality	INJECT	270 / 3004 (9.0)	285 / 3006 (9.5)	0.948	0.797 to 1.128
	COBALT	286 / 3585 (8.0)	270 / 3584 (7.5)	1.059	0.891 to 1.259
	GUSTO III	757 / 10138 (7.47)	356 / 4921 (7.24)	1.032	0.906 to 1.176
	ASSENT-2	523 / 8461 (6.2)	522 / 8488 (6.2)	1.005	0.887 to 1.139

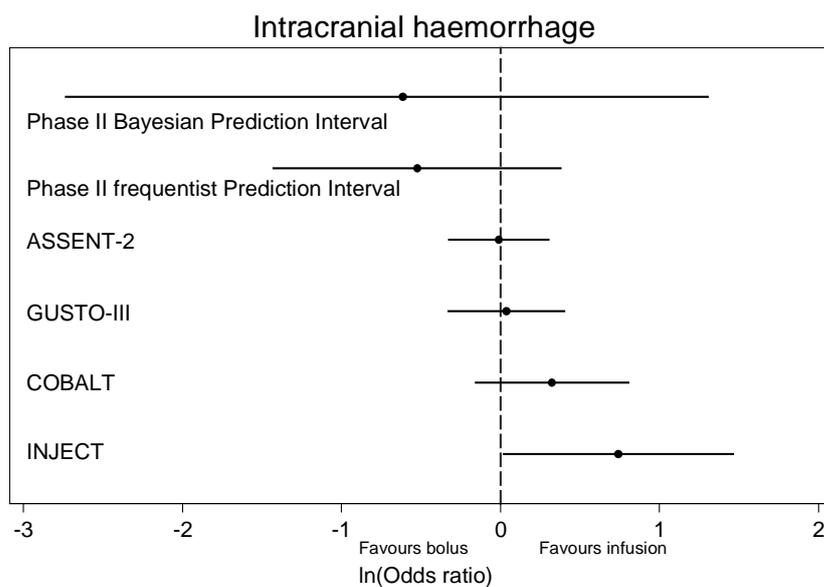


Figure 4.8: Comparison of prediction intervals for θ_{new} from a phase II meta-analysis with the estimate and 95% CI for the In(OR) from phase III trials for ICH.

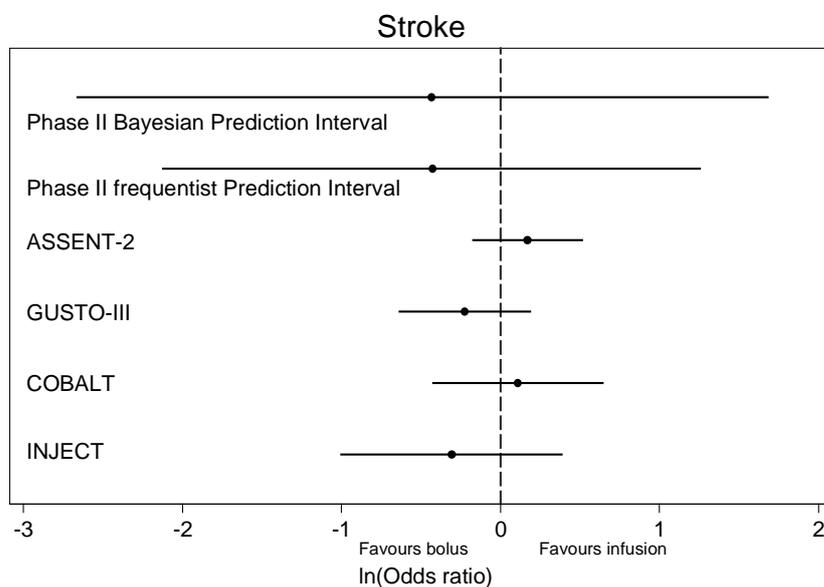


Figure 4.9: Comparison of prediction intervals for θ_{new} from a phase II meta-analysis with the estimate and 95% CI for the In(OR) from phase III trials for stroke.

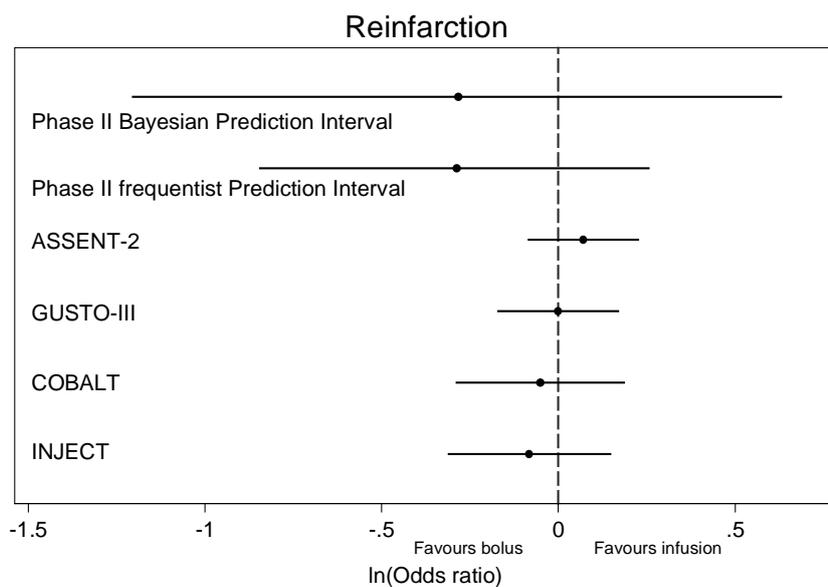


Figure 4.10: Comparison of prediction intervals for θ_{inew} from a phase II meta-analysis with the estimate and 95% CI for ln(OR) from phase III trials for reinfarction.

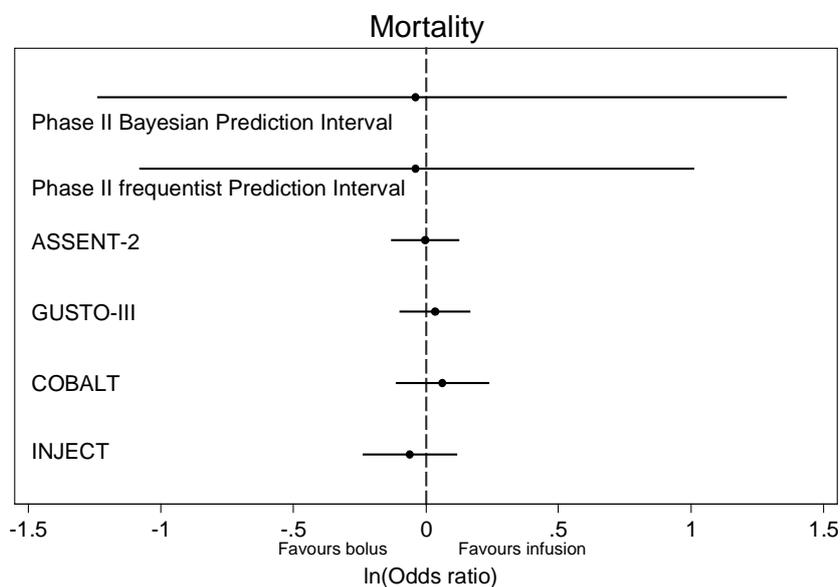


Figure 4.11: Comparison of prediction intervals for θ_{inew} from a phase II meta-analysis with the estimate and 95% CI for the ln(OR) from phase III trials for mortality.

4.9 Results III: Assessment of publication bias

A concern is that phase II trials are prone to bias, especially in the reporting of their results. Publication bias is an issue that occurs if studies with more favourable results are more likely to be published than those with less favourable results. Other types of bias related to publication bias can also occur, such as language bias and selective outcome reporting bias. These forms of bias can lead to potentially distorted results.²⁰² The threat of publication bias can be explored with funnel plots of study standard error of $\log(\text{OR})$ versus $\log(\text{OR})$ for each outcome. If there is no publication bias the assumption is that the studies should be symmetrically distributed about the summary estimate of the effect size in a funnel-like shape. Eikelboom et al. suggest that for the AMI trials there is evidence of publication bias shown in the funnel plot in Figure 4.12. However, this plot includes the study estimates from the phase II and phase III trials all together. This may not be appropriate if the summary effect estimate is different for the phase II and phase III trial meta-analyses since the funnel plot may suggest there is asymmetry, but this is due to the heterogeneity and not publication bias.²⁰³

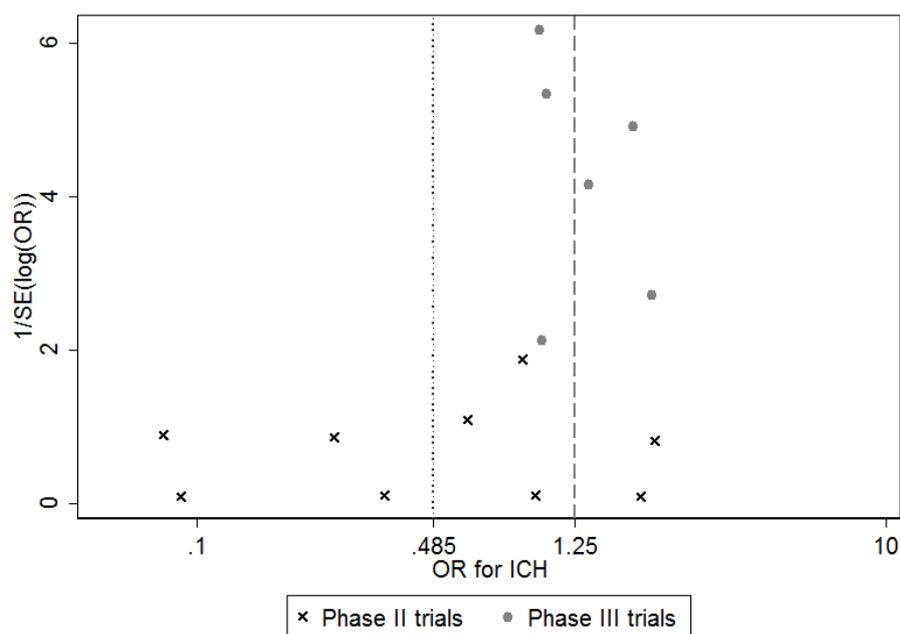


Figure 4.12: Funnel plot of $1/SE(\log(OR))$ versus OR for ICH for phase II and phase III trials.

(Note: The dotted line represents the summary treatment effect estimate for ICH from the Bayesian random-effects logistic regression analysis of the phase II trials; the dashed line indicates the summary treatment effect estimate for ICH by Eikelboom et al. for the phase III trials. The OR axis is shown on the log-scale.)

Therefore, separate funnel plots are now shown for the phase II trials only for each outcome in Figure 4.13 to Figure 4.16. They all suggest there is no clear evidence of publication bias since the studies with high standard error are observed equally in both directions of the summary effect estimate (the summary estimate shown in the figures corresponds to the Bayesian random-effects logistic regression analysis of the phase II trials). Therefore, these results cast doubt on the conclusion by Eikelboom et al. that publication bias could be a contributing factor towards the disagreement between the results of the phase II and phase III trials.

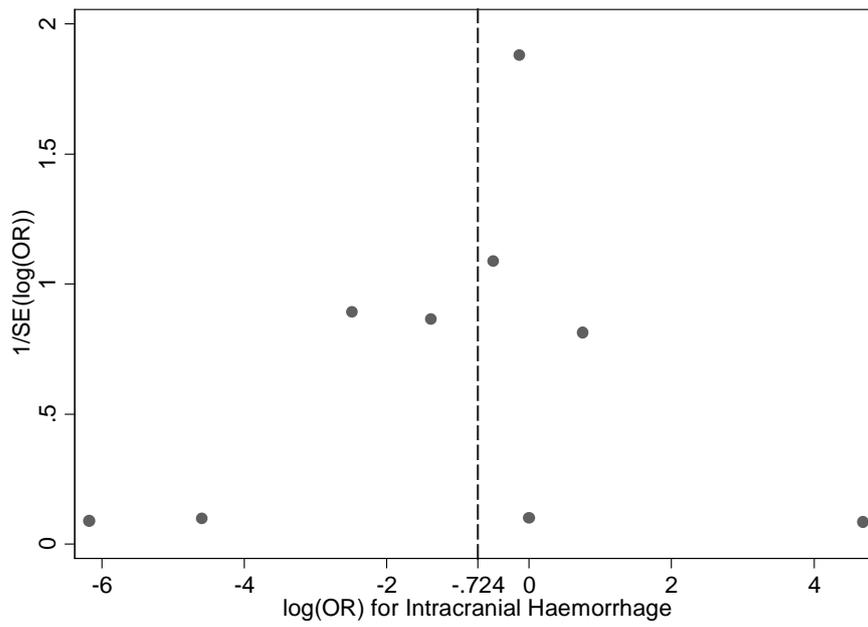


Figure 4.13: Funnel plot of $1/SE(\log(OR))$ vs. $\log(OR)$ for ICH in phase II trials.

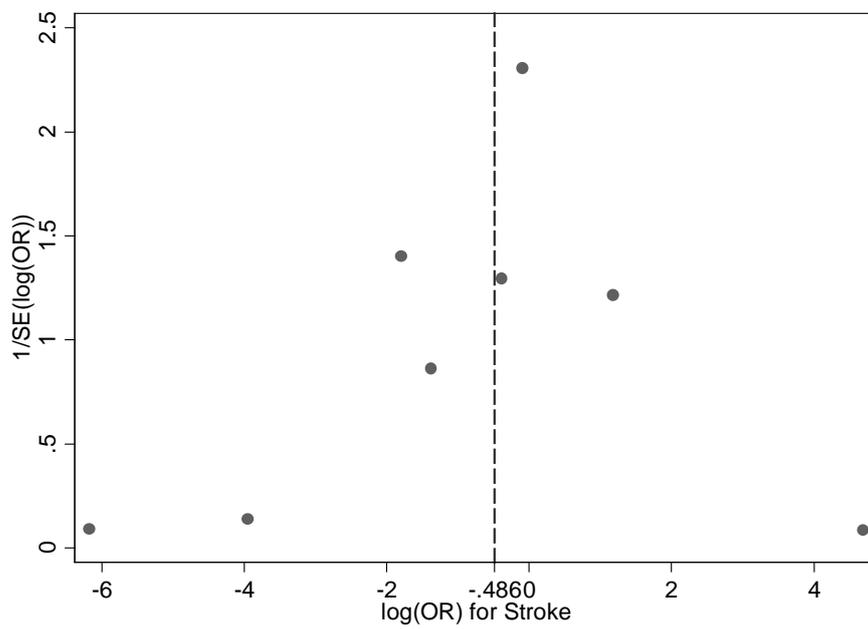


Figure 4.14: Funnel plot of $1/SE(\log(OR))$ vs. $\log(OR)$ for stroke in phase II trials.

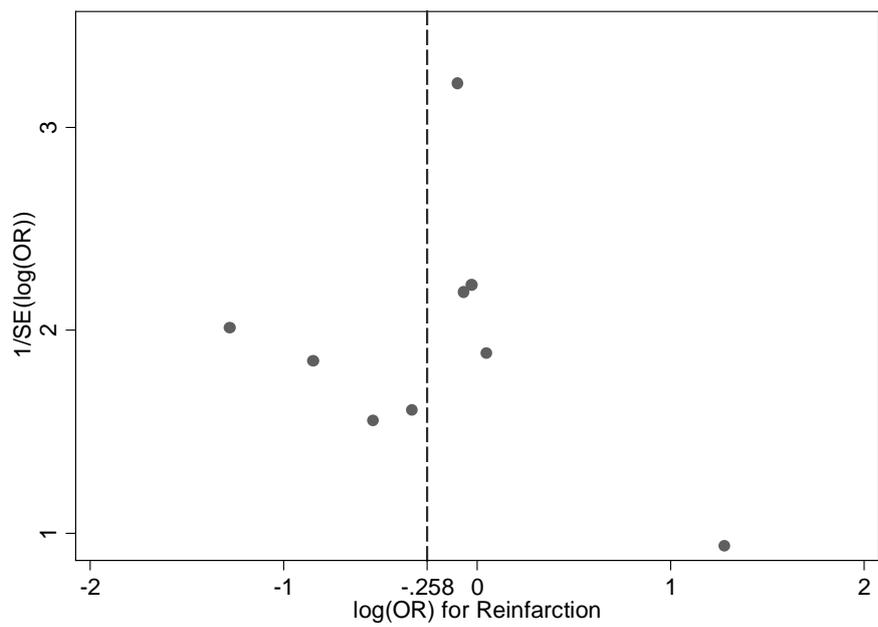


Figure 4.15: Funnel plot of $1/SE(\log(OR))$ vs. $\log(OR)$ for reinfarction in phase II trials.

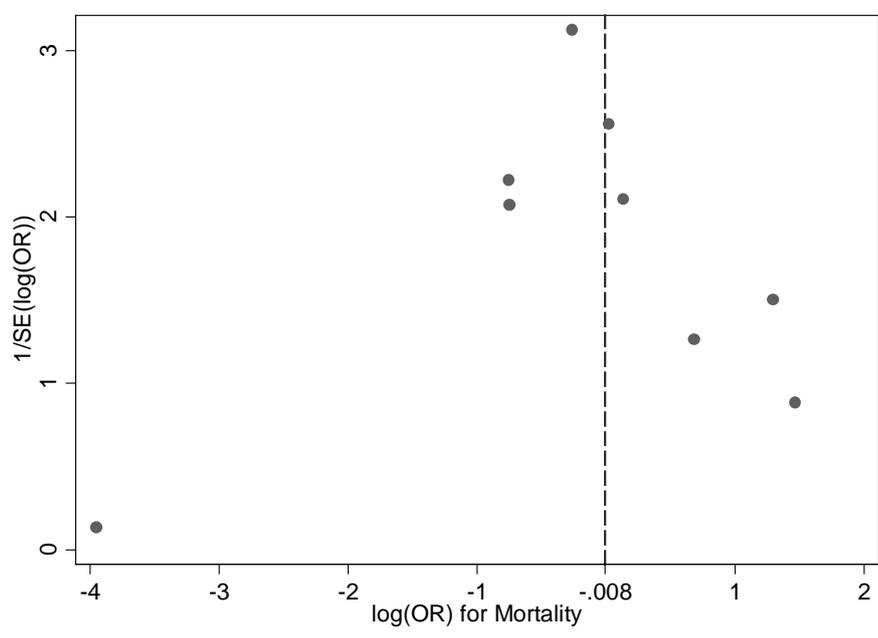


Figure 4.16: Funnel plot of $1/SE(\log(OR))$ vs. $\log(OR)$ for mortality in phase II trials.

4.10 Results IV: What if we thought the phase II results have occurred by chance?

Although section 4.9 suggests there is no clear evidence of publication bias, the phase II trial results may still be optimistic compared to the results in the subsequent phase III trials. The results of the individual phase III trials in Table 4.16 show that the treatment effect estimates are lower, with ORs closer to one, than those from the phase II trials (Table 4.2). Hypothetically, if the difference in risk between the treatment arms in the phase II trials has occurred by chance, it may be important to account for this optimism in the predictions of phase III results. Further, even if they are always published, phase II trials may be more prone to bias in their design, execution, analysis, and reporting, which may also cause optimistic summary meta-analysis results for phase II trials.

It is possible to limit the potential optimism in phase II trials by using a Bayesian meta-analysis and estimating the predictive probability that the true OR in a new trial will be less than 0.9 or 0.8, rather than less than 1, which is essentially a more stringent criterion for success. The results for reinfarction, as an example, are shown in Table 4.17.

The optimism can also be accounted for with a sceptical prior distribution on the pooled treatment effect that does not allow for large treatment effects.²² In a paper by Higgins & Spiegelhalter²⁰⁴, they considered a sceptical prior distribution on the pooled log OR such that there is little chance (5%) that the new treatment would reduce the odds of myocardial infarction by more than 25%. This results in a normal distribution for the log OR which is centred on zero and has variance 0.03. The results for reinfarction, assuming this sceptical prior distribution, are also shown in Table 4.17.

Table 4.17: Accounting for potential optimism in phase II trials in Bayesian meta-analysis: results for reinfarction.

Method	Summary OR from phase II meta-analysis	95% prediction interval new trial OR	Prob (new trial OR <1)	Prob (new trial OR <0.9)	Prob (new trial OR <0.8)
Inverse variance					
Vague prior	0.753	0.300 to 1.879	0.814	0.719	0.580
Sceptical prior	0.882	0.391 to 2.109	0.701	0.548	0.362
Logistic regression					
Vague prior	0.773	0.295 to 2.054	0.787	0.692	0.545
Sceptical prior	0.893	0.381 to 2.218	0.682	0.531	0.348

Probability of an OR<0.9

As expected, the estimated probability that the pooled OR is less than 0.9 or less than 0.8 is lower compared to the estimated probability that the OR is less than 1. For example, for the logistic regression model with random-effects and using the vague prior distributions (model (4.19)), the estimated probability that the true OR in a new trial is less than one is 0.787 compared to 0.692 for the probability that the true OR in a new trial is less than 0.9. If the conclusion to progress to Phase III is based on this probability of a larger summary treatment effect then this may alter the decision here.

Prediction interval following sceptical prior

The 95% prediction intervals for the true OR in a new trial, assuming the sceptical prior distribution for log OR, are closer to one compared to the results that assume the vague prior. For example, for the logistic regression model assuming the vague prior distribution, the 95% prediction interval for the true OR is 0.295 to 2.054. Whereas, assuming the sceptical prior distribution, the prediction interval is 0.381 to 2.218. Also, the predictive probabilities that the true OR is less than 1, 0.9, and 0.8, are lower for the sceptical prior distribution analysis compared to the vague prior analysis. For example, for the vague prior logistic regression analysis, the predictive probabilities are 0.787, 0.692, and 0.545, for the probability that the true OR is less than 1, 0.9, and 0.8, respectively. However, these probability estimates are 0.682, 0.531, and 0.348 when the sceptical prior distribution for the log OR is assumed. This again may alter the decision to proceed to phase III. The posterior distributions for the true treatment effect in a new study, using the vague and sceptical prior distributions for the pooled treatment effect, and the logistic regression model, are shown in Figure 4.17 as an illustration.

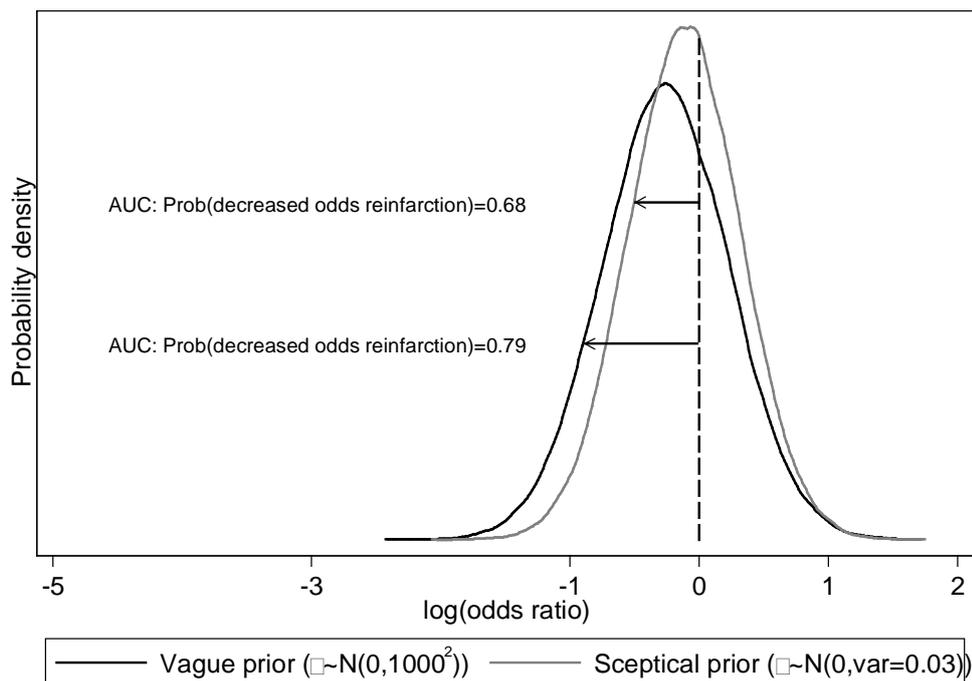


Figure 4.17: Posterior distributions for the treatment effect (log(OR)) in a new trial for reinfarction assuming a sceptical and vague prior distribution for the pooled average treatment effect.

Summary

This example shows that potential bias and optimism can be considered in the inference of results from a meta-analysis from phase II trials. The Bayesian framework can limit such bias through sceptical prior distributions and by estimating probabilities of a larger treatment effectiveness.

4.11 Results V: choice of prior distribution for between-study variance

As there are only nine trials in the phase II meta-analysis, a concern is the potential influence of prior distributions in the Bayesian analyses. So far, all the Bayesian analyses have used a prior distribution for the between-study standard deviation of $N(0,1)/(0,)$. The results of the sensitivity analysis for the prior distribution for the between-study variance, using the random-effects logistic regression model, are shown in Table 4.18 for the pooled results, and Table 4.19 for the prediction results, for mortality. Table 4.20 and Table 4.21 show the corresponding results for ICH. For mortality in the logistic regression model using $\tau \sim N(0,1)/(0,)$, $\hat{\tau}^2$ is 0.213 (95% CrI: 0.000 to 1.588), whereas for ICH $\hat{\tau}^2$ is 0.435 (95% CrI: 0.002 to 3.662). The sensitivity analysis considers 14 different prior distributions for the between-study variance (Table 4.5).

For mortality, the summary results for the log OR are similar for all prior distributions. The standard deviation of the pooled log OR, and the estimate of τ^2 , vary slightly for the different priors, which results in different CrI for the summary log(OR). For example, the standard deviation of the pooled log OR is 0.432 for prior 4a, whereas it is 0.230 for prior 3a. The median estimate of τ^2 for prior 4a is 0.663 (95% CrI 0.043 to 5.632), whereas for prior 3a, the median is 0.018 (95% CrI 0.000 to 1.157). The 95% CrI for the summary OR is 0.462 to 2.521 for prior 4a, whereas the CrI is 0.636 to 1.578 for prior 3a.

For ICH, the summary results are also similar for the log OR. Figure 4.18 shows the posterior distribution of the true treatment effect in a new trial (θ_{new}) for priors 2 and 6b. However, compared to the results for mortality, the differences in the estimated standard error of log(OR) and $\hat{\tau}^2$ between the 14 prior distributions are larger. For example, for prior 3a the standard deviation of the pooled log OR is 0.505, whereas for prior 1b it is 0.917. The

median estimate of τ^2 for prior 3a is 0.028 (95% CrI 0.000 to 5.606), whereas for prior 1b, the median is 1.421 (95% CrI 0.007 to 26.53). The 95% CrI for the summary OR is 0.175 to 1.216 for prior 3a, whereas the CrI is 0.056 to 2.039 for prior 1b. The prior distribution also has a large effect on the prediction interval for ICH. For prior 3a the 95% prediction interval for the true OR in a new trial is 0.065 to 2.945, whereas for prior 1b the interval is 0.004 to 30.27.

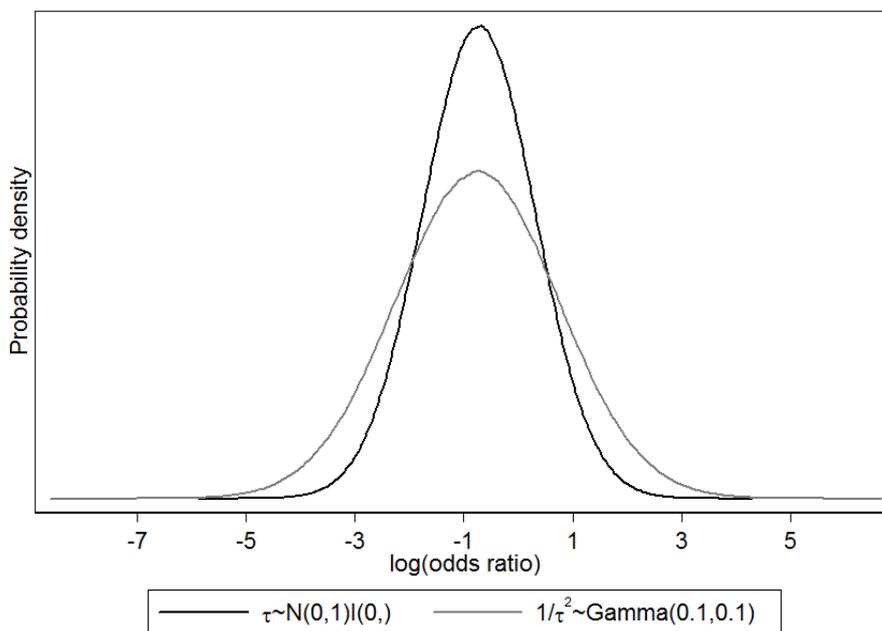


Figure 4.18: Posterior distribution for the true treatment effect in a new trial (θ_{new}) for ICH assuming two different prior distributions for the heterogeneity parameter.

The differences in the standard error and between-study variance estimates are partly due to the varying degree of information in the prior distribution because there is always some information, even in a supposedly vague distribution. The differences also arise because the prior distribution is placed on different scales, which include the standard deviation, variance, log(variance) and precision. The results for ICH and mortality show that it may also arise due

to the size of $\hat{\tau}^2$ and the uncertainty of this estimate, as the results are more sensitive for ICH compared to mortality.

Perhaps most importantly, for either mortality or ICH, the probability that the average OR and the true OR in a new trial are less than one is similar for most of the prior distributions for both outcomes. The probabilities for Prior 3a and 3b and 7a (or 7b) are slightly higher compared to the remaining prior distributions. Empirical priors 7a and 7b were, however, developed for use in meta-analysis of phase III trials, and thus may not reflect the greater magnitude of heterogeneity that is expected in phase II trials.

In summary, although the prior distribution is intended to be vague and thus not influence the results, Table 4.18 to Table 4.21 show that a vague prior distribution may still contribute substantially to the meta-analysis results when there is only a small amount of data,²⁹ which is the case in a meta-analysis of phase II trials. The prior distribution for the between-study uncertainty must be selected with caution and a sensitivity analysis is advisable to ensure the results are not influenced by the choice this distribution. Prediction probabilities for the true treatment effect in a new trial in this example are reasonably robust to the choice of prior.

Table 4.18: Results of sensitivity analysis on prior distribution for between-study variance for pooled results for mortality using random-effects logistic regression model.

Prior distribution for τ	Log(OR)	SE(log(OR))	Summary OR	95% CrI OR	$\hat{\tau}^2$ (95% CrI)	Prob summary OR<1
1a: $\tau \sim \text{Unif}(0,10)$	0.008	0.322	1.008	0.565 to 2.044	0.278 (0.001 to 2.734)	0.488
1b: $\tau \sim \text{Unif}(0,100)$	0.008	0.322	1.008	0.565 to 2.045	0.278 (0.001 to 2.739)	0.488
1c: $\tau \sim \text{Unif}(0,2)$	-0.002	0.313	0.998	0.565 to 1.995	0.278 (0.001 to 2.352)	0.502
2: $1/\tau^2 \sim \text{Gamma}(0.1, 0.1)$	0.018	0.307	1.018	0.578 to 1.961	0.295 (0.047 to 1.961)	0.474
3a: $\text{Log}(\tau^2) \sim \text{Unif}(-10,10)$	-0.039	0.230	0.961	0.636 to 1.578	0.018 (0.000 to 1.157)	0.577
3b: $\text{Log}(\tau^2) \sim \text{Unif}(-10,1.386)$	-0.043	0.229	0.958	0.644 to 1.585	0.020 (0.000 to 1.113)	0.585
4a: $\tau^2 \sim \text{Unif}(0.001,1000)$	0.036	0.432	1.037	0.462 to 2.521	0.663 (0.043 to 5.632)	0.456
4b: $\tau^2 \sim \text{Unif}(0.001,4)$	0.037	0.379	1.038	0.504 to 2.350	0.614 (0.040 to 3.172)	0.453
5a: $1/\tau^2 \sim \text{Pareto}(1,0.001)$	0.039	0.421	1.040	0.465 to 2.508	0.641 (0.043 to 5.379)	0.453
5b: $1/\tau^2 \sim \text{Pareto}(1,0.25)$	0.040	0.382	1.041	0.501 to 2.370	0.614 (0.041 to 3.165)	0.451
6a: $\tau \sim \text{N}(0,100) \parallel [0,]$	0.010	0.324	1.010	0.562 to 2.044	0.285 (0.001 to 2.750)	0.484
6b: $\tau \sim \text{N}(0,1) \parallel [0,]^*$	-0.008	0.283	0.992	0.599 to 1.843	0.213 (0.000 to 1.588)	0.513
7a: $\tau^2 \sim \text{logN}(-4.06,1.45^2)$	-0.050	0.192	0.952	0.664 to 1.411	0.021 (0.001 to 0.300)	0.604

* Primary prior distribution for between-study variance.

Table 4.19: Results of sensitivity analysis on prior distribution for between-study variance for prediction intervals for true treatment effect for mortality using random-effects logistic regression model.

Prior distribution for τ	95% prediction interval new trial OR	Prob (new trial OR<1)
1a: $\tau \sim \text{Unif}(0,10)$	0.205 to 5.786	0.504
1b: $\tau \sim \text{Unif}(0,100)$	0.205 to 5.800	0.504
1c: $\tau \sim \text{Unif}(0,2)$	0.208 to 5.518	0.516
2: $1/\tau^2 \sim \text{Gamma}(0.1, 0.1)$	0.231 to 5.035	0.493
3a: $\text{Log}(\tau^2) \sim \text{Unif}(-10,10)$	0.396 to 2.765	0.577
3b: $\text{Log}(\tau^2) \sim \text{Unif}(-10,1.386)$	0.396 to 2.725	0.585
4a: $\tau^2 \sim \text{Unif}(0.001,1000)$	0.102 to 11.80	0.486
4b: $\tau^2 \sim \text{Unif}(0.001,4)$	0.134 to 9.041	0.485
5a: $1/\tau^2 \sim \text{Pareto}(1,0.001)$	0.102 to 11.44	0.484
5b: $1/\tau^2 \sim \text{Pareto}(1,0.25)$	0.132 to 9.057	0.483
6a: $\tau \sim \text{N}(0,100) \parallel [0,]$	0.202 to 5.854	0.501
6b: $\tau \sim \text{N}(0,1) \parallel [0,]^*$	0.261 to 4.293	0.520
7a: $\tau^2 \sim \text{logN}(-4.06,1.45^2)$	0.536 to 1.797	0.585

* Primary prior distribution for between-study variance.

Table 4.20: Results of sensitivity analysis on prior distribution for between-study variance for pooled results for ICH using random-effects logistic regression model.

Prior distribution for τ	Log(OR)	SE(log(OR))	Summary OR	95% CrI OR	$\hat{\tau}^2$ (95% CrI)	Prob summary OR<1
1a: $\tau \sim \text{Unif}(0,10)$	-0.785	0.877	0.456	0.059 to 2.055	1.380 (0.004 to 24.78)	0.892
1b: $\tau \sim \text{Unif}(0,100)$	-0.785	0.917	0.456	0.056 to 2.039	1.421 (0.007 to 26.53)	0.893
1c: $\tau \sim \text{Unif}(0,2)$	-0.755	0.579	0.470	0.135 to 1.403	0.838 (0.004 to 3.685)	0.923
2: $1/\tau^2 \sim \text{Gamma}(0.1, 0.1)$	-0.766	0.657	0.465	0.114 to 1.472	0.681 (0.060 to 9.896)	0.919
3a: $\text{Log}(\tau^2) \sim \text{Unif}(-10,10)$	-0.638	0.505	0.529	0.175 to 1.216	0.028 (0.000 to 5.608)	0.946
3b: $\text{Log}(\tau^2) \sim \text{Unif}(-10,1.386)$	-0.695	0.436	0.499	0.201 to 1.086	0.019 (0.000 to 2.735)	0.960
4a: $\tau^2 \sim \text{Unif}(0.001,1000)$	-0.916	1.723	0.400	0.009 to 6.323	5.001 (0.193 to 128.8)	0.834
4b: $\tau^2 \sim \text{Unif}(0.001,4)$	-0.800	0.652	0.449	0.113 to 1.525	1.581 (0.086 to 3.839)	0.908
5a: $1/\tau^2 \sim \text{Pareto}(1,0.001)$	-0.911	1.586	0.402	0.012 to 5.289	4.507 (0.170 to 107.2)	0.840
5b: $1/\tau^2 \sim \text{Pareto}(1,0.25)$	-0.806	0.657	0.447	0.113 to 1.560	1.587 (0.086 to 3.842)	0.904
6a: $\tau \sim \text{N}(0,100) \parallel [0,]$	-0.798	0.914	0.450	0.060 to 2.031	1.383 (0.008 to 26.87)	0.898
6b: $\tau \sim \text{N}(0,1) \parallel [0,]^*$	-0.724	0.524	0.485	0.155 to 1.266	0.435 (0.002 to 3.662)	0.937
7b: $\tau^2 \sim \text{logN}(-3.02,1.85^2)$	-0.652	0.371	0.521	0.244 to 1.051	0.052 (0.009 to 0.295)	0.965

* Primary prior distribution for between-study variance.

Table 4.21: Results of sensitivity analysis on prior distribution for between-study variance for prediction intervals for true treatment effect for ICH using random-effects logistic regression model.

Prior distribution for τ	95% prediction interval new trial OR	Prob (new trial OR<1)
1a: $\tau \sim \text{Unif}(0,10)$	0.004 to 28.95	0.753
1b: $\tau \sim \text{Unif}(0,100)$	0.004 to 30.27	0.749
1c: $\tau \sim \text{Unif}(0,2)$	0.032 to 5.751	0.788
2: $1/\tau^2 \sim \text{Gamma}(0.1, 0.1)$	0.021 to 7.938	0.787
3a: $\text{Log}(\tau^2) \sim \text{Unif}(-10,10)$	0.065 to 2.945	0.888
3b: $\text{Log}(\tau^2) \sim \text{Unif}(-10,1.386)$	0.094 to 2.097	0.907
4a: $\tau^2 \sim \text{Unif}(0.001,1000)$	0.000 to 1788.0	0.671
4b: $\tau^2 \sim \text{Unif}(0.001,4)$	0.020 to 8.652	0.741
5a: $1/\tau^2 \sim \text{Pareto}(1,0.001)$	0.000 to 970.6	0.677
5b: $1/\tau^2 \sim \text{Pareto}(1,0.25)$	0.020 to 8.496	0.742
6a: $\tau \sim \text{N}(0,100) \parallel [0,]$	0.004 to 29.93	0.753
6b: $\tau \sim \text{N}(0,1) \parallel [0,]^*$	0.049 to 3.809	0.824
7b: $\tau^2 \sim \text{logN}(-3.02,1.85^2)$	0.203 to 1.260	0.929

* Primary prior distribution for between-study variance.

4.12 Discussion

This chapter has considered many possible statistical methods for meta-analysis of phase II trials, and shown how the results may be used to make predictions and thus inform decisions about new phase III trials. Key findings of the work are summarized in Figure 4.19, and it has raised a number of interesting questions, which are now discussed. The work has recently been published in the journal *Trials*²⁷ (Appendix B3).

4.12.1 Is a meta-analysis useful to the phase III trial decision?

The decision to progress to phase III is based on all existing evidence, which includes information other than the results of phase II trials, such as cost and feasibility. However, if multiple phase II trials exist, a meta-analysis of the phase II trials should be considered important. The work in this chapter shows how univariate meta-analysis of phase II trials can inform the decision by summarizing existing evidence and predicting treatment effects in new trials. The Bayesian approach appears most natural for this as it accounts for all parameter uncertainty in the predictions and estimation.

A meta-analysis of phase II trials can also reveal the large uncertainty upon which the phase III trial decision is based, even after the results of the individual trials have been pooled. For example, for stroke, the median OR is 0.615 with a 95% CrI 0.201 to 1.578 (using a Bayesian random-effects logistic regression model). The 95% prediction interval for the true treatment effect in a new trial, using the same estimation method, is wider than the CrI from 0.052 to 5.659. The small sample sizes in phase II trials, and the rare event rate in these particular trials, combined with large between-trial heterogeneity in treatment effects, are the key contributing factors to the large uncertainty in this prediction.

Figure 4.19: Key findings.

- It is difficult to examine and quantify heterogeneity in a meta-analysis of phase II trials due to the small number of studies and the small number of patients within studies.
- Bayesian credible intervals (and prediction intervals) are wider compared to the equivalent frequentist intervals because the Bayesian model accounts for more uncertainty compared to the frequentist approach.
- Prediction intervals for the true (or estimated) treatment effect in a new study should be derived to help consider whether to proceed to phase III rather than just considering the 95% CI of the summary pooled estimate.
- Bayesian meta-analysis additionally allows us to estimate probabilities regarding parameters of interest, which can contribute to phase III trial decisions, such as the probability the treatment is truly effective in a new trial, or the probability, in a new trial with a given sample size, that the CrI of the estimated treatment effect will not contain the null value.
- Bayesian meta-analysis methods can limit potential bias and optimism in the prediction intervals from phase II results through sceptical prior distributions and by estimating probabilities of more stringent or large treatment effectiveness.
- Prior distributions for heterogeneity in Bayesian meta-analysis can influence the summary results when there are few phase II trials, and so sensitivity analysis to the choice of prior is recommended and consideration of empirical prior distributions for the size of heterogeneity in previous phase II meta-analyses.

The analysis in Section 4.8 has shown that 95% prediction intervals for the true treatment effect in a new trial may help the decision process to proceed to phase III because the intervals consider the potential effect of the new treatment when it is applied in a new individual study, i.e. a phase III trial, rather than the uncertainty of the average treatment effect that is calculated in a 95% CI. For example, for reinfarction, the 95% CIs for the OR are narrow and only just include one (95% CI OR using REML: 0.546 to 1.041), which suggests strong evidence that the bolus therapy reduces the risk of reinfarction compared to infusion therapy on average. However, the 95% prediction intervals include values larger

than one, so although on average the bolus therapy seems effective, it may not always be beneficial in an individual trial or population (95% prediction interval for true OR using MM: 0.463 to 1.214).

4.12.2 Advantages of Bayesian meta-analysis compared to meta-analysis in a frequentist framework

Bayesian probability estimates can aid the decision to proceed to phase III by properly quantifying the uncertainty in the results. For example, for reinfarction, the probability that the true OR in a new trial will be less than one is estimated at 0.814 (using the inverse variance method). This large probability may encourage the decision to proceed to phase III. Likewise, the probability estimates can also reveal the uncertainty in the treatment effect estimates, such as for mortality. The probability that the true OR is less than one for this outcome is estimated to be 0.554 (using the inverse variance method), which suggests that there is large uncertainty as to whether the new treatment will reduce the risk compared to the standard treatment. Decision-makers and pharmaceutical companies may be less inclined to spend investments when probabilities are this low.

The Bayesian framework naturally incorporates uncertainty in the summary results and the between-trial variance. This leads to prediction intervals that can account for more of the uncertainty, and thus are wider, compared to the frequentist equivalent, which assumes that the between-trial variance is fixed. Frequentist methods that inflate the standard error of the pooled effect estimate to account for the uncertainty in the between-study variance still have narrower prediction intervals compared to the Bayesian prediction intervals.

The Bayesian framework can also account for any potential bias or optimism by adopting methods such as estimating probabilities of large treatment effectiveness, or assuming

sceptical prior distributions when calculating prediction intervals. The quality of the phase II trials was not formally evaluated in this chapter, but this could also be done to identify the higher quality phase II trials. Funnel plots can also be used to check for small study effects and potential publication bias, but this did not appear to be an issue for the AMI phase II trials available.

4.12.3 Quantifying heterogeneity in phase II meta-analysis

The estimates of the Q-statistic, I^2 and τ^2 show that it is difficult to quantify and examine the potential heterogeneity in a meta-analysis of phase II trials due to the small sample sizes, which causes low power and large within-study variation. I^2 is always likely to be small when within-study variances are large and p-values from the Q-test are unlikely to reach significance given less than 10 trials, even when heterogeneity truly exists. Since these phase II trials have small patient numbers and have been conducted separately it is likely that there is heterogeneity that needs to be accounted for, and so researchers may decide, a priori, that a random-effects model will be used.

4.12.4 Recommendations for undertaking phase II meta-analyses of binary outcomes

The choice of meta-analysis methods can influence the decision about whether to proceed to phase III. Table 4.22 summarises the recommendations for good practice within meta-analysis of phase II trials with binary outcomes. The logistic regression approach is recommended because there is no need to add a continuity correction given a zero count in one treatment, and it avoids the need to approximate the observed data by a normal distribution,^{184,205,206} which is unlikely to be appropriate when the event rate is rare.

Table 4.22: Recommendations for improved meta-analysis of phase II trials of binary outcomes.

Issue	Recommendation
Framework	Use a logistic regression model to model the binomial distribution of the data within studies, and to avoid continuity corrections given a zero event in one arm.
Choice of model	Do not make decisions to use a fixed-effect or a random-effects model based on I^2 or tests for heterogeneity.
Heterogeneity	State <i>a priori</i> that a random-effects model will be used to account for heterogeneity in treatment effects.
Prediction intervals	Report 95% prediction intervals as they reveal the potential treatment effect in a new population, and inform subsequent phase III decisions.
Bias	Use sceptical prior distributions for the treatment effect if there is evidence to suggest the phase II trials may be biased in favour of the treatment.

4.12.5 Relevance of this work to recent meta-analyses of phase II trials

This chapter has focussed on the meta-analysis of phase II trials conducted by Eikelboom et al.¹⁶⁹ in which they ignored heterogeneity by using a fixed-effect model. There are other examples, in more recent years, where the method for meta-analysing phase II studies could be improved similarly. In particular, the decision to use a fixed-effect or a random-effects model is often based on the p-value derived from the Q statistic chi-squared test for heterogeneity,⁴¹ and/or the I^2 statistic.²⁰⁷⁻²¹⁰ The fixed-effect model is selected if the p-value is not statistically significant and/or the I^2 is low. However, if there are few studies there is very low power to detect heterogeneity, and therefore a significant p-value is unlikely in the meta-analysis of phase II trials. Thus, genuine heterogeneity may be ignored. Similarly, the example in AMI in this chapter showed low values of I^2 are also potentially misleading for phase II meta-analysis.

There are two known meta-analyses of phase II trials where the authors decided *a priori* that a random-effects model was more appropriate because of the expectation that the studies would estimate different, yet related, treatment effects.^{211,212} However, in these and other articles using a random-effects model, the conclusions only focussed on the pooled estimate of treatment effect, and the prediction interval for the treatment effect in a new trial was not considered.²⁰⁷⁻²¹² Thus, the full uncertainty of the potential treatment effect in new populations (or phase III trials) is often ignored.

4.12.6 Further work

This chapter has explored the use of univariate meta-analysis of the comparison between bolus therapy and infusion therapy, which assumes that the treatment effect is independent for the four outcomes. However, the effect of the treatment may be correlated for the outcomes and a multivariate meta-analysis can account for this. A multivariate approach allows the joint synthesis of all the outcomes, which can use more statistical information, and joint inferences can be made for multiple outcomes. Therefore, Chapter 5 is an exploration of the use of multivariate meta-analysis on the same dataset in AML.

In the sensitivity analysis for the prior distribution for the between-study variance in the random-effects meta-analysis model, two empirical prior distributions were considered that were derived by Turner et al.³⁶ for a future meta-analysis with a binary outcome. Empirical prior distributions have been derived for several health-care settings, for both binary³⁶ and continuous outcomes.³⁸ However, they have not been considered specifically in the context of phase II trial meta-analysis, therefore further work is needed to consider whether the existing empirical prior distributions are appropriate in this setting.

4.13 Conclusion

Univariate meta-analysis of phase II trials can inform phase III trial decisions. A frequentist or Bayesian framework can be applied, a fixed-effect or random-effects approach can be assumed, and multiple estimation methods can be adopted. These choices can influence the decision about whether to proceed to phase III and thus need to be clearly documented and investigated whenever a phase II meta-analysis is performed.

Eikelboom et al. originally applied a fixed-effect Mantel-Haenszel meta-analysis of phase II trials and compared the results to a meta-analysis of subsequent phase III trials. They concluded that there were conflicting results between the two meta-analyses for ICH. However, in this chapter, random-effects meta-analysis with estimated prediction intervals from phase II trials shows that the results are not necessarily contradictory.

CHAPTER 5: MULTIVARIATE META-ANALYSIS OF PHASE II TRIALS WITH MULTIPLE OUTCOMES

5.1 Context and objective

The previous chapter investigated the use of univariate meta-analysis when there are multiple phase II studies in order to account for all available evidence when planning a phase III trial. However, if there are multiple outcomes in these trials, a multivariate meta-analysis approach may be more statistically efficient as it can account for correlation between the outcomes.²¹³ For example, overall survival is likely to be correlated to progression-free survival; therefore a bivariate meta-analysis of both outcomes is potentially more efficient than separate meta-analyses, as the correlation provides additional information to the estimation of both outcomes.^{214,215} However, although appealing, multivariate meta-analyses are often rejected in favour of separate univariate meta-analyses, due to the added complexity of the multivariate approach and the requirement to specify correlations, which are often unknown.

The objective of this chapter is to consider the use of multivariate meta-analysis in the context of Phase II trials with multiple outcomes. In particular, to examine how results from a Bayesian multivariate meta-analysis can be used to estimate the probability that the treatment is effective for each outcome separately or across combinations of the outcomes. Furthermore, to evaluate whether the multivariate approach adds value over the univariate approach from Chapter 4, in regards to predicting treatment efficacy in new (phase III) trials.

Thus, this chapter evaluates the benefits and limitations of multivariate meta-analysis of phase II trials, and shows how they can be useful toward phase III decisions. The chapter will also explore how applicable the multivariate meta-analysis approach is in the setting of

phase II trials. The focus here is on bivariate meta-analysis approaches, which is a joint meta-analysis of two outcomes. The bivariate results will be compared to the univariate meta-analysis results, of Chapter 4, to determine the impact of borrowing strength and accounting for potential correlation between outcomes, and to determine: (i) whether the conclusions remain the same if the correlations are accounted for; (ii) how the posterior probabilities for each outcome's treatment effect are affected jointly; and (iii) whether joint probability estimates across both outcomes can further aid the decision to conduct a phase III trial.

5.2 Clinical data and multiple outcomes of interest

The clinical example in this chapter is the same nine randomised controlled phase II trials that were used to explore univariate meta-analysis methods in Chapter 4.¹⁶⁹ These phase II trials evaluated the efficacy of bolus thrombolytic therapy compared with standard infusion therapy for the in-hospital treatment of acute myocardial infarction (AMI). The trials are used to examine multivariate meta-analysis methods for combining small phase II trials. They are also used to predict phase III trial results. The clinical objective of the meta-analysis is to assess if there is any difference in the risk of adverse event outcomes for patients who receive bolus therapy compared to those patients that receive infusion therapy. The hope is that bolus therapy reduces odds of the adverse event compared to those that receive infusion therapy, and if so, further phase III research might be funded to demonstrate this formally. As seen in Chapter 4, for all results in this chapter, an odds ratio (OR) less than one indicates that those patients who receive bolus therapy have lower odds of the event compared to those that receive infusion therapy.

The multiple outcomes of interest in this chapter are three serious adverse events that were examined in Chapter 4: stroke, reinfarction, and mortality. The primary objective is to explore

the use of a bivariate meta-analysis of stroke and mortality. These outcomes are selected as the primary focus because treatment effects for stroke are likely to be correlated with treatment effects for mortality; also, if a patient has had a stroke they may have a higher risk of mortality.²¹⁶ Further, stroke is missing for one study, which allows one to utilise correlation by borrowing strength from the data for mortality when estimating the treatment effect for stroke. A secondary analysis explores the bivariate meta-analysis of reinfarction and mortality. The data from the original nine papers of the phase II randomised trials is shown in Table 4.2 in Chapter 4 and in Table 5.1 below.

Table 5.1: Incidence of stroke, reinfarction, and mortality in nine phase II trials of bolus vs. infusion thrombolytic therapy in AMI.¹⁷¹⁻¹⁷⁹

Trial	Sample size, N		Outcome	No. of events (%)		Log(OR)	Var(log(OR))	OR	95% CI OR
	Bolus	Infusion		Bolus	Infusion				
RAPID, ¹⁷¹ 1995	452	154	Stroke	1 (0.7)	6 (3.9)	-1.80	0.51	0.17	0.04 to 0.67
			Reinfarction	20 (4.4)	7 (4.5)	-0.03	0.20	0.97	0.40 to 2.35
			Mortality	20 (4.4)	6 (3.9)	0.13	0.23	1.14	0.45 to 2.90
RAPID-II, ¹⁷² 1996	169	155	Stroke	3 (1.8)	4 (2.6)	-0.38	0.60	0.68	0.15 to 3.10
			Reinfarction	8 (4.7)	7 (4.5)	0.05	0.28	1.05	0.37 to 2.97
			Mortality	7 (4.1)	13 (8.4)	-0.75	0.23	0.47	0.18 to 1.22
Kawai et al., ¹⁷³ 1997	97	102	Stroke	-	-	-	-	-	-
			Reinfarction	4 (4.1)	7 (6.9)	-0.54	0.41	0.58	0.17 to 2.06
			Mortality	4 (4.1)	1 (1.0)	1.470	1.27	4.34	0.48 to 39.57
Vanderschueren et al., ¹⁷⁴ 1997	50	52	Stroke	0 (0)	1 (1.9)	-1.08	2.71	0.34	0.01 to 8.54
			Reinfarction	5 (10.0)	7 (13.4)	-0.34	0.39	0.71	0.21 to 2.42
			Mortality	0 (0)	1 (1.9)	-1.08	2.71	0.34	0.01 to 8.54
BASE, ¹⁷⁵ 1998	139	53	Stroke	2 (1.4)	0 (0)	0.67	2.43	1.95	0.09 to 41.19
			Reinfarction	9 (6.5)	1 (1.9)	1.28	1.14	3.60	0.44 to 29.13
			Mortality	10 (7.2)	2 (3.8)	0.68	0.63	1.98	0.42 to 9.34
DOUBLE, ¹⁷⁶ 1998	224	237	Stroke	6 (2.7)	2 (0.8)	1.17	0.68	3.23	0.65 to 16.19
			Reinfarction	5 (2.2)	12 (5.1)	-0.85	0.29	0.43	0.15 to 1.24
			Mortality	10 (4.5)	3 (1.3)	1.29	0.44	3.65	0.99 to 13.42

Table 5.1: continued.

Trial	Sample size, N		Outcome	No. of events (%)		Log(OR)	Var(log(OR))	OR	95% CI OR
	Bolus	Infusion		Bolus	Infusion				
InTIME, ¹⁷⁷ 1998	478	124	Stroke	0 (0)	1 (0.8)	-2.45	2.68	0.09	0.00 to 2.12
			Reinfarction	9 (1.9)	8 (6.5)	-1.28	0.25	0.28	0.11 to 0.74
			Mortality	15 (3.1)	8 (6.5)	-0.76	0.20	0.47	0.19 to 1.13
TIMI-10B, ¹⁷⁸ 1998	540	316	Stroke	14 (2.6)	9 (2.8)	-0.10	0.19	0.91	0.39 to 2.12
			Reinfarction	28 (5.2)	18 (5.7)	-0.10	0.10	0.91	0.49 to 1.66
			Mortality	24 (4.4)	18 (5.7)	-0.26	0.10	0.77	0.41 to 1.44
TIMIKO, ¹⁷⁹ 1998	350	268	Stroke	1 (0.3)	3 (1.1)	-1.37	1.34	0.25	0.03 to 2.45
			Reinfarction	11 (3.1)	9 (3.4)	-0.07	0.21	0.93	0.38 to 2.29
			Mortality	16 (4.6)	12 (4.5)	0.02	0.15	1.02	0.48 to 2.20

5.3 Multivariate random-effects meta-analysis methods using a Bayesian approach

In contrast to Chapter 4, where frequentist and Bayesian methods were compared, this chapter assumes a Bayesian approach only. This decision is for a number of reasons; one of which is that Chapter 4 highlighted several advantages of the Bayesian approach compared to the frequentist approach, such as the ability to easily account for uncertainty in parameters of interest. Also, many of the comparisons between the Bayesian and frequentist methods are likely to be similar for the analyses in this chapter, such as wider Bayesian credible intervals compared to frequentist confidence intervals for pooled treatment effects, due to the additional uncertainty that is accounted for in the Bayesian approach.

This chapter focuses on the use of bivariate meta-analysis, which is a joint meta-analysis of two outcomes; however the methods that are explored can be extended to more than two outcomes. The methodology for multivariate meta-analysis within a Bayesian framework has recently been developed although it remains underexplored.^{17,18} The potential advantages of this approach, compared to the frequentist equivalent, include the ability to make joint probability statements regarding quantities of interest, and the ability to make predictive statements, conditional on the current state of knowledge. Prior distributions can also be specified to quantify the evidence prior to the meta-analysis data, and to account for uncertainty in the parameters, which may not be easily accounted for in the frequentist framework, such as the uncertainty in the estimate of between-study heterogeneity.

5.3.1 Random-effects multinomial regression analysis

Chapter 4 highlighted that the meta-analysis of binary data should ideally model the binomial distribution of the data exactly, and this is especially important when there are small numbers

of patients and rare event rates. If this is not possible (e.g. if only odds ratios are available rather than two-by-two tables) an approximation can be made by assuming that the treatment effect estimates in each trial are normally distributed. The extension to the logistic regression model for a single outcome to when there are multiple binary outcomes is a multinomial regression analysis. In order to use this approach, the responses must be mutually exclusive and exhaustive.²¹⁷ For example, the four response categories for stroke and mortality are shown in Table 5.2, where π_{ikl} denotes the probability of response $k=1,\dots,4$ in trial i ($i=1,\dots,9$) for those in treatment group l ($l=0$ or 1).

Table 5.2: Probability of observed event for four mutually exclusive and exhaustive responses in study, i , for treatment group, l .

Prob (Outcome 1): Stroke, dead	Prob (Outcome 2): No stroke, dead	Prob (Outcome 3): Stroke, alive	Prob (Outcome 4): No stroke, alive
π_{i1l}	π_{i2l}	π_{i3l}	$1 - \pi_{i1l} - \pi_{i2l} - \pi_{i3l}$

In order to implement a multinomial model one would need to know the number of patients with response k in trial i for treatment group l .²¹⁷ Then from a multinomial model, the OR for the treatment effect in each of the categories for stroke and mortality can be derived. Unfortunately, without individual patient data (IPD) this model is not possible to fit as the number of events for each category in Table 5.2 is usually unknown; one just knows the number of strokes and the number of deaths, but not the number with *both* stroke and death. Thus, in this chapter, an approximate and more general method is necessary to undertake a joint analysis of these outcomes.

5.3.2 A general bivariate normal random-effects meta-analysis model

Since there is no IPD, an approximation assumes the two estimated treatment effects (one for each outcome) in each study are from a bivariate normal distribution, which leads to a

random-effects bivariate normal meta-analysis model on a logistic scale. This is the general framework to bivariate meta-analysis, and therefore can be applied to effects that relate to any outcomes (binary/continuous/survival, etc.). The approach considers $i=1$ to n studies, which examine the effect of treatment on two outcomes ($j=1$ or 2), and the bivariate model is written in model (5.1) where two observed treatment effects in study i , Y_{i1} and Y_{i2} (the log ORs for outcomes 1 and 2), have a bivariate normal distribution with means, θ_{i1} and θ_{i2} , respectively, and known within-study covariance matrix, \mathbf{S}_i . The within-study variance in trial i for outcome, j , is represented by s_{ij}^2 (the variance of logOR, Y_{ij}). The within-study covariances are represented by $\rho_{W_i} s_{i1} s_{i2}$, where ρ_{W_i} is the within-study correlation between Y_{i1} and Y_{i2} in study i .

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \mathbf{S}_i \right) \quad (5.1)$$

$$\mathbf{S}_i = \begin{pmatrix} s_{i1}^2 & \rho_{W_i} s_{i1} s_{i2} \\ \rho_{W_i} s_{i1} s_{i2} & s_{i2}^2 \end{pmatrix}$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \mathbf{D} \right)$$

$$\mathbf{D} = \begin{pmatrix} \tau_1^2 & \rho_B \tau_1 \tau_2 \\ \rho_B \tau_1 \tau_2 & \tau_2^2 \end{pmatrix}$$

Priors: $\beta_j \sim N(0, 1000^2); j=1,2$, $\tau_j \sim N(0,1) / (0,)$, $\rho_B \sim \text{Uniform}(-1,1)$

This bivariate random-effects model assumes that the true underlying effects in the i th study, θ_{i1} and θ_{i2} , arise from a bivariate normal distribution with average treatment effects, β_1 and β_2 , and between-study covariance matrix, \mathbf{D} . The diagonal elements of \mathbf{D} represent the between-study variances, τ_j^2 , for the treatment effect for outcome j , which represent the heterogeneity in θ_{i1} and θ_{i2} across studies. The off-diagonal elements, $\rho_B \tau_1 \tau_2$, are the between-study covariances, where ρ_B represents the between-study correlation for θ_{i1} and θ_{i2} . When

$\rho_{W_i}=\rho_B=0$, i.e. when the outcomes are assumed independent, the model reverts to two separate univariate meta-analyses; one for each outcome. The interpretation of within and between study correlation is explained further in section 5.3.3.

The Bayesian model (5.1) also specifies prior distributions for the average treatment effects, β_1 and β_2 , which are given vague normal prior distributions. The prior distribution for the between-study standard deviation is a normal distribution truncated at zero ($N(0,1)/(0,)$), and a prior uniform(-1,1) distribution is specified for the between-study correlation parameter, ρ_B . Chapter 4 showed that the choice of prior distribution for between-study variance parameters can have an impact on conclusions, and this will remain valid in the multivariate setting. However, in this chapter the focus is on whether the multivariate model (5.1) adds value over the univariate model of each outcome separately (i.e. where $\rho_B=\rho_{W_i}=0$ in model (5.1)), and so only the prior distributions given in model (5.1) will be considered in this chapter.

Where there are zero event rates for stroke and death in some of the studies, the ‘treatment arm’ continuity correction is applied.¹⁸¹ This continuity correction method was explored in Chapter 4, and it is an appropriate continuity correction if there is imbalance in the sample sizes in the treatment groups.

5.3.3 Within-study and between-study correlation

The within-study correlations for study i , ρ_{W_i} , and between-study correlation, ρ_B , are two measures of correlation in a multivariate meta-analysis model.²¹⁵ The within-study correlation for study i , ρ_{W_i} , is a measure of the association between the treatment effect estimates, Y_{i1} and Y_{i2} .²¹⁸ In the phase II trials in this chapter, potential within-study correlation arises from the fact that each individual in a study contributes information toward each of the multiple outcomes. This induces correlation in the effect estimates. For example, a patient who

experienced a stroke is at higher risk of death, and therefore individuals with a stroke in the dataset are often those who are dead. The within-study correlations in model (5.1) are assumed known.²¹⁵

The between-study correlation, ρ_B , is generally unknown and must be estimated in the meta-analysis model. This is a measure of how the true underlying outcome effects, θ_{i1} and θ_{i2} , are related across studies, which may have occurred because of between-study differences in the patient characteristics, such as age²¹⁵. For example, those studies with lower average age may have higher treatment effects on both stroke and mortality.

5.3.4 Unknown within-study correlations

In the bivariate meta-analysis model (5.1), the within-study correlations must be specified in addition to the s_{ij}^2 needed in a univariate meta-analysis. However, they are rarely reported within the original published articles and are often difficult to obtain.²¹³ There have been various suggestions for how to adequately deal with the unknown within-study correlations.^{219,220} The use of IPD has been described as the gold standard approach in which bootstrapping methods can be used to obtain estimates of the within-study correlations.^{213,221,222} If the IPD is not available for all studies, the within-study correlation estimates from the IPD in one study can be used to inform the likely value of the within-study correlation for the remaining studies.²¹³ Another option is to impute correlations over the entire range of values from -1 to 1 as a sensitivity analysis to determine how the correlations alter the pooled estimates.^{213,223}

The within-study correlation estimates were not reported within any of the nine phase II trials, and there is no available IPD. If the individual studies have large sample sizes, the impact of incorrectly specifying the within-study correlation is small because the within-study variation

is relatively small compared to between-study variation.²¹⁵ However, in general, phase II trials have small sample sizes, and so the within-study variations are relatively large. Thus, in this chapter, a number of different options are investigated to determine the impact of the within-study correlation values on the pooled estimates.

Option (1): Fixed values for within-study correlations

Within-study correlations lie between -1 and 1. An estimate of the overall correlation for the log ORs (which is an amalgam of within-study and between-study correlation²¹⁵) can be obtained by estimating the Pearson correlation coefficients using the summary estimates from each trial, and the correlation estimates are shown in Table 5.3. These estimates suggest that the overall correlations are positive between each pair of outcomes. This is also clinically sensible. The estimate of overall correlation is particularly strong for stroke and mortality (0.659), as expected, and this is also shown in Figure 5.1. Therefore, the within-study correlation estimates might also be assumed positive for these two outcomes. Indeed, Kirkham et al. suggest using the observed overall correlation to approximate the within-study correlation.²²¹ Therefore, positive within-study correlations are assumed here for stroke and mortality, and a range of different values from zero to 0.9 are investigated for all studies. Note that the correlation is also positive, but only mildly so, between mortality and reinfarction. For brevity, although there is evidence of correlation between reinfarction and stroke (correlation estimate of 0.402), a bivariate analysis of these two outcomes is not considered in this chapter.

Table 5.3: Pearson correlation coefficients for the estimates of log OR for stroke, reinfarction, and mortality.

	Stroke	Reinfarction	Mortality
Stroke	1		
Reinfarction	0.4024	1	
Mortality	0.6586	0.0879	1

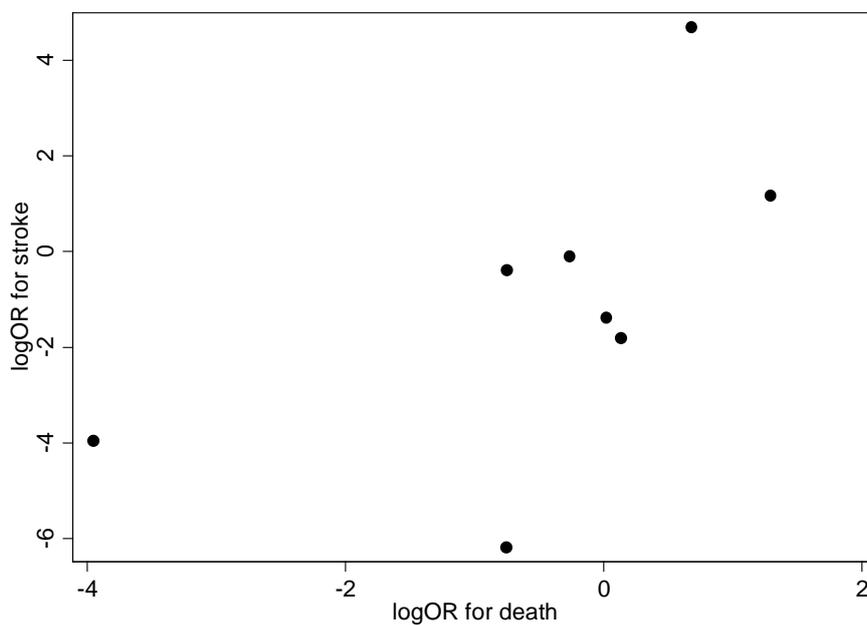


Figure 5.1: Scatter plot of treatment effect estimates (logOR) for stroke and mortality in each trial.

(Note: there are only eight observations as there is no information for stroke in one trial.)

Option (2): Prior distributions for within-study correlations

The Bayesian multivariate meta-analysis model can account for the unknown within-study correlations by assuming a prior distribution for these parameters.¹⁷ A positive uniform(0,1) distribution for the within-study correlations between stroke and mortality allows for all possible positive values for the within-study correlations, equally. The positive uniform

distribution is chosen since the estimate of overall correlation in option (1) is positive, and clinically this is sensible, as discussed. A second prior distribution is also considered, based on external but related clinical data leading to a uniform(0.02, 0.29) prior distribution. This is based on IPD from an original meta-analysis of ten trials in patients with hypertension, which investigated the use of anti-hypertensive drugs against placebo or no treatment for the prevention of death and stroke, among other outcomes.^{224,225} The effect estimates in the published meta-analysis are log hazard ratios, and the within-study correlations were estimated using bootstrapping methods for a bivariate meta-analysis.²²⁵ The within-study correlation estimates between the log hazard ratios for stroke and mortality are: 0.02, 0.20, 0.06, 0.11, 0.23, 0.29, 0.15, 0.05, 0.27, and 0.23. Although the effect estimates are log hazard ratios, which is not the same measure as the log odds ratio estimates in this analysis, the uniform(0.02,0.29) prior distribution is considered a useful and sensible prior distribution to use because the context is related in terms of the primary outcomes and the patient population having cardiovascular disease. The prior distribution also agrees with the vague uniform(0,1) prior distribution in the sense that there are only positive correlation values, but it is narrower in terms of the potential values for the within-study correlations.

Option (3): Alternative overall correlation method (Riley method)

The impact of incorrectly specifying the unknown within-study correlation will be larger if the trials are small, due to relatively large within-study variability, and this is the case in the phase II trials in AMI. An alternative method of dealing with unknown within-study correlations is to estimate the overall correlation, which is a combination of the within-study and between-study correlations.²¹⁵ The advantage of this method is it avoids the need to specify the unknown within-study correlations, or to estimate the between-study correlations. In the case of the phase II trials in AMI, this method is especially convenient because there

are no available estimates of the within-study correlations, and there is little data (<10 studies) to estimate the between-study correlation.

Model (5.2) provides this alternative specification of the bivariate random-effects model.²¹⁵ In this alternative model, the overall variation is still partitioned into within-study variance and between-study variance, like in model (5.1). However, the overall correlation is not partitioned into within-study and between-study correlation, but a single parameter, ρ_o , models the overall correlation directly. The within-study variances are represented by s_{ij}^2 and are specified and known. The additional variation beyond sampling error is represented by ψ_j^2 , which is not equivalent to τ_j^2 from model (5.1), although these two parameters will be similar the within-study variation is small.²¹⁵ This is because as the additional variation increases, the within-study variance becomes relatively small, and the covariance in Φ_i tends to $\rho_o\psi_1\psi_2$ and the covariance in \mathbf{D} tends to $\rho_B\tau_1\tau_2$. The covariances are then likely to be similar because the estimate of ρ_o will be based mainly on ρ_B . Thus, τ_j and ψ_j will also be similar when the within-study parameters are relatively small. On the other hand, when the within-study variance is relatively large, the alternative model and the general model are less likely to be similar. Riley et al. show, however, that the pooled estimates of μ_1 and μ_2 are unbiased and have better statistical properties than separate univariate meta-analyses.²¹⁵

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \Phi_i \right\} \quad (5.2)$$

$$\Phi_i = \begin{pmatrix} \psi_1^2 + s_{i1}^2 & \rho_o \sqrt{(\psi_1^2 + s_{i1}^2)(\psi_2^2 + s_{i2}^2)} \\ \rho_o \sqrt{(\psi_1^2 + s_{i1}^2)(\psi_2^2 + s_{i2}^2)} & \psi_2^2 + s_{i2}^2 \end{pmatrix}$$

$$\beta_j \sim N(0, 100); j=1, 2 \quad \psi_j \sim N(0, 1) / (0, \infty), \quad \rho_o \sim \text{Uniform}(-1, 1)$$

Vague normal prior distributions are assumed for the mean pooled treatment effects, β_j . A normal distribution, truncated at zero, is assumed for the additional standard deviation, ψ_j . A uniform(-1,1) distribution is specified for the overall correlation, ρ_o .

5.3.5 Probability statements

The Bayesian approach enables posterior probabilities of interest to be estimated, which may be useful in the decision to progress to a phase III trial, as discussed in Chapter 4. Following models (5.1) or (5.2), the probability that the OR is less than one for each outcome separately can be estimated. Also, in a multivariate meta-analysis model, it is possible to make *joint* probability statements, such as the probability that, on average, the treatment will reduce both the odds of mortality *and* stroke by at least 10%, i.e. the probability that the OR for mortality is less than 0.9 *and* the OR for stroke is less than 0.9. The ability to make joint probability estimates is a potential advantage of the multivariate approach compared to the univariate approach, which analyses the outcomes separately. The joint probability estimates in this chapter are therefore compared to the probability estimates for the outcomes separately to determine whether these joint probabilities further aid the clinical decision process.

5.3.6 Estimation and reporting of models

The summary estimates of interest in this chapter are the median of the marginal posterior distributions for the summary log ORs, the median of the marginal posterior distributions for the summary ORs, the posterior standard deviation of the log ORs, the 95% posterior credible intervals (CrI) for the summary ORs, the posterior probability that the individual ORs are less than one, the posterior probability that both ORs are less than one, and the median and CrI of the marginal posterior distribution of the between-study variances. Also of interest are the estimates of the between-study correlation for the general model (5.1), or an estimate of the overall correlation in the alternative model (5.2). Posterior estimates of parameters of interest are obtained using the Gibbs Sampler Markov chain Monte Carlo (MCMC) method,¹⁸⁹ which is implemented in WinBUGS version 1.4.3.³² Example code for random-effects bivariate meta-analysis is given in Appendix C1 and C2. In this chapter, the analyses were performed with 300,000 iterations after allowing for a 100,000 iteration burn in, and the samples were thinned by 10 to reduce any concerns of auto-correlation. The convergence of parameters was checked using history and trace plots.

5.3.7 Prediction intervals following a Bayesian bivariate random-effects meta-analysis

As shown in Chapter 4, when there is heterogeneity in treatment effects, the effect in a new trial, $\theta_{ij_{new}}$, may be very different to the average effect, β_j . Thus, to inform the decision to proceed to a phase III trial following the general meta-analysis model (5.1), the predictive distribution for $\theta_{i1_{new}}$ and $\theta_{i2_{new}}$ (the true intervention effects (log odds ratios) in a new trial) is calculated assuming model (5.3):

$$\begin{pmatrix} \theta_{i1_{new}} \\ \theta_{i2_{new}} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \mathbf{D} \right) \quad (5.3)$$

The 95% prediction interval for $\theta_{i1_{new}}$ and $\theta_{i2_{new}}$ can be obtained by taking the 2.5% and 97.5% values of the posterior samples drawn from model (5.3) during the MCMC process. These intervals will also account for the uncertainty in β_1 and β_2 , and the uncertainty in the between-study covariance matrix, \mathbf{D} .

It is not possible to derive the prediction intervals for the log odds ratios in a new trial from the alternative model (5.2). This is because the alternative model does not partition the correlation into within and between-study correlation, and therefore the between-study covariance matrix, \mathbf{D} , in model (5.3) cannot be specified.

The prediction distribution for $\theta_{i1_{new}}$ and $\theta_{i2_{new}}$ can also be used to calculate the probability that the intervention will be truly effective in a new trial for each outcome, either at all (i.e. probability(new OR<1)) or by some clinically relevant amount, such as the odds being reduced by at least 10% (i.e. probability(new OR<0.9)). Additionally, in the multivariate setting it is possible to use the prediction distributions to compute joint probability estimates, such as the probability that the intervention will be effective in a new trial for both outcomes (i.e. probability(OR for stroke in a new trial<1 and OR for mortality in a new trial<1)).

5.4 Application of the multivariate meta-analysis methods

The methods described in section 5.3 are now applied, which focus mainly on the bivariate meta-analysis of stroke and mortality. Later in section 5.4.2 a bivariate meta-analysis of reinfarction and mortality are briefly considered.

5.4.1 Primary analysis: mortality and stroke

Table 5.4 displays the summary results from the bivariate meta-analyses for stroke and mortality, respectively, for the models introduced earlier. The results from the Bayesian univariate random-effects model assuming normality (i.e. model (5.1) with $\rho_B=\rho_W=0$) are also shown in Table 5.4 for both outcomes. The key findings are now discussed.

Estimates of correlation

The estimate of overall correlation in model (5.2) is quite large since the median of the posterior distribution for ρ_O is 0.425. This indicates why the bivariate model is of interest, over and above two separate univariate meta-analyses that assume correlations are zero. The 95% CrI for the overall correlation is narrower (-0.207 to 0.809) compared to the CrI for the between-study correlation, such as -0.912 to 0.961 for model (5.1) with a uniform(0.02,0.29) prior distribution for the within-study correlations. Also, as the fixed value of the within-study correlation gets too large in relation to the overall correlation, the estimate of between-study correlation in model (5.1) adjusts for this with a negative estimate. For example, when the within-study correlation is 0.9 the between-study correlation is -0.018 (95% CrI -0.944 to 0.837).

Summary of pooled results

In general, when compared to the univariate analysis, the bivariate analyses reveal that there is a slightly larger average treatment effect for both outcomes, but this increase is only marginal. For example, for stroke, the pooled OR is 0.649 (95% CrI 0.219 to 1.747) from a univariate analysis whereas the pooled OR is 0.625 (95% CrI 0.254 to 1.416) from bivariate model (5.1) with a uniform(0.02,0.29) prior distribution for the within-study correlations (i.e. the prior based on external evidence).

Table 5.4: Univariate and bivariate random-effects meta-analysis results for bolus versus infusion therapy for stroke and mortality.

	Model	Assumption	Pooled log(OR)	SE(Pooled log(OR))	Pooled OR (95% CrI)	$\hat{\tau}$ (95% CrI)	Prob (Pooled OR<1)	$\hat{\rho}_B$ (95% CrI)
STROKE	Univariate	$\rho_B = \rho_W = 0$	-0.432	0.515	0.649 (0.219 to 1.747)	0.704 (0.048 to 1.818)	0.829	-
	Model (5.1)	$\rho_{W_i} = 0$	-0.468	0.443	0.627 (0.248 to 1.454)	0.659 (0.042 to 1.683)	0.877	0.291 (-0.894 to 0.967)
		$\rho_{W_i} = 0.2$	-0.470	0.430	0.625 (0.254 to 1.419)	0.607 (0.037 to 1.617)	0.884	0.201 (-0.919 to 0.960)
		$\rho_{W_i} = 0.5$	-0.464	0.422	0.629 (0.260 to 1.404)	0.595 (0.037 to 1.557)	0.884	0.082 (-0.936 to 0.938)
		$\rho_{W_i} = 0.9$	-0.379	0.415	0.684 (0.292 to 1.528)	0.613 (0.042 to 1.560)	0.839	-0.018 (-0.944 to 0.837)
		$\rho_{W_i} \sim \text{Unif}(0,1)$	-0.456	0.414	0.634 (0.266 to 1.383)	0.566 (0.033 to 1.530)	0.884	0.093 (-0.935 to 0.940)
		$\rho_{W_i} \sim \text{Unif}(0.02,0.29)$	-0.471	0.430	0.625 (0.254 to 1.416)	0.613 (0.036 to 1.625)	0.884	0.219 (-0.912 to 0.961)
	Model (5.2)	$\rho_O \sim \text{Unif}(-1,1)$	-0.437	0.450	0.646 (0.254 to 1.545)	0.700* (0.051 to 1.684)	0.854	0.425 [†] (-0.207 to 0.809)
MORTALITY	Univariate	$\rho_B = \rho_W = 0$	-0.038	0.276	0.962 (0.594 to 1.780)	0.405 (0.022 to 1.201)	0.565	-
	Model (5.1)	$\rho_{W_i} = 0$	-0.058	0.267	0.944 (0.585 to 1.704)	0.396 (0.021 to 1.172)	0.596	0.291 (-0.894 to 0.967)
		$\rho_{W_i} = 0.2$	-0.051	0.256	0.950 (0.599 to 1.665)	0.353 (0.018 to 1.106)	0.589	0.201 (-0.919 to 0.960)
		$\rho_{W_i} = 0.5$	-0.025	0.252	0.976 (0.619 to 1.690)	0.355 (0.019 to 1.072)	0.544	0.082 (-0.936 to 0.938)
		$\rho_{W_i} = 0.9$	0.117	0.286	1.124 (0.646 to 2.026)	0.588 (0.147 to 1.235)	0.325	-0.018 (-0.944 to 0.837)
		$\rho_{W_i} \sim \text{Unif}(0,1)$	-0.032	0.252	0.969 (0.617 to 1.686)	0.359 (0.019 to 1.078)	0.555	0.093 (-0.935 to 0.940)
		$\rho_{W_i} \sim \text{Unif}(0.02,0.29)$	-0.054	0.258	0.948 (0.598 to 1.677)	0.359 (0.018 to 1.118)	0.592	0.219 (-0.912 to 0.961)
	Model (5.2)	$\rho_O \sim \text{Unif}(-1,1)$	-0.004	0.283	0.996 (0.613 to 1.904)	0.442* (0.027 to 1.207)	0.507	0.425 [†] (-0.207 to 0.809)

ρ_{W_i} is the within-study correlation in model (5.1); τ is the between-study standard deviation; * this is $\hat{\psi}$ - the additional variation in the alternative

bivariate model (5.2); ρ_B is the between-study correlation in model (5.1); [†] this is ρ_O - the overall correlation in model (5.2).

Change in standard error of pooled summary effects

In comparison to the univariate analysis for stroke, the standard errors of the pooled log OR are generally smaller for the bivariate meta-analyses since there is missing data for one trial for this outcome, and thus strength can be borrowed from the complete data for mortality. For example, for stroke the univariate standard error is 0.515, and the standard error is 0.414 for bivariate model (5.1) with a uniform(0,1) prior distribution for ρ_{wi} .

Between-study standard deviation

The estimates of between-study standard deviation tend to be smaller, with narrower credible intervals, for model (5.1) for both outcomes, compared to the univariate results. For example, for stroke, $\hat{\tau}$ is 0.704 (95% CrI 0.048 to 1.818) for the univariate model, whereas $\hat{\tau}$ is 0.613 (95% CrI 0.036 to 1.625) for the bivariate model (assuming a uniform(0.02,0.29) prior distribution for the within-study correlations).

The difference in parameter uncertainty and, in particular, estimates of between-study variance has an effect on the 95% prediction intervals and probability estimates, which are shown in Table 5.5 for stroke and mortality. The 95% prediction intervals for the true odds ratio in a new study are narrower for bivariate model (5.1) compared to the univariate analyses. For example, for stroke, the univariate 95% prediction interval is 0.070 to 5.367, whereas the 95% prediction interval is 0.089 to 3.843 for the bivariate analysis that assumes a uniform(0.02,0.29) prior distribution for the within-study correlations. The probability that the true OR is less than one in a new study is larger for the bivariate analyses compared to the univariate analyses. For example, for stroke the probability estimate is 0.762 for the bivariate analysis assuming a uniform(0.02,0.29) prior distribution for the within-study correlations compared to an estimate of 0.721 in the univariate analysis.

However, in model (5.2) the estimate of additional variation, $\hat{\psi}$, is larger compared to the estimate of τ in the univariate analyses for mortality. For mortality, $\hat{\psi}$ is 0.442 (95% CrI 0.027 to 1.207) whereas $\hat{\tau}$ is 0.405 (95% CrI 0.022 to 1.201) for the univariate analysis.

Larger probability estimates

Pooled ORs

The probability that the pooled OR is less than one is larger for the bivariate model (5.1) when using the uniform(0.02,0.29) prior for both outcomes when compared to the univariate analyses. For example, in Table 5.4 the probability that the pooled OR is less than one for stroke is 0.829 for the univariate model, compared to 0.884 for the bivariate model assuming a uniform(0.02,0.29) prior distribution for the within-study correlations.

The alternative Riley model (5.2) obtains markedly different probability estimates about the pooled effect than model (5.1) using the uniform(0.02,0.29) prior distribution, potentially due to the larger variances estimated and thus wider credible intervals from this method. For mortality, the probability that the pooled OR is less than one is lower for the alternative model with an estimate of 0.507, compared to an estimate of 0.592 in model (5.1) (Table 5.4). This is because the credible interval for the pooled OR is wider from the Riley model, which is due to a larger standard error for the pooled log OR and a larger estimate of additional variation, $\hat{\psi}$.

Table 5.5: 95% prediction intervals and posterior probability estimates for stroke and mortality from model (5.1).

	Assumption	95% prediction interval for true OR in a new trial	Prob(new trial true OR<1)
Stroke	$\rho_B=\rho_{W_i}=0$	0.070 to 5.367	0.721
	$\rho_{W_i}=0$	0.082 to 4.281	0.751
	$\rho_{W_i}=0.2$	0.091 to 3.867	0.762
	$\rho_{W_i}=0.5$	0.097 to 3.679	0.762
	$\rho_{W_i}=0.9$	0.108 to 4.131	0.720
	$\rho_{W_i}\sim\text{Uniform}(0,1)$	0.102 to 3.484	0.884
	$\rho_{W_i}\sim\text{Uniform}(0.02,0.29)$	0.089 to 3.843	0.762
Mortality	$\rho_B=\rho_{W_i}=0$	0.290 to 3.889	0.554
	$\rho_{W_i}=0$	0.284 to 3.654	0.574
	$\rho_{W_i}=0.2$	0.308 to 3.362	0.569
	$\rho_{W_i}=0.5$	0.325 to 3.352	0.541
	$\rho_{W_i}=0.9$	0.256 to 5.163	0.421
	$\rho_{W_i}\sim\text{Uniform}(0,1)$	0.326 to 3.386	0.549
	$\rho_{W_i}\sim\text{Uniform}(0.02,0.29)$	0.305 to 3.382	0.572

ORs in a new trial

The probability estimates for the predicted true ORs in a new trial being less than one, from the univariate model and bivariate model (5.1), are shown in Table 5.5. All probability estimates are larger from the bivariate model compared to the univariate model for both outcomes, other than when the within-study correlation is fixed at 0.9. For example, for stroke, the predicted probability that the true treatment effect in a new trial will be less than one is 0.721 from the univariate analysis, whereas this estimate is 0.762 assuming a $\text{uniform}(0.02,0.29)$ prior distribution for ρ_{W_i} in bivariate model (5.1).

If the 95% prediction interval for the true OR in a new trial is narrower with an upper bound closer to one, the probability that the true OR will be less than one is larger. For example, for stroke when $\rho_W=0.2$, the 95% prediction interval is 0.091 to 3.867 and the probability that the true OR will be less than one is 0.762. However, when $\rho_W \sim \text{uniform}(0,1)$, the 95% prediction interval is narrower (0.102 to 3.484) and the probability estimate is 0.884.

Joint probabilistic inferences

Additionally, the *joint* probability estimates for the pooled and predicted ORs are shown in Table 5.6 for the univariate analyses and the bivariate analyses using model (5.1). There is positive correlation for stroke and mortality and, as a consequence, the joint probability estimates are greater than the product of the individual probability estimates for each outcome. For example, for the $\text{uniform}(0.02,0.29)$ prior distribution for the within-study correlations, the probability that both pooled ORs are less than one is 0.537. However, the product of the individual probabilities is 0.468 (the product of 0.829 and 0.565). Thus, there appears a greater prospect of bolus therapy improving both outcomes, once the correlation has been properly accounted for.

Table 5.6: Joint posterior probability estimates for stroke and mortality.

Model	Assumption	Prob(both summary ORs<1)	Prob(both new trial true ORs<1)	Between-study correlation (95% CrI)
	$\rho_B = \rho_{W_i} = 0$	0.468	0.399	-
	$\rho_{W_i} = 0$	0.533	0.458	0.291 (-0.894 to 0.967)
	$\rho_{W_i} = 0.2$	0.535	0.457	0.201 (-0.919 to 0.960)
Model (5.1)	$\rho_{W_i} = 0.5$	0.507	0.434	0.082 (-0.936 to 0.938)
	$\rho_{W_i} = 0.9$	0.307	0.319	-0.018 (-0.944 to 0.837)
	$\rho_{W_i} \sim \text{Unif}(0,1)$	0.519	0.446	0.093 (-0.935 to 0.940)
	$\rho_{W_i} \sim \text{Unif}(0.02,0.29)$	0.537	0.460	0.219 (-0.912 to 0.961)
Model (5.2)	$\rho_o \sim \text{Unif}(-1,1)$	0.469	-	0.425 (-0.207 to 0.809)*

* an estimate of the overall correlation in the alternative model

5.4.2 Secondary analysis of reinfarction and mortality

The pooled results for the secondary analysis of reinfarction and mortality are shown in Table 5.7. The 95% prediction intervals and the joint posterior probabilities are shown in Table C1 and Table C2, respectively, in Appendix C3.

Impact of correlation on parameter estimates

In this example, the overall correlation estimate, $\widehat{\rho}_o$, in the alternative model (5.2) is only 0.059 (95% CrI -0.501 to 0.575) and thus there is little extra information here. Therefore, there is little impact on the parameters estimates compared to the primary analysis of stroke and mortality, where there was stronger correlation between the outcomes. The bivariate meta-analysis results are thus similar to those seen in the separate univariate meta-analyses. The point estimates for the pooled ORs are in favour of the bolus therapy compared to infusion therapy for both outcomes (i.e. ORs are less than one), although there is still large uncertainty in the estimates. For example, for mortality the univariate pooled OR

is 0.962 (95% CrI 0.594 to 1.780) and in the alternative bivariate meta-analysis model (5.2) this estimate is 0.947 (95% CrI 0.577 to 1.767).

Table 5.7: Univariate and bivariate random-effects meta-analysis results for bolus versus infusion therapy for reinfarction and mortality.

	Model	Assumption	Pooled log(OR)	SE(Pooled log(OR))	Pooled OR (95% CrI)	$\hat{\tau}$ (95% CrI)	Prob (Pooled OR<1)	$\hat{\rho}_B$ (95% CrI)
REINFARCTION	Univariate	$\rho_B=\rho_{W_i}=0$	-0.284	0.213	0.753 (0.491 to 1.139)	0.257 (0.013 to 0.871)	0.922	-
		$\rho_{W_i}=0$	-0.288	0.214	0.750 (0.491 to 1.144)	0.251 (0.012 to 0.883)	0.920	0.046 (-0.941 to 0.952)
		$\rho_{W_i}=0.2$	-0.297	0.214	0.743 (0.486 to 1.133)	0.258 (0.012 to 0.891)	0.926	-0.025 (-0.949 to 0.941)
	Model (5.1)	$\rho_{W_i}=0.5$	-0.316	0.226	0.729 (0.464 to 1.135)	0.318 (0.017 to 0.973)	0.930	-0.195 (-0.962 to 0.893)
		$\rho_{W_i}=0.9$	-0.335	0.256	0.715 (0.424 to 1.180)	0.471 (0.042 to 1.157)	0.919	-0.453 (-0.976 to 0.653)
		$\rho_{W_i}\sim\text{Unif}(0,1)$	-0.286	0.223	0.751 (0.477 to 1.156)	0.299 (0.015 to 0.948)	0.914	-0.157 (-0.961 to 0.917)
	Model (5.2)	$\rho_o\sim\text{Unif}(-1,1)$	-0.294	0.220	0.745 (0.481 to 1.150)	0.279* (0.014 to 0.934)	0.919	0.059 [†] (-0.501 to 0.575)
MORTALITY	Univariate	$\rho_B=\rho_{W_i}=0$	-0.038	0.276	0.962 (0.594 to 1.780)	0.405 (0.022 to 1.201)	0.565	-
		$\rho_{W_i}=0$	-0.062	0.265	0.940 (0.585 to 1.689)	0.382 (0.020 to 1.164)	0.604	0.046 (-0.941 to 0.952)
		$\rho_{W_i}=0.2$	-0.065	0.265	0.937 (0.583 to 1.684)	0.381 (0.020 to 1.171)	0.608	-0.025 (-0.949 to 0.941)
	Model (5.1)	$\rho_{W_i}=0.5$	-0.057	0.283	0.945 (0.571 to 1.771)	0.462 (0.028 to 1.256)	0.590	-0.195 (-0.962 to 0.893)
		$\rho_{W_i}=0.9$	-0.027	0.325	0.973 (0.538 to 1.975)	0.648 (0.080 to 1.474)	0.537	-0.453 (-0.976 to 0.653)
		$\rho_{W_i}\sim\text{Unif}(0,1)$	-0.080	0.273	0.923 (0.571 to 1.704)	0.402 (0.020 to 1.230)	0.630	-0.157 (-0.961 to 0.917)
	Model (5.2)	$\rho_o\sim\text{Unif}(-1,1)$	-0.055	0.279	0.947 (0.577 to 1.767)	0.431* (0.024 to 1.232)	0.587	0.059 [†] (-0.501 to 0.575)

ρ_{W_i} is the within-study correlation in model (5.1); τ is the between-study standard deviation; * this is $\hat{\varphi}$ - the additional variation in the alternative

bivariate model (5.2); ρ_B is the between-study correlation in model (5.1); [†] this is ρ_o - the overall correlation in model (5.2).

5.5 Discussion

This chapter explored the use of multivariate meta-analysis in small phase II trials with multiple outcomes, which extends the univariate meta-analysis of phase II trials in Chapter 4. The key findings of the work are summarized in Figure 5.2, and it has raised a number of interesting questions, which are now discussed.

Figure 5.2: Key findings.

- The multivariate approach for meta-analysis of phase II trials can alter inferences for each outcome on their own by borrowing strength from the other outcome; this is especially apparent given missing outcomes in some trials.
- The multivariate approach additionally allows us to estimate joint probabilities. When there is positive correlation between outcomes the joint probability estimates are larger compared to the product of the probabilities from the univariate analyses.
- The value of the within-study correlations can have a large impact on multivariate model results, including pooled estimates, credible intervals, and posterior joint probabilities, especially when the within-study variances are large relative to the between-study variances.
- Within-study correlations will often be unavailable, and so clinical guidance or external information (e.g. from a different but related setting) is needed to help specify a prior distribution for the missing ρ_{wi} .
- The alternative model (5.2) is useful to estimate pooled effects when within-study correlations are unknown since it does not require within-study correlations to be specified; however the model does not allow predictions for true effects in new trials and is thus not especially useful for phase II meta-analysis.
- The 95% credible interval from the posterior distribution of the between-study correlation will usually be very wide due to the small number of studies in a phase II meta-analysis.

5.5.1 What are the additional benefits of multivariate meta-analysis of phase II trials in comparison to univariate methods?

Multivariate meta-analysis methods have been promoted when there are multiple outcomes since the multivariate methods are statistically more efficient by accounting for correlation between the outcomes.²¹³ Bayesian methods have also been promoted for multivariate meta-analysis^{17,18} as existing information can be incorporated using prior distributions. The work in this chapter shows how Bayesian multivariate meta-analysis methods can be used in the setting of phase II trials, and identified the pros and cons.

Firstly, the chapter showed that unknown within-study correlation estimates are an issue when no IPD are available. Multiple methods were considered to account for the unknown within-study correlations, such as using fixed values and a number of prior distributions, and an alternative model where the correlation is not partitioned into within-study and between-study correlations. The pooled results were compared to assess the impact of these various choices.

The primary analysis results show that the estimates and standard errors of the pooled log ORs, the between-study variance estimates and predicted probabilities can differ in the multivariate meta-analysis approach compared to the univariate approach. For example, for stroke, the univariate standard error of the pooled log OR is 0.515. For the bivariate model (5.1) the standard error is 0.414. For the same outcome in the bivariate analysis, assuming a uniform(0.02, 0.29) prior distribution for the within-study correlations, the estimate of τ is 0.613 (95% CrI 0.036 to 1.625). This estimate is lower compared to univariate analysis where $\hat{\tau}$ is 0.704 (95% CrI 0.048 to 1.818).

The greatest advantage of the multivariate Bayesian approach for phase II trials is that it enables joint probability estimates, such as the probability that both pooled ORs are less than the value of no difference, or lower than some clinically relevant difference. For example, the joint probability estimate that both pooled ORs for stroke and mortality are less than one is 0.537, assuming a uniform(0.02,0.29) prior distribution for the within-study correlations. This probability is considerably higher than the joint probability from the univariate analysis, which is 0.468. This is because the bivariate model borrows strength by modelling the correlation between the outcomes, whereas the joint probability estimate from the univariate analysis is misleading because it does not account for the correlation between the outcomes. These joint probabilities from the multivariate analyses could be additionally useful in terms of clinical decision making.

5.5.2 Is the multivariate approach appropriate in the setting of phase II trials?

If there are multiple phase II trials with several correlated outcomes then multivariate meta-analysis should be considered to inform phase III trial decisions, rather than using separate univariate meta-analyses, because the multivariate approach utilises the correlation and thus more information. Clearly, those making decisions about phase III trials want to use all the information available, and so it seems inappropriate to ignore the correlation and only perform univariate analyses, thereby disregarding useful information. Therefore, in order to use a multivariate approach it would be advantageous to have the IPD available to enable multinomial modelling or, if the general multivariate normal approach to bivariate analysis is preferred (e.g. for non-binary outcomes), the calculation of within-study correlations in each trial.²²⁵

However, without IPD it is difficult to specify the within-study correlations. In this situation, bivariate model (5.1) is recommended because it can be implemented with prior distributions specified for the within-study correlations, which allow for uncertainty, instead of simply using fixed values. Previous evidence could be used to inform this prior distribution, such as the uniform(0.02,0.29) prior distribution that was explored in the example in acute myocardial infarction in this chapter. Or, there may be clinical guidance that could restrict the within-study correlation to a range of values, such as a uniform(0,1) prior that was also used in this chapter because stroke and death are likely to be positively correlated.

Bivariate model (5.1) is also more appropriate than the alternative model (5.2) because model (5.1) can derive prediction intervals for the true treatment effects in a new trial, whereas the alternative model cannot do this due to an inappropriate hierarchical structure.

5.5.3 Limitations and further work

This chapter has illustrated the use of multivariate meta-analysis, but there are limitations to the work. In particular, the findings are specific to outcomes from one dataset. Further evaluation in other phase II datasets are needed. An advantage of the multivariate approach compared to the univariate method is when there is missing data since information can be borrowed from the other outcomes.²¹³ However, in the example in AMI, there is missing data for stroke in just one trial and there is complete data for reinfarction and mortality. Therefore, the benefit in the multivariate approach for phase II trials with missing data was not fully explored, although this could be done in further research by artificially removing some of the data.

This chapter only explored the use of the methods for the multivariate meta-analysis of two correlated outcomes, although there may be more than two correlated outcomes in a different setting. There may be additional difficulties when synthesising more than two

outcomes. For more than two outcomes there would need to be further specification of prior distributions for within-study correlations and there would be multiple between-study correlations (rather than just one in bivariate meta-analysis). Additionally, the model specification becomes more difficult in terms of separating the parameters in the variance-covariance matrices.

A uniform(-1,1) prior distribution was used for the between-study correlation for all model (5.1) analyses in this chapter. All prior distributions must be chosen with caution^{21,29} but the impact of this uniform(-1,1) prior distribution on the estimates of the other parameters was not evaluated. This will be explored formally in the next chapter through a simulation study.

If there was IPD it would be possible to estimate the pooled treatment effects with a multinomial regression model.²¹⁷ A logistic regression model was used in Chapter 4 as well as models that assumed normality of treatment effect estimates. In Chapter 4, the logistic regression model was considered to be more appropriate due to the modelling of the binomial distribution of the data and the low event rates. However, this chapter could not explore the potential benefits of the extension to this model.

5.6 Conclusion and recommendations

Multivariate meta-analysis of phase II trials can be used when there are multiple outcomes instead of the more commonly used univariate meta-analysis approach for each outcome separately. The example in this chapter showed that the utilisation of correlation in a Bayesian framework has the potential to produce added information from a meta-analysis of phase II trials. In particular, joint probability estimates can be derived, which could further inform phase III trial decisions, and information about missing outcomes can be gathered using the known information from other correlated outcomes that are available, which may change inferences. However, there are several challenges involved in using multivariate

meta-analysis, such as lack of IPD forcing a general multivariate normal framework rather than multinomial modelling. This is only an approximation and is also challenging in terms of specifying unknown within-study correlations. The issue of unknown within-study correlations is further complicated in the phase II trial setting because the effect of incorrectly specifying the within-study correlations is worsened when the within-study variance is relatively large, and this tends to be the case in phase II trials due to the small sample sizes. The bivariate model (5.1) is perhaps more appropriate than the alternative bivariate model (5.2), because this allows joint predictions for true effects in new trials and allows for uncertainty in unknown within-study correlations by specifying prior distributions, which can be informed by previous evidence. Previous work has also advocated the use of prior distributions for unknown within-study correlations, such as a multivariate meta-analysis of mixed outcomes by Bujkiewicz et al., which used IPD from external studies to derive informative prior distributions.¹⁷ Nam et al. conducted a sensitivity analysis for the prior distributions for the within-study and between-study correlations in a multivariate meta-analysis due to the lack of external evidence to derive informative prior distributions.²²⁶ In a slightly different context, Abrams et al. also considered the use of plausible prior distributions for missing within-subject correlation between baseline and follow-up measurements in a meta-analysis of randomised controlled trials.²²⁷

In conclusion, the recommendations for future meta-analysis of phase II trials with multiple correlated outcomes are given in Figure 5.3.

Figure 5.3: Key recommendations.

- Where possible, multivariate meta-analysis of phase II trials should be considered to inform phase III decisions as it utilises more of the evidence compared to univariate meta-analysis.
- Researchers should provide IPD to facilitate multivariate meta-analysis approaches, such as multinomial modelling or to derive within-study correlations for a multivariate normal approach.
- When IPD is not available, a Bayesian approach with prior distributions for the missing within-study correlations is the preferred approach, for example, based on external evidence or clinical rationale.
- Further research is needed to assess the impact of the prior distributions for between-study variance-covariance parameters in multivariate meta-analysis, especially for the between-study correlation.

CHAPTER 6: A SIMULATION STUDY TO ASSESS THE CHOICE OF PRIOR DISTRIBUTION FOR THE BETWEEN-STUDY CORRELATION IN A BAYESIAN BIVARIATE META-ANALYSIS

6.1 Introduction and aims

Chapters 4 and 5 have illustrated several advantages and methodological issues of a Bayesian framework for meta-analysis of small trials. The Bayesian approach is a natural way to account for all parameter uncertainty, to make predictions regarding the possible treatment effect estimate in a new trial, and to derive joint probability estimates in multivariate meta-analysis. Chapter 4 explored the impact of the prior distribution for the between-study variance in a univariate Bayesian meta-analysis and agreed with those from a previous simulation study by Lambert et al.²⁹: the pooled treatment effect estimates can be particularly sensitive to the choice of this prior distribution. In Chapter 5, the Bayesian framework was extended to a bivariate meta-analysis with two correlated outcomes. However, an unanswered question remains: what is the impact of the choice of prior distribution for the between-study correlation parameter? This question motivates the research in this chapter.

In a Bayesian multivariate meta-analysis, a conjugate prior for the entire between-study covariance matrix is the inverse-Wishart prior distribution. For example, Riley et al.²²⁸ and Zwinderman and Bossuyt²²⁹ use an inverse-Wishart prior for the between-study covariance matrix in a bivariate meta-analysis of test accuracy studies. However, Wei and Higgins suggest it is preferable to place separate priors on each component of the between-study covariance matrix,^{18,230,231} because the Wishart prior can be very influential toward the

posterior estimates of the between-study variances; the Wishart prior is a generalisation of the gamma prior for some parameters, which Lambert et al.²⁹ shows is influential in univariate meta-analysis when the true between-study variances are close to zero. Separation of the between-study covariance matrix also allows more flexibility in the choice of priors for each component, and is thus the approach considered here. Thus in a bivariate meta-analysis, prior distributions will be required for the between-study variances and the between-study correlation.

There are many different prior distributions that could be chosen for the between-study correlation, depending on whether external evidence is available or not. In this chapter, the focus is mainly on situations where there is no prior information and so vague prior distributions are desired. Often, in this situation, researchers use a uniform(-1,1) prior distribution for the between-study correlation. This gives equal weight to all possible values of the between-study correlation, and thus may be perceived to be suitable when there is no previous evidence about the correlation's magnitude. However, as discussed, meta-analysis results are known to be sensitive to seemingly vague priors for the between-study variance²⁹ – thus it seems plausible that results also be sensitive to this seemingly vague uniform(-1,1) prior for the correlation. Despite this, many researchers do not conduct sensitivity analyses to check whether their prior distribution for the between-study correlation is appropriate. For example, Nam et al.²²⁶ use a uniform(-1,1) prior distribution for the between-study correlation in a Bayesian bivariate meta-analysis that investigated the impact of passive smoking on two correlated outcomes of asthma and lower respiratory disease; no sensitivity analysis was undertaken for the choice of this prior. Another previous study¹⁸ of the choice of prior distribution for the between-study covariance matrix in a Bayesian multivariate meta-analysis explored various separation strategies that allow prior distributions to be placed on different components of the matrix, including the between-study correlation. They also selected a uniform(-1,1) prior distribution for the between-study correlation but did not explore other

choices of prior for this parameter.¹⁸ Riley et al.²¹³ investigated the influence of ignoring within-study correlation in a Bayesian meta-analysis with correlated outcomes and also focussed only on a uniform(-1,1) prior distribution for the between-study correlation.

Thus, it seems that the choice of prior distribution for the between-study correlation has been somewhat neglected in the multivariate meta-analysis literature. Therefore, in this chapter an extensive simulation study will be performed to assess the choice of a range of vague prior distributions for the between-study correlation in a bivariate random-effects meta-analysis model. In particular, this will explore the impact of the prior distribution on the posterior pooled treatment effects, both for complete and missing outcome data, and on the posterior estimates of the between-study correlation itself. It will also examine if the priors for the between-study variance impact on the posterior between-study correlation estimates (and vice-versa), since variance and correlation parameters are not independent.

The remainder of the chapter is structured as follows. Section 6.2 introduces the bivariate random-effects meta-analysis model and the prior distributions for the between-study correlation that will be used throughout this chapter. Section 6.3 details two motivating example meta-analysis datasets and results for the various prior distributions for the between-study correlation. The findings from the motivating examples are used to inform the selection of scenarios in section 6.4, which is a simulation study that is used to determine whether the results can be generalised. This chapter concludes with discussion of the findings.

6.2 Bivariate random-effects meta-analysis model

As mentioned, the focus here is on the prior distributions in a bivariate normal random-effects meta-analysis model. The extension to more than two outcomes will be discussed briefly in the Discussion. To ensure clarity, the bivariate model (previously given as model (5.1)) is

given again in (6.1) below. For two outcomes, the treatment effect estimates, Y_{i1} and Y_{i2} , in each trial, i , are assumed to be normally distributed.

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N\left\{\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \mathbf{S}_i\right\} \quad (6.1)$$

$$\mathbf{S}_i = \begin{pmatrix} s_{i1}^2 & \rho_{W_i} s_{i1} s_{i2} \\ \rho_{W_i} s_{i1} s_{i2} & s_{i2}^2 \end{pmatrix}$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N\left\{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \mathbf{D}\right\}$$

$$\mathbf{D} = \begin{pmatrix} \tau_1^2 & \rho_B \tau_1 \tau_2 \\ \rho_B \tau_1 \tau_2 & \tau_2^2 \end{pmatrix}$$

The pooled treatment effects for outcomes, $j=1,2$, are denoted by β_1 and β_2 , which are the average effects of treatment across all studies. The within-study covariance matrix, \mathbf{S}_i , contains within-study variances for each trial, i , for outcome $j=1,2$, s_{i1}^2 and s_{i2}^2 (assumed known and estimated by the trial data), and within-study covariances, $\rho_{W_i} s_{i1} s_{i2}$, where ρ_{W_i} represents the within-study correlation (also assumed known). The between-study covariance matrix, \mathbf{D} contains the unknown between-study standard deviations, τ_j , and between-study correlation, which is represented by ρ_B . A vague normal prior distribution, centred on zero with variance 1000, is used throughout this chapter for the pooled treatment effects, β_j .

6.2.1 Prior distributions for between-study correlation

In this chapter, a range of prior distributions are considered to account for varying levels of hypothetical prior knowledge that may inform the prior distribution for the between-study

correlation. Below are five possible “vague” priors in which priors 1 to 3 allow the between-study correlation to be positive or negative, and priors 4 and 5 only allow the between-study correlation to be positive. The five prior distributions are shown in Figure 6.1.

Prior 1

$$\rho_B \sim \text{Uniform}(-1,1)$$

This prior distribution gives equal weight to all possible positive and negative values of correlation. This distribution is often considered appropriate when there is no prior information regarding the true value of the between-study correlation. As mentioned, it is commonly used in Bayesian multivariate meta-analysis applications, such as Nam et al.²²⁶ This distribution has a mean of zero and 95% CrI -0.95 to 0.95.

Prior 2

$$z = \frac{1}{2} \log \left(\frac{1 + \rho_B}{1 - \rho_B} \right) \sim N(0, \text{sd}=0.4)$$

This prior distribution is referred to as a Fisher prior¹⁷ and it is similar to prior 1 with the same mean and allows both positive and negative values. However, there is more weight around the mean and less weight at the extremes, and it has a 95% prior CrI for ρ_B of -0.655 to 0.655. This prior distribution could be applied if there is external evidence suggesting that the correlation is unlikely to be either very highly negative or positive, so values between -0.6 to +0.6 are most plausible (but the direction of the correlation is not known). It could also be useful if the prior belief was that the between-study correlation is zero. Riley et al.²¹⁸ show that the between-study correlations are often poorly estimated at +1 or -1 when there is little

data to estimate them, and so this prior distribution would potentially restrain this issue by giving very little weight to values close to +1 or -1.

Prior 3

$$\frac{(\rho_B+1)}{2} \sim \text{Beta}(1.5,1.5)$$

Similar to priors 1 and 2, the Beta prior distribution 3 also allows for positive and negative values of the between-study correlation. It is similar to prior 1 in that it is relatively flat across the range of values, with the exception that values at the extreme ends of the distribution are considered extremely unlikely. This prior distribution could be considered practical when there is very little external information to inform the true value of the between-study correlation; however, it may be clinically sensible to suggest that values close to -1 and +1 are implausible. The scale and shape parameter values of 1.5 are chosen here to ensure a prior that is noticeably different to both priors 1 and 2. It gives a 95% prior CrI for ρ_B of -0.878 to 0.878, which is wider than prior 2 but narrower than prior 1.

Prior 4

$$\rho_B \sim \text{Uniform}(0,1)$$

This prior distribution gives equal weight to all possible *positive* values of correlation. This distribution could be applied if there is external evidence to suggest a positive correlation but there is very little evidence to suggest the true positive value. An example is treatment effects for nested outcomes, such as disease free survival and overall survival, which by definition should be positively correlated but the magnitude is often unknown. This prior distribution is centred on 0.5 with 95% CrI 0.025 to 0.975.

Prior 5

$$\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$$

Similar to prior 4, this logit prior distribution allows only positive values, however, more weight is given around the mean and less weight is given in the tails of the distribution (Figure 6.1). Therefore, this prior is also centred on 0.5 but the 95% CrI is 0.173 to 0.827, which is narrower than for prior 4. This prior could be selected if there is prior evidence that the true value of the correlation is likely to be moderately high.

Many other choices of prior distributions could, of course, be specified, but these five are chosen as they reflect a sufficient range of different and realistic prior distributions for evaluation here onwards.

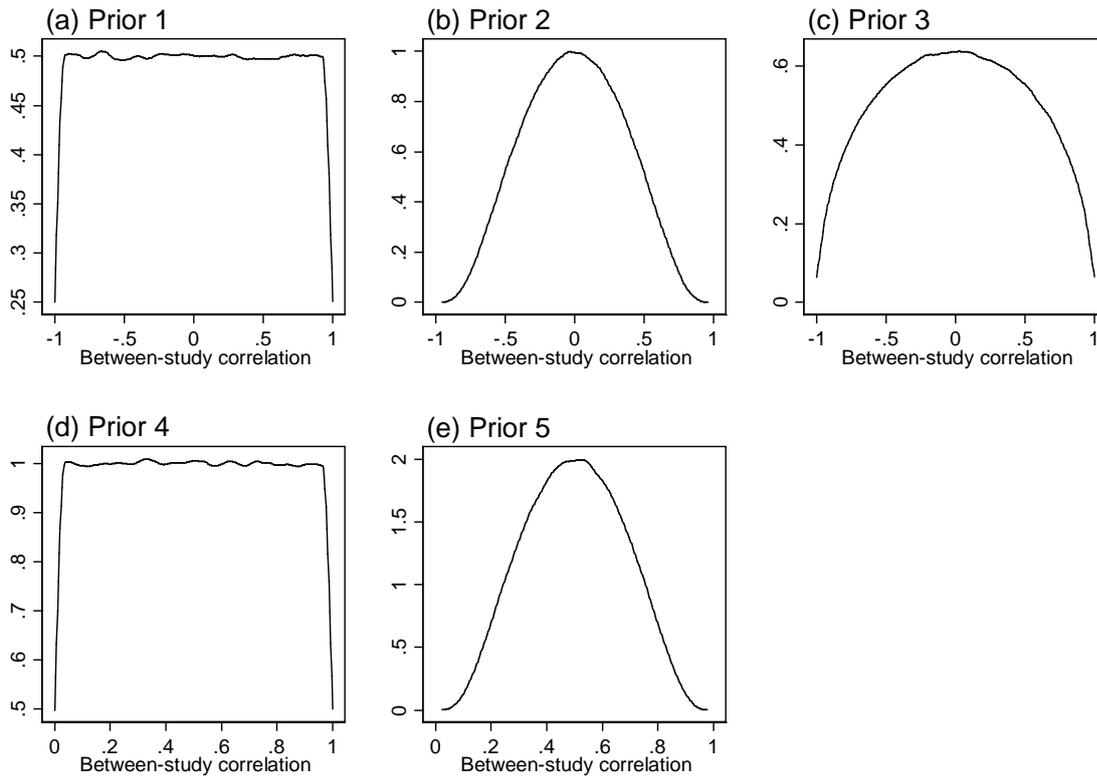


Figure 6.1: Density plots for prior distributions for between-study correlation:

(a) $\rho_B \sim \text{Uniform}(-1,1)$ (Prior 1); (b) $\frac{1}{2} \log \left(\frac{1+\rho_B}{1-\rho_B} \right) \sim \text{N}(0, \text{sd}=0.4)$ (Prior 2);

(c) $\frac{(\rho_B+1)}{2} \sim \text{Beta}(1.5,1.5)$ (Prior 3); (d) $\rho_B \sim \text{Uniform}(0,1)$ (Prior 4);

(e) $\text{logit}(\rho_B) \sim \text{N}(0, \text{sd}=0.8)$ (Prior 5).

6.3 Motivating examples

Before the simulation study is conducted it is useful to obtain some preliminary findings from real-life examples. Therefore, this section will explore the prior distributions described above in two motivating example datasets from previous bivariate meta-analyses. The results of these analyses will give an initial idea of the impact of the choice of prior distribution on the pooled results and inform how the simulation study should proceed. Since this is a Bayesian

analysis, prior distributions must also be specified for the between-study variances. Previous simulation studies revealed that the choice of prior distribution for the between-study variance is important and must be selected with caution.²⁹ In these motivating examples, the amount of between-study variance is larger in example A compared to example B, therefore it is possible to explore the impact of the choice of prior distribution for the between-study correlation on this parameter itself, and also how the prior distributions for the between-study variances affect the estimate of the between-study correlation and the pooled results.

6.3.1 Motivating example A

The dataset for the first motivating example is from an individual participant data (IPD) meta-analysis of hypertension trials by Wang et al.²²⁴, which investigated whether hypertension treatments reduce systolic blood pressure (SBP) and diastolic blood pressure (DBP) more than control. The summary results for the 10 trials are shown in Table 6.1 as taken from Riley et al.²²⁵ The within-study correlations were obtained via non-parametric bootstrapping.

In a previous frequentist analysis of this data, using restricted maximum likelihood, the estimate of between-study correlation was 0.78, and estimates of the between-study standard deviations were 2.71 and 1.48 for the effect of hypertension treatment on SBP and DBP, respectively.²²⁵

Table 6.1: Summary results for the 10 trials in the meta-analysis by Wang et al.²²⁴ for systolic blood pressure and diastolic blood pressure.

Trial name*	Control	Treatment	SBP	DBP	Within-study correlations SBP & DBP (from bootstrap)
			Mean difference (var)	Mean difference (var)	
ATMH	750	780	-6.66 (0.72)	-2.99 (0.27)	0.79
HEP	199	150	-14.17 (4.73)	-7.87 (1.44)	0.50
EWPHE	82	90	-12.88 (10.31)	-6.01 (1.77)	0.59
HDFP	2371	2427	-8.71 (0.30)	-5.11 (0.10)	0.77
MRC-1	3445	3546	-8.70 (0.14)	-4.64 (0.05)	0.64
MRC-2	1337	1314	-10.60 (0.58)	-5.56 (0.18)	0.50
SHEP	2371	2365	-11.36 (0.30)	-3.98 (0.075)	0.48
STOP	131	137	-17.93 (5.82)	-6.54 (1.31)	0.59
Sy-Chi	1139	1252	-6.55 (0.41)	-2.08 (0.11)	0.45
Sy-Eur	2297	2398	-10.26 (0.20)	-3.49 (0.04)	0.48

* Trial names are consistent with Wang et al.,²²⁴ where further details and trial publications can be found.

Prior distribution for between-study variance

The impact of the prior distribution for the between-study correlation must be viewed together with the choice of prior for the between-study variances, τ_j^2 . A previous simulation study of the impact of the prior distribution for between-study variance in univariate meta-analysis,²⁹ and the sensitivity analysis conducted in Chapter 4, both highlighted that the pooled results can be particularly sensitive to the choice of prior distribution for the between-study variance. This may be even more of a concern in bivariate meta-analysis because priors for the between-study variance might also affect the posterior estimate of the between-study correlation, which in turn will affect the pooled treatment effect estimates (especially if there is missing data for either outcome²³²) and their credibility intervals. Therefore, several prior distributions for the between-study variances are also investigated in these motivating examples, and also later in the simulation studies as follows:

Variance prior 1a

$$\tau_j \sim N(0,1)I(0,)$$

This specifies a half normal distribution for τ_j , where $I(0,)$ denotes the distribution is truncated at zero. The variance of this distribution is small so that the posterior estimate of between-study variance is not unreasonably large.

Variance prior 1b

$$\tau_j \sim N(0,2)I(0,)$$

Given the frequentist estimates of between-study variance in the hypertension example, a slightly more conservative prior distribution is also considered.

Variance prior 2a

$$\frac{1}{\tau_j^2} \sim \text{Gamma}(0.1, 0.1)$$

This is a common prior distribution used for a transformation of the variance.^{22,29} This distribution is approximately uniform for most of the range, but has a ‘spike’ of probability mass close to zero.²⁹

Variance prior 2b

$$\frac{1}{\tau_j^2} \sim \text{Gamma}(1, 1)$$

Prior distribution 2a allows very large values of τ_j , therefore, for a sensitivity analysis, the same prior distribution with shape and scale parameters of one is also explored.

Variance prior 3

$$\tau_j \sim \text{Uniform}(0,2)$$

This is a prior distribution for the between-study standard deviation, which contains plausible values of the between-study variance for these examples, rather than a wide vague prior distribution that will allow large estimates of the variance parameters.

Variance prior 4

$$\log(\tau_j^2) \sim \text{Uniform}(-10, 1.386)$$

This prior distribution for the log of the variance is fairly informative since it bounds τ_j^2 between zero and four. The choice of the value of the upper bound is subjective here, however it may be considered implausible that the between-study variance could be larger than four.²⁹

Variance prior 5

$$\log(\tau_j^2) \sim t(-3.44, 2.59^2, 5)$$

This is an empirical prior distribution for the log of the between-study variance predicted for a new meta-analysis of randomised trials with a continuous outcome, as derived empirically from previous meta-analyses of randomised trials with continuous outcomes by Rhodes et al.³⁸ This student's t -distribution was recommended by Rhodes et al. for use in a general

setting and has a median of 0.03 and a 95% range of 0.0002 to 5.14 on the untransformed scale (τ_j).

6.3.2 Motivating example B

The second example dataset is from the same IPD meta-analysis in motivating example A, but is concerned with whether smoking is a prognostic factor for stroke, where smoking is a binary variable by yes (current smoker) or no (not current smoker). There are two prognostic effects for smoking: a partially adjusted log hazard ratio (HR), which is adjusted for treatment, and a fully adjusted log HR, which is adjusted for treatment, age and BMI. The summary results are shown in Table 6.2. There is missing information for age and BMI in five out of 10 trials, and so only partially adjusted HR estimates are available in these. However, in the remaining five trials, there is information to estimate both fully and partially adjusted log HRs and these prognostic effect estimates are highly correlated with the within-study correlations (derived from bootstrapping) close to +1.²²⁵

In addition to the priors described in Section 6.3.1, another prior distribution considered for the between-study variance in this example is an empirical prior distribution that was derived by Turner et al.³⁶

Variance prior 6:

$$\tau_j^2 \sim \text{lognormal}(-2.89, 1.91^2)$$

Turner et al. derived predictive distributions for the between-study variance for future meta-analyses with a binary outcome where there may be few studies to estimate this parameter within the model.³⁶ This prior distribution is for non-pharmacological interventions with semi-objective outcomes (an objective outcome that is not all-cause mortality). The median is

0.056 and a 95% prior interval is 0.001 to 2.35 for τ_j^2 . This prior distribution is not an exact match as these are prognostic effects rather than intervention effects, and the outcome is survival rather than binary. However, the event (stroke) is rare in this example and hazard ratios and odds ratios are often similar in this setting,^{233,234} therefore this empirical prior distribution is considered suitable for application in this example.

Table 6.2: Summary results for the 10 trials in the meta-analysis by Riley et al.²²⁴ for partially-adjusted and fully-adjusted log hazard ratios (log HR).

Trial name*	Control	Treatment	Partially-adjusted log HR (var)	Fully-adjusted log HR (var)	Within-study correlations (from bootstrap)
ATMH	750	780	0.216 (0.752)	0.173 (0.754)	0.992
HEP	199	150	1.238 (0.182)	1.477 (0.223)	0.893
EWPHE	82	90	-1.038 (1.080)	-0.667 (1.125)	0.988
HDFP	2371	2427	0.884 (0.072)	0.894 (0.074)	0.985
MRC-1	3445	3546	1.232 (0.119)	1.209 (0.120)	0.986
MRC-2	1337	1314	0.379 (0.039)	-	-
SHEP	2371	2365	0.399 (0.027)	-	-
STOP	131	137	1.203 (1.256)	-	-
Sy-Chi	1139	1252	0.633 (0.042)	-	-
Sy-Eur	2297	2398	0.156 (0.100)	-	-

* Trial names are consistent with Wang et al.,²²⁴ where further details and trial publications can be found.

Examples A and B provide a helpful contrast. Example B relates to time-to-event outcomes, whilst example A relates to continuous outcomes. Example B represents a situation where bivariate meta-analysis results may differ considerably to those from a univariate meta-analysis ($\rho_W = \rho_B = 0$ in model (6.1)). There are large within-study correlations, which are accounted for in the bivariate meta-analysis to borrow strength between partially and fully-adjusted results,^{232,235} which is considerably important here as five studies do not provide the fully-adjusted result. By contrast, example A has complete data for both outcomes and so

univariate and bivariate results are likely to be similar.²³² Example B also has smaller estimates of the between-study variances compared to example A ($\hat{\tau}_1$ is 0.063 and $\hat{\tau}_2$ is 0.033 using REML in the original paper). Therefore, it is possible to examine the impact of prior distributions for the between-study correlation for different sizes of true between-study variation across the two examples. It is also likely that a prior distribution for the between-study variance, which is appropriate for the data in example A, may not be appropriate for the data in example B. Hence examples A and B provide contrasting examples.

The posterior estimates of the parameters in model (6.1) for each dataset are obtained using the Gibbs Sampler MCMC method, which is implemented in SAS 9.3 using the PROC MCMC procedure.²³⁶ The analyses are performed with 100000 iterations after allowing for a 100000 iteration burn in and the samples are thinned by 10 to reduce concerns of autocorrelation. The convergence of parameters was checked using history and trace plots.

6.3.3 Summary of key findings from motivating examples

Table 6.3 shows the pooled results for example A using the various prior distributions for the between-study correlation, combined with the choices of prior distributions for between-study variance. Table 6.4 shows the corresponding results for example B. The key findings as observed in these examples are now discussed.

Key finding (i): The choice of prior distribution for ρ_B influences the posterior estimates for ρ_B .

As expected, in both these examples as the prior distribution for ρ_B is varied, the posterior mean of ρ_B and the 95% CrI change. In example B, with a $N(0,2)/(0,)$ prior for τ_j , $\hat{\rho}_B$ is 0.074 (95% CrI -0.608 to 0.700) using a Fisher prior distribution for transformation of ρ_B . This is much smaller than the estimate of ρ_B when a logit transformation prior distribution is used for

ρ_B , which is 0.531 (95% CrI 0.184 to 0.847). However, these large changes in the between-study correlation estimates do not greatly affect the pooled treatment effect estimates, and the conclusions remain the same. For example, for the same two prior distributions for ρ_B , the pooled estimate of the partially-adjusted log HR is 0.586 (95% CrI 0.352 to 0.845) compared to 0.587 (95% CrI 0.346 to 0.866). The pooled estimate of the fully-adjusted log HR is 0.717 (95% CrI 0.373 to 1.121) (Fisher prior distribution) compared to 0.696 (95% CrI 0.356 to 1.082) (logit prior distribution). Thus, it appears that the between-study correlation has little impact in example B in regard to the pooled estimates; this is also apparent in example A.

Key finding (ii): The prior distribution for τ_j influences the posterior results for ρ_B .

In both examples, as the estimates of the between-study variances increase, $\hat{\rho}_B$ also increases, even when the prior distribution for ρ_B remains the same. For example, in motivating example B, when $\rho_B \sim \text{uniform}(-1,1)$, $\hat{\rho}_B$ is 0.226 (95% CrI -0.923 to 0.981) if a $N(0,1)/(0,)$ prior distribution is applied for τ_j , and 0.842 (95% CrI -0.644 to 0.999) when a $\text{gamma}(0.1,0.1)$ prior distribution is used for $1/\tau_j^2$. This is because the posterior estimates of τ_j are larger if the prior distribution for τ_j is wider ($\hat{\tau}_1^2$ is 0.092 and $\hat{\tau}_2^2$ is 0.142 with a $N(0,1)/(0,)$ prior distribution, compared to $\hat{\tau}_1^2$, which is 4.508 and $\hat{\tau}_2^2$, which is 6.821 with a $\text{gamma}(0.1,0.1)$ prior). Therefore, these examples illustrate that the choice of prior distribution for the between-study variances can impact considerably upon the posterior distribution for ρ_B .

Table 6.3: Motivating example A - summary results from bivariate meta-analysis for various prior distributions for ρ_B and τ_j .

<u>Prior for τ</u> Prior for ρ_B	Pooled mean difference in systolic BP (95% Crl)	Pooled mean difference in diastolic BP (95% Crl)	$\hat{\tau}_1^2$ (95% Crl)	$\hat{\tau}_2^2$ (95% Crl)	$\hat{\rho}_B$ (95% Crl)	Prob(reduction in SBP>9 & reduction in DBP>4)
<u>$\tau_j \sim N(0,1)I(0,)$</u>						
$\rho_B = \rho_{W1} = 0$	-9.906 (-11.491 to -8.523)	-4.540 (-5.580 to -3.593)	4.033 (1.984 to 8.890)	2.068 (0.678 to 4.916)	-	0.438
$\rho_B \sim \text{Uniform}(-1,1)$	-9.831 (-11.327 to -8.476)	-4.448 (-5.376 to -3.569)	3.754 (1.174 to 8.101)	1.653 (0.557 to 3.880)	0.483 (-0.096 to 0.842)	0.492
Fisher z $\sim N(0, \text{sd}=0.4)$	-9.844 (-11.291 to -8.520)	-4.436 (-5.367 to -3.579)	3.509 (1.108 to 7.809)	1.577 (0.520 to 3.729)	0.349 (-0.155 to 0.724)	0.478
$(\rho_B + 1)/2 \sim \text{Beta}(1.5, 1.5)$	-9.805 (-11.307 to -8.470)	-4.443 (-5.357 to -3.583)	3.682 (1.177 to 8.121)	1.607 (0.559 to 3.689)	0.453 (-0.114 to 0.822)	0.490
$\rho_B \sim \text{Uniform}(0,1)$	-9.832 (-11.324 to -8.491)	-4.456 (-5.410 to -3.596)	3.744 (1.199 to 8.082)	1.637 (0.567 to 3.838)	0.515 (0.074 to 0.845)	0.500
Logit(ρ_B) $\sim N(0, \text{sd}=0.8)$	-9.823 (-11.333 to -8.476)	-4.450 (-5.364 to -3.576)	3.693 (1.223 to 7.942)	1.618 (0.579 to 3.691)	0.528 (0.224 to 0.791)	0.502
<u>$\tau_j \sim N(0,2)I(0,)$</u>						
$\rho_B = \rho_{W1} = 0$	-9.991 (-11.749 to - 8.421)	-4.567 (-5.702 to -3.531)	5.502 (1.556 to 13.085)	2.473 (0.736 to 6.210)	-	0.422
$\rho_B \sim \text{Uniform}(-1,1)$	-9.894 (-11.564 to - 8.322)	-4.476 (-5.479 to -3.549)	5.064 (1.505 to 11.976)	2.010 (0.630 to 4.899)	0.523 (-0.073 to 0.872)	0.497
Fisher z $\sim N(0, \text{sd}=0.4)$	-9.919 (-11.598 to - 8.430)	-4.469 (-5.492 to -3.528)	4.718 (1.397 to 11.177)	1.911 (0.604 to 4.896)	0.365 (-0.162 to 0.753)	0.471

Table 6.3: continued.

<i>Prior for τ</i> Prior for ρ_B	Pooled mean difference in systolic BP (95% Crl)	Pooled mean difference in diastolic BP (95% Crl)	$\hat{\tau}_1^2$ (95% Crl)	$\hat{\tau}_2^2$ (95% Crl)	$\hat{\rho}_B$ (95% Crl)	Prob(reduction in SBP>9 & reduction in DBP>4)
$(\rho_B+1)/2 \sim \text{Beta}(1.5, 1.5)$	-9.889 (-11.581 to - 8.393)	-4.470 (-5.473 to -3.555)	5.037 (1.493 to 11.823)	1.972 (0.624 to 4.877)	0.483 (-0.093 to 0.854)	0.485
$\rho_B \sim \text{Uniform}(0, 1)$	-9.916 (-11.611 to - 8.383)	-4.478 (-5.487 to -3.526)	5.183 (1.540 to 12.248)	2.209 (0.639 to 4.950)	0.552 (0.085 to 0.870)	0.504
Logit(ρ_B)$\sim N(0, \text{sd}=0.8)$	-9.917 (-11.637 to -8.432)	-4.473 (-5.458 to -3.540)	4.978 (1.538 to 11.505)	1.932 (0.629 to 4.569)	0.547 (0.234 to 0.808)	0.505
<i>$1/\hat{\tau}_j^2 \sim \text{Gamma}(1, 1)$</i>						
$\rho_B = \rho_{w_i} = 0$	-10.028 (-11.970 to -8.374)	-4.542 (-5.657 to -3.569)	6.409 (1.440 to 19.304)	2.251 (0.699 to 6.047)	-	0.431
$\rho_B \sim \text{Uniform}(-1, 1)$	-9.899 (-11.707 to -8.344)	-4.467 (-5.493 to -3.539)	5.450 (1.377 to 14.886)	1.901 (0.638 to 4.747)	0.523 (-0.075 to 0.872)	0.487
Fisher z$\sim N(0, \text{sd}=0.4)$	-9.926 (-11.683 to -8.420)	-4.458 (-5.402 to -3.586)	4.857 (1.249 to 13.802)	1.704 (0.588 to 4.259)	0.367 (-0.152 to 0.748)	0.472
$(\rho_B+1)/2 \sim \text{Beta}(1.5, 1.5)$	-9.897 (-11.665 to -8.362)	-4.468 (-5.433 to -3.553)	5.283 (1.333 to 15.200)	1.836 (0.604 to 4.595)	0.479 (-0.098 to 0.851)	0.483
$\rho_B \sim \text{Uniform}(0, 1)$	-9.906 (-11.690 to -8.328)	-4.467 (-5.459 to -3.540)	5.393 (1.383 to 14.769)	1.891 (0.617 to 4.995)	0.545 (0.083 to 0.875)	0.507
Logit(ρ_B)$\sim N(0, \text{sd}=0.8)$	-9.885 (-11.555 to -8.365)	-4.463 (-5.406 to -3.551)	5.151 (1.350 to 13.897)	1.793 (0.627 to 4.360)	0.545 (0.232 to 0.807)	0.502

Table 6.3: continued.

<i>Prior for τ</i> Prior for ρ_B	Pooled mean difference in systolic BP (95% CrI)	Pooled mean difference in diastolic BP (95% CrI)	$\hat{\tau}_1^2$ (95% CrI)	$\hat{\tau}_2^2$ (95% CrI)	$\hat{\rho}_B$ (95% CrI)	Prob(reduction in SBP>9 & reduction in DBP>4)
<i>$1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$</i>						
$\rho_B = \rho_{W_i} = 0$	-10.298 (-12.893 to -7.958)	-4.657 (-6.311 to -3.027)	13.846 (2.608 to 31.881)	6.219 (1.582 to 13.380)	-	0.388
$\rho_B \sim \text{Uniform}(-1, 1)$	-10.096 (-12.766 to -7.551)	-4.554 (-6.152 to -2.981)	16.352 (4.508 to 47.418)	6.089 (2.152 to 16.434)	0.712 (0.036 to 0.956)	0.494
Fisher z $\sim \text{N}(0, \text{sd}=0.4)$	-10.108 (-12.481 to -7.827)	-4.569 (-6.081 to -3.075)	12.678 (3.808 to 34.851)	5.316 (1.959 to 13.619)	0.414 (-0.249 to 0.825)	0.435
$(\rho_B + 1)/2 \sim \text{Beta}(1.5, 1.5)$	-10.111 (-12.718 to -7.545)	-4.565 (-6.120 to -2.997)	15.037 (4.397 to 40.828)	5.701 (2.045 to 14.301)	0.652 (-0.051 to 0.944)	0.479
$\rho_B \sim \text{Uniform}(0, 1)$	-10.122 (-12.855 to -7.395)	-4.567 (-6.125 to -2.932)	17.279 (4.684 to 50.817)	6.226 (2.168 to 16.668)	0.739 (0.191 to 0.959)	0.497
Logit(ρ_B) $\sim \text{N}(0, \text{sd}=0.8)$	-10.124 (-12.629 to -7.725)	-4.576 (-6.087 to -3.067)	13.486 (4.068 to 36.220)	5.300 (1.986 to 13.069)	0.630 (0.269 to 0.881)	0.482
<i>$\tau_j \sim \text{Uniform}(0, 2)$</i>						
$\rho_B = \rho_{W_i} = 0$	-9.815 (-11.098 to -8.539)	-4.566 (-5.591 to -3.540)	2.980 (1.613 to 4.000)	2.241 (0.869 to 3.941)	-	0.434
$\rho_B \sim \text{Uniform}(-1, 1)$	-9.812 (-11.125 to -8.562)	-4.483 (-5.442 to -3.588)	2.985 (1.430 to 3.960)	1.808 (0.618 to 3.648)	0.482 (-0.072 to 0.828)	0.506
Fisher z $\sim \text{N}(0, \text{sd}=0.4)$	-9.753 (-11.029 to -8.535)	-4.430 (-5.357 to -3.567)	2.909 (1.306 to 3.956)	1.750 (0.581 to 3.640)	0.343 (-0.165 to 0.721)	0.462

Table 6.3: continued.

<i>Prior for τ</i> Prior for ρ_B	Pooled mean difference in systolic BP (95% Crl)	Pooled mean difference in diastolic BP (95% Crl)	$\hat{\tau}_1^2$ (95% Crl)	$\hat{\tau}_2^2$ (95% Crl)	$\hat{\rho}_B$ (95% Crl)	Prob(reduction in SBP>9 & reduction in DBP>4)
$(\rho_B+1)/2 \sim \text{Beta}(1.5, 1.5)$	-9.773 (-11.042 to -8.553)	-4.461 (-5.418 to -3.547)	2.952 (1.343 to 3.958)	1.813 (0.605 to 3.711)	0.446 (-0.120 to 0.810)	0.479
$\rho_B \sim \text{Uniform}(0, 1)$	-9.792 (-11.101 to -8.566)	-4.459 (-5.413 to -3.564)	2.981 (1.439 to 3.961)	1.808 (0.611 to 3.646)	0.506 (0.070 to 0.829)	0.504
Logit(ρ_B)$\sim N(0, \text{sd}=0.8)$	-9.773 (-11.042 to -8.524)	-4.452 (-5.412 to -3.540)	2.971 (1.385 to 3.959)	1.796 (0.632 to 3.652)	0.526 (0.220 to 0.783)	0.491
<i><u>Log(τ_i^2)$\sim \text{Uniform}(-10, 1.386)$</u></i>						
$\rho_B = \rho_{w_i} = 0$	-9.806 (-11.101 to -8.567)	-4.550 (-5.574 to -3.591)	2.874 (1.470 to 3.999)	2.031 (0.669 to 3.793)	-	0.432
$\rho_B \sim \text{Uniform}(-1, 1)$	-9.792 (-11.066 to -8.550)	-4.451 (-5.339 to -3.594)	2.829 (1.153 to 3.954)	1.599 (0.538 to 3.493)	0.464 (-0.083 to 0.822)	0.506
Fisher z$\sim N(0, \text{sd}=0.4)$	-9.718 (-11.006 to -8.502)	-4.408 (-5.322 to -3.581)	2.755 (1.143 to 3.942)	1.552 (0.500 to 3.485)	0.335 (-0.179 to 0.715)	0.477
$(\rho_B+1)/2 \sim \text{Beta}(1.5, 1.5)$	-9.755 (-11.038 to -8.523)	-4.428 (-5.329 to -3.566)	2.828 (1.232 to 3.942)	1.594 (0.527 to 3.488)	0.438 (-0.118 to 0.804)	0.484
$\rho_B \sim \text{Uniform}(0, 1)$	-9.751 (-11.045 to -8.501)	-4.432 (-5.334 to -3.555)	2.852 (1.228 to 3.951)	1.609 (0.552 to 3.483)	0.491 (0.068 to 0.823)	0.500
Logit(ρ_B)$\sim N(0, \text{sd}=0.8)$	-9.759 (-11.108 to -8.540)	-4.433 (-5.356 to -3.590)	2.840 (1.237 to 3.942)	1.601 (0.539 to 3.470)	0.516 (0.215 to 0.779)	0.495

Table 6.3: continued.

Prior for τ Prior for ρ_B	Pooled mean difference in systolic BP (95% CrI)	Pooled mean difference in diastolic BP (95% CrI)	$\hat{\tau}_1^2$ (95% CrI)	$\hat{\tau}_2^2$ (95% CrI)	$\hat{\rho}_B$ (95% CrI)	Prob(reduction in SBP>9 & reduction in DBP>4)
<i>$\log(\tau_i^2) \sim t(-3.44, 2.59^2, 5)$</i>						
$\rho_B = \rho_{Wi} = 0$	-10.088 (-12.187 to -8.366)	-4.548 (-5.686 to -3.570)	7.147 (1.491 to 22.128)	2.311 (0.626 to 6.597)	-	0.431
$\rho_B \sim \text{Uniform}(-1, 1)$	-9.982 (-11.946 to -8.388)	-4.492 (-5.527 to -3.575)	6.111 (1.286 to 18.818)	1.954 (0.536 to 5.405)	0.521 (-0.083 to 0.885)	0.512
Fisher z $\sim N(0, \text{sd}=0.4)$	-9.868 (-11.638 to -8.320)	-4.439 (-5.388 to -3.569)	5.208 (1.184 to 15.481)	1.671 (0.485 to 4.474)	0.359 (-0.166 to 0.748)	0.465
$(\rho_B + 1)/2 \sim \text{Beta}(1.5, 1.5)$	-9.909 (-11.750 to -8.232)	-4.451 (-5.465 to -3.520)	5.906 (1.271 to 17.634)	1.897 (0.546 to 5.345)	0.486 (-0.095 to 0.857)	0.485
$\rho_B \sim \text{Uniform}(0, 1)$	-9.916 (-11.765 to -8.263)	-4.459 (-5.455 to -3.511)	6.041 (1.338 to 18.137)	1.956 (0.532 to 5.567)	0.550 (0.077 to 0.885)	0.499
Logit(ρ_B) $\sim N(0, \text{sd}=0.8)$	-9.912 (-11.675 to -8.358)	-4.457 (-5.418 to -3.556)	5.588 (1.280 to 16.212)	1.786 (0.537 to 4.809)	0.544 (0.227 to 0.814)	0.508

Table 6.4: Motivating example B - summary results from bivariate meta-analysis for various prior distributions for ρ_B and τ_j .

<i>Prior for τ</i> Prior for ρ_B	Pooled partially- adjusted log HR (95% CrI)	Pooled fully- adjusted log HR (95% CrI)	$\hat{\tau}_1^2$ (95% CrI)	$\hat{\tau}_2^2$ (95% CrI)	$\hat{\rho}_B$ (95% CrI)	Prob (Partially- adjusted logHR>0.405 & fully-adjusted logHR>0.405)
<u>$\tau_j \sim N(0,1)I(0,)$</u>						
$\rho_B = \rho_{w_i} = 0$	0.587 (0.281 to 0.910)	0.928 (0.162 to 1.536)	0.112 (0.000 to 0.549)	0.311 (0.000 to 1.859)	-	0.656
$\rho_B \sim \text{Uniform}(-1,1)$	0.587 (0.345 to 0.863)	0.708 (0.379 to 1.100)	0.092 (0.000 to 0.298)	0.142 (0.000 to 0.410)	0.226 (-0.923 to 0.981)	0.754
Fisher z $\sim N(0, \text{sd}=0.4)$	0.583 (0.356 to 0.833)	0.710 (0.389 to 1.099)	0.047 (0.000 to 0.232)	0.060 (0.000 to 0.398)	0.071 (-0.619 to 0.702)	0.736
$(\rho_B + 1)/2 \sim \text{Beta}(1.5, 1.5)$	0.587 (0.349 to 0.854)	0.708 (0.383 to 1.093)	0.055 (0.000 to 0.272)	0.064 (0.000 to 0.391)	0.161 (-0.832 to 0.930)	0.694
$\rho_B \sim \text{Uniform}(0,1)$	0.591 (0.337 to 0.871)	0.695 (0.353 to 1.063)	0.073 (0.000 to 0.338)	0.078 (0.000 to 0.460)	0.585 (0.039 to 0.987)	0.695
Logit(ρ_B) $\sim N(0, \text{sd}=0.8)$	0.587 (0.337 to 0.865)	0.700 (0.368 to 1.077)	0.066 (0.000 to 0.299)	0.069 (0.000 to 0.415)	0.529 (0.188 to 0.848)	0.740
<u>$\tau_j \sim N(0,2)I(0,)$</u>						
$\rho_B = \rho_{w_i} = 0$	0.591 (0.289 to 0.935)	0.924 (0.071 to 1.616)	0.124 (0.000 to 0.607)	0.447 (0.001 to 2.726)	-	0.648
$\rho_B \sim \text{Uniform}(-1,1)$	0.588 (0.342 to 0.857)	0.707 (0.372 to 1.102)	0.065 (0.000 to 0.331)	0.076 (0.000 to 0.520)	0.221 (-0.923 to 0.979)	0.729

Table 6.4: continued.

<i>Prior for τ</i> Prior for ρ_B	Pooled partially- adjusted log HR (95% CrI)	Pooled fully- adjusted log HR (95% CrI)	$\hat{\tau}_1^2$ (95% CrI)	$\hat{\tau}_2^2$ (95% CrI)	$\hat{\rho}_B$ (95% CrI)	Prob (Partially- adjusted logHR>0.405 & fully-adjusted logHR>0.405)
Fisher $z \sim N(0, sd=0.4)$	0.586 (0.352 to 0.845)	0.717 (0.373 to 1.121)	0.053 (0.000 to 0.254)	0.071 (0.000 to 0.467)	0.074 (-0.608 to 0.700)	0.768
$(\rho_B+1)/2 \sim \text{Beta}(1.5, 1.5)$	0.589 (0.353 to 0.857)	0.717 (0.383 to 1.104)	0.060 (0.000 to 0.297)	0.070 (0.000 to 0.451)	0.166 (-0.832 to 0.931)	0.783
$\rho_B \sim \text{Uniform}(0, 1)$	0.594 (0.335 to 0.888)	0.698 (0.351 to 1.095)	0.079 (0.000 to 0.378)	0.088 (0.000 to 0.528)	0.589 (0.044 to 0.987)	0.689
Logit(ρ_B) $\sim N(0, sd=0.8)$	0.587 (0.346 to 0.866)	0.696 (0.356 to 1.082)	0.063 (0.000 to 0.289)	0.079 (0.000 to 0.477)	0.531 (0.184 to 0.847)	0.716
<hr/> <i>$1/\hat{\tau}_i^2 \sim \text{Gamma}(1, 1)$</i>						
$\rho_B = \rho_{W_i} = 0$	0.599 (0.091 to 1.102)	0.835 (-0.224 to 1.813)	0.474 (0.160 to 1.245)	0.938 (0.210 to 3.395)	-	0.426
$\rho_B \sim \text{Uniform}(-1, 1)$	0.607 (0.080 to 1.126)	0.724 (0.072 to 1.415)	0.480 (0.170 to 1.180)	0.586 (0.179 to 1.730)	0.723 (-0.532 to 0.998)	0.555
Fisher $z \sim N(0, sd=0.4)$	0.618 (0.128 to 1.098)	0.823 (0.021 to 1.622)	0.411 (0.149 to 1.028)	0.647 (0.174 to 2.057)	0.242 (-0.525 to 0.807)	0.495
$(\rho_B+1)/2 \sim \text{Beta}(1.5, 1.5)$	0.612 (0.117 to 1.114)	0.763 (0.037 to 1.532)	0.452 (0.162 to 1.118)	0.617 (0.177 to 1.880)	0.569 (-0.620 to 0.986)	0.504
$\rho_B \sim \text{Uniform}(0, 1)$	0.609 (0.083 to 1.123)	0.710 (0.068 to 1.333)	0.489 (0.169 to 1.215)	0.583 (0.187 to 1.654)	0.813 (0.163 to 0.998)	0.558

Table 6.4: continued.

<i>Prior for τ</i> Prior for ρ_B	Pooled partially- adjusted log HR (95% CrI)	Pooled fully- adjusted log HR (95% CrI)	$\hat{\tau}_1^2$ (95% CrI)	$\hat{\tau}_2^2$ (95% CrI)	$\hat{\rho}_B$ (95% CrI)	Prob (Partially- adjusted logHR>0.405 & fully-adjusted logHR>0.405)
Logit(ρ_B)~N(0,sd=0.8)	0.610 (0.124 to 1.094)	0.749 (0.040 to 1.449)	0.424 (0.152 to 1.043)	0.616 (0.178 to 1.905)	0.606 (0.234 to 0.895)	0.520
<i>$1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$</i>						
$\rho_B = \rho_{W_i} = 0$	0.553 (-0.667 to 1.779)	0.645 (-2.409 to 3.588)	3.512 (1.262 to 9.138)	10.999 (2.000 to 46.298)	-	0.283
$\rho_B \sim \text{Uniform}(-1, 1)$	0.575 (-0.811 to 1.938)	0.674 (-1.146 to 2.434)	4.508 (1.570 to 11.341)	6.821 (1.924 to 22.742)	0.842 (-0.644 to 0.999)	0.446
Fisher z~N(0,sd=0.4)	0.580 (-0.658 to 1.819)	0.741 (-2.061 to 3.507)	3.475 (1.266 to 8.940)	10.201 (1.929 to 42.271)	0.143 (-0.647 to 0.804)	0.322
$(\rho_B + 1)/2 \sim \text{Beta}(1.5, 1.5)$	0.572 (-0.750 to 1.885)	0.676 (-1.493 to 2.902)	3.963 (1.391 to 10.333)	8.083 (1.883 to 31.501)	0.629 (-0.765 to 0.998)	0.457
$\rho_B \sim \text{Uniform}(0, 1)$	0.581 (-0.818 to 2.006)	0.666 (-1.044 to 2.396)	4.642 (1.598 to 11.856)	6.423 (1.884 to 21.295)	0.932 (0.414 to >0.999)	0.504
Logit(ρ_B)~N(0,sd=0.8)	0.559 (-0.672 to 1.770)	0.666 (-1.768 to 3.082)	3.488 (1.297 to 8.919)	8.955 (1.922 to 34.985)	0.622 (0.234 to 0.908)	0.401
<i>$\tau_j \sim \text{Uniform}(0, 2)$</i>						
$\rho_B = \rho_{W_i} = 0$	0.582 (0.297 to 0.899)	0.871 (-0.036 to 1.586)	0.102 (0.000 to 0.570)	0.435 (2.18E-6 to 2.784)	-	0.640
$\rho_B \sim \text{Uniform}(-1, 1)$	0.588 (0.336 to 0.864)	0.705 (0.353 to 1.117)	0.067 (0.000 to 0.343)	0.090 (0.000 to 0.607)	0.256 (-0.893 to 0.982)	0.774

Table 6.4: continued.

<i>Prior for τ</i> Prior for ρ_B	Pooled partially- adjusted log HR (95% CrI)	Pooled fully- adjusted log HR (95% CrI)	$\hat{\tau}_1^2$ (95% CrI)	$\hat{\tau}_2^2$ (95% CrI)	$\hat{\rho}_B$ (95% CrI)	Prob (Partially- adjusted logHR>0.405 & fully-adjusted logHR>0.405)
Fisher z~N(0,sd=0.4)	0.585 (0.351 to 0.855)	0.717 (0.387 to 1.131)	0.055 (0.000 to 0.265)	0.083 (0.000 to 0.594)	0.083 (-0.617 to 0.709)	0.740
$(\rho_B+1)/2$~Beta(1.5,1.5)	0.595 (0.357 to 0.864)	0.723 (0.376 to 1.135)	0.061 (0.000 to 0.308)	0.081 (0.000 to 0.551)	0.166 (-0.813 to 0.932)	0.760
ρ_B~Uniform(0,1)	0.590 (0.329 to 0.883)	0.693 (0.335 to 1.086)	0.081 (0.000 to 0.395)	0.098 (0.000 to 0.594)	0.591 (0.036 to 0.983)	0.723
Logit(ρ_B)~N(0,sd=0.8)	0.585 (0.322 to 0.876)	0.696 (0.306 to 1.081)	0.069 (0.000 to 0.321)	0.095 (0.000 to 0.596)	0.530 (0.185 to 0.853)	0.753
<hr/> <i>Log(τ_i^2)~Uniform(-10,1.386)</i>						
$\rho_B=\rho_{W_i}=0$	0.578 (0.367 to 0.801)	0.967 (0.462 to 1.398)	0.035 (5.8E-5 to 0.256)	0.089 (5.7E-5 to 0.826)	-	0.798
ρ_B~Uniform(-1,1)	0.566 (0.378 to 0.773)	0.638 (0.395 to 0.947)	0.016 (0.000 to 0.105)	0.013 (0.000 to 0.105)	0.041 (-0.947 to 0.951)	0.849
Fisher z~N(0,sd=0.4)	0.564 (0.369 to 0.769)	0.634 (0.393 to 0.935)	0.014 (0.000 to 0.097)	0.011 (0.000 to 0.086)	0.020 (-0.651 to 0.674)	0.854
$(\rho_B+1)/2$~Beta(1.5,1.5)	0.569 (0.383 to 0.777)	0.641 (0.417 to 0.935)	0.015 (0.000 to 0.099)	0.013 (0.000 to 0.091)	0.028 (-0.878 to 0.887)	0.838
ρ_B~Uniform(0,1)	0.570 (0.375 to 0.783)	0.635 (0.400 to 0.934)	0.017 (0.000 to 0.112)	0.015 (0.000 to 0.113)	0.515 (0.026 to 0.977)	0.897

Table 6.4: continued.

<i>Prior for τ</i> Prior for ρ_B	Pooled partially- adjusted log HR (95% CrI)	Pooled fully- adjusted log HR (95% CrI)	$\hat{\tau}_1^2$ (95% CrI)	$\hat{\tau}_2^2$ (95% CrI)	$\hat{\rho}_B$ (95% CrI)	Prob (Partially- adjusted logHR>0.405 & fully-adjusted logHR>0.405)
Logit(ρ_B)~N(0,sd=0.8)	0.572 (0.372 to 0.790)	0.640 (0.396 to 0.947)	0.017 (0.000 to 0.116)	0.013 (0.000 to 0.102)	0.505 (0.174 to 0.830)	0.884
$\tau_i^2 \sim \text{lognormal}(-$ $2.89, 1.91^2)$						
$\rho_B = \rho_{W_i} = 0$	0.585 (0.335 to 0.852)	0.978 (0.447 to 1.449)	0.057 (0.000 to 0.272)	0.120 (0.001 to 0.965)	-	0.730
$\rho_B \sim \text{Uniform}(-1,1)$	0.581 (0.362 to 0.821)	0.692 (0.399 to 1.031)	0.036 (0.001 to 0.155)	0.035 (0.001 to 0.200)	0.199 (-0.917 to 0.974)	0.775
Fisher z~N(0,sd=0.4)	0.580 (0.367 to 0.812)	0.701 (0.410 to 1.037)	0.033 (0.001 to 0.143)	0.031 (0.001 to 0.161)	0.069 (-0.618 to 0.695)	0.795
$(\rho_B+1)/2 \sim \text{Beta}(1.5,1.5)$	0.581 (0.364 to 0.817)	0.696 (0.410 to 1.027)	0.036 (0.001 to 0.160)	0.036 (0.001 to 0.160)	0.029 (0.001 to 0.150)	0.768
$\rho_B \sim \text{Uniform}(0,1)$	0.584 (0.359 to 0.826)	0.681 (0.404 to 0.995)	0.042 (0.001 to 0.177)	0.037 (0.001 to 0.187)	0.561 (0.035 to 0.983)	0.771
Logit(ρ_B)~N(0,sd=0.8)	0.581 (0.360 to 0.821)	0.682 (0.400 to 1.001)	0.038 (0.001 to 0.149)	0.035 (0.001 to 0.177)	0.522 (0.183 to 0.841)	0.772

Key finding (iii): Bivariate meta-analysis impacts upon the pooled estimates more when there are missing data.

The pooled results from the univariate meta-analysis (where $\rho_{Wi}=\rho_B=0$ in model (6.1)) for both examples are also shown in Table 6.3 and Table 6.4. In example A, the pooled treatment effects are very similar regardless of using bivariate or univariate meta-analysis. Riley et al.²³² identify this is due to complete data in all studies (i.e. both treatment effects are known in all 10 studies). However, in example B, the univariate pooled estimate of the fully-adjusted log HR is vastly different to the bivariate pooled estimate due to additional borrowing of strength in those five studies where the fully-adjusted result was missing. For example, the univariate meta-analysis has a pooled fully-adjusted log HR of 0.928 (95% CrI 0.162 to 1.536) using a prior of $\tau_j \sim N(0,1)/(0,.)$. In the bivariate meta-analysis with the same prior distribution for τ_j , and $\rho_B \sim \text{uniform}(-1,1)$, the pooled log HR is 0.708 (95% CrI 0.379 to 1.100). Although the conclusions based on these two pooled estimates would be the same, the pooled effect is much smaller with a narrower credible interval in the bivariate meta-analysis due to the borrowing of strength across outcomes.

Also, the joint probability estimates are affected in the bivariate model compared to the univariate model. For example, in the univariate analysis where $\tau_j \sim N(0,1)/(0,.)$ the joint probability that partially adjusted log HR > 0.405 and fully adjusted log HR > 0.405 is 0.656. This probability is higher for the bivariate meta-analysis model for all prior distributions for ρ_B . In particular, when $\rho_B \sim \text{uniform}(-1,1)$, this joint probability is 0.754.

Key finding (iv): Bivariate meta-analysis also impacts upon the posterior results for τ_j .

The estimates of the between-study variances are also different between the univariate and bivariate meta-analyses. This is especially evident in example B. For example, in the univariate analysis using a prior of $\tau_j \sim N(0,1)/(0,.)$, $\hat{\tau}_1^2$ is 0.112 (95% CrI 0.000 to 0.549);

however, in the bivariate analysis using priors of $\rho_B \sim \text{uniform}(-1,1)$ and $\tau_j \sim N(0,1)/(0,)$, $\hat{\tau}_1^2$ is 0.092 (95% CrI 0.000 to 0.298). Similarly, $\hat{\tau}_2^2$ is 0.311 (95% CrI 0.000 to 1.859) in the univariate analysis, compared to 0.142 (95% CrI 0.000 to 0.410) in the bivariate analysis.

Key finding (v): The prior distribution for ρ_B also influences the posterior estimates for τ_j .

The prior distribution for ρ_B can alter the posterior estimates of the between-study variance for the same choice of prior distribution for τ_j . For example, in example B when using priors of $1/\tau_j^2 \sim \text{gamma}(0.1,0.1)$ and $\rho_B \sim \text{uniform}(-1,1)$ the posterior estimate of τ_1^2 is 4.508 (95% CrI 1.570 to 11.341) and for τ_2^2 is 6.821 (95% CrI 1.924 to 22.742). In this analysis, the posterior estimate of ρ_B is 0.842 (95% CrI: -0.644 to 0.999) However, when the Fisher transformation prior is used for ρ_B the posterior estimate of τ_1^2 is 3.475 (95% CrI 1.266 to 8.940) and the posterior estimate of τ_2^2 is 10.201 (95% CrI 1.929 to 42.271). In this analysis, the posterior estimate of ρ_B is 0.143 (95% CrI: -0.647 to 0.804).

Key finding (vi): The gamma prior distribution for $1/\tau_j^2$ appears inappropriate.

The gamma prior distribution for the between-study variance appears particularly inappropriate because the posterior estimates of τ_j^2 are much larger than other prior distributions, and this increases the posterior estimate of ρ_B , which affects the joint probability estimates. For example, in motivating example B with a $\text{uniform}(-1,1)$ prior for ρ_B the posterior estimates of τ_1^2 and τ_2^2 are 4.508 (95% CrI 1.570 to 11.341) and 6.821 (95% CrI 1.924 to 22.742), respectively when using a $\text{gamma}(0.1,0.1)$ prior distribution for the between-study variance parameters. The posterior estimate of ρ_B is 0.842 (95% CrI -0.644 to 0.999) and the joint probability of interest is 0.446. In contrast, the corresponding estimates, with perhaps a more realistically vague $N(0,2)/(0,)$ prior for τ_j , are 0.065 (95% CrI 0.000 to

0.331) for τ_1^2 and 0.076 (95% CrI 0.000 to 0.520) for τ_2^2 . The posterior estimate of ρ_B is 0.221 (95% CrI -0.923 to 0.979) and the joint probability of interest is 0.729, both dramatically different to when the gamma prior was used. This finding agrees with those already determined by Lambert et al.²⁹ and Wei et al.²³⁷ about the influential impact of a gamma (Wishart) prior on the between-study variances; in addition, the above results reveal it is also influential toward the between-study correlation and joint inferences.

Summary

In summary, the two motivating examples highlight six key findings. Clearly, the choice of prior distributions for the between-study correlation and between-study variances are important. These can all influence posterior results and joint inferences, sometimes dramatically so when the chosen priors are inappropriate. This is especially important when there is large borrowing of strength in a bivariate meta-analysis as in example B. Thus, the examples indicate that all prior distributions must be specified with due consideration in a bivariate meta-analysis and that seemingly vague priors may actually be influential. Of course, these are merely two examples that motivate further inquiry, and thus a simulation study will now be conducted to explore whether these results are found more generally.

6.4 Methods of the simulation study

This section introduces a simulation study of the impact of prior distributions for the between-study correlation in a Bayesian estimation of bivariate meta-analysis model (6.1). The simulation focuses mainly on $N=10$ studies per meta-analysis, as this represents a fairly typical size of meta-analysis.²³⁸ Section 6.7 will briefly consider alternative N . Both complete data (both outcomes are available in all studies) and missing data (some studies only provide one outcome) are considered.

Given below is an outline of the procedure used to simulate the data. The description of the procedure is brief because this data was generated for a previous simulation study by Riley et al.,²¹⁸ where the aim was to evaluate the benefits of model (6.1) in comparison to univariate meta-analysis models (in a frequentist framework). Riley et al. generated meta-analysis data for N=50 studies per meta-analysis dataset, which consisted of an estimate for each outcome (Y_{i1} and Y_{i2}) in each study, and their variances (s_{i1}^2 and s_{i2}^2) and a chosen within-study correlation. Complete data were generated for both outcomes and the full dataset contained 1000 simulated meta-analyses.

6.4.1 How Riley et al. generated bivariate meta-analysis data for 50 studies

To obtain the data, Riley et al.²²⁵ generated data at summary level rather than for the individual patient level, for 50 studies as follows. Two *true* mean (pooled) treatment effects were chosen to reflect contrasting effects, so $\beta_1=0$ (e.g. no difference in blood pressure) to reflect no clinical benefit, and $\beta_2=2$ (e.g. 2 mmHg difference in blood pressure) to reflect potentially useful clinical benefit. The 50 within-study variances (i.e. s_{i1}^2 and s_{i2}^2 from model (6.1)) were sampled from the following distribution: $\begin{pmatrix} \ln(s_{i1}^2) \\ \ln(s_{i2}^2) \end{pmatrix} \sim N\left(\begin{pmatrix} 0.25 \\ 0.25 \end{pmatrix}, \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}\right)$, and were assumed known and kept the same across all the 1000 simulated meta-analysis datasets. Using the known 50 s_{i1}^2 and 50 s_{i2}^2 , and letting $\tau_1=0.5$ and $\tau_2=0.5$, 50 Y_{i1} and 50 Y_{i2} were generated 1000 times for each of the four following settings.

1. $\rho_{W_i}=0; \rho_B=0; \beta_1=0; \beta_2=2, \tau_1=0.5, \tau_2=0.5$ and 50 s_{i1}^2 and 50 s_{i2}^2
2. $\rho_{W_i}=0; \rho_B=0.8; \beta_1=0; \beta_2=2, \tau_1=0.5, \tau_2=0.5$ and 50 s_{i1}^2 and 50 s_{i2}^2
3. $\rho_{W_i}=0.8; \rho_B=0; \beta_1=0; \beta_2=2, \tau_1=0.5, \tau_2=0.5$ and 50 s_{i1}^2 and 50 s_{i2}^2

4. $\rho_{W_i}=0.8$; $\rho_B=0.8$; $\beta_1=0$; $\beta_2=2$, $\tau_1=0.5$, $\tau_2=0.5$ and 50 s_{i1}^2 and 50 s_{i2}^2

These four settings vary according to the magnitude of within-study and between-study correlations. Only non-negative within-study and between-study correlations are considered, as bivariate meta-analyses are similar regardless of whether positive or negative correlations are considered²¹⁸ in terms of their impact on pooled treatment effect estimates.

For each of the four settings, 50 θ_{i1} and 50 θ_{i2} were randomly generated 1000 times from the corresponding bivariate distribution at the study-level. Using setting 1 as an example, where there is no between-study correlation between the treatment effects, the 50 θ_{i1} and 50 θ_{i2} were simulated assuming the distribution in (6.2).

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix} \right\} \quad (6.2)$$

For settings 2 and 4, equation (6.2) was changed so that the covariance in \mathbf{D} was 0.2, to ensure the 50 θ_{i1} and θ_{i2} were generated according to a between-study correlation of 0.8.

Then, the next step was to generate the Y_{i1} and Y_{i2} . In each setting separately, and for each set of the 50 θ_{i1} and 50 θ_{i2} , obtained, 50 Y_{i1} and 50 Y_{i2} were generated using the 50 θ_{i1} and 50 θ_{i2} together with the 50 s_{i1}^2 and 50 s_{i2}^2 . Using setting 1 as an example, where there is no within-study correlation, 50 Y_{i1} and 50 Y_{i2} were simulated from the bivariate distribution in (6.3).

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \mathbf{S}_i = \begin{pmatrix} s_{i1}^2 & 0 \\ 0 & s_{i2}^2 \end{pmatrix} \right\} \quad (6.3)$$

In settings 3 and 4, the covariance in equation (6.3) was changed to ensure the within-study correlation was 0.8. In this way, for each of the 1000 simulations in each setting, 50 Y_{i1} and 50 Y_{i2} were generated.

6.4.2 Step-by-step guide to new simulation study

The simulation procedure used in this chapter is now described.

Step 1: Data set-up

The data generated by Riley et al. was used in this new simulation study to evaluate the impact of prior distributions in each of settings 1 to 4. As described, 50 study estimates were available per outcome in each meta-analysis dataset, but only a proportion of these are used here. For example, if there are 10 studies per meta-analysis, the first 10 studies are selected and the rest are discarded.

Step 2: Analysis

In each setting and for each number of studies per meta-analysis, to each of the 1000 meta-analysis datasets a bivariate meta-analysis model (6.1) was fitted with a particular set of chosen prior distributions. This was then repeated for each different set of prior distributions of interest. The posterior estimates of the parameters in model (6.1) for each dataset were obtained using the Gibbs Sampler Markov chain Monte Carlo (MCMC) method, which was implemented in SAS 9.3 using the PROC MCMC procedure.²³⁶ For each dataset, the analyses were performed with 300000 iterations after allowing for a 200000 iteration burn in and the samples were thinned by 100 to reduce concerns of autocorrelation (see Appendix D1 for SAS code). The convergence of parameters was checked using history and trace plots. The mean, median, standard deviation, and 95% credible interval from the posterior

distribution for each of the parameters in model (6.1), and the individual and joint probabilities were stored from the results of each of the 1000 datasets.

Step 3: Comparison of results

To compare the choice of prior distribution for the between-study correlation, the following were calculated from the set of 1000, for each combination of priors:

- The average posterior mean estimate of pooled treatment effects ($\hat{\beta}_j$) across all simulations, the average (mean/median) posterior between-study standard deviation estimates ($\hat{\tau}_j$) across all simulations; and the average (mean/median) posterior between-study correlation estimate ($\hat{\rho}_B$) across all simulations (to check for bias),
- The mean and median standard error of the posterior pooled treatment effects, $\hat{\beta}_1$ and $\hat{\beta}_2$ across simulations,
- The mean-squared error (MSE) of the pooled treatment effects, calculated by the average squared deviation from the true value across the 1000 simulated datasets,
- The coverage performance of the 95% credible intervals for the pooled treatment effects, calculated by the percentage of the 95% credible intervals that contain the true treatment effect,
- The average marginal probabilities that $\theta_{i1new} > 0$ and $\theta_{i2new} > 2$, and the average joint probability that $\theta_{i1new} > 0$ and $\theta_{i2new} > 2$.

Recall that, based on bivariate model (6.1), the predictive distribution of treatment effects in a new trial is assumed to be:

$$\begin{pmatrix} \theta_{i1_{new}} \\ \theta_{i2_{new}} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \mathbf{D} \right) \quad (6.4)$$

When drawing $\theta_{i1_{new}}$ and $\theta_{i2_{new}}$ samples from this distribution during the MCMC process, the uncertainty in the pooled average treatment effects, β_1 and β_2 , and the uncertainty in the between-study covariance matrix, \mathbf{D} , will also be accounted for. In settings 1 and 3, where $\rho_B=0$, the true marginal probabilities are 0.5 for both treatment effects, and the true joint probability that $\theta_{i1_{new}} > 0$ and $\theta_{i2_{new}} > 2$ is 0.25. When $\rho_B=0.8$ in settings 2 and 4, the true joint probability is 0.4.

These simulation results (e.g. bias, MSE, coverage) are essentially a frequentist evaluation of a Bayesian analysis, which some may argue may not be appropriate. In particular, Senn²³⁹ previously suggested that it is perhaps philosophically incorrect to conduct a simulation study to assess the performance of Bayesian prior distributions because it is “irrelevant to any Bayesian who truly believed what the prior distribution represented”. However, the rationale for a simulation study to assess the various prior distributions here is similar to that of Lambert et al. in their response to this article. They say that, “if a statistician desires to have a model with good bias and coverage properties, but needs/wants to use Bayesian methods, then we believe that simulation is a very good way of establishing this.”²⁴⁰

6.4.3 Choice of prior distribution for between-study correlation

The prior distributions for the between-study correlation considered in the simulation study were the same as those used in motivating Examples A and B described in section 6.3. The distributions are repeated below:

- *Prior 1:* $\rho_B \sim \text{uniform}(-1,1)$,
- *Prior 2:* $z = \frac{1}{2} \log \left(\frac{1+\rho_B}{1-\rho_B} \right) \sim N(0, \text{sd}=0.4)$,
- *Prior 3:* $\frac{(\rho_B+1)}{2} \sim \text{Beta}(1.5,1.5)$,
- *Prior 4:* $\rho_B \sim \text{uniform}(0,1)$,
- *Prior 5:* $\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$.

6.4.4 Choice of prior distribution for between-study variance

The motivating examples in section 6.3 highlight that the choice of prior distribution for the between-study variance does affect the between-study correlation and the pooled results. Rather than observing the effect of the prior distribution for the correlation parameter, any potential observations of this choice of prior may be mistaken for those that actually arise from the choice of prior distribution for the between-study variances. Alternatively, it may be that the prior distribution for the between-study correlation may be inappropriate due to the choice of prior distribution for the between-study variance since these parameters are not independent.¹⁷ Thus, it is important to initially determine an appropriate prior distribution for the between-study variance parameters, τ_j^2 , to use in the simulation study. To do this, a univariate meta-analysis approach was applied first to the simulated datasets, where $\rho_{Wi}=\rho_B=0$ in model (6.1), in order to determine an appropriate prior distribution for τ_j . The following prior distributions for the between-study variance as introduced in section 6.3 were evaluated. No empirical prior distributions are used in this simulation study.

- $\tau_j \sim N(0,1) / (0,)$ and $\tau_j \sim N(0,2) / (0,)$,
- $1/\tau_j^2 \sim \text{gamma}(1,1)$ and $1/\tau_j^2 \sim \text{gamma}(0.1,0.1)$,
- $\tau_j \sim \text{uniform}(0,2)$,
- $\log(\tau_j^2) \sim \text{uniform}(-10, 1.386)$.

The results are shown in Appendix D2 and indicate that, for this simulation study where the true between-study standard deviations are 0.5, the $\tau_j \sim N(0,2) / (0,)$ prior distribution was most suitable giving appropriate posterior estimates and coverage for all parameters, including the between-study standard deviations, which had posterior means close to the true value of $\tau_j=0.5$. The gamma priors for $1/\tau_j^2$ were by far the poorest. Thus, in the bivariate simulations that follow, the simulations are repeated for both the $N(0,2) / (0,)$ prior distribution (a good prior) and, for contrast, a $\text{gamma}(0.1,0.1)$ prior distribution (a poor prior).

6.4.5 Summary of all the settings and prior distributions to be evaluated in the simulation study of bivariate meta-analysis

To summarise, the four following settings are assessed in this simulation study:

1. $\rho_{W_i}=0; \rho_B=0; \beta_1=0; \beta_2=2, \tau_1=0.5, \tau_2=0.5$ and 50 s_{i1}^2 and 50 s_{i2}^2
2. $\rho_{W_i}=0; \rho_B=0.8; \beta_1=0; \beta_2=2, \tau_1=0.5, \tau_2=0.5$ and 50 s_{i1}^2 and 50 s_{i2}^2
3. $\rho_{W_i}=0.8; \rho_B=0; \beta_1=0; \beta_2=2, \tau_1=0.5, \tau_2=0.5$ and 50 s_{i1}^2 and 50 s_{i2}^2
4. $\rho_{W_i}=0.8; \rho_B=0.8; \beta_1=0; \beta_2=2, \tau_1=0.5, \tau_2=0.5$ and 50 s_{i1}^2 and 50 s_{i2}^2

Within each setting, all 10 different combinations of the chosen prior distributions shown below in Table 6.5 for the between-study correlation and the between-study variance parameters are evaluated.

Table 6.5: All combinations of prior distributions for between-study correlation and between-study variance.

Combination	Prior distribution for ρ_B	Prior distribution for τ_j
(i)	$\rho_B \sim \text{uniform}(-1, 1)$	$\tau_j \sim N(0, 2) / (0,)$
(ii)	$z = \frac{1}{2} \log \left(\frac{1 + \rho_B}{1 - \rho_B} \right) \sim N(0, \text{sd}=0.4)$	$\tau_j \sim N(0, 2) / (0,)$
(iii)	$\frac{(\rho_B + 1)}{2} \sim \text{Beta}(1.5, 1.5)$	$\tau_j \sim N(0, 2) / (0,)$
(iv)	$\rho_B \sim \text{uniform}(0, 1)$	$\tau_j \sim N(0, 2) / (0,)$
(v)	$\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$	$\tau_j \sim N(0, 2) / (0,)$
(vi)	$\rho_B \sim \text{uniform}(-1, 1)$	$1/\tau_j^2 \sim \text{gamma}(0.1, 0.1)$
(vii)	$z = \frac{1}{2} \log \left(\frac{1 + \rho_B}{1 - \rho_B} \right) \sim N(0, \text{sd}=0.4)$	$1/\tau_j^2 \sim \text{gamma}(0.1, 0.1)$
(viii)	$\frac{(\rho_B + 1)}{2} \sim \text{Beta}(1.5, 1.5)$	$1/\tau_j^2 \sim \text{gamma}(0.1, 0.1)$
(ix)	$\rho_B \sim \text{uniform}(0, 1)$	$1/\tau_j^2 \sim \text{gamma}(0.1, 0.1)$
(x)	$\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$	$1/\tau_j^2 \sim \text{gamma}(0.1, 0.1)$

6.5 Results of the bivariate simulation study with complete data

6.5.1 Results when using prior distribution for between-study variance of $\tau_j \sim N(0, 2) / (0,)$

The results of the simulation study are now described, focussing firstly on a meta-analysis of 10 studies with complete data (i.e. both outcomes available in all studies). Table 6.6 displays the simulation results for setting 1 for the various prior distributions for the between-study correlation, where the prior distribution for τ_j is always $N(0, 2) / (0,)$. The equivalent results for settings 2, 3, and 4 are presented in Table 6.7, Table 6.8, and Table 6.9, respectively. The key findings are now discussed.

Table 6.6: Simulation results for 10 studies with complete data (setting 1). The within-study correlation, ρ_{wi} was zero and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/ median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/ median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/ median $\hat{\tau}_1$	Mean/ median $\hat{\tau}_2$	Mean/ median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.0	0.25
ρ_B -Uniform(-1,1)	-0.0020 (0.1955)	0.2185/ 0.2159	0.0382	956 (95.6)	0.4969	2.0011 (0.2198)	0.2616/ 0.2583	0.0483	966 (96.6)	0.4989	0.5006/ 0.4985	0.5344/ 0.5280	0.0070/ 0.0045	0.2483
Fisher z~ N(0,sd=0.4)	-0.0021 (0.1952)	0.2168/ 0.2136	0.0381	958 (95.8)	0.4966	2.0011 (0.2193)	0.2606/ 0.2569	0.0480	966 (96.6)	0.4989	0.4965/ 0.4953	0.5293/ 0.5281	0.0026/ -0.0012	0.2478
$(\rho_B+1)/2$ ~ Beta(1.5,1.5)	-0.0025 (0.1957)	0.2179/ 0.2149	0.0383	956 (95.6)	0.4965	2.0014 (0.2195)	0.2614/ 0.2600	0.0481	966 (96.6)	0.5000	0.4995/ 0.4996	0.5327/ 0.5295	0.0049/ 0.0055	0.2480
ρ_B -Uniform(0,1)	-0.0020 (0.1954)	0.2190/ 0.2160	0.0382	955 (95.5)	0.4963	2.0022 (0.2203)	0.2616/ 0.2579	0.0485	964 (96.4)	0.4994	0.5006/ 0.5050	0.5326/ 0.5302	0.4121/ 0.4114	0.2958
Logit(ρ_B)~ N(0,sd=0.8)	-0.0019 (0.1955)	0.2198/ 0.2175	0.0382	955 (95.5)	0.4955	2.0017 (0.2204)	0.2619/ 0.2573	0.0485	965 (96.5)	0.4991	0.5033/ 0.5041	0.5359/ 0.5285	0.4682/ 0.4738	0.3012

MSE is mean-square error; Crl is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.7: Simulation results for 10 studies with complete data (setting 2). The within-study correlation, ρ_{wi} was zero and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% Crls for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% Crls for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
$\rho_B \sim \text{Uniform}(-1,1)$	-0.0053 (0.1969)	0.2197/ 0.2153	0.0388	944 (94.4)	0.4913	1.9929 (0.2300)	0.2543/ 0.2533	0.0529	963 (96.3)	0.4945	0.5117/ 0.5081	0.5205/ 0.5163	0.4666/ 0.5047	0.3107
Fisher $z \sim N(0, \text{sd}=0.4)$	-0.0057 (0.1974)	0.2177/ 0.2150	0.0389	951 (95.1)	0.4920	1.9956 (0.2232)	0.2540/ 0.2506	0.0498	959 (95.9)	0.4950	0.5026/ 0.5021	0.5115/ 0.5117	0.2036/ 0.2046	0.2799
$(\rho_B+1)/2 \sim \text{Beta}(1.5,1.5)$	-0.0053 (0.1970)	0.2188/ 0.2154	0.0388	946 (94.6)	0.4916	1.9934 (0.2303)	0.2541/ 0.2514	0.0530	960 (96.0)	0.4939	0.5073/ 0.5059	0.5171/ 0.5129	0.3712/ 0.3869	0.2994
$\rho_B \sim \text{Uniform}(0,1)$	-0.0056 (0.1971)	0.2200/ 0.2159	0.0388	950 (95.0)	0.4916	1.9946 (0.2197)	0.2536/ 0.2496	0.0483	962 (96.2)	0.4940	0.5126/ 0.5093	0.5240/ 0.5172	0.6203/ 0.6172	0.3308
Logit(ρ_B) $\sim N(0, \text{sd}=0.8)$	-0.0052 (0.1971)	0.2177/ 0.2144	0.0388	945 (94.5)	0.4924	1.9936 (0.2285)	0.2520/ 0.2494	0.0522	964 (96.4)	0.4948	0.5057/ 0.5034	0.5176/ 0.5114	0.5432/ 0.5377	0.3201

MSE is mean-square error; Crl is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.8: Simulation results for 10 studies with complete data (setting 3). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.0	0.25
$\rho_B \sim \text{Uniform}(-1,1)$	0.0041 (0.1999)	0.2214/ 0.2202	0.0400	955 (95.5)	0.5046	1.9976 (0.2204)	0.2598/ 0.2585	0.0485	963 (96.3)	0.4987	0.5318/ 0.5343	0.5361/ 0.5405	-0.0352/ -0.0546	0.2637
Fisher z~ $N(0, \text{sd}=0.4)$	0.0047 (0.1999)	0.2189/ 0.2178	0.0399	951 (95.1)	0.5051	1.9990 (0.2191)	0.2578/ 0.2570	0.0480	964 (96.4)	0.4993	0.5257/ 0.5280	0.5298/ 0.5387	0.0017/ -0.0069	0.2666
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	0.0041 (0.1997)	0.2209/ 0.2195	0.0399	953 (95.3)	0.5052	1.9982 (0.2198)	0.2592/ 0.2592	0.0483	964 (96.4)	0.4981	0.5304/ 0.5331	0.5349/ 0.5409	-0.0138/ -0.0317	0.2651
$\rho_B \sim \text{Uniform}(0,1)$	0.0040 (0.2006)	0.2333/ 0.2337	0.0402	961 (96.1)	0.5039	1.9977 (0.2227)	0.2729/ 0.2743	0.0496	966 (96.6)	0.4991	0.5734/ 0.5853	0.5825/ 0.5954	0.3754/ 0.3441	0.3123
Logit(ρ_B)~ $N(0, \text{sd}=0.8)$	0.0039 (0.2004)	0.2355/ 0.2358	0.0401	964 (96.4)	0.5051	1.9980 (0.2236)	0.2757/ 0.2749	0.0499	966 (96.6)	0.4989	0.5843/ 0.5931	0.5952/ 0.6107	0.4560/ 0.4505	0.3215

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.9: Simulation results for 10 studies with complete data (setting 4). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)I(0,)$ and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% Crls for $\hat{\beta}_1$ including β_1 (%)	Mean prob $(\hat{\theta}_{i1new}>0)$	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% Crls for $\hat{\beta}_2$ including β_2 (%)	Mean prob $(\hat{\theta}_{i2new}>2)$	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob $(\hat{\theta}_{i1new}>0 \& \hat{\theta}_{i2new}>2)$
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
ρ_B-Uniform(-1,1)	-0.0091 (0.1850)	0.2081/ 0.2085	0.0343	956 (95.6)	0.4962	2.0009 (0.2029)	0.2319/ 0.2320	0.0411	963 (96.3)	0.4973	0.5134/ 0.5203	0.4965/ 0.5019	0.5160/ 0.5770	0.3279
Fisher z~N(0,sd=0.4)	-0.0088 (0.1848)	0.2055/ 0.2070	0.0342	955 (95.5)	0.4968	2.0002 (0.2055)	0.2329/ 0.2301	0.0422	960 (96.0)	0.4974	0.5047/ 0.5150	0.4813/ 0.4829	0.2363/ 0.2452	0.2915
$(\rho_B+1)/2$~Beta(1.5,1.5)	-0.0088 (0.1846)	0.2071/ 0.2081	0.0341	956 (95.6)	0.4963	2.0012 (0.2034)	0.2315/ 0.2301	0.0413	963 (96.3)	0.4975	0.5094/ 0.5179	0.4893/ 0.4933	0.4226/ 0.4582	0.3159
ρ_B-Uniform(0,1)	-0.0090 (0.1844)	0.2074/ 0.2067	0.0341	953 (95.3)	0.4958	1.9994 (0.2116)	0.2299/ 0.2283	0.0448	962 (96.2)	0.4975	0.5105/ 0.5153	0.5036/ 0.5056	0.6458/ 0.6562	0.3460
Logit(ρ_B)~N(0,sd=0.8)	-0.0092 (0.1844)	0.2045/ 0.2045	0.0341	954 (95.4)	0.4957	2.0012 (0.2026)	0.2285/ 0.2270	0.0410	959 (95.9)	0.4975	0.5021/ 0.5081	0.4891/ 0.4908	0.5545/ 0.5538	0.3312

MSE is mean-square error; Crl is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Impact of prior distribution for ρ_B on the posterior estimates for ρ_B

In all settings the prior distribution for ρ_B is informative of the posterior estimate of ρ_B , which is consistent with the finding in the motivating examples in section 6.3.3. For the prior distributions for ρ_B that are centred on zero, in all settings the average estimate of ρ_B across simulations is close to zero, regardless of the true value of ρ_B . However, when the prior distribution is centred on a different value (in this case, those priors centred on 0.5), the average posterior mean of ρ_B is close to 0.5. This is expected since there are only 10 studies per meta-analysis, which means there is little data to estimate this parameter and thus the posterior mean is similar to the prior mean. For example, in setting 1 (where $\rho_{WI}=\rho_B=0$) where $\rho_B\sim\text{uniform}(-1,1)$, the posterior mean of ρ_B across simulations is 0.007. When $\rho_B\sim\text{uniform}(0,1)$, the posterior mean of ρ_B across simulations is 0.412. A similar result is observed in settings 2-4. In setting 2 and setting 4, the true ρ_B is 0.8, however, none of the selected priors led to estimates of this parameter close to its true value. For example, in setting 4 ($\rho_{WI}=\rho_B=0.8$) where $\rho_B\sim\text{uniform}(0,1)$, the posterior mean of ρ_B is only 0.646.

Impact of prior distribution for ρ_B on marginal and joint inferences about treatment effect

The average marginal probability estimates are close to the true value of 0.5 for both treatment effects for all settings. However, the prior distribution for ρ_B has an impact on the *joint* probability estimates of interest. The average joint probabilities are slightly higher for the positive prior distributions for ρ_B even when the true between-study correlation is zero. In setting 1, where the true joint probability of interest is 0.25, the average estimate of this probability is 0.296 when $\rho_B\sim\text{uniform}(0,1)$. When $\rho_B\sim\text{uniform}(-1,1)$, the average estimate of this joint probability is 0.248. Also, since no prior distribution leads to estimates of the true between-study correlation close to 0.8 in settings 2 and 4, the joint probability is lower than the true value of 0.4. In setting 4 (Table 6.9), where $\rho_B\sim\text{uniform}(-1,1)$, the posterior average

joint probability is 0.328 compared to the true value of 0.4. With a uniform(0,1) prior distribution for ρ_B , the average joint probability is closest to 0.4 compared to the other prior distribution analyses with an estimate of 0.346. Given that the joint probability is underestimated this will be investigated further for larger numbers of studies per meta-analysis in Section 6.7.

Impact of prior distribution for ρ_B on pooled treatment effect estimates

Although the prior distribution for ρ_B affects the estimate of ρ_B , the posterior means of β_1 and β_2 across all simulations are very close to the true values. For example, in setting 1 (Table 6.6) when $\rho_B \sim \text{uniform}(-1,1)$, the mean β_1 is -0.002 compared to $\beta_1=0$, and the mean β_2 is 2.001 compared to the true $\beta_2=2$. The performance of the 95% credible intervals is also close to 95% for the pooled estimates.

Impact of prior distribution for ρ_B on between-study standard deviations

The average between-study standard deviation across simulations is estimated close to the true value for all settings ($\tau_j=0.5$). The estimates of between-study variability tend to be slightly larger when the estimate of between-study correlation is larger. For example, in setting 3 (Table 6.8), the posterior mean τ_1 across simulations is 0.584 when $\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$, and the posterior mean of ρ_B is 0.456. However, the posterior mean τ_1 is 0.532 when $\rho_B \sim \text{uniform}(-1,1)$ and the posterior mean of ρ_B is lower (-0.035).

6.5.2 Results when prior distribution for between-study variance is $1/\tau_j^2 \sim \text{gamma}(0.1,0.1)$

Table 6.10 displays the simulation results for setting 1 for the various prior distributions for the between-study correlation, where the prior distribution for $1/\tau_j^2$ is gamma(0.1,0.1). The

equivalent results for settings 2, 3, and 4 are presented in Table 6.11, Table 6.12, and Table 6.13, respectively. The key findings for these scenarios are now discussed.

The mean pooled treatment effect estimates remain unbiased for all prior distributions for ρ_B for settings 1 to 4. However, as observed in the univariate meta-analysis simulation study in section 6.4.4 and Appendix D2, the gamma prior distribution leads to estimates of between-study variance that are much larger than the true value for both outcomes. Therefore, the average standard errors of the pooled treatment effect estimates are much larger than those when $\tau_j \sim N(0,2)I(0,)$. Thus, the credible intervals for the pooled effect estimates are much wider, which means the number of credible intervals that contain the true value is 100% for all prior distributions in all settings.

The motivating examples highlighted that if the between-study variance is larger, this is likely to increase the estimate of ρ_B . This is evident again as the mean estimates of ρ_B are larger than in the corresponding analyses where the prior distribution of $\tau_j \sim N(0,2)I(0,)$ was used. For example, where $1/\tau_j^2 \sim \text{gamma}(0.1,0.1)$ and $\rho_B \sim \text{uniform}(-1,1)$ in setting 4 (Table 6.13, $\rho_W = \rho_B = 0.8$), the posterior mean ρ_B across simulations is 0.809. However, the posterior mean ρ_B is only 0.516 with the same prior for ρ_B , but where $\tau_j \sim N(0,2)I(0,)$ (Table 6.9). This is because the average estimates of τ_j are much higher with the gamma prior (posterior mean $\tau_1 = 1.954$, posterior mean $\tau_2 = 2.126$) compared to the normal prior (posterior mean $\tau_1 = 0.513$, posterior mean $\tau_2 = 0.497$).

Table 6.10: Simulation results for 10 studies with complete data (setting 1). The within-study correlation, ρ_{wi} was zero and the same for each study. The prior distribution for $1/\tau_j^2$ is gamma(0.1,0.1) and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.0	0.25
$\rho_B \sim \text{Uniform}(-1,1)$	-0.0019 (0.1976)	0.6792/ 0.6743	0.0390	1000 (100)	0.4994	1.9999 (0.2408)	0.7942/ 0.7850	0.0579	1000 (100)	0.4998	1.9125/ 1.9031	2.1467/ 2.1258	0.0146/ 0.1259	0.2513
Fisher z~ N(0,sd=0.4)	-0.0013 (0.1987)	0.6028/ 0.6002	0.0395	1000 (100)	0.5000	1.9982 (0.2526)	0.6976/ 0.6905	0.0637	1000 (100)	0.4993	1.7142/ 1.7053	1.8832/ 1.8686	0.0002/ 0.0060	0.2496
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.0022 (0.1974)	0.6498/ 0.6468	0.0389	1000 (100)	0.4995	1.9992 (0.2433)	0.7582/ 0.7518	0.0591	1000 (100)	0.4997	1.8363/ 1.8292	2.0499/ 2.0357	0.0130/ 0.0459	0.2511
$\rho_B \sim \text{Uniform}(0,1)$	-0.0021 (0.1970)	0.6777/ 0.6728	0.0388	1000 (100)	0.4995	2.0011 (0.2425)	0.7888/ 0.7791	0.0587	1000 (100)	0.5002	1.9048/ 1.8903	2.1274/ 2.1067	0.8429/ 0.8703	0.3982
Logit(ρ_B)~ N(0,sd=0.8)	-0.0021 (0.1979)	0.6194/ 0.6167	0.0391	1000 (100)	0.4994	1.9996 (0.2479)	0.7173/ 0.7101	0.0614	1000 (100)	0.5000	1.7614/ 1.7534	1.9455/ 1.9304	0.6419/ 0.6458	0.3534

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.11: Simulation results for 10 studies with complete data (setting 2). The within-study correlation, ρ_{Wi} was zero and the same for each study. The prior distribution for $1/\tau_j^2$ is gamma(0.1,0.1) and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
$\rho_B \sim \text{Uniform}(-1,1)$	-0.0068 (0.1976)	0.6851/ 0.6806	0.0390	1000 (100)	0.4985	1.9904 (0.2362)	0.7868/ 0.7785	0.0558	1000 (100)	0.4985	1.9386/ 1.9236	2.1519/ 2.1348	0.7863/ 0.9393	0.3781
Fisher z~ N(0,sd=0.4)	-0.0060 (0.1973)	0.6021/ 0.5993	0.0389	1000 (100)	0.4988	1.9901 (0.2559)	0.6936/ 0.6883	0.0655	1000 (100)	0.4978	1.7127/ 1.7041	1.8731/ 1.8598	0.1934/ 0.2050	0.2741
(ρ_B+1)/2~ Beta(1.5,1.5)	-0.0066 (0.1973)	0.6535/ 0.6505	0.0389	1000 (100)	0.4987	1.9899 (0.2396)	0.7537/ 0.7476	0.0575	1000 (100)	0.4984	1.8532/ 1.8431	2.0530/ 2.0387	0.7035/ 0.8550	0.3494
$\rho_B \sim \text{Uniform}(0,1)$	-0.0068 (0.1972)	0.6871/ 0.6836	0.0389	1000 (100)	0.4986	1.9908 (0.2330)	0.7854/ 0.7798	0.0543	1000 (100)	0.4984	1.9444/ 1.9313	2.1546/ 2.1360	0.9414/ 0.9548	0.4306
Logit(ρ_B)~ N(0,sd=0.8)	-0.0064 (0.1967)	0.6098/ 0.6068	0.0387	1000 (100)	0.4987	1.9899 (0.2454)	0.7029/ 0.6978	0.0603	1000 (100)	0.4981	1.7392/ 1.7293	1.9121/ 1.8973	0.6957/ 0.7003	0.3610

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.12: Simulation results for 10 studies with complete data (setting 3). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for $1/\tau_j^2$ is gamma(0.1,0.1) and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.0	0.25
$\rho_B \sim \text{Uniform}(-1,1)$	0.0024 (0.1999)	0.6813/ 0.6785	0.0399	1000 (100)	0.5008	1.9923 (0.2412)	0.7940/ 0.7822	0.0582	1000 (100)	0.4986	1.9255/ 1.9173	2.1574/ 2.1207	0.6047/ 0.8646	0.3419
Fisher z~ $N(0, \text{sd}=0.4)$	0.0031 (0.2006)	0.5985/ 0.5968	0.0402	1000 (100)	0.5010	1.9937 (0.2506)	0.6925/ 0.6886	0.0628	1000 (100)	0.4986	1.7049/ 1.6961	1.8709/ 1.8571	0.2378/ 0.2482	0.2811
$(\rho_B+1)/2 \sim$ $\text{Beta}(1.5,1.5)$	0.0029 (0.1997)	0.6511/ 0.6491	0.0399	1000 (100)	0.5008	1.9923 (0.2418)	0.7595/ 0.7513	0.0584	1000 (100)	0.4992	1.8459/ 1.8391	2.0611/ 2.0399	0.6045/ 0.7817	0.3338
$\rho_B \sim \text{Uniform}(0,1)$	0.0023 (0.2005)	0.6822/ 0.6770	0.0401	1000 (100)	0.5005	1.9931 (0.2446)	0.7980/ 0.7863	0.0598	1000 (100)	0.4988	1.9280/ 1.9166	2.1711/ 2.1359	0.8858/ 0.8982	0.4132
Logit(ρ_B)~ $N(0, \text{sd}=0.8)$	0.0029 (0.2013)	0.6077/ 0.6049	0.0405	1000 (100)	0.5010	1.9941 (0.2480)	0.7095/ 0.7037	0.0615	1000 (100)	0.4995	1.7370/ 1.7305	1.9309/ 1.9167	0.6942/ 0.6970	0.3637

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.13: Simulation results for 10 studies with complete data (setting 4). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for $1/\tau_j^2$ is gamma(0.1,0.1) and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
$\rho_B \sim \text{Uniform}(-1,1)$	-0.0102 (0.1914)	0.6866/ 0.6825	0.0367	1000 (100)	0.4983	1.9951 (0.2132)	0.7693/ 0.7611	0.0454	1000 (100)	0.4993	1.9535/ 1.9410	2.1255/ 2.1009	0.8089/ 0.9556	0.3865
Fisher z~ $N(0, \text{sd}=0.4)$	-0.0079 (0.1926)	0.5979/ 0.5958	0.0371	1000 (100)	0.4983	1.9938 (0.2389)	0.6868/ 0.6820	0.0571	1000 (100)	0.4987	1.7047/ 1.6997	1.8594/ 1.8453	0.1802/ 0.1980	0.2729
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.0097 (0.1908)	0.6534/ 0.6508	0.0365	1000 (100)	0.4981	1.9946/ 0.2190	0.7405/ 0.7350	0.0479	1000 (100)	0.4989	1.8598/ 1.8527	2.0335/ 2.0171	0.7342/ 0.8911	0.3574
$\rho_B \sim \text{Uniform}(0,1)$	-0.0106 (0.1902)	0.6886/ 0.6866	0.0362	1000 (100)	0.4976	1.9953 (0.2104)	0.7628/ 0.7553	0.0442	1000 (100)	0.4988	1.9617/ 1.9465	2.1200/ 2.0941	0.9544/ 0.9650	0.4382
Logit(ρ_B)~ $N(0, \text{sd}=0.8)$	-0.0093 (0.1930)	0.6050/ 0.6022	0.0373	1000 (100)	0.4979	1.9940 (0.2256)	0.6880/ 0.6832	0.0509	1000 (100)	0.4984	1.7295/ 1.7226	1.8837/ 1.8704	0.7063/ 0.7115	0.3620

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Also, when the gamma prior distribution is used for the between-study variance parameters, and thus the between-study variance is overestimated, there can be a huge upward bias in the posterior estimate of the between-study correlation even if the uniform(-1,1) prior distribution is selected for ρ_B . For example, in Table 6.13 (setting 3: $\rho_W=0.8$, $\rho_B=0$) and $\rho_B \sim \text{uniform}(-1,1)$ the mean posterior ρ_B is 0.605. This highlights the need to conduct sensitivity analyses for the prior distributions on both between-study variance and between-study correlation, especially where robust external information is unavailable *a priori* for these parameters.

The marginal probability estimates for θ_{i1new} and θ_{i2new} are close to the true values of 0.5 for all settings with the gamma prior. However, as expected, the joint probability($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$) is influenced by the estimate of correlation between the outcomes (the true joint probability is 0.4 if $\rho_B=0.8$, and the true joint probability is 0.25 if $\rho_B=0$). For example, in setting 4 ($\rho_B=0.8$, Table 6.13), the joint probability estimate is 0.273 for the analysis with a Fisher prior distribution for ρ_B , whereas this probability is 0.387 when $\rho_B \sim \text{uniform}(-1,1)$. These probabilities are also different to those estimated with variance prior $\tau_j \sim N(0,2)/(0,)$ in which the mean joint probability with the Fisher prior distribution for ρ_B is 0.292 and this joint probability estimate is 0.328 when $\rho_B \sim \text{uniform}(-1,1)$. These examples highlight that if the prior distribution is incorrectly specified for the between-study variances and between-study correlation, this can lead to incorrect joint inferences.

6.6 Impact of prior distribution for ρ_B when there is missing outcome data

The simulation studies in this chapter so far have assessed the various prior distributions for unknown parameters when the data is complete for both outcomes. One important finding is that although the prior for the between-study correlation is informative toward the between-

study correlation estimates, it has little impact on the pooled results of interest. However, these findings may be different if there is missing data for one outcome because the role of the between-study correlation becomes more important. In particular, a multivariate meta-analysis should reduce the MSE in the pooled estimate for the outcome with missing data compared to a univariate meta-analysis.^{17,200,213,215,225,232} But, an incorrect choice of prior for the between-study correlation may alter this. In order to explore this, a dataset with missing data (missing not at random) was generated by selectively removing the treatment effect estimate for outcome 1 (Y_{i1}) if it is larger than the unweighted mean of Y_{i1} within each set of 10 trials, i.e.:

$$\text{Remove } Y_{i1} \text{ if } Y_{i1} > \frac{1}{10} \sum_{i=1}^{10} Y_{i1}$$

On average, this will remove half of the treatment effect estimates and their standard errors for outcome 1 in each simulated dataset. Thus, it is of particular interest whether the prior distributions affect the outcome 1 results in this setting. The intention to selectively remove the data for outcome 1 in this way is to represent a situation similar to selective outcome reporting bias and to explore the impact of the choice of prior distribution in this setting. Both the $N(0,2)/(0,)$ prior distribution for τ_j and the $\text{gamma}(0.1,0.1)$ prior distribution for $1/\tau_j^2$ are used but only for settings 3 and 4 where there is within-study correlation of 0.8.

6.6.1 Results with missing data when prior distribution for τ_j is $N(0,2)/(0,)$

The simulation results for this missing data scenario are shown in Table 6.14 for setting 3 ($\beta_1=0, \beta_2=2, \tau_1=\tau_2=0.5, \rho_W=0.8, \rho_B=0$) and in Table 6.15 for setting 4 ($\beta_1=0, \beta_2=2, \tau_1=\tau_2=0.5, \rho_W=\rho_B=0.8$). As expected, due to the selective missingness, the average pooled treatment effect estimate for β_1 is consistently lower than the true value for all prior distributions, and in all settings. For example, in setting 4 (Table 6.15), the posterior mean β_1 is -0.432 (s.e. is

0.250) where $\rho_B \sim \text{uniform}(-1,1)$. However, if the posterior mean ρ_B is higher, the bias in the pooled posterior estimate of β_1 is lower. For example, in the same set of results, if $\rho_B \sim \text{uniform}(0,1)$ then the posterior mean ρ_B is 0.545 and the mean β_1 is -0.390. The estimate of treatment effect for outcome two remains unbiased across all settings as there is complete data for this outcome.

The posterior mean pooled β_1 across simulations is closer to the true value for all prior distributions for ρ_B compared to the univariate model. For example, in Table 6.15 ($\rho_w = \rho_B = 0.8$) where $\rho_B \sim \text{uniform}(-1,1)$, the posterior mean β_1 is -0.436 (s.e. is 0.258) whereas this estimate from the univariate analysis is -0.483 (s.e. is 0.251). The MSE of β_1 is lower in the bivariate model compared to the univariate model, for all prior distributions for ρ_B . In the same scenario, the MSE of β_1 is 0.249 when $\rho_B \sim \text{uniform}(-1,1)$ and the equivalent statistic for the univariate analysis is higher with an estimate of 0.296. Further, if a more appropriate prior distribution is used for ρ_B , the MSE is reduced compared to the other prior distributions. For example, in the same setting, where $\rho_B \sim \text{uniform}(0,1)$ (posterior mean ρ_B is 0.545), the MSE of β_1 is 0.210. This is smaller than when $\rho_B \sim \text{uniform}(-1,1)$ above, and it is also considerably lower than that from the univariate analysis. The more appropriate prior distributions for ρ_B also lead to increased numbers of 95% Crls that contain the true value. Where $\rho_B \sim \text{uniform}(0,1)$, the number of 95% Crls that contain β_1 is 73.5%, compared to 67.2% when $\rho_B \sim \text{uniform}(-1,1)$, and just 61.2% in the univariate analysis.

Table 6.14: Simulation results for 10 studies with missing data for outcome 1 (setting 3). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)/I(0,)$ and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/ median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/ median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/ median $\hat{\tau}_1$	Mean/ median $\hat{\tau}_2$	Mean/ median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.0	0.25
Univariate	-0.4826 (0.2513)	0.2820/ 0.2553	0.2960	612 (61.2)	0.1501	2.0009 (0.2185)	0.2593/ 0.2580	0.0477	965 (96.5)	0.4989	0.2890/ 0.2485	0.5249/ 0.5293	-	0.0749
$\rho_B \sim$Uniform(-1,1)	-0.4304 (0.2601)	0.2831/ 0.2593	0.2528	694 (69.4)	0.1892	2.0305 (0.2301)	0.2739/ 0.2701	0.0538	965 (96.5)	0.5195	0.3298/ 0.2866	0.6338/ 0.6240	-0.0201/ -0.0241	0.1077
Fisher z~ N(0,sd=0.4)	-0.4302 (0.2590)	0.2801/ 0.2578	0.2521	684 (68.4)	0.1876	2.0307 (0.2298)	0.2718/ 0.2664	0.0537	970 (97.0)	0.5199	0.3238/ 0.2820	0.6268/ 0.6136	-0.0056/ -0.0012	0.1063
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.4307 (0.2598)	0.2824/ 0.2572	0.2530	689 (68.9)	0.1877	2.0302 (0.2295)	0.2728/ 0.2688	0.0535	967 (96.7)	0.5206	0.3280/ 0.2846	0.6307/ 0.6185	-0.0133/ -0.0112	0.1068
$\rho_B \sim$Uniform(0,1)	-0.3977 (0.2626)	0.2799/ 0.2564	0.2271	702 (70.2)	0.2087	2.0309 (0.2302)	0.2808/ 0.2766	0.0539	972 (97.2)	0.5198	0.3525/ 0.2967	0.6568/ 0.6437	0.4674/ 0.4699	0.1509
Logit(ρ_B)~ N(0,sd=0.8)	-0.3975 (0.2620)	0.2795/ 0.2560	0.2266	718 (71.8)	0.2087	2.0311 (0.2299)	0.2800/ 0.2754	0.0538	971 (97.1)	0.5203	0.3550/ 0.3006	0.6562/ 0.6422	0.4898/ 0.4931	0.1517

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.15: Simulation results for 10 studies with missing data for outcome 1 (setting 4). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)/I(0,)$ and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
Univariate	-0.4826 (0.2513)	0.2820/ 0.2553	0.2960	612 (61.2)	0.1501	2.0009 (0.2185)	0.2593/ 0.2580	0.0477	965 (96.5)	0.4989	0.2890/ 0.2485	0.5249/ 0.5293	-	0.0749
$\rho_B \sim$Uniform(-1,1)	-0.4324 (0.2496)	0.2787/ 0.2535	0.2492	672 (67.2)	0.1800	2.0041 (0.2324)	0.2615/ 0.2596	0.0540	952 (95.2)	0.5031	0.3216/ 0.2801	0.6108/ 0.6085	0.1552/ 0.1531	0.1129
Fisher z~ N(0,sd=0.4)	-0.4438 (0.2506)	0.2727/ 0.2468	0.2597	641 (64.1)	0.1714	2.0085 (0.2243)	0.2602/ 0.2579	0.0503	956 (95.6)	0.5040	0.3126/ 0.2654	0.6048/ 0.6029	0.0497/ 0.0406	0.0987
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.4364 (0.2494)	0.2756/ 0.2506	0.2526	664 (66.4)	0.1765	2.0050 (0.2329)	0.2608/ 0.2584	0.0542	956 (95.6)	0.5040	0.3176/ 0.2719	0.6092/ 0.6068	0.1109/ 0.1052	0.1061
$\rho_B \sim$Uniform(0,1)	-0.3920 (0.2370)	0.2718/ 0.2480	0.2098	737 (73.7)	0.1948	2.0005 (0.2225)	0.2604/ 0.2582	0.0495	958 (95.8)	0.5000	0.3210/ 0.2819	0.6112/ 0.6101	0.5450/ 0.5452	0.1453
Logit(ρ_B)~ N(0,sd=0.8)	-0.3965 (0.2367)	0.2695/ 0.2480	0.2132	735 (73.5)	0.1918	2.0017 (0.2226)	0.2589/ 0.2569	0.0495	957 (95.7)	0.5004	0.3161/ 0.2762	0.6078/ 0.6073	0.5132/ 0.5132	0.1397

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Although there are reductions in the standard errors and MSE of the average treatment effects with the bivariate meta-analysis model compared to the univariate model, the prior distribution for the between-study correlation is even more influential on its posterior distribution compared to the complete case simulation study. This is due to fewer studies that provide both outcomes. Therefore, in these simulation settings all prior distributions lead to underestimated between-study correlation. In Table 6.15, where the true ρ_B is 0.8, and for $\rho_B \sim \text{uniform}(-1,1)$, the mean $\hat{\rho}_B$ is 0.155. This has an impact on the joint probability estimates. For example, for the same scenario, the joint probability estimate that $\hat{\theta}_{i1new} > 0$ and $\hat{\theta}_{i2new} > 2$ is 0.113 (true joint probability is 0.4). For $\rho_B \sim \text{uniform}(0,1)$, the posterior mean ρ_B across simulations is 0.545 and, as a result, the joint probability is slightly higher with an estimate of 0.145.

6.6.2 Results with missing data when prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$

The results of the missing data scenario when the prior distribution for $1/\tau_j^2$ is $\text{gamma}(0.1,0.1)$ are shown in Table 6.16 ($\rho_W=0.8, \rho_B=0$) and Table 6.17 ($\rho_W=\rho_B=0.8$). The results here are similar to those from the complete data simulation study with this prior distribution for the between-study variances. The posterior estimates of τ_j are hugely overestimated, which means that the posterior mean estimates of ρ_B are higher for all prior distributions for ρ_B (compared to the $N(0,2)/(0,)$ prior for τ_j). This increases the average standard deviation of the posterior distribution for β_1 for the bivariate analyses compared to the univariate analysis and thus the number of 95% CrIs that contain the true β_1 is 100%. For example, in Table 6.17 ($\rho_W=\rho_B=0.8$) where $\rho_B \sim \text{uniform}(-1,1)$ the posterior mean of τ_1 is 2.585 ($\tau_1=0.5$) and the mean standard deviation of the posterior distribution for β_1 is 1.421 compared to 0.282 in a univariate analysis.

Table 6.16: Simulation results for 10 studies with missing data for outcome 1 (setting 3). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for $1/\tau_i^2$ is gamma(0.1,0.1) and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.0	0.25
Univariate	-0.4826 (0.2513)	0.2820/ 0.2553	0.2960	612 (61.2)	0.2865	2.0009 (0.2185)	0.2593/ 0.2580	0.0477	965 (96.5)	0.4993	0.2890/ 0.2485	0.5249/ 0.5293	-	0.1430
$\rho_B \sim \text{Uniform}(-1,1)$	-0.4368 (0.2785)	1.4114/ 1.2527	0.2682	1000 (100)	0.4373	2.0211 (0.2718)	0.7181/ 0.7062	0.0742	1000 (100)	0.5035	2.5701/ 2.4859	1.9659/ 1.9342	0.3142/ 0.3422	0.2525
Fisher z~ $N(0, \text{sd}=0.4)$	-0.4646 (0.2588)	1.5083/ 1.3434	0.2828	1000 (100)	0.4323	2.0138 (0.2767)	0.6731/ 0.6644	0.0767	1000 (100)	0.5028	2.5163/ 2.4345	1.8468/ 1.8224	0.0359/ 0.0318	0.2237
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.4468 (0.2660)	1.4481/ 1.2749	0.2703	1000 (100)	0.4355	2.0179 (0.2736)	0.6957/ 0.6862	0.0751	1000 (100)	0.5030	2.5451/ 2.4613	1.9088/ 1.8792	0.2064/ 0.1732	0.2418
$\rho_B \sim \text{Uniform}(0,1)$	-0.2671 (0.3660)	1.3470/ 1.1876	0.2052	1000 (100)	0.4561	2.0225 (0.2711)	0.7229/ 0.7104	0.0740	1000 (100)	0.5039	2.5514/ 2.4701	1.9784/ 1.9436	0.8053/ 0.8200	0.3634
Logit(ρ_B)~ $N(0, \text{sd}=0.8)$	-0.3009 (0.3316)	1.4180/ 1.2592	0.2004	1000 (100)	0.4512	2.0167 (0.2733)	0.6807/ 0.6718	0.0749	1000 (100)	0.5031	2.5059/ 2.4249	1.8716/ 1.8468	0.5888/ 0.5854	0.3194

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.17: Simulation results for 10 studies with missing data for outcome 1 (setting 4). The within-study correlation, ρ_{wi} was 0.8 and the same for each study. The prior distribution for $1/\tau_j^2$ is gamma(0.1,0.1) and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
Univariate	-0.4826 (0.2513)	0.2820/ 0.2553	0.2960	612 (61.2)	0.2865	2.0009 (0.2185)	0.2593/ 0.2580	0.0477	965 (96.5)	0.4993	0.2890/ 0.2485	0.5249/ 0.5293	-	0.1430
$\rho_B \sim \text{Uniform}(-1,1)$	-0.4364 (0.2580)	1.4210/ 1.2413	0.2570	1000 (100)	0.4366	1.9743 (0.2614)	0.7128/ 0.7042	0.0689	1000 (100)	0.4956	2.5848/ 2.4855	1.9563/ 1.9358	0.2764/ 0.3095	0.2471
Fisher z~ N(0,sd=0.4)	-0.4764 (0.2539)	1.5122/ 1.3402	0.2914	1000 (100)	0.4312	1.9844 (0.2672)	0.6710/ 0.6641	0.0716	1000 (100)	0.4976	2.5190/ 2.4361	1.8425/ 1.8184	0.0154/ 0.0130	0.2180
$(\rho_B+1)/2 \sim \text{Beta}(1.5,1.5)$	-0.4560 (0.2505)	1.4509/ 1.2794	0.2706	1000 (100)	0.4335	1.9788 (0.2640)	0.6922/ 0.6836	0.0701	1000 (100)	0.4962	2.5531/ 2.4649	1.9019/ 1.8795	0.1399/ 0.1096	0.2325
$\rho_B \sim \text{Uniform}(0,1)$	-0.2529 (0.3231)	1.3604/ 1.1996	0.1682	1000 (100)	0.4579	1.9695 (0.2593)	0.7109/ 0.7032	0.0681	1000 (100)	0.4946	2.5760/ 2.4965	1.9578/ 1.9336	0.8331/ 0.8557	0.3675
Logit(ρ_B)~ N(0,sd=0.8)	-0.3085 (0.2914)	1.4410/ 1.2724	0.1800	1000 (100)	0.4496	1.9792 (0.2630)	0.6752/ 0.6674	0.0695	1000 (100)	0.4959	2.5198/ 2.4319	1.8607/ 1.8359	0.5869/ 0.5856	0.3148

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

6.7 The number of trials per meta-analysis and the influence of the prior distribution for the between-study correlation on its posterior distribution

The results from the simulations so far illustrate that the prior distribution for the between-study correlation is highly informative towards its posterior distribution and this is largely because there are only 10 studies per meta-analysis to estimate this parameter (or five studies per meta-analysis in the missing data scenarios). In settings 2 and 4, where there is strong true between-study correlation ($\rho_B=0.8$), most of the prior distributions for ρ_B result in this parameter being underestimated. This is a problem because the estimate of between-study correlation affects the joint inferences, such as joint probability estimates and therefore it is important to estimate this parameter accurately. Thus, the following section investigates the number of studies per meta-analysis that result in a more accurate estimate of the between-study correlation. This simulation study is only investigated for complete case data in setting 4, where $\beta_1=0$, $\beta_2=2$, $\tau_1=\tau_2=0.5$, $\rho_W=0.8$ and $\rho_B=0.8$, and a $N(0,2)/(0,)$ prior distribution is selected again for τ_j . The numbers of studies per meta-analysis which are explored are 25 and 50.

Table 6.18: Simulation results for 25 studies with complete data (setting 4). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)/0$ and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
Univariate	-0.0053 (0.1367)	0.1404/ 0.1401	0.0187	951 (95.1)	0.4956	1.9943 (0.1487)	0.1522/ 0.1520	0.0221	945 (94.5)	0.4937	0.5043/ 0.5043	0.4787/ 0.4860	-	0.2447
$\rho_B \sim \text{Uniform}(-1,1)$	-0.0057 (0.1217)	0.1294/ 0.1287	0.0148	955 (95.5)	0.4956	1.9981 (0.1324)	0.1367/ 0.1367	0.0175	945 (94.5)	0.4963	0.5072/ 0.5052	0.4873/ 0.4918	0.6660/ 0.7161	0.3547
Fisher z~ $N(0, \text{sd}=0.4)$	-0.0086 (0.1234)	0.1284/ 0.1282	0.0153	952 (95.2)	0.4955	1.9982 (0.1341)	0.1366/ 0.1364	0.0180	945 (94.5)	0.4958	0.4958/ 0.4940	0.4709/ 0.4789	0.4208/ 0.4432	0.3134
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.0055 (0.1218)	0.1293/ 0.1287	0.0149	954 (95.4)	0.4955	1.9980 (0.1326)	0.1368/ 0.1369	0.0176	944 (94.4)	0.4961	0.5040/ 0.5018	0.4832/ 0.4893	0.6020/ 0.6432	0.3428
$\rho_B \sim \text{Uniform}(0,1)$	-0.0057 (0.1213)	0.1293/ 0.1287	0.0147	953 (95.3)	0.4954	1.9982 (0.1324)	0.1365/ 0.1364	0.0175	945 (94.5)	0.4969	0.5069/ 0.5028	0.4883/ 0.4931	0.6934/ 0.7254	0.3608
Logit(ρ_B)~ $N(0, \text{sd}=0.8)$	-0.0074 (0.1231)	0.1280/ 0.1273	0.0152	946 (94.6)	0.4947	2.0000 (0.1329)	0.1358/ 0.1360	0.0176	947 (94.7)	0.4960	0.4980/ 0.4945	0.4767/ 0.4833	0.5971/ 0.6025	0.3412

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

Table 6.19: Simulation results for 50 studies with complete data (setting 4). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. The prior distribution for τ_j is $N(0,2)/0$ and for β_j is $N(0,1000^2)$.

Prior for ρ_B	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
Univariate	-0.0032 (0.1013)	0.1053/ 0.1056	0.0103	956 (95.6)	0.4965	2.000 (0.1070)	0.1080/ 0.1073	0.0114	951 (95.1)	0.4994	0.4993/ 0.4989	0.4855/ 0.4850	-	0.2480
$\rho_B \sim \text{Uniform}(-1,1)$	-0.0027 (0.0888)	0.0947/ 0.0943	0.0079	966 (96.6)	0.4981	2.0010 (0.0936)	0.0960/ 0.0956	0.0088	960 (96.0)	0.5019	0.5013/ 0.5007	0.4916/ 0.4921	0.7344/ 0.7617	0.3770
Fisher z~ $N(0, \text{sd}=0.4)$	-0.0029 (0.0892)	0.0944/ 0.0942	0.0080	965 (96.5)	0.4975	2.0007 (0.0938)	0.0961/ 0.0957	0.0088	960 (96.0)	0.5009	0.4917/ 0.4909	0.4807/ 0.4800	0.5538/ 0.5754	0.3394
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.0028 (0.0889)	0.0945/ 0.0944	0.0079	965 (96.5)	0.4974	2.0009 0.0936	0.0961/ 0.0956	0.0087	957 (95.7)	0.5010	0.4987/ 0.4979	0.4887/ 0.4885	0.6927/ 0.7194	0.3670
$\rho_B \sim \text{Uniform}(0,1)$	-0.0029 (0.0892)	0.0946/ 0.0946	0.0080	964 (96.4)	0.4978	1.9995 (0.1062)	0.0959/ 0.0958	0.0113	958 (95.8)	0.5013	0.5012/ 0.5012	0.4925/ 0.4926	0.7384/ 0.7622	0.3776
Logit(ρ_B)~ $N(0, \text{sd}=0.8)$	-0.0030 (0.0889)	0.0945/ 0.0943	0.0079	966 (96.6)	0.4970	2.0007 (0.0935)	0.0959/ 0.0956	0.0087	958 (95.8)	0.5003	0.4949/ 0.4936	0.4842/ 0.4848	0.6473/ 0.6582	0.3558

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

As expected, as the number of studies per meta-analysis increases, the posterior mean of ρ_B is closer to the true value. For example, when there are 10 studies (Table 6.9) and $\rho_B \sim \text{uniform}(-1,1)$, the posterior mean ρ_B across simulations is 0.516, compared to an estimate of 0.734 when there are 50 studies. Interestingly, the average posterior estimate of ρ_B is still not especially close to the true value of 0.8 for any of the priors for ρ_B . Therefore, the choice of prior distribution for this parameter is very important because it is still influential on the posterior estimate of ρ_B even when there are 50 studies. Also, this improvement in the estimate of ρ_B has an effect on the joint probability estimates, but they are still lower than the true value of 0.4 for all prior distributions for ρ_B even when $N=50$.

Also, as the number of studies per meta-analysis increases, the MSE and average standard errors for the two pooled treatment effect estimates reduce. For example, when there are 50 studies and $\rho_B \sim \text{uniform}(-1,1)$, the MSE of β_1 is 0.008, compared to 0.038 when there are 10 studies. Also, the mean standard error of the posterior distribution of β_1 is 0.095 with 50 studies compared to 0.219 with 10 studies.

6.8 What if the between-study correlation is known?

There is still a downward bias in the joint probability estimates even with large numbers of studies. This is, in part, due to the estimate of the between-study correlation, which is still not estimated particularly well even when there are 50 studies. However, it is not clear whether there may be another influence on the joint probability estimates. Therefore, setting 4 was repeated for a number of scenarios where the true between-study correlation is assumed known and equal to 0.8, but the other parameters are either fixed or assumed unknown (with complete data and 10 studies). Therefore, the following scenarios are investigated:

1. $\rho_B=0.8, \tau \sim N(0,2)I(0,), \beta_j \sim N(0,1000^2)$

2. $\rho_B=0.8, \tau_f=0.5, \beta_f \sim N(0, 1000^2)$
3. $\rho_B=0.8, \tau_f \sim N(0, 2) | I(0, \cdot), \beta_1=0, \beta_2=2$
4. $\rho_B=0.8, \tau_f=0.5, \beta_1=0, \beta_2=2.$

The results are shown in Table 6.20. For scenario 1, the joint probability is still underestimated with an average estimate across simulations of 0.372 compared to 0.4. Given that the between-study correlation is fixed at the true value, it appears that the uncertainty in the parameter estimates has an effect on the joint inferences. This is confirmed in scenario 4 where the mean posterior joint probability estimate is 0.397. Therefore, it would seem that it is not just the downward bias of the between-study correlation that has been affecting the downward bias in the joint probability inferences throughout the simulations in this chapter. In other words, unless the meta-analysis has a very large number of studies, the uncertainty in the estimates of the pooled treatment effects, the between-study variances and the between-study correlation, all lead to joint probabilistic inferences that are lower than expected if these parameters were known.

This finding can perhaps be considered comparable to the use of the t -distribution for the derivation of prediction intervals for θ_{new} by Higgins et al. in a frequentist framework.¹⁹⁰ Here, the t -distribution is used instead of the normal distribution to account for the uncertainty in the between-study variance. This can be extended to a bivariate setting. If 2,000,000 samples of x and y are drawn from a bivariate t -distribution (with 8 degrees of freedom since the number of trials is 10) with means zero and two, respectively, variances equal to 0.25, and correlation equal to 0.8, then the joint probability that $x > 0$ and $y > 2$ is just 0.366. This is similar to the estimate of 0.372 for scenario 1 in Table 6.20. The joint probability is only equal to 0.4 when the bivariate normal distribution is assumed. If 2,000,000 x and y are sampled from the bivariate normal distribution, with the same parameter values as those used above, then the resulting probability is very close to 0.4.

Table 6.20: Simulation results for fixed values of parameters with 10 studies with complete data (setting 4). The within-study correlation, ρ_{Wi} was 0.8 and the same for each study. In all analyses, ρ_B is fixed equal to 0.8.

Scenario	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i1new}>0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i2new}>2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$	Mean/median $\hat{\rho}_B$	Mean prob ($\hat{\theta}_{i1new}>0$ & $\hat{\theta}_{i2new}>2$)
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
1. $\rho_B=0.8$, $\tau_j \sim N(0,2)/(0,)$, $\beta_j \sim N(0,1000^2)$	-0.0099 (0.1840)	0.2085/ 0.2085	0.0339	954 (95.4)	0.4958	2.0018 (0.2022)	0.2286/ 0.2271	0.0408	961 (96.1)	0.4994	0.5163/ 0.5226	0.5122/ 0.5170	0.8/0.8	0.3715
2. $\rho_B=0.8$, $\tau_j=0.5$, $\beta_j \sim N(0,1000^2)$	-0.0097 (0.1833)	0.1836/ 0.1836	0.0337	959 (95.9)	0.4927	1.9991 (0.1972)	0.1955/ 0.1955	0.0388	951 (95.1)	0.4993	0.5/0.5	0.5/0.5	0.8/0.8	0.3820
3. $\rho_B=0.8$, $\beta_1=0$, $\beta_2=2$, $\tau_j \sim N(0,2)/(0,)$	0	0	0	1000	0.5004	2	0	0	1000	0.5001	0.5148/ 0.5174	0.5168/ 0.5261	0.8/0.8	0.3978
4. $\rho_B=0.8$, $\beta_1=0$, $\beta_2=2$, $\tau_j=0.5$	0	0	0	1000	0.4999	2	0	0	1000	0.4994	0.5/0.5	0.5/0.5	0.8/0.8	0.3972

MSE is mean-square error; CrI is credible interval; s.e. is standard error; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

6.9 Discussion

This chapter explored the choice of prior distribution for the between-study correlation parameter in a bivariate random-effects meta-analysis model within a simulation study. This extends the univariate and multivariate meta-analysis of phase II trials in Chapter 4 and Chapter 5. The key findings of the work are summarized in Figure 6.2, which are now discussed.

6.9.1 Overview of key findings

The prior distribution for the between-study correlation is informative of its posterior distribution.

As expected, the prior distribution for the between-study correlation is informative of its posterior distribution, especially when there are few studies in the meta-analysis, or missing outcome data. Even with large numbers of studies, such as 50 studies in Table 6.19, the prior distribution still influences the posterior estimates. In this scenario, if a uniform(-1,1) prior distribution is used, the mean posterior estimate of ρ_B is 0.734, whereas this estimate is 0.554 with a Fisher transformation prior for ρ_B (true value is 0.8 in this scenario). Therefore, the prior distribution for the correlation must be selected carefully.

Figure 6.2: Key findings.

- The prior distribution for the between-study correlation is informative of the posterior distribution of this parameter, in particular when there are few studies in a meta-analysis.
- If the data is complete, the prior distribution for between-study correlation has a minimal impact on the pooled treatment effect estimates. However, if the posterior estimate of the between-study correlation is incorrect, this can lead to incorrect joint inferences, such as the joint probability estimates.
- If there is missing data, the choice of the prior distribution for the between-study correlation has more influence on its posterior distribution as the number of studies with both outcomes is reduced. In particular, if the true between-study correlation is large and there are few studies, several vague prior distributions underestimate this parameter.
- The choice of prior distribution for the between-study variance impacts the posterior distribution of the between-study correlation. Further, if the between-study variance posterior estimate is upwardly biased this increases the estimate of the between-study correlation.
- If there is selectively missing data for outcome j , a Bayesian bivariate meta-analysis reduces the bias in the posterior mean estimate of β_j and increases the number of 95% CrIs that contain the true parameter value, compared to a univariate meta-analysis, for all prior distributions for the between-study correlation. What's more, a narrower prior distribution that is centred closer to the true value for the correlation parameter further reduces the bias in the posterior mean estimate of β_j and further increases the number of 95% CrIs that contain the true value.
- Even when the true between-study correlation is zero, inappropriate prior distributions (e.g. gamma) for between-study variances lead to dramatically upwardly biased posterior estimates of between-study correlation, and thus a potentially misleading impact of between-study correlation.
- Even if the between-study correlation is estimated perfectly, the joint probabilities are still affected by the uncertainty in the other parameters. In the scenarios in this simulation study, the joint predictions are downwardly biased. This is akin to joint probabilities being lower from a bivariate t-distribution compared to a bivariate normal distribution.

There is minimal impact of the between-study correlation on the estimates of pooled treatment effects with complete data.

A key finding from all the analyses is that, if the data is complete and an appropriate prior distribution is chosen for the between-study variance, the choice of prior distribution for the between-study correlation, from a range of various vague prior distributions, has little impact on the pooled treatment effect estimates and their standard errors. This result remains true even when the posterior distribution for the between-study correlation is centred far from the true value of the correlation. The scenario in Table 6.7 is used for illustration here where there is complete data and the true values for the correlation parameters are $\rho_W=0$, $\rho_B=0.8$, and $\tau_j \sim N(0,2) \parallel (0,.)$. If $\rho_B \sim \text{uniform}(-1,1)$, the posterior mean ρ_B across simulations is 0.467, the mean β_1 across simulations is -0.005 (s.e. of mean 0.197) (true β_1 is zero) and the posterior mean estimate of β_2 is 1.993 (s.e. of mean 0.230) (true β_2 is two). In contrast, if $\rho_B \sim \text{uniform}(0,1)$, the mean estimate of ρ_B is much higher (and closer to the true value of 0.8) at 0.620, but the mean treatment effect estimates are almost unchanged with the posterior mean β_1 equal to -0.006 (s.e. of mean 0.197) and the mean β_2 equal to 1.995 (s.e. of mean 0.220). Therefore, in the setting with complete case data there is very little impact of any of the prior distributions for the between-study correlation (used in this chapter) on the pooled treatment effect estimates, and therefore there is only minimal impact on any decision-making on the basis of these results.

The choice of prior distribution for the between-study correlation is more important if there is missing data.

If there is missing data, the prior distribution for the between-study correlation must be selected more carefully. The prior distribution is more informative of the posterior distribution for this parameter since there is less data to estimate the between-study correlation with less trials reporting both outcomes. Also, a prior distribution that is centred closer to the true value

of the correlation will reduce the average standard error and the MSE of the pooled treatment effect estimate for the outcome with missing data compared to other prior distributions. For example, observe the missing data scenario given in Table 6.15, where the true parameter values are $\beta_1=0$, $\beta_2=2$, $\rho_W=0.8$, $\rho_B=0.8$, $\tau_1=\tau_2=0.5$ and outcome one is selectively missing for half the studies (Y_{ij} is missing if it is greater than the un-weighted mean of the Y_{ij} within the meta-analysis). If $\rho_B \sim \text{uniform}(0,1)$, the posterior mean ρ_B is 0.545 and the mean standard error of β_1 is 0.237 and the MSE of β_1 is 0.210. However, if $\rho_B \sim \text{uniform}(-1,1)$, which is centred on zero, the mean ρ_B is 0.155 and so the mean standard error of β_1 is larger (0.250) and the MSE of β_1 is larger (0.249).

The difference in the posterior estimate of the between-study correlation by the choice of prior distribution also affects the joint probability estimates since the mean joint probability across simulations is 0.113 with a $\text{uniform}(-1,1)$ prior distribution for ρ_B , but this is higher with a $\text{uniform}(0,1)$ prior since the mean joint probability estimate is 0.145.

Additionally, if the prior distribution for the between-study variance parameters is inappropriate within a missing data scenario, the posterior estimate of ρ_B across simulations can be hugely overestimated when the posterior estimates of τ_j are overestimated. As a result, the mean standard error of β_j is much greater, which increases the width of the 95% CrIs from the posterior distributions of β_j and the performance of the 95% CrI becomes 100%. For example, in Table 6.16 ($\rho_B=0$) when $1/\tau_j^2 \sim \text{gamma}(0.1,0.1)$ and $\rho_B \sim \text{uniform}(-1,1)$, the mean τ_j across simulations is 2.570, the mean ρ_B is greater than zero with an estimate of 0.314 and as a result, the mean standard error of β_1 is 1.411 and so the number of 95% CrI that contain the true β_1 is 100%.

The choice of prior distribution for τ_j impacts the posterior distribution of ρ_B .

The choice of prior distribution for the between-study correlation and the between-study variance are not independent, and therefore wise choices must be made for both parameters in the bivariate meta-analysis model. Where Chapter 4 illustrates how important the prior distribution is for the between-study variance in a univariate meta-analysis, the simulation studies in this chapter reveal this is perhaps even truer for a bivariate meta-analysis. If an inappropriate prior distribution is selected for the between-study variance, this not only has an impact on the posterior estimates of the between-study variances themselves, but also on the posterior estimate of between-study correlation and the standard error of the pooled treatment effect estimates.

For illustration, the scenario in Table 6.12 is used where the data is complete, the true $\rho_{WI}=0.8$, $\rho_B=0$, and the prior distribution for the between-study variation is $1/\tau_j^2 \sim \text{gamma}(0.1, 0.1)$, which overestimates the true $\tau_j=0.5$. In this scenario, if $\rho_B \sim \text{uniform}(-1, 1)$, the mean estimates of between-study standard deviation are too large since the posterior mean τ_1 is 1.926, the mean τ_2 is 2.157, and as a result, the mean posterior estimate of between-study correlation across simulations is 0.605 (true $\rho_B=0$). The average standard errors of the pooled treatment effect estimates are larger with the gamma prior for the between-study variance compared to those in the corresponding analyses where $\tau_j \sim N(0, 2)/(0,)$, and this increases the width of the 95% credible intervals around the pooled treatment effect estimates. In the same scenario with the gamma prior, the mean standard error of β_1 is 0.681 and the mean standard error of β_2 is 0.794. In the corresponding scenario with a normal prior for τ_j , the mean standard error of β_1 is 0.221 and the mean standard error of β_2 is 0.260. However, even though all these parameters are poorly estimated with a gamma prior for the between-study heterogeneity, the mean pooled treatment effect estimates across simulations are unbiased with complete data. In the same scenario in Table

6.12 with a gamma prior for the between-study variance parameters, the posterior mean of β_1 is 0.0024 (s.e. of mean 0.200) and the posterior mean of β_2 is 1.992 (s.e. of mean 0.241).

The joint probability is affected by the uncertainty in the other parameters as well as the estimate of the between-study correlation.

The results from all scenarios show that the posterior estimate of the between-study correlation affects the estimate of the joint probability. However, in addition, the uncertainty in the other parameters also affects the joint probability, which is observed when the between-study correlation is fixed at 0.8. For the analysis that estimates ρ_B perfectly (Scenario 1 in Table 6.20), the average posterior estimate of the between-study correlation is only 0.372, compared to the true value of 0.4. If all parameters are fixed at their true values, the average joint probability is very close to the true value (0.398). This can be considered akin to the probability being lower from a bivariate t -distribution compared to a bivariate normal distribution. The t -distribution is used to estimate prediction intervals in a frequentist meta-analysis to account for uncertainty in the between-study variance estimates.

6.9.2 Recommendations for the choice of prior distributions in future bivariate meta-analyses

The key findings have led to the derivation of a number of recommendations for future Bayesian bivariate meta-analyses, which are shown in Figure 6.3 and are now discussed.

Figure 6.3: Key recommendations.

When adopting a Bayesian approach for multivariate meta-analysis, researchers should consider the following:

- The prior distributions for between-study variances need to be chosen sensibly as they impact on parameter estimates including the between-study correlation. For this purpose, empirical prior distributions may be useful, such as those by Turner et al.³⁶ and Rhodes et al.³⁸
- The prior distribution for between-study correlation(s) also need to be chosen sensibly. A uniform(-1,1) prior distribution for ρ_B is not always vague and thus should not be routinely used without due thought, especially when the number of studies is small.
- Further research should establish empirical prior distributions for between-study correlations, if possible, for particular pairs of outcomes across a range of disease fields.
- Without empirical evidence, clinical or biological rationale might provide external evidence to inform the prior distribution for between-study correlations (for example, a uniform(0,1) prior distribution could be specified if only positive values are plausible).
- Sensitivity analysis for the choice of prior distributions is needed, especially when external evidence to inform the prior distributions is not available.

If possible, previous evidence should be considered to derive prior distributions for between-study variances, such as those previously derived by Turner et al.,³⁶ and Rhodes et al.,³⁸ due to the dependency of the between-study variance-covariance parameters. Additionally, the prior distribution for the between-study correlation needs to be chosen sensibly. Some of the scenarios considered in this simulation study revealed that the uniform(-1,1) prior distribution for ρ_B is not as “vague” as it is sometimes considered. This is largely due to the dependency of ρ_B and τ_j on each other and the need for an appropriate prior distribution for τ_j . Therefore, the uniform(-1,1) prior distribution for ρ_B should not be used without careful thought, in particular if there are small numbers of studies.

It would be useful to derive empirical prior distributions for the between-study correlation for future studies. However, this is likely to be difficult until multivariate meta-analysis becomes routine practice for the meta-analysis of multiple outcomes and the practicalities of this challenge are discussed further in Chapter 7. Until empirical prior distributions have been derived, clinical guidance should be sought to develop plausible prior distributions, such as a uniform(0,1) prior distribution that only allows positive values. If there is no previous evidence to inform the prior distribution then the uniform(-1,1) prior distribution is recommended, but a sensitivity analysis for this choice is important, especially when there are few studies.

6.9.3 Limitations

This chapter presents several important findings of the impact of the choice of the prior distribution for the between-study correlation. However, the findings may only apply to the specific settings in which the simulation studies have been conducted. Whilst there are some key findings that have been described above, there are, of course, numerous other prior distributions for the between-study correlation that could be used in a bivariate random-effects meta-analysis, but were not considered in this chapter; in particular prior distributions with a very narrow range of possible values.

Also, the simulation is specifically for bivariate meta-analysis, but of course there may be more than two correlated outcomes. In this case, there are several between-study correlation parameters that require prior distributions.

Importantly, this simulation study only considered one set of true values for the between-study standard deviations ($\tau=0.5$). For these true values, the normal(0,2)/(0,) prior distribution was most appropriate out of those considered. However, this would not necessarily be the case for different values of the between-study standard deviations. Also, the relative size of the within-study and between-study variances alters the impact of the

between-study variances and therefore different relationships of these variances need to be explored.

6.10 Conclusion

The motivating examples and the simulation study in this chapter reveal that the choice of prior distribution for the between-study correlation in a Bayesian bivariate random-effects meta-analysis is important and must be chosen with caution, and in conjunction with the choice of prior distribution for the between-study variance. The prior distribution for the between-study correlation is highly informative for the posterior estimate of this parameter when there are few studies per meta-analysis, and impacts upon inferences about the treatment effects for each outcome, in particular when there is missing outcome data for some of the studies. In conclusion, empirical studies are recommended to identify suitable prior distributions for the between-study correlation for particular pairs of outcomes across a range of disease fields that can be used for future bivariate (and multivariate) meta-analyses when there are few studies to estimate this parameter.

CHAPTER 7: DISCUSSION

7.1 Overview of thesis

Randomised clinical trials are the best method to obtain evidence about the effect of an intervention, and to thereby update current knowledge and to support health care decision making. The key aims of a clinical trial are to explore whether a treatment, device or strategy is safe and effective in a population of interest, and a crucial part of the evaluation of a clinical trial is the statistical design and analysis, and then synthesis with results of other trials. Traditionally, the analysis and synthesis of clinical trials has been undertaken using a frequentist framework. However, this thesis has explored the use of Bayesian methods and compared these with a frequentist approach in several settings to determine key differences, advantages and limitations.

In Chapter 2, the literature review of Bayesian methods in randomised trials in general medical journals in 2012 suggested that Bayesian methods may be advantageous in clinical trials where the sample size is small because the Bayesian approach can formally incorporate external evidence via the prior distributions. Therefore, the Bayesian framework was investigated in Chapter 3 in a randomised trial that failed to recruit its target sample size (PLUTO). In this trial, for the condition of interest there was elicited prior information derived from specialist clinicians and this was formally incorporated into the analysis of the trial. The findings suggested that there was evidence that the intervention may be beneficial, and the additional probability estimates from the Bayesian analyses could be used to inform guideline updates for clinical practice. However, long term follow-up is still required to assess quality of life.

Chapter 2 also identified that very few randomised phase III trials (8% of articles reviews in the Lancet in 2012) used results from previous or related phase II trials to inform their design in terms of sample size. Hence the use of meta-analysis of phase II trials was explored in Chapters 4 and 5 to explore the issues relating to meta-analysis of small trials and how to incorporate informative prior distributions within a Bayesian framework. Firstly, the methods were considered in the context of univariate meta-analysis of phase II trials (Chapter 4) where the purpose was to obtain an estimate of treatment effect from all available evidence to inform whether further research in a large-scale phase III trial should be recommended. Several Bayesian and frequentist methods were applied and compared in an existing published meta-analysis in acute myocardial infarction for illustration. The Bayesian approach was shown to easily account for all parameter uncertainty, which could be quantified by prediction intervals for the estimated treatment effect in a new phase III trial. Another possible advantage of the Bayesian approach, which was seen in an article in the literature review in Chapter 2, was that direct probability estimates were useful for decision making. This was useful in meta-analysis of phase II trials where probability estimates regarding the pooled or predicted treatment effect could be derived. However, a limitation of the Bayesian approach was the need to specify an appropriate prior distribution for the between-study variance parameter and thus empirical prior distributions must be sought, or a sensitivity analysis for this prior distribution is important.

Previous work highlighted the benefits of a *multivariate* meta-analysis if there are several correlated outcomes; in particular this approach has advantages over separate univariate meta-analyses if there is missing outcome data.^{200,213} Bayesian methods were therefore explored in this setting (Chapter 5) in the context of synthesising phase II trials with multiple correlated outcomes. The Bayesian approach was shown to help reduce the common issue of unknown within-study correlations through the use of prior distributions for this parameter. Also, the Bayesian framework allowed joint predictive probabilities to be derived, which are

potentially useful for clinical guidance that makes joint decisions in the context of multiple possible outcomes. The same example dataset in acute myocardial infarction was used for illustration in this chapter and showed that the within-study correlations can impact on the model estimates. Therefore, external information from a related setting can be useful for the prior distributions for the within-study correlations.

Throughout this thesis, an important message is that an appropriate choice of prior distributions is key to any Bayesian analysis, especially when sample sizes are small within a trial (or when there are small numbers of studies within a meta-analysis), and therefore sensitivity to the choice of prior distributions was routinely conducted. In particular, Chapter 3 showed that the findings of the Bayesian analysis of the under-recruited trial were heavily influenced by the choice of prior distribution for the treatment effect. Chapter 4 also showed that sensitivity to the choice of prior distribution for the between-study variance parameter in a random-effects univariate meta-analysis was important because it is difficult to select a non-informative prior distribution.²⁹ In Chapter 6, the impact of prior distributions was explored in the context of a bivariate meta-analysis, with an extensive simulation study undertaken to assess the impact of various prior distributions for the between-study correlation and variance parameters. This explored several different scenarios including complete and missing outcome data, and identified a number of novel findings. In particular, an inappropriate choice of prior distribution for the between-study variance can adversely impact the posterior influences about the between-study correlation. Further, the routine use of a uniform(-1,1) prior distribution for the between-study correlation should be avoided, where possible, as they are not necessarily vague; rather, a more informative prior distribution should be incorporated, for example restricting values to be either positive or negative as indicated by clinical or prior knowledge.

In the remainder of this final chapter, the main advantages and limitations of Bayesian methods for clinical trial analysis and synthesis are discussed in light of the findings in this thesis. Overarching recommendations are made, and suggestions for further research are presented to tackle the remaining challenges left unaddressed by the research.

7.2 Key findings and recommendations

In this section the key findings from all chapters are discussed alongside the pivotal recommendations, which are displayed in Figure 7.1.

7.2.1 The use of Bayesian methods for the analysis of randomised clinical trials was rare in general medical journal articles

There have been vast developments of Bayesian methods for the analysis of clinical trials, which have been promoted in various settings; in particular for interim monitoring of large randomised trials²⁴¹ and adaptive trials,^{241,242} because the Bayesian approach has several potential advantages compared to the frequentist approach. First, the Bayesian can estimate direct probabilities of interest such as the probability that treatment A is better than treatment B. In contrast, the frequentist can only interpret the p-value as the probability of observing the estimate of effect or one more extreme, if the null hypothesis of no difference is true. Secondly, the Bayesian can formally incorporate external, related information in the prior distribution, whereas the frequentist can only make inferences from the data. This advantage is also the Bayesian's downfall because the derivation of prior distributions can be extremely difficult, and thus – even where vague prior distributions are desired - sensitivity to the prior distribution should always be advocated. The Bayesian approach can also be more flexible than the frequentist in terms of updating the inferences based on accumulating data, which is why the methods are so useful for interim monitoring.¹⁴² However, in 2012 in seven general medical journals, only 3 out of 314 published randomised trials adopted a Bayesian approach

for the analysis of all, or even part, of their trial (Chapter 2). Evidently, although there are these potential advantages of the Bayesian framework, they are still not being widely used in research that is deemed important by general medical journals for their readers.

7.2.2 Clinical implications were identified in some specific Bayesian analyses

Though much of the thesis concentrated on methodology development, a number of real applications were made that have important clinical value. Bayesian methods were explored in Chapter 3 for the analysis of the PLUTO trial, which investigated a rare disease. In this setting, several elicited prior distributions had already been derived from specialist clinicians with the hope that there would be increased precision in the conclusions of the trial if the prior distributions were combined with the data, and, in turn, this would help the decision process. The frequentist analyses were inconclusive; yet, Bayesian posterior probability estimates highlighted the high probability of a benefit of the intervention in the PLUTO trial. However, there were consistently low probabilities of a clinically important treatment effect (as defined by a HR estimate lower than 0.646), which helped to inform the updated guidelines of clinical practice in the future. The recommendations were that there may be some benefit of the intervention in some settings, but that the mother must be advised as to the risk of the procedure to the unborn baby, and that long-term outcomes had not yet been assessed.

Chapters 4 and 5 investigated univariate and multivariate meta-analysis using a previously published meta-analysis of phase II trials in acute myocardial infarction for illustration. The authors of the original meta-analysis raised concerns that the results from the phase II meta-analysis were contradictory to the results from a meta-analysis of the subsequent phase III trials.¹⁶⁹ They also suggested that this was due to publication bias of the phase II trials.

However, when a Bayesian approach was used within a random-effects meta-analysis model, and prediction intervals for the true treatment effect in a new trial were derived, the meta-analysis results showed that the subsequent phase III trial results were entirely plausible based on the phase II evidence, and that publication bias was not necessarily an issue. The wide prediction intervals for the true treatment effect in a new phase III trial highlighted the large uncertainty upon which the decision to conduct the phase III trials was actually based. It also highlighted that the causes of heterogeneity in treatment effects should ideally be identified to ensure that phase III trials can be better targeted to those populations where the treatment effect is likely to be higher.

7.2.3 Bayesian methods were identified as important in the meta-analysis of phase II trials to inform phase III trial decisions

The decision to proceed to a phase III trial should be based on all existing evidence, including results of earlier phase II trials, alongside their potential cost and feasibility. If there are multiple phase II trials then a meta-analysis of these results should be conducted to help inform phase III decisions. However, due to the small numbers of phase II trials, and the likely small sample sizes within these trials, there are several methodological issues that arise, as follows.

It is difficult to examine and quantify the heterogeneity in a phase II meta-analysis.

The small sample sizes within phase II trials can cause low power and large within-study variation, which make it difficult to assess the amount of between-study heterogeneity. It has previously been argued that the I^2 statistic or tests for heterogeneity should not be used to determine whether to use a fixed-effect or random-effects model to pool treatment effects,¹⁹⁹ however they are still often used in practice.²⁰⁷⁻²¹² In the phase II meta-analysis example in acute myocardial infarction in Chapter 4, where there were small studies combined with rare

event rates, the I^2 statistic often suggested little or no evidence of heterogeneity, whereas direct estimates of this parameter, particularly in a Bayesian framework, suggested that heterogeneity was present. In the frequentist framework, this is a general issue for hypothesis testing when the result of a significance test is used to make the assumption that no evidence of an effect is the same as evidence of no effect, which is not appropriate. Therefore, a key recommendation is to not rely on the I^2 statistic for making modelling decisions in a meta-analysis of phase II trials. It is likely that heterogeneity will exist in a meta-analysis of phase II trials and therefore researchers are better off deciding, *a priori*, that a random-effects model is more appropriate.

The derivation of prediction intervals for the true treatment effect in a new trial is important for the phase III decision.

When informing the phase III decision it is crucial that decision-makers understand the statistical uncertainty in the phase II trial result before making the decision to proceed. Often the decision is based on the average pooled effect estimates,³⁹ as done in the meta-analysis by Eikelboom et al.,¹⁶⁹ however it is perhaps more important to derive prediction intervals for the true intervention effect (or the estimated treatment effect with a given sample size) in a new trial because this may be different to the average effect. Prediction intervals derived from the posterior distribution of true (or estimated) treatment effects in new studies will, in the Bayesian framework, naturally account for all parameter uncertainty, and thus allow for more appropriate dissemination of the uncertainty of evidence for those making phase III decisions (section 4.6). This is important because researchers need to decide to progress to phase III for interventions with potential clinical benefit, but they also need to design the phase III trial appropriately in terms of its sample size. The number of phase III trials that achieve statistical significance can be low in certain health-care areas; in particular a report in 2012 reviewed 235 recently published phase III randomised clinical trials in oncology, and

only 38% achieved results with statistical significance.²⁴³ This is often considered to be due to an incorrect sample size calculation, which has been designed with an overly optimistic estimate of treatment effect.²⁴⁴

Bayesian meta-analysis methods also lead naturally to direct probability statements regarding the treatment's potential effectiveness in a new trial and this can be estimated for varying sample sizes of a potential new trial to further aid decisions regarding the design of the phase III trial (section 4.6.3). Additionally, the researcher can incorporate an analysis with a sceptical or cautious prior distribution for the pooled treatment effect if there is reason to believe that there may be optimism in the results (section 4.10, for example, as a result of publication bias).

The choice of prior distribution for the between-study variances is important.

The choice of prior distribution for the between-study variance parameter in a Bayesian meta-analysis can influence the results, and an inappropriately selected prior distribution for this parameter can be highly informative towards the posterior distribution of the pooled treatment effect estimates and the prediction intervals²⁹ (section 4.11). Therefore, empirical prior distributions for the between-study variance should be used where possible in practice.^{36,38} The empirical prior distributions that have been derived by Turner et al.³⁵ and Rhodes et al.⁴⁰ for binary and continuous outcomes are particularly useful for a future Bayesian random-effects meta-analysis because the number of studies in the meta-analysis is often low and therefore there is little data to estimate this parameter well. However, empirical prior distributions have not been derived specifically for meta-analyses of small phase II trials and therefore, until these are available, a sensitivity analysis for the prior distribution for the between-study variance should always be conducted.

It is beneficial to use a Bayesian multivariate meta-analysis model when within-study correlations are unknown.

Where possible, phase II multivariate meta-analysis for multiple correlated outcomes should be considered to inform phase III trial decisions as the approach utilises more of the available evidence compared to univariate meta-analysis.^{200,215,218,225,232} The multivariate approach accounts for the correlation amongst outcomes, and thereby can ‘borrow strength’ to gain more information toward the meta-analysis results, especially when some outcomes are missing in some trials. Although multivariate meta-analysis has been growing in importance, they had not been previously considered for synthesis of phase II trials. This application identified an important issue: within-study correlations are often unknown and cannot be derived without individual participant data (IPD) (Chapter 5). Researchers should provide IPD to facilitate multivariate approaches, such as multinomial modelling, or to derive within-study correlation estimates for a multivariate normal approach.

If IPD are not available, one option to deal with this is to use the alternative Riley model, which does not require the user to specify the unknown within-study correlations.²¹⁵ In a meta-analysis of phase III trials where the focus is often on the pooled result, this may be adequate; however, Chapter 4 showed that a crucial element of meta-analysis of phase II trials is the derivation of prediction intervals to account for all parameter uncertainty, and to estimate the true treatment effect in a new trial. Such prediction intervals cannot be derived from the alternative model by Riley et al.,²¹⁵ as it does not correctly model the between-study variance. Therefore, in Chapter 5 the Bayesian approach to multivariate meta-analysis was proposed in order to specify a range of possible values for within-study correlations using prior distributions. In particular, upon application to meta-analysis of phase II trials in acute myocardial infarction, a prior distribution was derived for missing within-study correlations based on external evidence from trials and outcomes in a related context, where such

correlations had been reported. The Bayesian approach was also shown to easily estimate joint posterior probabilities for the true treatment effects in a new trial, which are useful to inform whether to proceed to a phase III trial based on the evidence for multiple outcomes, rather than single outcomes (section 5.3.7). But, further work is required to develop prior distributions for the within-study correlations of pairs of outcomes in future multivariate meta-analyses.

The prior distributions for between-study covariance parameters in a bivariate meta-analysis must be chosen carefully.

Previous methodological and applied articles that used Bayesian methods for multivariate meta-analysis revealed that there does not appear to be a consensus for the choice of prior distribution for the between-study correlation parameter in a Bayesian bivariate random-effects meta-analysis. Often the uniform(-1,1) prior distribution is selected without a sensitivity analysis. Bujkiewicz et al. also demonstrate that if prior distributions are specified for the individual components of the between-study covariance matrix (rather than for the matrix itself) then caution must be taken to ensure that all prior distributions are appropriate, since these parameters are not independent.¹⁷ A simulation study of the choice of prior distributions for the between-study covariance parameters in Chapter 6 revealed that when adopting a Bayesian approach for multivariate meta-analysis, researchers need to ensure that the prior distribution for the between-study correlation is chosen sensibly. This is because the prior distribution for the between-study correlation is informative of its posterior distribution, in particular when there are few studies in the meta-analysis. This has little impact on the pooled treatment effect estimates if the data is complete, however it does alter the joint inferences, such as the joint probability estimates, which are a key benefit of the Bayesian approach for decision-making. Therefore, even in this setting, the choice of prior distribution for the between-study correlation is important. More importantly, if there is

missing outcome data, the choice of prior distribution for the between-study correlation might even alter the pooled treatment effect estimates, and so the choice of prior distribution is even more important in this situation (section 6.6). Importantly, a uniform(-1,1) prior distribution is not always vague and should not be used routinely without careful consideration, especially when the number of studies is small. Therefore, if possible, a key recommendation is to use a more informative prior distribution that is based on prior clinical or external information. For example, if the outcomes were disease-free and overall survival, or systolic and diastolic blood pressure then a uniform(0,1) prior distribution may be considered appropriate for the between-study correlation. If the between-study correlation is considered to be negative then a uniform(-1,0) prior distribution may be appropriate, for example for the meta-analysis of sensitivity and specificity from test-accuracy studies.

Interestingly, even if the between-study correlation is estimated perfectly, the joint probabilities are still affected by the uncertainty in the other parameters (section 6.8). In the scenarios in this simulation study, the joint predictions are downwardly biased (that is, under a frequentist long-run perspective, the joint probabilities on average are lower than the true joint probability). This merely reflects that uncertainty in parameter estimates makes it hard to make joint inferences appropriately. This is akin to joint probabilities being lower from a bivariate t -distribution compared to a bivariate normal distribution, unless the degrees of freedom are considerably large.

Also, it has already been discussed in detail throughout this thesis that the pooled results can be sensitive to the choice of prior distribution for between-study variance parameters.²⁹ However, in a bivariate meta-analysis setting, the variance parameters additionally affect the posterior distribution of the between-study correlation. If the posterior estimates of the between-study variances are upwardly biased this increases the estimates of the between-study correlation, which in turn alters the pooled estimates if there is missing data, and

changes the joint inferences. Further research should be considered to establish empirical prior distributions for between-study correlations for particular pairs of outcomes across a range of disease fields. If there is a lack of empirical evidence, clinical guidance should be used where possible to inform the prior distribution for between-study correlations. In all meta-analyses with a Bayesian approach, sensitivity analysis for the choice of prior distribution should be considered, especially when external evidence to inform the prior distributions is not available.

Figure 7.1: Key methodological recommendations.

- In trials with small patient numbers, either by design or due to failure to recruit, frequentist analyses will invariably be underpowered, or the results will likely be inconclusive. An alternative Bayesian analysis should be considered that can generate probabilistic statements, which may aid decision-making.
- Meta-analysis of phase II trials should be conducted to inform the phase III trial decision.
- The decision to use a fixed-effect or random-effects univariate meta-analysis model should not be based on the I^2 statistic or tests for heterogeneity, especially in trials with small sample sizes (for example, phase II trials).
- Prediction intervals for the true treatment effect in a new phase III trial should be derived following a meta-analysis of phase II trials, ideally in a Bayesian framework, to account for all parameter uncertainty.
- In a Bayesian univariate or multivariate meta-analysis model, empirical prior distributions for between-study variances based on previous evidence should be implemented.
- Where possible, phase II multivariate meta-analysis for multiple correlated outcomes should be considered to gain more information toward phase III trial decisions, and enable inferences based on multiple outcomes jointly.
- Researchers should provide IPD to facilitate multivariate meta-analysis in order to use a multinomial model for correlated binary outcomes, or to derive within-study correlation estimates for an approximate multivariate normal approach.
- If IPD are not available for multivariate meta-analysis, a Bayesian approach with prior distributions for the unknown within-study correlations should be used, ideally based on external evidence.
- The prior distribution for between-study correlations in a Bayesian multivariate meta-analysis model must be chosen sensibly. A uniform(-1,1) prior distribution should not be used without clinical guidance because it is not always vague.
- In all Bayesian meta-analyses, sensitivity analysis for the choice of prior distribution should be considered, especially when external evidence to inform the prior distributions is not available.

7.3 Bayesian methods in the context of current clinical research: should they be used?

7.3.1 Where is the Bayesian approach most justifiable?

An overarching question that remains is whether there is a rationale for the use of Bayesian methods for the design, analysis or synthesis of clinical trials, instead of an approach within a frequentist framework? It is acknowledged that there is a lack of Bayesian methods in clinical trials, despite the methodological advances, ethical advantages, and alternative designs for randomised clinical trials.¹⁴² However, based on the findings of this thesis, there are key reasons why Bayesian methods are specifically advocated:

Incorporation of external evidence

Bayesian methods are especially encouraged when one wishes to formally incorporate external evidence via the prior distributions. In particular, in Chapter 3 there were elicited prior distributions derived from specialist clinicians that could be combined with the small data from a trial that failed to recruit its target sample size within a rare disease setting. It was useful to see whether there would be different, or even stronger, conclusions with the addition of this previous clinical evidence, as conclusions based solely on the data were not strong.

In univariate random-effects meta-analysis, which was explored in detail in Chapter 4, it was shown that prior distributions were useful for the unknown between-study variance because very often there is little information to estimate this parameter well from the data alone. It is also difficult to specify truly *vague* prior distributions for variance parameters²⁹ therefore, empirical prior distributions are particularly useful to specify a reasonable distribution for the between-study variance.^{36,38}

Another setting where informative prior distributions were useful was for unknown within-study correlations within bivariate meta-analysis. Chapter 5 discussed the benefits of using multivariate meta-analysis when there are correlated outcomes of interest.^{17,18,200,213,214,218,226}

A common issue with this approach is that the within-study correlations must be specified but they are unknown. Prior distributions for the unknown within-study correlations enable the meta-analyst to define a range of values and account for the uncertainty in these parameters.¹⁷ This approach is more beneficial if there is external evidence to inform the likely range of values, for example if it is known clinically that the correlation is positive.

Accounting for all parameter uncertainty

The Bayesian approach can easily account for all parameter uncertainty, which is important for clinical decision making. In particular, this is important when decision-makers decide whether to conduct a phase III trial based on the results of a meta-analysis of phase II trials. In a random-effects meta-analysis, the Bayesian approach can easily estimate the uncertainty in the between-study variance parameter, which is important when there are few studies per meta-analysis because the between-study variance parameter is often poorly estimated²⁴⁵ (section 4.6.2). In terms of making decisions about further research, a Bayesian value-of-information analysis could then be used to formally incorporate other information, such as costs and adverse events, as well as information on treatment effect. These methods consider the value of additional information that could be collected against the costs of uncertainty surrounding a clinical decision.²⁴⁶ The “expected costs of uncertainty are determined by the probability that a treatment decision based on existing (prior) information will be wrong and by the consequences if the wrong decision is made (loss function).”²⁴⁶ Value-of-information modelling may be used to predict the potential impact of studies that are being considered in terms of reduced uncertainty in parameters, such as treatment effect, and prioritise where additional investment may have maximal benefit.

Making direct probability statements

The Bayesian approach is also advocated for the benefit of being able to make direct probability statements. For example, in Chapter 3 in the analysis of the PLUTO trial, it was useful to estimate the probability of a benefit of the intervention, and also to estimate the probability that the intervention was clinically important. This highlighted that there was a high probability that the intervention improved survival compared to the standard care, however, the probability of a clinically meaningful difference in survival rates was low. The use of Bayesian methods in trials with rare diseases has previously been suggested as “a way out of a conundrum” because small trials “are expected to provide useful information, even when a definitive answer is unlikely in prospect.”¹⁴⁸ However, the Bayesian approach can give probability estimates that the clinical effect lies within a range of values. Probability estimates were also estimated for pooled and predicted treatment effects in meta-analysis of phase II trials. In particular, in multivariate meta-analysis (Chapter 5), joint probability estimates were derived, which incorporated the correlation between outcomes.

7.3.2 Will there be acceptance of Bayesian results if they depend on the choice of prior distributions?

A counter argument to the advantage of using informative prior distributions is whether the research community will ever truly accept results that depend on the choice of prior distributions. It would perhaps be more difficult to find acceptance of results that combine data with subjective prior distributions. In the PLUTO trial in Chapter 3, several of the prior distributions were subjective. The results from these analyses were largely in agreement with the frequentist results and therefore additional strength can perhaps be gained from the Bayesian approach in general, such as probability estimates. However, if the results using the subjective prior distributions were contradictory, either to each other, or to the frequentist results, it is unlikely that decisions would be made based purely on these subjective prior

analyses. Perhaps in this scenario more weight would be given to those conclusions using a frequentist approach, or a Bayesian analysis with vague prior distributions.

In random-effects meta-analysis (univariate or multivariate) empirical prior distributions for variance parameters appear useful.^{36,38} But, it is difficult to know how similar the current meta-analysis is to the meta-analyses from which the empirical prior distribution was derived. The variance parameters are highly influential in the estimation of pooled effects and subsequent prediction intervals and it is very difficult to know whether the selected prior distributions are appropriate. Therefore, the results of such an analysis may not be accepted by some parts of the medical community if they disagree with the meta-analysts reasons for adopting a particular set of prior distributions.

When there is uncertainty in the choice of prior distributions, and thus 'vague' prior distributions are desired, a sensitivity analysis is often advocated, especially for the between-study variance parameters. As a consequence, there will be several results to inform one decision, and this makes the decision process more difficult. This is even more difficult if there is no evidence to inform the prior distributions because the results could be more diverse. In Chapter 4, a sensitivity analysis for the prior distribution for the between-study variance highlighted the difference between the prediction intervals for the true treatment effect in a new trial, depending on the choice of this prior distribution. This begs the question as to whether a Bayesian approach should be adopted at all if there is no evidence to inform the prior distributions that, as a consequence, can be very informative to the results.

7.3.3 What if a Bayesian approach is not chosen?

If the arguments above convince researchers not to use Bayesian methods then instead a frequentist approach will be adopted. In a clinical trial setting where there is a large dataset, for example a large-scale phase III trial, which is powered to detect clinically meaningful

differences, there may be no additional benefits from using a Bayesian approach that outweigh the difficulties in using the Bayesian framework, such as specifying prior distributions. Perhaps this is the reason why there is very low prevalence of Bayesian methods in randomised trials in the general medical journal literature in 2012 (see Chapter 2), because these journals are more likely to publish the results from larger trials, rather than the more challenging trial settings, such as rare diseases. Or perhaps researchers are not recognising that external evidence exists and that there may be benefits from incorporating this into a Bayesian analysis as an alternative to using a frequentist framework.

There are several settings in this thesis that have demonstrated where the frequentist approach struggles in comparison to the Bayesian, and perhaps where a Bayesian approach could more commonly be adopted. For example, the data may not be sufficient to answer the clinical question, such as the small dataset in the PLUTO trial (Chapter 3). All frequentist analyses were statistically non-significant, and it is difficult to make decisions about clinical practice based on results such as these. Whereas the Bayesian can additionally quantify their uncertainty with direct probability estimates, such as the probability that the new treatment is clinically significantly different to standard care.

The frequentist approach also struggles to estimate certain parameters, such as the between-study variances in both univariate and multivariate random-effects meta-analysis, because there are often few studies. But, as already mentioned, this parameter is highly influential to pooled results and subsequent predictions, and so it needs to be estimated well. Also, in multivariate meta-analysis, sometimes the frequentist cannot estimate between-study correlation; in particular, if there are few studies, or if the within-study variance is relatively large compared to the between-study variance, the maximum likelihood estimator routinely estimates the between-study correlation as -1 or +1.²¹⁸ Further, the frequentist approach cannot estimate pooled estimates from a multivariate meta-analysis without

specifying the within-study correlations, which are often unknown. Thus, the frequentist could either fix the within-study correlations and these may not be reliable, or they could use the alternative Riley model,²¹⁵ which cannot be used to make predictions.

An alternative proposal for the statistical approach could be to use a frequentist analysis but interpret the results as if they were obtained from a Bayesian analysis with vague prior distributions. Therefore, confidence intervals would be interpreted as Bayesian credible intervals. For example, the results from the PLUTO trial were very similar for the frequentist analyses compared to those with a vague prior distribution for the unknown treatment effect. However, this would only be acceptable if the prior distributions were truly vague and as discussed earlier, it is not easy to specify a vague prior distribution for variance-covariance parameters. The concept of using a frequentist analysis followed by a Bayesian interpretation is what led to the recommendation to use a t -distribution for the prediction interval of the true treatment effect in a new trial following a meta-analysis.^{39,190} However, this t -distribution is an approximation of the unknown underlying distribution, but a Bayesian analysis using MCMC does not need to make this assumption because it can sample from the posterior distribution of true effects even when it does not know its true form. Therefore, it may still be better to use a Bayesian approach in this setting.

7.4 Future research needs

This thesis has made important contributions to the Bayesian analysis of clinical trials, however there are several research needs that remain unaddressed and these are listed in Figure 7.2. Firstly, in Chapter 2, the literature review obtained an overview of the prevalence of Bayesian methods in randomised trials, and what methods have recently been adopted in practice. As previously mentioned it would be beneficial to know what the prevalence is in

specialised journals (e.g. in cancer) that may be more willing to publish alternative statistical methodology in order to see whether the prevalence is the same in these settings.

In Chapter 5, it was suggested that the best method of conducting a multivariate meta-analysis of two or more binary outcomes is to use a multinomial model because this method avoids the need to specify within-study correlations, which are often unknown. However, this model could not be explored for the example in acute myocardial infarction because there was no available IPD. Therefore, Bayesian methods for the multinomial model need to be explored and compared to the (approximate) multivariate normal model.

Figure 7.2: Key future research needs.

- The prevalence of Bayesian methods in randomised controlled trials in specialised medical journals needs to be reviewed, for example, in cancer.
- The use of a multinomial model for the meta-analysis of two or more binary outcomes in phase II trials needs to be explored and compared to the approximate multivariate normal model.
- Further research should establish empirical prior distributions for between-study variances in meta-analysis of phase II trials, and for between-study correlations for particular pairs of outcomes across a range of health-care settings.
- The multivariate meta-analysis methods of Chapters 5 and 6 needs to be extended to the synthesis of *more* than two outcomes.
- Additional scenarios need to be considered to fully assess the impact of prior distributions in a bivariate meta-analysis. In particular, prior distributions for between-study variance-covariance parameters need to be assessed for different sizes and ratios of within and between-study variances.

Empirical prior distributions need to be derived, if possible, for between-study variances for meta-analysis of phase II trials and between-study correlations for a range of different

disease areas, rather than using vague prior distributions for these parameters, which result in the need for several sensitivity analyses. Empirical prior distributions have previously been derived for between-study variances for several health-care settings, for both continuous³⁸ and binary outcomes.³⁶ These are particularly useful because they provide realistic prior distributions for a future meta-analysis without the need to specify a vague prior distribution and a sensitivity analysis for this choice. However, it is unclear whether these are applicable to the setting of meta-analysis of phase II trials since there are often less phase II trials and they may be more heterogeneous than phase III trials. If empirical prior distributions are not available, it may be possible to elicit subjective beliefs. The methods involved for the elicitation of subjective prior beliefs have been presented previously by O'Hagan.^{26,247} The elicitation of subjective beliefs to form probability distributions remains a difficult task because there may be a wide range of research findings that inform an expert's opinion, and this makes it difficult to ensure that the expert's judgement has been captured accurately and reliably.²⁶

The recommendation to derive empirical priors for between-study correlations is, in theory, a necessary development. However, it is important to consider what this would involve practically and whether this is realistic at this stage. The methods to derive empirical priors would likely be similar to those of Turner et al.³⁶ and Rhodes et al.³⁸ In their development of empirical prior distributions they collected thousands of meta-analyses and used hierarchical models to analyse the data, while investigating the effects of meta-analysis characteristics on the level of between-study heterogeneity.^{36,38} Thus, for the empirical prior distributions for between-study correlations to be valid for future meta-analyses there would need to be a substantial number of applied multivariate meta-analyses from which to derive them. In theory, the use of multivariate meta-analysis has been advocated;^{17,200,213,218,221} however, the application in practice still seems to be limited. There is certainly a greater uptake of multivariate meta-analysis for sensitivity and specificity than other settings;²⁴⁸ but there

needs to be application of the methods in all disease areas, for various outcomes, and different types of interventions, before empirical prior distributions for between-study correlations can be derived that would be applicable to a wide range of future multivariate meta-analyses. Thus, for now, it is unlikely that empirical prior distributions can be derived for between-study correlations.

The use of a Bayesian framework for multivariate meta-analysis within Chapters 5 and 6 looked specifically at bivariate meta-analysis. Of course, it may be clinically important to investigate the joint synthesis of *more* than two outcomes and this has not been explored in either of these chapters, but there may be additional difficulties when synthesising more than two outcomes. For more than two outcomes there would need to be further specification of prior distributions for within-study correlations and there would be multiple between-study correlations (rather than just one in bivariate meta-analysis). Additionally, the model specification becomes more difficult in terms of separating the parameters in the variance-covariance matrices.

The simulation study in Chapter 6 did not assess all possible scenarios for bivariate meta-analysis. Importantly, this only focussed on five prior distributions for the between-study correlation but there are numerous other possible choices that could have been explored. Also, for all simulations, the true between-study variance was 0.5 for both outcomes, which is fairly comparable to the within-study variances. However, it is likely that the impact of the prior distribution for the between-study correlation would differ in alternative settings. Large and small between-study variances relative to the within-study variances, and between-study variances that are different to each other, should be assessed in future research.

7.5 Final conclusions

In conclusion, there are clearly many arguments for and against the use of Bayesian methods for the analysis and synthesis of clinical trials. This debate has existed for a long time and despite the important methodological findings made, the work in this thesis is perhaps only a small step toward more use of Bayesian methods in applied research. Perhaps in clinical research, Bayesian methods will always remain those that are adopted as a consequence of difficulties faced by the frequentist approach, and as a result, maybe the prevalence of Bayesian methods will always remain low. However, in certain applications, such as random-effects (multivariate) meta-analysis that requires estimation of between-study variance-covariance parameters, the Bayesian approach has the potential for much benefit compared to a frequentist approach, especially if empirical prior distributions are deemed sensible for use. Also, for a long time, the Bayesian approach was not adopted because it was too computationally complex to implement. However, this is no longer the case and, in fact, simulation methods such as MCMC allow highly complex analyses, which make the Bayesian approach more attractive.²⁴⁹ But, perhaps there is still some way to go before Bayesian methods will be accepted by the wider research community as the default method in any setting. Breslow previously stated “it seems clear that Bayesian and frequentist approaches each will have a role to play in biostatistical applications in the years to come.”²⁵⁰ It has been suggested that possibly the first step in introducing Bayesian methods more commonly into randomised trials is to compliment a frequentist analysis with a Bayesian analysis²⁵¹, like that of the PLUTO trial. But, despite the current challenges for the application of Bayesian methods in trials, this thesis has shown that the Bayesian approach has more to offer the clinical trial community than is perhaps being realised.

APPENDIX

APPENDIX A: Total articles included in the review (314 articles)

Annals of Internal Medicine – 23 articles

- Bakker MF, Jacobs JW, Welsing PM, Verstappen SM, Tekstra J, Ton E et al. Low-Dose Prednisone Inclusion in a Methotrexate-Based, Tight Control Strategy for Early Rheumatoid Arthritis A Randomized Trial. *Annals of Internal Medicine* 2012; 156(5):329-U138.
- Bronfort G, Evans R, Anderson AV, Svendsen KH, Bracha Y, Grimm RH. Spinal Manipulation, Medication, or Home Exercise With Advice for Acute and Subacute Neck Pain A Randomized Trial. *Annals of Internal Medicine* 2012; 156(1):1-U47.
- Cohen SP, White RL, Kurihara C, Larkin TM, Chang A, Griffith SR et al. Epidural Steroids, Etanercept, or Saline in Subacute Sciatica A Multicenter, Randomized Trial. *Annals of Internal Medicine* 2012; 156(8):551-U154.
- Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N et al. Inviting Patients to Read Their Doctors' Notes: A Quasi-experimental Study and a Look Ahead. *Annals of Internal Medicine* 2012; 157(7):461-U36.
- Fan VS, Gaziano J, Lew R, Bourbeau J, Adams SG, Leatherman S et al. A Comprehensive Care Management Program to Prevent Chronic Obstructive Pulmonary Disease Hospitalizations A Randomized, Controlled Trial. *Annals of Internal Medicine* 2012; 156(10):673-U46.
- Gallagher J, Sai A, Templin T, II, Smith L. Dose Response to Vitamin D Supplementation in Postmenopausal Women A Randomized Trial. *Annals of Internal Medicine* 2012; 156(6):425-U76.
- Hashkes PJ, Spalding SJ, Giannini EH, Huang B, Johnson A, Park G et al. Riloncept for Colchicine-Resistant or -Intolerant Familial Mediterranean Fever A Randomized Trial. *Annals of Internal Medicine* 2012; 157(8):533-+.
- Horton MR, Santopietro V, Mathew L, Horton KM, Polito AJ, Liu MC et al. Thalidomide for the Treatment of Cough in Idiopathic Pulmonary Fibrosis A Randomized Trial. *Annals of Internal Medicine* 2012; 157(6):398-U49.
- Kimball AB, Kerdel F, Adams D, Mrowietz U, Gelfand JM, Gniadecki R et al. Adalimumab for the Treatment of Moderate to Severe Hidradenitis Suppurativa A Parallel Randomized Trial. *Annals of Internal Medicine* 2012; 157(12):846-U180.

- Kripalani S, Roumie CL, Dalal AK, Cawthon C, Businger A, Eden SK et al. Effect of a Pharmacist Intervention on Clinically Important Medication Errors After Hospital Discharge A Randomized Trial. *Annals of Internal Medicine* 2012; 157(1):1-U35.
- Lehouck A, Mathieu C, Carremans C, Baeke F, Verhaegen J, Van Eldere J et al. High Doses of Vitamin D to Reduce Exacerbations in Chronic Obstructive Pulmonary Disease. *Annals of Internal Medicine* 2012; 156(2):105-W20.
- Long JA, Jahnle EC, Richardson DM, Loewenstein G, Volpp KG. Peer Mentoring and Financial Incentives to Improve Glucose Control in African American Veterans A Randomized Trial. *Annals of Internal Medicine* 2012; 156(6):416-U60.
- Naidoo K, Yende-Zuma N, Padayatchi N, Naidoo K, Jithoo N, Nair G et al. The Immune Reconstitution Inflammatory Syndrome After Antiretroviral Therapy Initiation in Patients With Tuberculosis: Findings From the SAPiT Trial. *Annals of Internal Medicine* 2012; 157(5):313-U48.
- Perkins GD, Kimani PK, Bullock I, Clutton-Brock T, Davies RP, Gale M et al. Improving the Efficiency of Advanced Life Support Training A Randomized, Controlled Trial. *Annals of Internal Medicine* 2012; 157(1):19-U57.
- Qureshi N, Armstrong S, Dhiman P, Saukko P, Middlemass J, Evans PH et al. Effect of Adding Systematic Family History Enquiry to Cardiovascular Disease Risk Assessment in Primary Care. *Annals of Internal Medicine* 2012; 156(4):253-U15.
- Robbins GK, Lester W, Johnson KL, Chang Y, Estey G, Surrao D et al. Efficacy of a Clinical Decision-Support System in an HIV Practice A Randomized Trial. *Annals of Internal Medicine* 2012; 157(11).
- Schaefer M, Sarkar R, Knop V, Effenberger S, Friebe A, Heinze L et al. Escitalopram for the Prevention of Peginterferon-alpha 2a-Associated Depression in Hepatitis C Virus-Infected Patients Without Previous Psychiatric Disease A Randomized Trial. *Annals of Internal Medicine* 2012; 157(2):94-+.
- Shorr RI, Chandler A, Mion LC, Waters TM, Liu M, Daniels MJ et al. Effects of an Intervention to Increase Bed Alarm Use to Prevent Falls in Hospitalized Patients A Cluster Randomized Trial. *Annals of Internal Medicine* 2012; 157(10):692-+.
- Spitzer M, Basaria S, Trivison TG, Davda MN, Paley A, Cohen B et al. Effect of Testosterone Replacement on Response to Sildenafil Citrate in Men With Erectile Dysfunction A Parallel, Randomized Trial. *Annals of Internal Medicine* 2012; 157(10):681-+.
- Thornton J, Alejandro-Rodriguez M, Leon JB, Albert JM, Baldeon EL, De Jesus LM et al. Effect of an iPod Video Intervention on Consent to Donate Organs A Randomized Trial. *Annals of Internal Medicine* 2012; 156(7):483-U47.

Wang C, Cao B, Liu Q, Zou Z, Liang Z, Gu L et al. Oseltamivir compared with the Chinese traditional therapy maxingshigan-yinqiaosan in the treatment of H1N1 influenza. A randomized trial. (vol 155, pg 217, 2011). *Annals of Internal Medicine* 2012; 156(1):71.

Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G. Do Physicians Understand Cancer Screening Statistics? A National Survey of Primary Care Physicians in the United States. *Annals of Internal Medicine* 2012; 156(5):340-U152.

Wilding JP, Woo V, Soler NG, Pahor A, Sugg J, Rohwedder K et al. Long-Term Efficacy of Dapagliflozin in Patients With Type 2 Diabetes Mellitus Receiving High Doses of Insulin A Randomized Trial. *Annals of Internal Medicine* 2012; 156(6):405-U50.

PLOS Medicine – 14 articles

Barasa EW, Ayieko P, Cleary S, English M. A Multifaceted Intervention to Improve the Quality of Care of Children in District Hospitals in Kenya: A Cost-Effectiveness Analysis. *Plos Medicine* 2012; 9(6).

Boily MC, Masse B, Alsallaq R, Padian NS, Eaton JW, Vesga JF et al. HIV Treatment as Prevention: Considerations in the Design, Conduct, and Analysis of Cluster Randomized Controlled Trials of Combination HIV Prevention. *Plos Medicine* 2012; 9(7).

Campbell TB, Smeaton LM, Kumarasamy N, Flanigan T, Klingman KL, Firnhaber C et al. Efficacy and Safety of Three Antiretroviral Regimens for Initial Treatment of HIV-1: A Randomized Clinical Trial in Diverse Multinational Settings. *Plos Medicine* 2012; 9(8).

Coburn KD, Marcantonio S, Lazansky R, Keller M, Davis N. Effect of a Community-Based Nursing Intervention on Mortality in Chronically Ill Older Adults: A Randomized Controlled Trial. *Plos Medicine* 2012; 9(7).

Cohen CR, Lingappa JR, Baeten JM, Ngayo MO, Spiegel CA, Hong T et al. Bacterial Vaginosis Associated with Increased Risk of Female-to-Male HIV-1 Transmission: A Prospective Cohort Analysis among African Couples. *Plos Medicine* 2012; 9(6).

Crowther CA, Dodd JM, Hiller JE, Haslam RR, Robinson JS. Planned Vaginal Birth or Elective Repeat Caesarean: Patient Preference Restricted Cohort with Nested Randomised Trial. *Plos Medicine* 2012; 9(3).

Gibb DM, Kizito H, Russell EC, Chidziva E, Zalwango E, Nalumenya R et al. Pregnancy and Infant Outcomes among HIV-Infected Women Taking Long-Term ART with and without Tenofovir in the DART Trial. *Plos Medicine* 2012; 9(5).

Grijzen ML, Steingrover R, Wit FW, Jurriaans S, Verbon A, Brinkman K et al. No Treatment versus 24 or 60 Weeks of Antiretroviral Treatment during Primary HIV Infection: The Randomized Primo-SHM Trial. *Plos Medicine* 2012; 9(3).

Huybregts L, Hougbe F, Salpeteur C, Brown R, Roberfroid D, Ait-Aissa M et al. The Effect of Adding Ready-to-Use Supplementary Food to a General Food Distribution on Child Nutritional Status and Morbidity: A Cluster-Randomized Controlled Trial. *Plos Medicine* 2012; 9(9).

Lockman S, Hughes M, Sawe F, Zheng Y, McIntyre J, Chipato T et al. Nevirapine- Versus Lopinavir/Ritonavir-Based Initial Therapy for HIV-1 Infection among Women in Africa: A Randomized Trial. *Plos Medicine* 2012; 9(6).

More NS, Bapat U, Das S, Alcock G, Patil S, Porel M et al. Community Mobilization in Mumbai Slums to Improve Perinatal Care and Outcomes: A Cluster Randomized Controlled Trial. *Plos Medicine* 2012; 9(7).

Senn N, Rarau P, Stanistic DI, Robinson L, Barnadas C, Manong D et al. Intermittent Preventive Treatment for Malaria in Papua New Guinean Infants Exposed to *Plasmodium falciparum* and *P. vivax*: A Randomized Controlled Trial. *Plos Medicine* 2012; 9(3).

Seward N, Osrin D, Li L, Costello A, Pulkki-Braennstroem AM, Houweling TA et al. Association between Clean Delivery Kit Use, Clean Delivery Practices, and Neonatal Survival: Pooled Analysis of Data from Three Sites in South Asia. *Plos Medicine* 2012; 9(2).

van der Ham DP, Vijgen SM, Nijhuis JG, van Beek JJ, Opmeer BC, Mulder AL et al. Induction of Labor versus Expectant Management in Women with Preterm Prelabor Rupture of Membranes between 34 and 37 Weeks: A Randomized Controlled Trial. *Plos Medicine* 2012; 9(4).

Nature - one article

Chen Z, Cheng K, Walton Z, Wang Y, Ebi H, Shimamura T et al. A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response. *Nature* 2012; 483(7391):613-617.

New England Journal of Medicine - 74 articles

Agus MS, Steil GM, Wypij D, Costello JM, Laussen PC, Langer M et al. Tight Glycemic Control versus Standard Care after Pediatric Cardiac Surgery. *New England Journal of Medicine* 2012; 367(13):1208-1219.

Anasetti C, Logan BR, Lee SJ, Waller EK, Weisdorf DJ, Wingard JR et al. Peripheral-Blood Stem Cells versus Bone Marrow from Unrelated Donors. *New England Journal of Medicine* 2012; 367(16):1487-1496.

Attal M, Lauwers-Cances V, Marit G, Caillot D, Moreau P, Facon T et al. Lenalidomide Maintenance after Stem-Cell Transplantation for Multiple Myeloma. *New England Journal of Medicine* 2012; 366(19):1782-1791.

- Aversano T, Lemmon CC, Liu L. Outcomes of PCI at Hospitals with or without On-Site Cardiac Surgery. *New England Journal of Medicine* 2012; 366(19):1792-1802.
- Baeten J, Donnell D, Ndase P, Mugo N, Campbell J, Wangisi J et al. Antiretroviral Prophylaxis for HIV Prevention in Heterosexual Men and Women. *New England Journal of Medicine* 2012; 367(5):399-410.
- Bart BA, Goldsmith SR, Lee KL, Givertz MM, O'Connor CM, Bull DA et al. Ultrafiltration in Decompensated Heart Failure with Cardiorenal Syndrome. *New England Journal of Medicine* 2012; 367(24):2296-2304.
- Baselga J, Campone M, Piccart M, Burris HA, Rugo HS, Sahmoud T et al. Everolimus in Postmenopausal Hormone-Receptor-Positive Advanced Breast Cancer. *New England Journal of Medicine* 2012; 366(6):520-529.
- Belshe RB, Leone PA, Bernstein DI, Wald A, Levin MJ, Stapleton JT et al. Efficacy Results of a Trial of a Herpes Simplex Vaccine. *New England Journal of Medicine* 2012; 366(1):34-43.
- Bischoff-Ferrari HA, Willett WC, Orav EJ, Lips P, Meunier PJ, Lyons RA et al. A Pooled Analysis of Vitamin D Dose Requirements for Fracture Prevention. *New England Journal of Medicine* 2012; 367(1):40-49.
- Buller HR, Prins MH, Lensing AW, Decousus H, Jacobson BF, Minar E et al. Oral Rivaroxaban for the Treatment of Symptomatic Pulmonary Embolism. *New England Journal of Medicine* 2012; 366(14):1287-1297.
- Coleman T, Cooper S, Thornton JG, Grainge MJ, Watts K, Britton J et al. A Randomized Trial of Nicotine-Replacement Therapy Patches in Pregnancy. *New England Journal of Medicine* 2012; 366(9):808-818.
- Collet JP, Cuisset T, Range G, Cayla G, Elhadad S, Pouillot C et al. Bedside Monitoring to Adjust Antiplatelet Therapy for Coronary Stenting. *New England Journal of Medicine* 2012; 367(22):2100-2109.
- Comi G, Jeffery D, Kappos L, Montalban X, Boyko A, Rocca MA et al. Placebo-Controlled Trial of Oral Laquinimod for Multiple Sclerosis. *New England Journal of Medicine* 2012; 366(11):1000-1009.
- Crook JM, O'Callaghan CJ, Duncan G, Dearnaley DP, Higano CS, Horwitz EM et al. Intermittent Androgen Suppression for Rising PSA Level after Radiotherapy. *New England Journal of Medicine* 2012; 367(10):895-903.
- De Benedetti F, Brunner HI, Ruperto N, Kenwright A, Wright S, Calvo I et al. Randomized Trial of Tocilizumab in Systemic Juvenile Idiopathic Arthritis. *New England Journal of Medicine* 2012; 367(25):2385-2395.

- De Bruyne B, Pijls NH, Kalesan B, Barbato E, Tonino PA, Piroth Z et al. Fractional Flow Reserve-Guided PCI versus Medical Therapy in Stable Coronary Disease. *New England Journal of Medicine* 2012; 367(11):991-1001.
- Devanand D, Mintzer J, Schultz SK, Andrews HF, Sultzer DL, de la Pena D et al. Relapse Risk after Discontinuation of Risperidone in Alzheimer's Disease. *New England Journal of Medicine* 2012; 367(16):1497-1507.
- Donnez J, Tomaszewski J, Vazquez F, Bouchard P, Lemieszczuk B, Baro F et al. Ulipristal Acetate versus Leuprolide Acetate for Uterine Fibroids. *New England Journal of Medicine* 2012; 366(5):421-432.
- Ebbeling CB, Feldman HA, Chomitz VR, Antonelli TA, Gortmaker SL, Osganian SK et al. A Randomized Trial of Sugar-Sweetened Beverages and Adolescent Body Weight. *New England Journal of Medicine* 2012; 367(15):1407-1416.
- Elmunzer B, Scheiman JM, Lehman GA, Chak A, Mosler P, Higgins PD et al. A Randomized Trial of Rectal Indomethacin to Prevent Post-ERCP Pancreatitis. *New England Journal of Medicine* 2012; 366(15):1414-1422.
- Farkouh ME, Domanski M, Sleeper LA, Siami FS, Dangas G, Mack M et al. Strategies for Multivessel Revascularization in Patients with Diabetes. *New England Journal of Medicine* 2012; 367(25):2375-2384.
- Finfer S, Liu B, Chittock DR, Norton R, Myburgh JA, McArthur C et al. Hypoglycemia and Risk of Death in Critically Ill Patients. *New England Journal of Medicine* 2012; 367(12):1108-1118.
- Furlan AJ, Reisman M, Massaro J, Mauri L, Adams H, Albers GW et al. Closure or Medical Therapy for Cryptogenic Stroke with Patent Foramen Ovale. *New England Journal of Medicine* 2012; 366(11):991-999.
- Gler MT, Skripconoka V, Sanchez-Garavito E, Xiao H, Cabrera-Rivero JL, Vargas-Vasquez DE et al. Delamanid for Multidrug-Resistant Pulmonary Tuberculosis. *New England Journal of Medicine* 2012; 366(23):2151-2160.
- Gold R, Kappos L, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K et al. Placebo-Controlled Phase 3 Study of Oral BG-12 for Relapsing Multiple Sclerosis. *New England Journal of Medicine* 2012; 367(12):1098-1107.
- Homma S, Thompson JL, Pullicino PM, Levin B, Freudenberger RS, Teerlink JR et al. Warfarin and Aspirin in Patients with Heart Failure and Sinus Rhythm. *New England Journal of Medicine* 2012; 366(20):1859-1869.
- Kallmes D. A Randomized Trial of Vertebroplasty for Osteoporotic Spinal Fractures (vol 361, pg 569, 2009). *New England Journal of Medicine* 2012; 366(10):970.

- Kang DH, Kim YJ, Kim SH, Sun BJ, Kim DH, Yun SC et al. Early Surgery versus Conventional Treatment for Infective Endocarditis. *New England Journal of Medicine* 2012; 366(26):2466-2473.
- Kerstjens HA, Engel M, Dahl R, Paggiaro P, Beck E, Vandewalker M et al. Tiotropium in Asthma Poorly Controlled with Standard Combination Therapy. *New England Journal of Medicine* 2012; 367(13):1198-1207.
- Kluin-Nelemans H, Hoster E, Hermine O, Walewski J, Trneny M, Geisler C et al. Treatment of Older Patients with Mantle-Cell Lymphoma. *New England Journal of Medicine* 2012; 367(6):520-531.
- Lamy A, Devereaux P, Prabhakaran D, Taggart DP, Hu S, Paolasso E et al. Off-Pump or On-Pump Coronary-Artery Bypass Grafting at 30 Days. *New England Journal of Medicine* 2012; 366(16):1489-1497.
- Lazarus JH, Bestwick JP, Channon S, Paradise R, Maina A, Rees R et al. Antenatal Thyroid Screening and Childhood Cognitive Function. *New England Journal of Medicine* 2012; 366(6):493-501.
- Lebwohl M, Swanson N, Anderson LL, Melgaard A, Xu Z, Berman B. Ingenol Mebutate Gel for Actinic Keratosis. *New England Journal of Medicine* 2012; 366(11):1010-1019.
- Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G et al. Olaparib Maintenance Therapy in Platinum-Sensitive Relapsed Ovarian Cancer. *New England Journal of Medicine* 2012; 366(15):1382-1392.
- Li F, Harmer P, Fitzgerald K, Eckstrom E, Stock R, Galver J et al. Tai Chi and Postural Stability in Patients with Parkinson's Disease. *New England Journal of Medicine* 2012; 366(6):511-519.
- Lok AS, Gardiner DF, Lawitz E, Martorell C, Everson GT, Ghalib R et al. Preliminary Study of Two Antiviral Agents for Hepatitis C Genotype 1. *New England Journal of Medicine* 2012; 366(3):216-224.
- Mallick U, Harmer C, Yap B, Wadsley J, Clarke S, Moss L et al. Ablation with Low-Dose Radioiodine and Thyrotropin Alfa in Thyroid Cancer. *New England Journal of Medicine* 2012; 366(18):1674-1685.
- Mega JL, Braunwald E, Wiviott SD, Bassand JP, Bhatt DL, Bode C et al. Rivaroxaban in Patients with a Recent Acute Coronary Syndrome. *New England Journal of Medicine* 2012; 366(1):9-19.
- Mehta RS, Barlow WE, Albain KS, Vandenberg TA, Dakhil SR, Tirumali NR et al. Combination Anastrozole and Fulvestrant in Metastatic Breast Cancer. *New England Journal of Medicine* 2012; 367(5):435-444.

- Meyer RM, Gospodarowicz MK, Connors JM, Pearcey RG, Wells WA, Winter JN et al. ABVD Alone versus Radiation-Based Therapy in Limited-Stage Hodgkin's Lymphoma. *New England Journal of Medicine* 2012; 366(5):399-408.
- Mingrone G, Panunzi S, De Gaetano A, Guidone C, Iaiconelli A, Leccesi L et al. Bariatric Surgery versus Conventional Medical Therapy for Type 2 Diabetes. *New England Journal of Medicine* 2012; 366(17):1577-1585.
- Myburgh JA, Finfer S, Bellomo R, Billot L, Cass A, Gattas D et al. Hydroxyethyl Starch or Saline for Fluid Resuscitation in Intensive Care. *New England Journal of Medicine* 2012; 367(20):1901-1911.
- Nager CW, Brubaker L, Litman HJ, Zyczynski HM, Varner R, Amundsen C et al. A Randomized Trial of Urodynamic Testing before Stress-Incontinence Surgery. *New England Journal of Medicine* 2012; 366(21):1987-1997.
- Nielsen-Saines K, Watts H, Veloso VG, Bryson YJ, Joao EC, Pilotto JH et al. Three Postpartum Antiretroviral Regimens to Prevent Intrapartum HIV Infection. *New England Journal of Medicine* 2012; 366(25):2368-2379.
- Nielsen JC, Johannessen A, Raatikainen P, Hindricks G, Walfridsson H, Kongstad O et al. Radiofrequency Ablation as Initial Therapy in Paroxysmal Atrial Fibrillation. *New England Journal of Medicine* 2012; 367(17):1587-1595.
- Palumbo A, Hajek R, Delforge M, Kropff M, Petrucci MT, Catalano J et al. Continuous Lenalidomide Treatment for Newly Diagnosed Multiple Myeloma. *New England Journal of Medicine* 2012; 366(19):1759-1769.
- Parsons M, Spratt N, Bivard A, Campbell B, Chung K, Miteff F et al. A Randomized Trial of Tenecteplase versus Alteplase for Acute Ischemic Stroke. *New England Journal of Medicine* 2012; 366(12):1099-1107.
- Perner A, Haase N, Guttormsen AB, Tenhunen J, Klemenzson G, Aneman A et al. Hydroxyethyl Starch 130/0.4 versus Ringer's Acetate in Severe Sepsis. *New England Journal of Medicine* 2012; 367(2):124-134.
- Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas A et al. Colonoscopy versus Fecal Immunochemical Testing in Colorectal-Cancer Screening. *New England Journal of Medicine* 2012; 366(8):697-706.
- Raghu G, Anstrom KJ, King TE, Lasky JA, Martinez FJ. Prednisone, Azathioprine, and N-Acetylcysteine for Pulmonary Fibrosis. *New England Journal of Medicine* 2012; 366(21):1968-1977.
- Ranieri V, Thompson B, Barie PS, Dhainaut JF, Douglas IS, Finfer S et al. Drotrecogin Alfa (Activated) in Adults with Septic Shock. *New England Journal of Medicine* 2012; 366(22):2055-2064.

- Roe MT, Armstrong PW, Fox KA, White HD, Prabhakaran D, Goodman SG et al. Prasugrel versus Clopidogrel for Acute Coronary Syndromes without Revascularization. *New England Journal of Medicine* 2012; 367(14):1297-1309.
- Roth EM, McKenney JM, Hanotin C, Asset G, Stein EA. Atorvastatin with or without an Antibody to PCSK9 in Primary Hypercholesterolemia. *New England Journal of Medicine* 2012; 367(20):1891-1900.
- Rueangweerayut R, Phyo AP, Uthaisin C, Poravuth Y, Tran QB, Tinto H et al. Pyronaridine-Artesunate versus Mefloquine plus Artesunate for Malaria. *New England Journal of Medicine* 2012; 366(14):1298-1309.
- Ruperto N, Brunner HI, Quartier P, Constantin T, Wulffraat N, Horneff G et al. Two Randomized Trials of Canakinumab in Systemic Juvenile Idiopathic Arthritis. *New England Journal of Medicine* 2012; 367(25):2396-2406.
- Sandborn WJ, Gasink C, Gao LL, Blank MA, Johanns J, Guzzo C et al. Ustekinumab Induction and Maintenance Therapy in Refractory Crohn's Disease. *New England Journal of Medicine* 2012; 367(16):1519-1528.
- Schauer PR, Kashyap SR, Wolski K, Brethauer SA, Kirwan JP, Pothier CE et al. Bariatric Surgery versus Intensive Medical Therapy in Obese Patients with Diabetes. *New England Journal of Medicine* 2012; 366(17):1567-1576.
- Schlumberger M, Catargi B, Borget I, Deandreis D, Zerdoud S, Bridji B et al. Strategies of Radioiodine Ablation in Patients with Low-Risk Thyroid Cancer. *New England Journal of Medicine* 2012; 366(18):1663-1673.
- Schoen RE, Pinsky PF, Weissfeld JL, Yokochi LA, Church T, Laiyemo AO et al. Colorectal-Cancer Incidence and Mortality with Screening Flexible Sigmoidoscopy. *New England Journal of Medicine* 2012; 366(25):2345-2357.
- Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V et al. Prostate-Cancer Mortality at 11 Years of Follow-up. *New England Journal of Medicine* 2012; 366(11):981-990.
- Schwartz GG, Olsson AG, Abt M, Ballantyne CM, Barter PJ, Brumm J et al. Effects of Dalcetrapib in Patients with a Recent Acute Coronary Syndrome. *New England Journal of Medicine* 2012; 367(22):2089-2099.
- Shankaran S, Pappas A, McDonald SA, Vohr BR, Hintz SR, Epi M et al. Childhood Outcomes after Hypothermia for Neonatal Encephalopathy. *New England Journal of Medicine* 2012; 366(22):2085-2092.
- Silbergleit R, Durkalski V, Lowenstein D, Conwit R, Pancioli A, Palesch Y et al. Intramuscular versus Intravenous Therapy for Prehospital Status Epilepticus. *New England Journal of Medicine* 2012; 366(7):591-600.

- Stalmans P, Benz MS, Gandorfer A, Kampik A, Girach A, Pakola S et al. Enzymatic Vitreolysis with Ocriplasmin for Vitreomacular Traction and Macular Holes. *New England Journal of Medicine* 2012; 367(7):606-615.
- Stein EA, Mellis S, Yancopoulos GD, Stahl N, Logan D, Smith WB et al. Effect of a Monoclonal Antibody to PCSK9 on LDL Cholesterol. *New England Journal of Medicine* 2012; 366(12):1108-1118.
- Tang JY, Mackay-Wiggan JM, Aszterbaum M, Yauch RL, Lindgren J, Chang K et al. Inhibiting the Hedgehog Pathway in Patients with the Basal-Cell Nevus Syndrome. *New England Journal of Medicine* 2012; 366(23):2180-2188.
- Thiele H, Zeymer U, Neumann FJ, Ferenc M, Olbrich HG, Hausleiter J et al. Intraaortic Balloon Support for Myocardial Infarction with Cardiogenic Shock. *New England Journal of Medicine* 2012; 367(14):1287-1296.
- Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE, Segolodi TM et al. Antiretroviral Preexposure Prophylaxis for Heterosexual HIV Transmission in Botswana. *New England Journal of Medicine* 2012; 367(5):423-434.
- Tricoci P, Huang Z, Held C, Moliterno DJ, Armstrong PW, Van de Werf F et al. Thrombin-Receptor Antagonist Vorapaxar in Acute Coronary Syndromes. *New England Journal of Medicine* 2012; 366(1):20-33.
- Van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J, Kapiga S et al. Preexposure Prophylaxis for HIV Infection among African Women. *New England Journal of Medicine* 2012; 367(5):411-422.
- Violari A, Lindsey JC, Hughes MD, Mujuru HA, Barlow-Mosha L, Kamthunzi P et al. Nevirapine versus Ritonavir-Boosted Lopinavir for HIV-Infected Children. *New England Journal of Medicine* 2012; 366(25):2380-2389.
- Visco AG, Brubaker L, Richter HE, Nygaard I, Paraiso MF, Menefee SA et al. Anticholinergic Therapy vs. OnabotulinumtoxinA for Urgency Urinary Incontinence. *New England Journal of Medicine* 2012; 367(19):1803-1813.
- Wei JT, Nygaard I, Richter HE, Nager CW, Barber MD, Kenton K et al. A Midurethral Sling to Reduce Incontinence after Vaginal Prolapse Repair. *New England Journal of Medicine* 2012; 366(25):2358-2367.
- Wilt TJ, Brawer MK, Jones KM, Barry MJ, Aronson WJ, Fox S et al. Radical Prostatectomy versus Observation for Localized Prostate Cancer. *New England Journal of Medicine* 2012; 367(3):203-213.

BMJ - 47 articles

- Beckett N, Peters R, Tuomilehto J, Swift C, Sever P, Potter J et al. Immediate and late benefits of treating very elderly people with hypertension: results from active treatment extension to Hypertension in the Very Elderly randomised controlled trial. *British Medical Journal* 2012; 344.
- Bergh S, Selbaek G, Engedal K. Discontinuation of antidepressants in people with dementia and neuropsychiatric symptoms (DESEP study): double blind, randomised, parallel group, placebo controlled trial. *British Medical Journal* 2012; 344.
- Bhandari N, Mazumder S, Taneja S, Sommerfelt H, Strand TA. Effect of implementation of Integrated Management of Neonatal and Childhood Illness (IMNCI) programme on neonatal and infant mortality: cluster randomised controlled trial. *British Medical Journal* 2012; 344.
- Bischoff EW, Akkermans R, Bourbeau J, van Weel C, Vercoulen JH, Schermer TR. Comprehensive self management and routine monitoring in chronic obstructive pulmonary disease patients in general practice: randomised controlled trial. *British Medical Journal* 2012; 345.
- Bowen A, Hesketh A, Patchick E, Young A, Davies L, Vail A et al. Effectiveness of enhanced communication therapy in the first four months after stroke for aphasia and dysarthria: a randomised controlled trial. *British Medical Journal* 2012; 345.
- Bucknall C, Miller G, Lloyd S, Cleland J, McCluskey S, Cotton M et al. Glasgow supported self-management trial (GSuST) for patients with moderate to severe COPD: randomised controlled trial. *British Medical Journal* 2012; 344.
- Butler CC, Simpson SA, Dunstan F, Rollnick S, Cohen D, Gillespie D et al. Effectiveness of multifaceted educational programme to reduce antibiotic dispensing in primary care: practice based randomised controlled trial. *British Medical Journal* 2012; 344.
- Chalder M, Wiles NJ, Campbell J, Hollinghurst SP, Haase AM, Taylor AH et al. Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. *British Medical Journal* 2012; 344.
- Chappell LC, Gurung V, Seed PT, Chambers J, Williamson C, Thornton JG. Ursodeoxycholic acid versus placebo, and early term delivery versus expectant management, in women with intrahepatic cholestasis of pregnancy: semifactorial randomised clinical trial. *British Medical Journal* 2012; 344.
- Clemson L, Singh MA, Bundy A, Cumming RG, Manollaras K, O'Loughlin P et al. Integration of balance and strength training into daily life activity to reduce rate of falls in older people (the LiFE study): randomised parallel trial. *British Medical Journal* 2012; 345.

- Costa ML, Achten J, Parsons NR, Edlin RP, Foguet P, Prakash U et al. Total hip arthroplasty versus resurfacing arthroplasty in the treatment of patients with arthritis of the hip joint: single centre, parallel group, assessor blinded, randomised controlled trial. *British Medical Journal* 2012; 344.
- Crawford MJ, Killaspy H, Barnes TR, Barrett B, Byford S, Clayton K et al. Group art therapy as an adjunctive treatment for people with schizophrenia: multicentre pragmatic randomised trial. *British Medical Journal* 2012; 344.
- Day C, Michelson D, Thomson S, Penney C, Draper L. Evaluation of a peer led parenting intervention for disruptive behaviour problems in children: community based randomised controlled trial. *British Medical Journal* 2012; 344.
- Ferguson J, Docherty G, Bauld L, Lewis S, Lorgelly P, Boyd KA et al. Effect of offering different levels of support and free nicotine replacement therapy via an English national telephone quitline: randomised controlled trial. *British Medical Journal* 2012; 344.
- Gringras P, Gamble C, Jones A, Wiggs L, Williamson P, Sutcliffe A et al. Melatonin for sleep problems in children with neurodevelopmental disorders: randomised double masked placebo controlled trial. *British Medical Journal* 2012; 345.
- Hollands GJ, Whitwell SC, Parker RA, Prescott NJ, Forbes A, Sanderson J et al. Effect of communicating DNA based risk assessments for Crohn's disease on smoking cessation: randomised controlled trial. *British Medical Journal* 2012; 345.
- Holmgren T, Hallgren HB, Oberg B, Adolfsson L, Johansson K. Effect of specific exercise strategy on need for surgery in patients with subacromial impingement syndrome: randomised controlled study. *British Medical Journal* 2012; 344.
- Ibler KS, Jemec GB, Diepgen TL, Gluud C, Hansen JL, Winkel P et al. Skin care education and individual counselling versus treatment as usual in healthcare workers with hand eczema: randomised clinical trial. *British Medical Journal* 2012; 345.
- Joura EA, Garland SM, Paavonen J, Ferris DG, Perez G, Ault KA et al. Effect of the human papillomavirus (HPV) quadrivalent vaccine in a subgroup of women with cervical and vulvar disease: retrospective pooled analysis of trial data. *British Medical Journal* 2012; 344.
- Khunti K, Gray LJ, Skinner T, Carey ME, Realf K, Dallosso H et al. Effectiveness of a diabetes education and self management programme (DESMOND) for people with newly diagnosed type 2 diabetes mellitus: three year follow-up of a cluster randomised controlled trial in primary care. *British Medical Journal* 2012; 344.
- Leinonen MK, Nieminen P, Lonnberg S, Malila N, Hakama M, Pokhrel A et al. Detection rates of precancerous and cancerous cervical lesions within one screening round of primary human papillomavirus DNA testing: prospective randomised trial in Finland. *British Medical Journal* 2012; 345.

- McMillan AM, Landorf KB, Gilheany MF, Bird AR, Morrow AD, Menz HB. Ultrasound guided corticosteroid injection for plantar fasciitis: randomised controlled trial. *British Medical Journal* 2012; 344.
- Meeuwssen EJ, Melis RJ, Van der Aa GC, Goluke-Willemse GA, De Leest BJ, Van Raak FH et al. Effectiveness of dementia follow-up care by memory clinics or general practitioners: randomised controlled trial. *British Medical Journal* 2012; 344.
- Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. *British Medical Journal* 2012; 344.
- Miller G, Luo R, Zhang L, Sylvia S, Shi Y, Foo P et al. Effectiveness of provider incentives for anaemia reduction in rural China: a cluster randomised trial. *British Medical Journal* 2012; 345.
- Morrison AP, French P, Stewart SL, Birchwood M, Fowler D, Gumley AI et al. Early detection and intervention evaluation for people at risk of psychosis: multisite randomised controlled trial. *British Medical Journal* 2012; 344.
- Morthorst B, Krogh J, Erlangsen A, Alberdi F, Nordentoft M. Effect of assertive outreach after suicide attempt in the AID (assertive intervention for deliberate self harm) trial: randomised controlled trial. *British Medical Journal* 2012; 345.
- Palmer D, Sullivan T, Gold M, Prescott S, Heddle R, Gibson R et al. Effect of n-3 long chain polyunsaturated fatty acid supplementation in pregnancy on infants' allergies in first year of life: randomised controlled trial. *British Medical Journal* 2012; 344.
- Roberts I, Perel P, Prieto-Merino D, Shakur H, Coats T, Hunt BJ et al. Effect of tranexamic acid on mortality in patients with traumatic bleeding: prespecified analysis of data from randomised controlled trial. *British Medical Journal* 2012; 345.
- Robling M, McNamara R, Bennert K, Butler CC, Channon S, Cohen D et al. The effect of the Talking Diabetes consulting skills intervention on glycaemic control and quality of life in children with type 1 diabetes: cluster randomised controlled trial (DEPICTED study). *British Medical Journal* 2012; 344.
- Rosario DJ, Lane J, Metcalfe C, Donovan JL, Doble A, Goodwin L et al. Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study. *British Medical Journal* 2012; 344.
- Ryan D, Price D, Musgrave SD, Malhotra S, Lee AJ, Ayansina D et al. Clinical and cost effectiveness of mobile phone supported self monitoring of asthma: multicentre randomised controlled trial. *British Medical Journal* 2012; 344.
- Schierbeck LL, Rejnmark L, Tofteng CL, Stilgren L, Eiken P, Mosekilde L et al. Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women: randomised trial. *British Medical Journal* 2012; 345.

- Schmitz J, West KP, Khattry SK, Wu L, LeClerq SC, Karna SL et al. Vitamin A supplementation in preschool children and risk of hearing loss as adolescents and young adults in rural Nepal: randomised trial cohort follow-up study. *British Medical Journal* 2012; 344.
- Stallard P, Sayal K, Phillips R, Taylor JA, Spears M, Anderson R et al. Classroom based cognitive behavioural therapy in reducing symptoms of depression in high risk adolescents: pragmatic cluster randomised controlled trial. *British Medical Journal* 2012; 345.
- Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S et al. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *British Medical Journal* 2012; 344.
- Stewart S, Carrington MJ, Swemmer CH, Anderson C, Kurstjens NP, Amerena J et al. Effect of intensive structured care on individual blood pressure targets in primary care: multicentre randomised controlled trial. *British Medical Journal* 2012; 345.
- van de Port IG, Wevers LE, Lindeman E, Kwakkel G. Effects of circuit training as alternative to usual physiotherapy after stroke: randomised controlled trial. *British Medical Journal* 2012; 344.
- van den Broek IV, van Bergen JE, Brouwers EE, Fennema JS, Gotz HM, Hoebe CJ et al. Effectiveness of yearly, register based screening for chlamydia in the Netherlands: controlled trial with randomised stepped wedge implementation. *British Medical Journal* 2012; 344.
- Vernooij J, Kaasjager H, van der Graaf Y, Wierdsma J, Grandjean H, Hovens MMC et al. Internet based vascular risk factor management for patients with clinically manifest vascular disease: randomised controlled trial. *British Medical Journal* 2012; 344.
- Walden M, Atroshi I, Magnusson H, Wagner P, Hagglund M. Prevention of acute knee injuries in adolescent female football players: cluster randomised controlled trial. *British Medical Journal* 2012; 344.
- Waldorff F, Buss D, V, Eckermann A, Rasmussen M, Keiding N, Rishoj S et al. Efficacy of psychosocial intervention in patients with mild Alzheimer's disease: the multicentre, rater blinded, randomised Danish Alzheimer Intervention Study (DAISY). *British Medical Journal* 2012; 345.
- Walsh JM, McGowan CA, Mahony R, Foley ME, McAuliffe FM. Low glycaemic index diet in pregnancy to prevent macrosomia (ROLO study): randomised control trial. *British Medical Journal* 2012; 345.
- Walter FM, Morris HC, Humphrys E, Hall PN, Prevost A, Burrows N et al. Effect of adding a diagnostic aid to best practice to manage suspicious pigmented lesions in primary care: randomised controlled trial. *British Medical Journal* 2012; 344.

Wen LM, Baur LA, Simpson JM, Rissel C, Wardle K, Flood VM. Effectiveness of home based early intervention on children's BMI at age 2: randomised controlled trial. *British Medical Journal* 2012; 344.

Wirostko BM, Tressler C, Hwang LJ, Burgess G, Laties AM. Ocular safety of sildenafil citrate when administered chronically for pulmonary arterial hypertension: results from phase III, randomised, double masked, placebo controlled trial and open label extension. *British Medical Journal* 2012; 344.

Yardley L, Barker F, Muller I, Turner D, Kirby S, Mullee M et al. Clinical and cost effectiveness of booklet based vestibular rehabilitation for chronic dizziness in primary care: single blind, parallel group, pragmatic, randomised controlled trial. *British Medical Journal* 2012; 344.

JAMA - 69 articles

Alberts SR, Sargent DJ, Nair S, Mahoney MR, Mooney M, Thibodeau SN et al. Effect of Oxaliplatin, Fluorouracil, and Leucovorin With or Without Cetuximab on Survival Among Patients With Resected Stage III Colon Cancer A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(13):1383-1393.

Angiolillo DJ, Firstenberg MS, Price MJ, Tummala PE, Hutyra M, Welsby IJ et al. Bridging Antiplatelet Therapy With Cangrelor in Patients Undergoing Cardiac Surgery A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(3):265-274.

Bakker OJ, van Santvoort HC, van Brunschot S, Geskus RB, Besselink MG, Bollen TL et al. Endoscopic Transgastric vs Surgical Necrosectomy for Infected Necrotizing Pancreatitis A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(10):1053-1061.

Barbe F, Duran-Cantolla J, Sanchez-de-la-Torre M, Martinez-Alonso M, Carmona C, Barcelo A et al. Effect of Continuous Positive Airway Pressure on the Incidence of Hypertension and Cardiovascular Events in Nonsleepy Patients With Obstructive Sleep Apnea A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(20):2161-2168.

Barry MJ. Effect of Increasing Doses of Saw Palmetto Extract on Lower Urinary Tract Symptoms: A Randomized Trial (vol 306, pg 1344, 2011). *Jama-Journal of the American Medical Association* 2012; 307(22):2374.

Berwanger O, Guimaraes HP, Laranjeira LN, Cavalcanti AB, Kodama AA, Zazula AD et al. Effect of a Multifaceted Intervention on Use of Evidence-Based Therapies in Patients With Acute Coronary Syndromes in Brazil The BRIDGE-ACS Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(19):2041-2049.

- Bhasin S, Travison TG, Storer TW, Lakshman K, Kaushik M, Mazer NA et al. Early Surgical Therapy for Drug-Resistant Temporal Lobe Epilepsy A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(9):931-939.
- Blumenthal JA, Babyak MA, O'Connor C, Keteyian S, Landzberg J, Howlett J et al. Effects of Exercise Training on Depressive Symptoms in Patients With Chronic Heart Failure The HF-ACTION Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(5):465-474.
- Bray GA, Smith SR, de Jonge L, Xie H, Rood J, Martin CK et al. Effect of Dietary Protein Content on Weight Gain, Energy Expenditure, and Body Composition During Overeating A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(1):47-55.
- Brunkhorst FM, Oppert M, Marx G, Bloos F, Ludewig K, Putensen C et al. Effect of Empirical Treatment With Moxifloxacin and Meropenem vs Meropenem on Sepsis-Related Organ Dysfunction in Patients With Severe Sepsis A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(22):2390-2399.
- Calhoun WJ, Ameredes BT, King TS, Icitovic N, Bleecker ER, Castro M et al. Comparison of Physician-, Biomarker-, and Symptom-Based Strategies for Adjustment of Inhaled Corticosteroid Therapy in Adults With Asthma The BASALT Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(10):987-997.
- Davies HE, Mishra EK, Kahan BC, Wrightson JM, Stanton AE, Guhan A et al. Effect of an Indwelling Pleural Catheter vs Chest Tube and Talc Pleurodesis for Relieving Dyspnea in Patients With Malignant Pleural Effusion The TIME2 Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(22):2383-2389.
- Davis CL, Pollock NK, Waller JL, Allison JD, Dennis B, Bassali R et al. Exercise Dose and Diabetes Risk in Overweight and Obese Children A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(11):1103-1112.
- Dieleman JM, Nierich AP, Rosseel PM, van der Maaten JM, Hofland J, Diephuis JC et al. Intraoperative High-Dose Dexamethasone for Cardiac Surgery A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(17):1761-1767.
- Dixon JB, Schachter LM, O'Brien PE, Jones K, Grima M, Lambert G et al. Surgical vs Conventional Therapy for Weight Loss Treatment of Obstructive Sleep Apnea A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(11):1142-1149.
- Engel J, Jr., McDermott MP, Wiebe S, Langfitt JT, Stern JM, Dewar S et al. Early Surgical Therapy for Drug-Resistant Temporal Lobe Epilepsy A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(9):922-930.

- Fergusson DA, Hebert P, Hogan DL, LeBel L, Rouvinez-Bouali N, Smyth JA et al. Effect of Fresh Red Blood Cell Transfusions on Clinical Outcomes in Premature, Very Low-Birth-Weight Infants The ARIPI Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(14):1443-1451.
- Franklin ME. Cognitive Behavior Therapy Augmentation of Pharmacotherapy in Pediatric Obsessive-Compulsive Disorder: The Pediatric OCD Treatment Study II (POTS II) Randomized Controlled Trial (vol 306, pg 1224, 2011). *Jama-Journal of the American Medical Association* 2012; 307(16):1694.
- Fried MW, Navarro VJ, Afdhal N, Belle SH, Wahed AS, Hawke RL et al. Effect of Silymarin (Milk Thistle) on Liver Disease in Patients With Chronic Hepatitis C Unsuccessfully Treated With Interferon Therapy A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(3):274-282.
- Gallagher TQ, Hill C, Ojha S, Ference E, Keamy DG, Williams M et al. Perioperative Dexamethasone Administration and Risk of Bleeding Following Tonsillectomy in Children A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(12):1221-1226.
- Garbutt JM, Banister C, Spitznagel E, Piccirillo JF. Amoxicillin for Acute Rhinosinusitis A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(7):685-692.
- Gaziano J, Sesso HD, Christen WG, Bubes V, Smith JP, MacFadyen J et al. Multivitamins in the Prevention of Cancer in Men The Physicians' Health Study II Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(18):1871-1880.
- Gossger N, Snape MD, Yu LM, Finn A, Bona G, Esposito S et al. Immunogenicity and Tolerability of Recombinant Serogroup B Meningococcal Vaccine Administered With or Without Routine Infant Vaccinations According to Different Immunization Schedules A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(6):573-582.
- Gregg EW, Chen H, Wagenknecht LE, Clark JM, Delahanty LM, Bantle J et al. Association of an Intensive Lifestyle Intervention With Remission of Type 2 Diabetes. *Jama-Journal of the American Medical Association* 2012; 308(23):2489-2496.
- Gurbel PA, Erlinge D, Ohman EM, Neely B, Neely M, Goodman SG et al. Platelet Function During Extended Prasugrel and Clopidogrel Therapy for Patients With ACS Treated Without Revascularization The TRILOGY ACS Platelet Function Substudy. *Jama-Journal of the American Medical Association* 2012; 308(17):1785-1794.
- Hare JM, Fishman JE, Gerstenblith G, Velazquez DL, Zambrano JP, Suncion VY et al. Comparison of Allogeneic vs Autologous Bone Marrow-Derived Mesenchymal Stem Cells Delivered by Transendocardial Injection in Patients With Ischemic Cardiomyopathy The

- POSEIDON Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(22):2369-2379.
- Holbrook JT, Wise RA, Gold BD, Blake K, Brown ED, Castro M et al. Lansoprazole for Children With Poorly Controlled Asthma A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(4):373-381.
- Hooton TM, Roberts PL, Stapleton AE. Cefpodoxime vs Ciprofloxacin for Short-Course Treatment of Acute Uncomplicated Cystitis A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(6):583-589.
- Isanaka S, Mugusi F, Hawkins C, Spiegelman D, Okuma J, Aboud S et al. Effect of High-Dose vs Standard-Dose Multivitamin Supplementation at the Initiation of HAART on HIV Disease Progression and Mortality in Tanzania A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(15):1535-1544.
- Jakicic JM, Tate DF, Lang W, Davis KK, Polzien K, Rickman AD et al. Effect of a Stepped-Care Intervention Approach on Weight Loss in Adults A Randomized Clinical Trial. *Jama-Journal of the American Medical Association* 2012; 307(24):2617-2626.
- Jakob SM, Ruukonen E, Grounds R, Sarapohja T, Garratt C, Pocock SJ et al. Dexmedetomidine vs Midazolam or Propofol for Sedation During Prolonged Mechanical Ventilation Two Randomized Controlled Trials. *Jama-Journal of the American Medical Association* 2012; 307(11):1151-1160.
- Joensuu H, Eriksson M, Hall KS, Hartmann JT, Pink D, Schuette J et al. One vs Three Years of Adjuvant Imatinib for Operable Gastrointestinal Stromal Tumor A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(12):1265-1272.
- Klevens J, Kee R, Trick W, Garcia D, Angulo FR, Jones R et al. Effect of Screening for Partner Violence on Women's Quality of Life A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(7):681-689.
- Koepke S, Muehlhauser I, Gerlach A, Haut A, Haastert B, Moehler R et al. Effect of a Guideline-Based Multicomponent Intervention on Use of Physical Restraints in Nursing Homes A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(20):2177-2184.
- Kulier R, Guelmezoglu AM, Zamora J, Nieves Plana M, Carroli G, Cecatti JG et al. Effectiveness of a Clinically Integrated e-Learning Course in Evidence-Based Medicine for Reproductive Health Training A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(21):2218-2225.
- Latshang TD, Nussbaumer-Ochsner Y, Henn RM, Ulrich S, Lo Cascio CM, Ledergerber B et al. Effect of Acetazolamide and AutoCPAP Therapy on Breathing Disturbances Among Patients With Obstructive Sleep Apnea Syndrome Who Travel to Altitude A Randomized

- Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(22):2390-2398.
- Lok CE, Moist L, Hemmelgarn BR, Tonelli M, Vazquez MA, Dorval M et al. Effect of Fish Oil Supplementation on Graft Patency and Cardiovascular Events Among Patients With New Synthetic Arteriovenous Hemodialysis Grafts A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(17):1809-1816.
- Lucas BP, Trick WE, Evans AT, Mba B, Smith J, Das K et al. Effects of 2-vs 4-Week Attending Physician Inpatient Rotations on Unplanned Patient Revisits, Evaluations by Trainees, and Attending Physician Burnout A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(21):2199-2207.
- McDermott M. Treadmill Exercise and Resistance Training in Patients With Peripheral Arterial Disease With and Without Intermittent Claudication: A Randomized Trial (vol 301, pg 165, 2009). *Jama-Journal of the American Medical Association* 2012; 307(16):1694.
- Mehta S, Burry L, Cook D, Fergusson D, Steinberg M, Granton J et al. Daily Sedation Interruption in Mechanically Ventilated Critically Ill Patients Cared for With a Sedation Protocol A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(19):1985-1992.
- Mesotten D, Gielen M, Sterken C, Claessens K, Hermans G, Vlasselaers D et al. Neurocognitive Development of Children 4 Years After Critical Illness and Treatment With Tight Glucose Control A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(16):1641-1650.
- Mills KL, Teesson M, Back SE, Brady KT, Baker AL, Hopwood S et al. Integrated Exposure-Based Therapy for Co-occurring Posttraumatic Stress Disorder and Substance Dependence A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(7):690-699.
- Mohr DC, Ho J, Duffecy J, Reifler D, Sokol L, Burns MN et al. Effect of Telephone-Administered vs Face-to-face Cognitive Behavioral Therapy on Adherence to Therapy and Depression Outcomes Among Primary Care Patients A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(21):2278-2285.
- Monson CM, Fredman SJ, Macdonald A, Pukay-Martin ND, Resick PA, Schnurr PP. Effect of Cognitive-Behavioral Couple Therapy for PTSD A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(7):700-709.
- Mozaffarian D, Marchioli R, Macchia A, Sillelta MG, Ferrazzi P, Gardner TJ et al. Fish Oil and Postoperative Atrial Fibrillation The Omega-3 Fatty Acids for Prevention of Postoperative Atrial Fibrillation (OPERA) Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(19):2001-2011.

- Murdoch DR, Slow S, Chambers ST, Jennings LC, Stewart AW, Priest PC et al. Effect of Vitamin D-3 Supplementation on Upper Respiratory Tract Infections in Healthy Adults The VIDARIS Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(13):1333-1339.
- Neoptolemos JP. Effect of Adjuvant Chemotherapy With Fluorouracil Plus Folinic Acid or Gemcitabine vs Observation on Survival in Patients With Resected Periapillary Adenocarcinoma: The ESPAC-3 Periapillary Cancer Randomized Trial (vol 308, pg 147, 2012). *Jama-Journal of the American Medical Association* 2012; 308(18):1861.
- Newman MF, Ferguson T, White JA, Ambrosio G, Koglin J, Nussmeier NA et al. Effect of Adenosine-Regulating Agent Acadesine on Morbidity and Mortality Associated With Coronary Artery Bypass Grafting The RED-CABG Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(2):157-164.
- Ogawa H. Low-Dose Aspirin for Primary Prevention of Atherosclerotic Events in Patients With Type 2 Diabetes: A Randomized Controlled Trial (vol 300, pg 2134, 2008). *Jama-Journal of the American Medical Association* 2012; 308(18):1861.
- Paton NI, Goodall RL, Dunn DT, Franzen S, Collaco-Moraes Y, Gazzard BG et al. Effects of Hydroxychloroquine on Immune Activation and Disease Progression Among HIV-Infected Patients Not Receiving Antiretroviral Therapy A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(4):353-361.
- Perin EC, Willerson JT, Pepine CJ, Henry TD, Ellis SG, Zhao DX et al. Effect of Transendocardial Delivery of Autologous Bone Marrow Mononuclear Cells on Functional Capacity, Left Ventricular Function, and Perfusion in Chronic Heart Failure The FOCUS-CCTRN Trial. *Jama-Journal of the American Medical Association* 2012; 307(16):1717-1726.
- Persson LA, Arifeen S, Ekstrom EC, Rasmussen KM, Frongillo EA, Yunus M. Effects of Prenatal Micronutrient and Early Food Supplementation on Maternal Hemoglobin, Birth Weight, and Infant Mortality Among Children in Bangladesh The MINIMat Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(19):2050-2059.
- Raeber L, Kelbaek H, Ostojic M, Baumbach A, Heg D, Tueller D et al. Effect of Biolimus-Eluting Stents With Biodegradable Polymer vs Bare-Metal Stents on Cardiovascular Events Among Patients With Acute Myocardial Infarction The COMFORTABLE AMI Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(8):777-787.
- Rice TW, Wheeler AP, Thompson B, Steingrub J, Hite R, Moss M et al. ONLINE FIRST Initial Trophic vs Full Enteral Feeding in Patients With Acute Lung Injury The EDEN Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(8):795-803.

- Rosenfeld M, Ratjen F, Brumback L, Daniel S, Rowbotham R, McNamara S et al. Inhaled Hypertonic Saline in Infants and Children Younger Than 6 Years With Cystic Fibrosis The ISIS Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(21):2269-2277.
- Schmidt B, Anderson PJ, Doyle LW, Dewey D, Grunau RE, Asztalos EV et al. Survival Without Disability to Age 5 Years After Neonatal Caffeine Therapy for Apnea of Prematurity. *Jama-Journal of the American Medical Association* 2012; 307(3):275-282.
- Selker HP, Beshansky JR, Sheehan PR, Massaro JM, Griffith JL, D'Agostino RB et al. Out-of-Hospital Administration of Intravenous Glucose-Insulin-Potassium in Patients With Suspected Acute Coronary Syndromes The IMMEDIATE Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(18):1925-1933.
- Sesso HD, Christen WG, Bubes V, Smith JP, MacFadyen J, Schvartz M et al. Multivitamins in the Prevention of Cardiovascular Disease in Men The Physicians' Health Study II Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(17):1751-1760.
- Statland JM, Bundy BN, Wang Y, Rayan DR, Trivedi JR, Sansone VA et al. Mexiletine for Symptoms and Signs of Myotonia in Nondystrophic Myotonia A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 308(13):1357-1365.
- Stockwell MS, Kharbanda EO, Martinez RA, Vargas CY, Vawdrey DK, Camargo S. Effect of a Text Messaging Intervention on Influenza Vaccination in an Urban, Low-Income Pediatric and Adolescent Population A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(16):1702-1708.
- Stone GW, Maehara A, Witzenbichler B, Godlewski J, Parise H, Dambrink JH et al. Intracoronary Abciximab and Aspiration Thrombectomy in Patients With Large Anterior Myocardial Infarction The INFUSE-AMI Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 307(17):1817-1826.
- Sullivan D, Olsson AG, Scott R, Kim JB, Xue A, GebSKI V et al. Effect of a Monoclonal Antibody to PCSK9 on Low-Density Lipoprotein Cholesterol Levels in Statin-Intolerant Patients The GAUSS Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(23):2497-2506.
- Tan J, Wu W, Xu X, Liao L, Zheng F, Messinger S et al. Induction Therapy With Autologous Mesenchymal Stem Cells in Living-Related Kidney Transplants A Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(11):1169-1177.
- Thadhani R, Appelbaum E, Pritchett Y, Chang Y, Wenger J, Tamez H et al. Vitamin D Therapy and Cardiac Structure and Function in Patients With Chronic Kidney Disease The PRIMO Randomized Controlled Trial. *Jama-Journal of the American Medical Association* 2012; 307(7):674-684.

Traverse JH, Henry TD, Pepine CJ, Willerson JT, Zhao DX, Ellis SG et al. Effect of the Use and Timing of Bone Marrow Mononuclear Cell Delivery on Left Ventricular Function After Acute Myocardial Infarction The TIME Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(22):2380-2389.

van de Water W, Markopoulos C, van de Velde CJ, Seynaeve C, Hasenburg A, Rea D et al. Association Between Age at Diagnosis and Disease-Specific Mortality Among Postmenopausal Women With Hormone Receptor-Positive Breast Cancer. *Jama-Journal of the American Medical Association* 2012; 307(6):590-597.

Volpp KG, Shea JA, Small DS, Basner M, Zhu J, Norton L et al. Effect of a Protected Sleep Period on Hours Slept During Extended Overnight In-hospital Duty Hours Among Medical Interns A Randomized Trial. *Jama-Journal of the American Medical Association* 2012; 308(21):2208-2217.

Yeates P, O'Neill P, Mann K, Eva KW. Effect of Exposure to Good vs Poor Medical Trainee Performance on Attending Physician Ratings of Subsequent Performances. *Jama-Journal of the American Medical Association* 2012; 308(21):2226-2232.

Zafonte RD, Bagiella E, Ansel BM, Novack TA, Friedewald WT, Hesdorffer DC et al. Effect of Citicoline on Functional and Cognitive Status Among Patients With Traumatic Brain Injury Citicoline Brain Injury Treatment Trial (COBRIT). *Jama-Journal of the American Medical Association* 2012; 308(19):1993-2000.

Lancet - 86 articles

Abraham W, Adamson P, Bourge R, Aaron M, Costanzo M, Stevenson L et al. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial (vol 377, pg 658, 2011). *Lancet* 2012; 379(9814):412.

Aschner P, Chan J, Owens DR, Picard S, Wang E, Dain MP et al. Insulin glargine versus sitagliptin in insulin-naive patients with type 2 diabetes mellitus uncontrolled on metformin (EASIE): a multicentre, randomised open-label trial. *Lancet* 2012; 379(9833):2262-2269.

Avery AJ, Rodgers S, Cantrill JA, Armstrong S, Cresswell K, Eden M et al. A pharmacist-led information technology intervention for medication errors (PINCER): a multicentre, cluster randomised, controlled trial and cost-effectiveness analysis. *Lancet* 2012; 379(9823):1310-1319.

Baird SJ, Garfein RS, McIntosh CT, Oezler B. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *Lancet* 2012; 379(9823):1320-1329.

Bang YJ, Kim YW, Yang HK, Chung HC, Park YK, Lee KH et al. Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. *Lancet* 2012; 379(9813):315-321.

- Baselga J, Bradbury I, Eidtmann H, Di Cosimo S, de Azambuja E, Aura C et al. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *Lancet* 2012; 379(9816):633-640.
- Bhatnagar S, Wadhwa N, Aneja S, Lodha R, Kabra SK, Natchu UCM et al. Zinc as adjunct treatment in infants aged between 7 and 120 days with probable serious bacterial infection: a randomised, double-blind, placebo-controlled trial. *Lancet* 2012; 379(9831):2072-2078.
- Biere SS, Henegouwen MI, Maas KW, Bonavina L, Rosman C, Roig Garcia J et al. Minimally invasive versus open oesophagectomy for patients with oesophageal cancer: a multicentre, open-label, randomised controlled trial. *Lancet* 2012; 379(9829):1887-1892.
- Bolla M, van Poppel H, Tombal B, Vekemans K, Da Pozzo L, de Reijke TM et al. Postoperative radiotherapy after radical prostatectomy for high-risk prostate cancer: long-term results of a randomised controlled trial (EORTC trial 22911). *Lancet* 2012; 380(9858):2018-2027.
- Budde K, Becker T, Arns W. Everolimus-based, calcineurin-inhibitor-free regimen in recipients of de-novo kidney transplants: an open-label, randomised, controlled trial (vol 377, pg 837, 2011). *Lancet* 2012; 380(9858):1994.
- Buller HR, Gallus AS, Pillion G, Prins MH, Raskob GE. Enoxaparin followed by once-weekly idrabiotaparinux versus enoxaparin plus warfarin for patients with acute symptomatic pulmonary embolism: a randomised, double-blind, double-dummy, non-inferiority trial. *Lancet* 2012; 379(9811):123-129.
- Burant CF, Viswanathan P, Marcinak J, Cao C, Vakilynejad M, Xie B et al. TAK-875 versus placebo or glimepiride in type 2 diabetes mellitus: a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet* 2012; 379(9824):1403-1411.
- Camenzind E, Wijns W, Mauri L, Kurowski V, Parikh K, Gao R et al. Stent thrombosis and major clinical events at 3 years after zotarolimus-eluting or sirolimus-eluting coronary stent implantation: a randomised, multicentre, open-label, controlled trial. *Lancet* 2012; 380(9851):1396-1405.
- Castaigne S, Pautas C, Terre C, Raffoux E, Bordessoule D, Bastie JN et al. Effect of gemtuzumab ozogamicin on survival of adult patients with de-novo acute myeloid leukaemia (ALFA-0701): a randomised, open-label, phase 3 study. *Lancet* 2012; 379(9825):1508-1516.
- Cima RF, Maes IH, Joore MA, Scheyen DJ, El Refaie A, Baguley DM et al. Specialised treatment based on cognitive behaviour therapy versus usual care for tinnitus: a randomised controlled trial. *Lancet* 2012; 379(9830):1951-1959.
- Cohen JA, Coles AJ, Arnold DL, Confavreux C, Fox EJ, Hartung HP et al. Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting

multiple sclerosis: a randomised controlled phase 3 trial. *Lancet* 2012; 380(9856):1819-1828.

Coles AJ, Twyman CL, Arnold DL, Cohen JA, Confavreux C, Fox EJ et al. Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. *Lancet* 2012; 380(9856):1829-1839.

Coovadia HM, Brown ER, Fowler MG, Chipato T, Moodley D, Manji K et al. Efficacy and safety of an extended nevirapine regimen in infant children of breastfeeding mothers with HIV-1 infection for prevention of postnatal HIV-1 transmission (HPTN 046): a randomised, double-blind, placebo-controlled trial. *Lancet* 2012; 379(9812):221-228.

Corbacioglu S, Cesaro S, Faraci M, Valteau-Couanet D, Gruhn B, Rovelli A et al. Defibrotide for prophylaxis of hepatic veno-occlusive disease in paediatric haemopoietic stem-cell transplantation: an open-label, phase 3, randomised controlled trial. *Lancet* 2012; 379(9823):1301-1309.

Crawley J, Sismanidis C, Goodman T, Milligan P. Effect of intermittent preventive treatment for malaria during infancy on serological responses to measles and other vaccines used in the Expanded Programme on Immunization: results from five randomised controlled trials. *Lancet* 2012; 380(9846):1001-1010.

Davalos A, Alvarez-Sabin J, Castillo J, Diez-Tejedor E, Ferro J, Martinez-Vila E et al. Citicoline in the treatment of acute ischaemic stroke: an international, randomised, multicentre, placebo-controlled study (ICTUS trial). *Lancet* 2012; 380(9839):349-357.

DeJesus E, Rockstroh JK, Henry K, Molina JM, Gathe J, Ramanathan S et al. Co-formulated elvitegravir, cobicistat, emtricitabine, and tenofovir disoproxil fumarate versus ritonavir-boosted atazanavir plus co-formulated emtricitabine and tenofovir disoproxil fumarate for initial treatment of HIV-1 infection: a randomised, double-blind, phase 3, non-inferiority trial. *Lancet* 2012; 379(9835):2429-2438.

Diacon AH, Dawson R, von Groote-Bidlingmaier F, Symons G, Venter A, Donald PR et al. 14-day bactericidal activity of PA-824, bedaquiline, pyrazinamide, and moxifloxacin combinations: a randomised trial. *Lancet* 2012; 380(9846):986-993.

El Arifeen S, Mullany LC, Shah R, Mannan I, Rahman SM, Talukder M et al. The effect of cord cleansing with chlorhexidine on neonatal mortality in rural Bangladesh: a community-based, cluster-randomised trial. *Lancet* 2012; 379(9820):1022-1028.

Elena Santolaya M, O'Ryan ML, Teresa Valenzuela M, Prado V, Vergara R, Munoz A et al. Immunogenicity and tolerability of a multicomponent meningococcal serogroup B (4CMenB) vaccine in healthy adolescents in Chile: a phase 2b/3 randomised, observer-blind, placebo-controlled study. *Lancet* 2012; 379(9816):617-624.

Enden T, Haig Y, Klow NE, Slagsvold CE, Sandvik L, Ghanima W et al. Long-term outcome after additional catheter-directed thrombolysis versus standard treatment for acute

- iliofemoral deep vein thrombosis (the CaVenT study): a randomised controlled trial. *Lancet* 2012; 379(9810):31-38.
- Engert A. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial (vol 379, pg 1791, 2012). *Lancet* 2012; 379(9828):1790.
- Fairall L, Bachmann MO, Lombard C, Timmerman V, Uebel K, Zwarenstein M et al. Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): a pragmatic, parallel, cluster-randomised trial. *Lancet* 2012; 380(9845):889-898.
- Fleshner NE, Lucia M, Egerdie B, Aaron L, Eure G, Nandy I et al. Dutasteride in localised prostate cancer management: the REDEEM randomised, double-blind, placebo-controlled trial. *Lancet* 2012; 379(9821):1103-1111.
- Gallwitz B, Rosenstock J, Rauch T, Bhattacharya S, Patel S, von Eynatten M et al. 2-year efficacy and safety of linagliptin compared with glimepiride in patients with type 2 diabetes inadequately controlled on metformin: a randomised, double-blind, non-inferiority trial. *Lancet* 2012; 380(9840):475-483.
- Gallwitz B, Guzman J, Dotta F, Guerci B, Simo R, Basson BR et al. Exenatide twice daily versus glimepiride for prevention of glycaemic deterioration in patients with type 2 diabetes with metformin failure (EUREXA): an open-label, randomised controlled trial. *Lancet* 2012; 379(9833):2270-2278.
- Garber AJ, King AB, Del Prato S, Sreenan S, Balci MK, Munoz-Torres M et al. Insulin degludec, an ultra-longacting basal insulin, versus insulin glargine in basal-bolus treatment with mealtime insulin aspart in type 2 diabetes (BEGIN Basal-Bolus Type 2): a phase 3, randomised, open-label, treat-to-target non-inferiority trial. *Lancet* 2012; 379(9825):1498-1507.
- Garcia PJ, Holmes KK, Carcamo CP, Garnett GP, Hughes JP, Campos PE et al. Prevention of sexually transmitted infections in urban communities (Peru PREVEN): a multicomponent community-randomised controlled trial. *Lancet* 2012; 379(9821):1120-1128.
- Gebre T, Ayele B, Zerihun M, Genet A, Stoller NE, Zhou Z et al. Comparison of annual versus twice-yearly mass azithromycin treatment for hyperendemic trachoma in Ethiopia: a cluster-randomised trial. *Lancet* 2012; 379(9811):143-151.
- Giugliano RP, Desai NR, Kohli P, Rogers WJ, Somaratne R, Huang F et al. Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 in combination with a statin in patients with hypercholesterolaemia (LAPLACE-TIMI 57): a randomised, placebo-controlled, dose-ranging, phase 2 study. *Lancet* 2012; 380(9858):2007-2017.

- Glazener C, Boachie C, Buckley B, Cochran C, Dorey G, Grant A et al. Urinary incontinence in men after formal one-to-one pelvic-floor muscle training following radical prostatectomy or transurethral resection of the prostate (MAPS): two parallel randomised controlled trials (vol 378, pg 328, 2011). *Lancet* 2012; 379(9814):412.
- Goya M, Pratcorona L, Merced C, Rodo C, Valle L, Romero A et al. Cervical pessary in pregnant women with a short cervix (PECEP): an open-label randomised controlled trial. *Lancet* 2012; 379(9828):1800-1806.
- Griffin S, Borch-Johnsen K, Davies M. Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): a cluster-randomised trial (vol 378, pg 156, 2011). *Lancet* 2012; 379(9818):804.
- Guelmezoglu A, Lumbiganon P, Landoulsi S, Widmer M, Abdel-Aleem H, Festin M et al. Active management of the third stage of labour with and without controlled cord traction: a randomised, controlled, non-inferiority trial. *Lancet* 2012; 379(9827):1721-1727.
- Hanney M, Prasher V, Williams N, Jones EL, Aarsland D, Corbett A et al. Memantine for dementia in adults older than 40 years with Down's syndrome (MEADOWS): a randomised, double-blind, placebo-controlled trial. *Lancet* 2012; 379(9815):528-536.
- Hauschild A, Grob JJ, Demidov LV, Jouary T, Gutzmer R, Millward M et al. Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *Lancet* 2012; 380(9839):358-365.
- Heller S, Buse J, Fisher M, Garg S, Marre M, Merker L et al. Insulin degludec, an ultra-longacting basal insulin, versus insulin glargine in basal-bolus treatment with mealtime insulin aspart in type 1 diabetes (BEGIN Basal-Bolus Type 1): a phase 3, randomised, open-label, treat-to-target non-inferiority trial. *Lancet* 2012; 379(9825):1489-1497.
- Jamieson DJ, Chasela CS, Hudgens MG, King CC, Kourtis AP, Kayira D et al. Maternal and infant antiretroviral regimens to prevent postnatal HIV-1 transmission: 48-week follow-up of the BAN randomised controlled trial. *Lancet* 2012; 379(9835):2449-2458.
- Johnston C, Saracino M, Kuntz S, Magaret A, Selke S, Huang MI et al. Standard-dose and high-dose daily antiviral therapy for short episodes of genital HSV-2 reactivation: three randomised, open-label, cross-over trials. *Lancet* 2012; 379(9816):641-647.
- Kirchhof P, Andresen D, Bosch R, Borggrefe M, Meinertz T, Parade U et al. Short-term versus long-term antiarrhythmic drug treatment after cardioversion of atrial fibrillation (Flec-SL): a prospective, randomised, open-label, blinded endpoint assessment trial. *Lancet* 2012; 380(9838):238-246.
- Kirsner RS, Marston WA, Snyder RJ, Lee TD, Cargill D, I, Slade HB. Spray-applied cell therapy with human allogeneic fibroblasts and keratinocytes for the treatment of chronic

- venous leg ulcers: a phase 2, multicentre, double-blind, randomised, placebo-controlled trial. *Lancet* 2012; 380(9846):977-985.
- Kobayashi T, Saji T, Otani T, Takeuchi K, Nakamura T, Arakawa H et al. Efficacy of immunoglobulin plus prednisolone for prevention of coronary artery abnormalities in severe Kawasaki disease (RAISE study): a randomised, open-label, blinded-endpoints trial. *Lancet* 2012; 379(9826):1613-1620.
- Koren MJ, Scott R, Kim JB, Knusel B, Liu T, Lei L et al. Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 as monotherapy in patients with hypercholesterolaemia (MENDEL): a randomised, double-blind, placebo-controlled, phase 2 study. *Lancet* 2012; 380(9858):1995-2006.
- Laharie D, Bourreille A, Branche J, Allez M, Bouhnik Y, Filippi J et al. Ciclosporin versus infliximab in patients with severe ulcerative colitis refractory to intravenous steroids: a parallel, open-label randomised controlled trial. *Lancet* 2012; 380(9857):1909-1915.
- Lopes RD, Al-Khatib SM, Wallentin L, Yang H, Ansell J, Cecilia Bahit M et al. Efficacy and safety of apixaban compared with warfarin according to patient risk of stroke and of bleeding in atrial fibrillation: a secondary analysis of a randomised controlled trial. *Lancet* 2012; 380(9855):1749-1758.
- Makkar RR, Smith RR, Cheng K, Malliaras K, Thomson LE, Berman D et al. Intracoronary cardiosphere-derived cells for heart regeneration after myocardial infarction (CADUCEUS): a prospective, randomised phase 1 trial. *Lancet* 2012; 379(9819):895-904.
- Manaseki-Holland S, Maroof Z, Bruce J, Mughal M, Masher MI, Bhutta ZA et al. Effect on the incidence of pneumonia of vitamin D supplementation by quarterly bolus dose to infants in Kabul: a randomised controlled superiority trial. *Lancet* 2012; 379(9824):1419-1427.
- Mees SM, Algra A, Vandertop W, van Kooten F, Kuijsten HA, Boiten J et al. Magnesium for aneurysmal subarachnoid haemorrhage (MASH-2): a randomised placebo-controlled trial. *Lancet* 2012; 380(9836):44-49.
- Mehilli J, Pache J, Abdel-Wahab M. Is Drug-Eluting-Stenting Associated with Improved Results in Coronary Artery Bypass Grafts? (ISAR-CABG) Investigators. Drug-eluting versus bare-metal stents in saphenous vein graft lesions (ISAR-CABG): a randomised controlled superiority trial (vol 378, pg 1071, 2011). *Lancet* 2012; 379(9811):122.
- Meirik O, Nguyen Thi MH, Piaggio G, Bergel E, von Hertzen H. Complications of first-trimester abortion by vacuum aspiration after cervical preparation with and without misoprostol: a multicentre randomised trial. *Lancet* 2012; 379(9828):1817-1824.
- Mitja O, Hays R, Ipai A, Penias M, Paru R, Fagaho D et al. Single-dose azithromycin versus benzathine benzylpenicillin for treatment of yaws in children in Papua New Guinea: an open-label, non-inferiority, randomised trial. *Lancet* 2012; 379(9813):342-347.

- Nijhof SL, Bleijenberg G, Uiterwaal CS, Kimpen JL, van de Putte EM. Effectiveness of internet-based cognitive behavioural treatment for adolescents with chronic fatigue syndrome (FITNET): a randomised controlled trial. *Lancet* 2012; 379(9824):1412-1418.
- Nogueira RG, Lutsep HL, Gupta R, Jovin TG, Albers GW, Walker GA et al. Trevo versus Merci retrievers for thrombectomy revascularisation of large vessel occlusions in acute ischaemic stroke (TREVO 2): a randomised trial. *Lancet* 2012; 380(9849):1231-1240.
- Papp K, Cather JC, Rosoph L, Sofen H, Langley RG, Matheson RT et al. Efficacy of apremilast in the treatment of moderate to severe psoriasis: a randomised controlled trial. *Lancet* 2012; 380(9843):738-746.
- Pavord ID, Korn S, Howarth P, Bleecker ER, Buhl R, Keene ON et al. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. *Lancet* 2012; 380(9842):651-659.
- Pickard R, Lam T, MacLennan G, Starr K, Kilonzo M, McPherson G et al. Antimicrobial catheters for reduction of symptomatic urinary tract infection in adults requiring short-term catheterisation in hospital: a multicentre randomised controlled trial. *Lancet* 2012; 380(9857):1927-1935.
- Rickard CM, Webster J, Wallis MC, Marsh N, McGrail MR, French V et al. Routine versus clinically indicated replacement of peripheral intravenous catheters: a randomised controlled equivalence trial. *Lancet* 2012; 380(9847):1066-1074.
- Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the JUPITER trial. *Lancet* 2012; 380(9841):565-571.
- Rini B, Escudier B, Tomczak P. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial (vol 378, pg 1931, 2011). *Lancet* 2012; 380(9856):1818.
- Roberts JD, Wells GA, Le May MR, Labinaz M, Glover C, Froeschl M et al. Point-of-care genetic testing for personalisation of antiplatelet treatment (RAPID GENE): a prospective, randomised, proof-of-concept trial. *Lancet* 2012; 379(9827):1705-1711.
- Ryan NM, Birring SS, Gibson PG. Gabapentin for refractory chronic cough: a randomised, double-blind, placebo-controlled trial. *Lancet* 2012; 380(9853):1583-1589.
- Sabate M, Cequier A, Iniguez A, Serra A, Hernandez-Antolin R, Mainar V et al. Everolimus-eluting stent versus bare-metal stent in ST-segment elevation myocardial infarction (EXAMINATION): 1 year results of a randomised controlled trial. *Lancet* 2012; 380(9852):1482-1490.
- Sabchareon A, Wallace D, Sirivichayakul C, Limkittikul K, Chanthavanich P, Suvannadabba S et al. Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue

- vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. *Lancet* 2012; 380(9853):1559-1567.
- Sandberg T, Skoog G, Hermansson AB, Kahlmeter G, Kuylenstierna N, Lannergard A et al. Ciprofloxacin for 7 days versus 14 days in women with acute pyelonephritis: a randomised, open-label and double-blind, placebo-controlled, non-inferiority trial. *Lancet* 2012; 380(9840):484-490.
- Sandercock P, Wardlaw JM, Lindley RI, Dennis M, Cohen G, Murray G et al. The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial. *Lancet* 2012; 379(9834):2352-2363.
- Saver JL, Jahan R, Levy EI, Jovin TG, Baxter B, Nogueira RG et al. Solitaire flow restoration device versus the Merci Retriever in patients with acute ischaemic stroke (SWIFT): a randomised, parallel-group, non-inferiority trial. *Lancet* 2012; 380(9849):1241-1249.
- Sax PE, DeJesus E, Mills A, Zolopa A, Cohen C, Wohl D et al. Co-formulated elvitegravir, cobicistat, emtricitabine, and tenofovir versus co-formulated efavirenz, emtricitabine, and tenofovir for initial treatment of HIV-1 infection: a randomised, double-blind, phase 3 trial, analysis of results after 48 weeks. *Lancet* 2012; 379(9835):2439-2448.
- Scirica BM, Bonaca MP, Braunwald E, De Ferrari GM, Isaza D, Lewis BS et al. Vorapaxar for secondary prevention of thrombotic events for patients with previous myocardial infarction: a prespecified subgroup analysis of the TRA 2 degrees P-TIMI 50 trial. *Lancet* 2012; 380(9850):1317-1324.
- Simmons RK, Echouffo-Tcheugui JB, Sharp SJ, Sargeant LA, Williams KM, Prevost A et al. Screening for type 2 diabetes and population mortality over 10 years (ADDITION-Cambridge): a cluster-randomised controlled trial. *Lancet* 2012; 380(9855):1741-1748.
- Smith FG, Perkins GD, Gates S, Young D, McAuley DF, Tunnicliffe W et al. Effect of intravenous beta-2 agonist treatment on clinical outcomes in acute respiratory distress syndrome (BALTI-2): a multicentre, randomised controlled trial. *Lancet* 2012; 379(9812):229-235.
- Smith MR, Saad F, Coleman R, Shore N, Fizazi K, Tombal B et al. Denosumab and bone-metastasis-free survival in men with castration-resistant prostate cancer: results of a phase 3, randomised, placebo-controlled trial. *Lancet* 2012; 379(9810):39-46.
- Solomon SD, Zile M, Pieske B, Voors AA, Shah A, Kraigher-Krainer E et al. The angiotensin receptor neprilysin inhibitor LCZ696 in heart failure with preserved ejection fraction: a phase 2 double-blind randomised controlled trial. *Lancet* 2012; 380(9851):1387-1395.
- Soofi S, Cousens S, Imdad A, Bhutto N, Ali N, Bhutta ZA. Topical application of chlorhexidine to neonatal umbilical cords for prevention of omphalitis and neonatal mortality in a rural

district of Pakistan: a community-based, cluster-randomised trial. *Lancet* 2012; 379(9820):1029-1036.

Soofi S, Ahmed S, Fox MP, MacLeod WB, Thea DM, Qazi SA et al. Effectiveness of community case management of severe pneumonia with oral amoxicillin in children aged 2-59 months in Matiari district, rural Pakistan: a cluster-randomised controlled trial. *Lancet* 2012; 379(9817):729-737.

Stein EA, Gipe D, Bergeron J, Gaudet D, Weiss R, Dufour R et al. Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a phase 2 randomised controlled trial. *Lancet* 2012; 380(9836):29-36.

Thiele H, Woehrle J, Hambrecht R, Rittger H, Birkemeyer R, Lauer B et al. Intracoronary versus intravenous bolus abciximab during primary percutaneous coronary intervention in patients with acute ST-elevation myocardial infarction: a randomised trial. *Lancet* 2012; 379(9819):923-931.

van der Graaf WT, Blay JY, Chawla SP, Kim DW, Binh B, Casali PG et al. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet* 2012; 379(9829):1879-1886.

van Vollenhoven RF, Geborek P, Forslind K, Albertsson K, Ernestam S, Petersson IF et al. Conventional combination treatment versus biological treatment in methotrexate-refractory early rheumatoid arthritis: 2 year follow-up of the randomised, non-blinded, parallel-group Swefot trial. *Lancet* 2012; 379(9827):1712-1720.

Wandt H, Schaefer-Eckart K, Wendelin K, Pilz B, Wilhelm M, Thalheimer M et al. Therapeutic platelet transfusion versus routine prophylactic transfusion in patients with haematological malignancies: an open-label, multicentre, randomised study. *Lancet* 2012; 380(9850):1309-1316.

Wong C, Jayaram L, Karalus N, Eaton T, Tong C, Hockey H et al. Azithromycin for prevention of exacerbations in non-cystic fibrosis bronchiectasis (EMBRACE): a randomised, double-blind, placebo-controlled trial. *Lancet* 2012; 380(9842):660-667.

Zinkstok SM, Roos YB. Early administration of aspirin in patients treated with alteplase for acute ischaemic stroke: a randomised controlled trial. *Lancet* 2012; 380(9843):731-737.

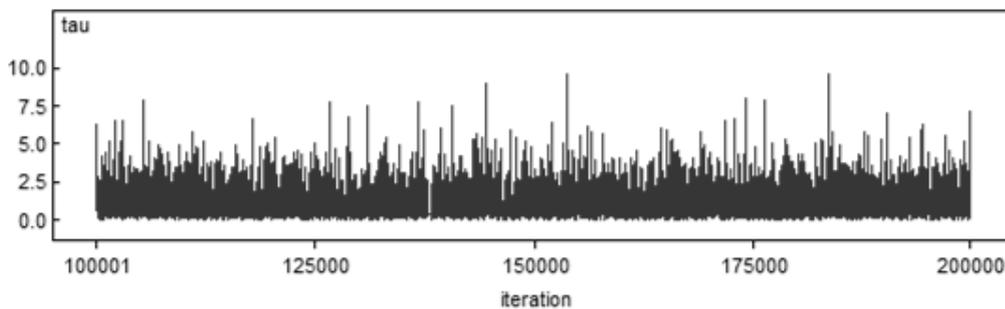
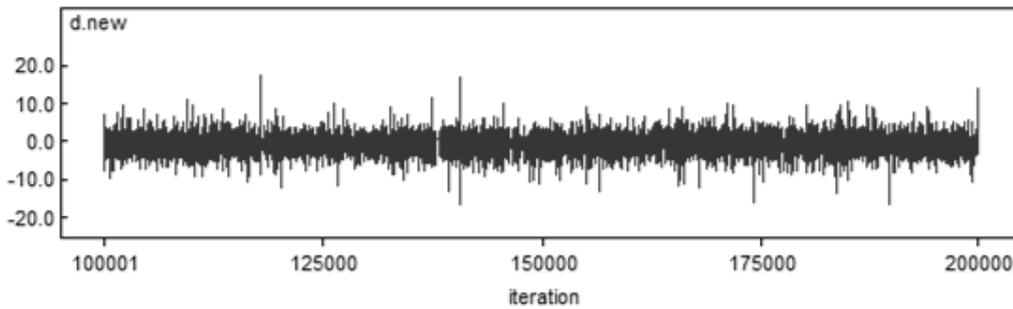
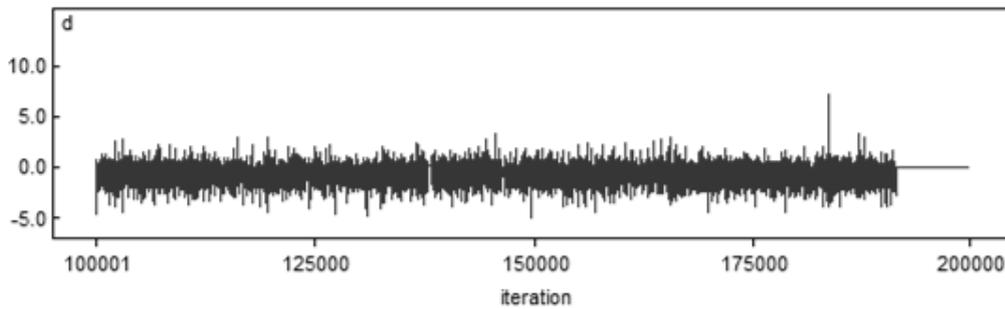
APPENDIX B

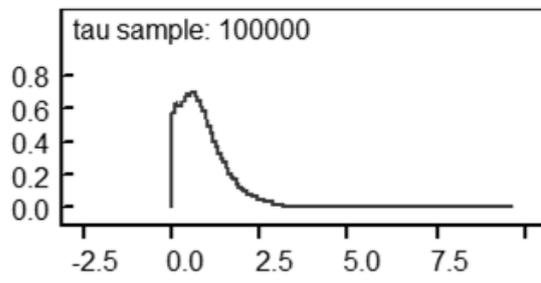
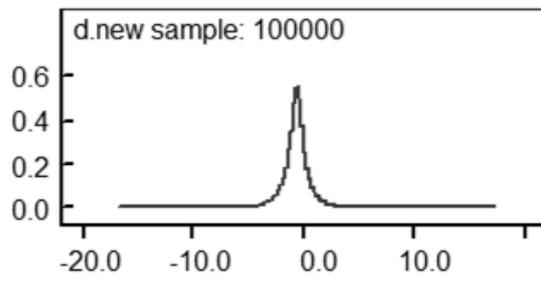
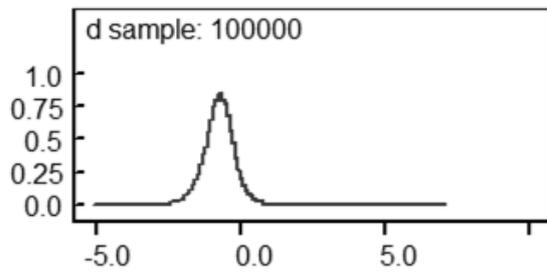
Appendix B1: WinBUGS code for univariate random-effects meta-analysis using a logistic regression model.

```
Model {  
  for( i in 1:Nstud ) {  
    rinf[i] ~ dbin(pinf[i], ninf[i])  
    rbol[i] ~ dbin(pbol[i], nbol[i])  
    logit(pinf[i]) <- mu[i]  
    logit(pbol[i]) <- mu[i] + delta[i]  
    mu[i] ~ dnorm(0.0, 1.0E-6)  
    delta[i] ~ dnorm(d, prec)  
  }  
  OR <- exp(d)  
  d ~ dnorm(0.0, 1.0E-6)  
  tau ~ dnorm(0, 1) | (0,)  
  tau.sq <- tau*tau  
  prec <- 1/(tau.sq)  
  d.new ~ dnorm(d, prec)  
  OR.new <- exp(d.new)  
  prob1 <- step(1-OR)  
  prob1.new <- step(1-OR.new)  
}
```

Appendix B2: History and trace plots to display convergence of parameters from a univariate random-effects logistic regression model.

In these plots, d is the pooled treatment effect, $d.new$ is the true treatment effect in a new study, and τ is the between-study standard deviation.





Appendix B3: Publication in *Trials* arising from the univariate meta-analysis of phase II trials that was explored in Chapter 4.

Burke et al. *Trials* 2014, **15**:346
<http://www.trialsjournal.com/content/15/1/346>



RESEARCH

Open Access

Meta-analysis of randomized phase II trials to inform subsequent phase III decisions

Danielle L Burke^{1*}, Lucinda J Billingham^{1,2}, Alan J Girling¹ and Richard D Riley³

Abstract

Background: If multiple Phase II randomized trials exist then meta-analysis is favorable to increase statistical power and summarize the existing evidence about an intervention's effect in order to help inform Phase III decisions. We consider some statistical issues for meta-analysis of Phase II trials for this purpose, as motivated by a real example involving nine Phase II trials of bolus thrombolytic therapy in acute myocardial infarction with binary outcomes.

Methods: We propose that a Bayesian random effects logistic regression model is most suitable as it models the binomial distribution of the data, helps avoid continuity corrections, accounts for between-trial heterogeneity, and incorporates parameter uncertainty when making inferences. The model also allows predictions that inform Phase III decisions, and we show how to derive: (i) the probability that the intervention will be truly beneficial in a new trial, and (ii) the probability that, in a new trial with a given sample size, the 95% credible interval for the odds ratio will be entirely in favor of the intervention. As Phase II trials are potentially optimistic due to bias in design and reporting, we also discuss how skeptical prior distributions can reduce this optimism to make more realistic predictions.

Results: In the example, the model identifies heterogeneity in intervention effect missed by an I-squared of 0%. Prediction intervals accounting for this heterogeneity are shown to support subsequent Phase III trials. The probability of success in Phase III trials increases as the sample size increases, up to 0.82 for intracranial hemorrhage and 0.79 for reinfarction outcomes.

Conclusions: The choice of meta-analysis methods can influence the decision about whether a trial should proceed to Phase III and thus need to be clearly documented and investigated whenever a Phase II meta-analysis is performed.

Keywords: Meta-analysis, Phase II and III, Prediction, Heterogeneity, Bayesian

Background

Phase III trials are rigorous evaluations of an intervention (such as a new drug or surgical technique), and are typically protocol-driven with large patient numbers, appropriate statistical power, and a suitable trial design and analysis plan. However, the decision to initiate a Phase III trial for a particular intervention is not straightforward and depends on many factors, such as costs, risks (to the trial funders and patients), and practicalities such as patient recruitment [1]. Perhaps the most pivotal factor is the intervention's likely effectiveness. Clearly, the more likely an intervention is to succeed, the more likely

funders will risk investment in a Phase III trial. To this end, before initiation of a Phase III trial funders will consider the existing evidence about an intervention's potential benefit, for example from earlier Phase trials.

The initial estimate of the intervention effect often arises from a Phase II randomized trial. These typically contain small patient numbers or events, and give an imprecise intervention effect estimate with a wide 95% confidence interval. However, sometimes multiple Phase II trials are conducted, for example in slightly different patient groups or by different (or competing) researchers (or companies) working on the same or similar interventions. In this situation, a meta-analysis is useful to increase statistical power [2] by combining the statistical estimates (such as odds ratios (ORs)) from the multiple trials and thereby summarizing the intervention effect

* Correspondence: d.burke@bham.ac.uk

¹Medical Research Council Midland Hub for Trials Methodology Research, School of Health and Population Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Full list of author information is available at the end of the article



© 2014 Burke et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

APPENDIX C

Appendix C1: WinBUGS code for bivariate random-effects normal meta-analysis model (5.1).

```
model {
  for (i in 1:N) {
    ws_covmat[i,1,1] <- sigmasq1[i]+tausq[1]
    ws_covmat[i,2,2] <- sigmasq2[i]+tausq[2]
    ws_covmat[i,1,2] <- ws_cov[i] + (rhob*sqrt(tausq[1]*tausq[2]))
    ws_covmat[i,2,1] <- ws_covmat[i,1,2]
    ws_cov[i]<-rhov[i]*sqrt(sigmasq1[i])*sqrt(sigmasq2[i])
    rhov[i]~dunif(0,1)
    ws_prec[i,1:2,1:2] <- inverse(ws_covmat[i, , ])
    Y[i,1:2] ~ dnorm(logOR[1:2],ws_prec[i,1:2,1:2])
  }
  tausq[1]<-tau[1]*tau[1]
  tausq[2]<-tau[2]*tau[2]
  tau[1]~dnorm(0,1)|(0,)
  tau[2]~dnorm(0,1)|(0,)
  bs_covmat[1,1]<-tausq[1]
  bs_covmat[1,2]<-rhob*sqrt(tausq[1]*tausq[2])
  bs_covmat[2,1]<-bs_covmat[1,2]
  bs_covmat[2,2]<-tausq[2]
  bs_prec[1:2,1:2]<-inverse(bs_covmat[,])
  logOR[1]~dnorm(0.0,1.0E-6)
  logOR[2]~dnorm(0.0,1.0E-6)
  rhob~dunif(-1,1)
```

```

OR[1]<-exp(logOR[1])
OR[2]<-exp(logOR[2])
logORnew[1:2]~dmnorm(logOR[1:2],bs_prec[1:2,1:2])
ORnew[1]<-exp(logORnew[1])
ORnew[2]<-exp(logORnew[2])
for(j in 1:2) {
  prob1[j]<-step(1-OR[j])
  prob1.new[j]<-step(1-ORnew[j])
}
prob1joint<-step((step(1-OR[1])+step(1-OR[2]))-1.5)
prob1jointnew<-step((step(1-ORnew[1])+step(1-ORnew[2]))-1.5)
}

```

Appendix C2: WinBUGS code for bivariate random-effects normal meta-analysis model (5.2).

```
model {
  for (i in 1:N) {
    covmat[i,1,1] <- sigmasq1[i]+psisq[1]
    covmat[i,2,2] <- sigmasq2[i]+psisq[2]
    covmat[i,1,2] <- rho*sqrt((psisq[1]+sigmasq1[i])*(psisq[2]+sigmasq2[i]))
    covmat[i,2,1] <- covmat[i,1,2]
    prec[i,1:2,1:2] <- inverse(covmat[i, , ])
    Y[i,1:2] ~ dnorm(logOR[1:2],prec[i,1:2,1:2])
  }
  rho~dunif(-1,1)
  psisq[1]<-psi[1]*psi[1]
  psisq[2]<-psi[2]*psi[2]
  psi[1]~dnorm(0,1)l(0,)
  psi[2]~dnorm(0,1)l(0,)
  pred_covmat[1,1]<-psisq[1]
  pred_covmat[1,2]<-rho*sqrt(psisq[1]*psisq[2])
  pred_covmat[2,1]<-pred_covmat[1,2]
  pred_covmat[2,2]<-psisq[2]
  pred_prec[1:2,1:2]<-inverse(pred_covmat[,])
  logOR[1]~dnorm(0.0,1.0E-6)
  logOR[2]~dnorm(0.0,1.0E-6)
  OR[1]<-exp(logOR[1])
  OR[2]<-exp(logOR[2])
  logORnew[1:2]~dmnorm(logOR[1:2],pred_prec[1:2,1:2])
  ORnew[1]<-exp(logORnew[1])
}
```

```
ORnew[2]<-exp(logORnew[2])
for(j in 1:2) {
  prob1[j]<-step(1-OR[j])
  prob1.new[j]<-step(1-ORnew[j])
}
prob1joint<-step((step(1-OR[1])+step(1-OR[2]))-1.5)
prob1jointnew<-step((step(1-ORnew[1])+step(1-ORnew[2]))-1.5)
}
```

Appendix C3: Results for secondary bivariate meta-analysis of reinfarction and mortality.

Table C1: 95% prediction intervals and posterior probability estimates for reinfarction and mortality for model (5.1).

	Assumption	95% prediction interval OR	Prob(new trial OR<1)
Reinfarction	$\rho_B=\rho_{W_i}=0$	0.300 to 1.879	0.814
	$\rho_{W_i}=0$	0.299 to 1.875	0.815
	$\rho_{W_i}=0.2$	0.291 to 1.874	0.819
	$\rho_{W_i}=0.5$	0.255 to 2.044	0.807
	$\rho_{W_i}=0.9$	0.188 to 2.634	0.762
	$\rho_{W_i}\sim\text{Unif}(0,1)$	0.268 to 2.026	0.795
Mortality	$\rho_B=\rho_{W_i}=0$	0.290 to 3.889	0.554
	$\rho_{W_i}=0$	0.286 to 3.566	0.579
	$\rho_{W_i}=0.2$	0.285 to 3.587	0.583
	$\rho_{W_i}=0.5$	0.250 to 4.139	0.567
	$\rho_{W_i}=0.9$	0.181 to 5.926	0.530
	$\rho_{W_i}\sim\text{Unif}(0,1)$	0.268 to 3.802	0.597

Table C2: Joint posterior probability estimates for reinfarction and mortality.

Model	Assumption	Prob(both OR<1)	Prob(both new trial OR<1)	Between-study correlation (95% CrI)
Model (5.1)	$(\rho_B=\rho_{W_i}=0)$	0.521	0.451	-
	$\rho_{W_i}=0$	0.557	0.476	0.046 (-0.941 to 0.952)
	$\rho_{W_i}=0.2$	0.569	0.481	-0.025 (-0.949 to 0.941)
	$\rho_{W_i}=0.5$	0.559	0.454	-0.195 (-0.962 to 0.893)
	$\rho_{W_i}=0.9$	0.504	0.373	-0.453 (-0.976 to 0.653)
	$\rho_{W_i}\sim\text{Unif}(0,1)$	0.589	0.476	-0.157 (-0.961 to 0.917)
Model (5.2)	$\rho_O\sim\text{Unif}(-1,1)$	0.543	-	0.059 (-0.501 to 0.575)

APPENDIX D

Appendix D1: SAS code to fit Bayesian bivariate random-effects meta-analysis on 1000 simulated datasets using MCMC

```
PROC MCMC data=sim.N10wc08bc08t105t205b10b22 nbi=200000
nmc=300000 ntu=1000 thin=100 stats=all simreport=10 seed=100
diagnostics=none monitor=(beta_A beta_B sdtau_A sdtau_B tau_A tau_B rhoB
cov_AB prob_thetanewA prob_thetanewB prob_thetanewjoint);

title 'Bayesian bivariate random-effects model
      sdtau~Normal(0,2)I(0,) rhoB ~ Uniform(0,1)';

by simid;

ods select Postsummaries PostIntervals ;

ods output PostSummaries=bivsim.b_ws08bs08tN02rh01_pool
PostIntervals=bivsim.b_ws08bs08tN02rh01_cred
PosteriorSample=bivsim.b_ws08bs08tN02rh01_post;

/* declarations and initializations */
array theta[2] theta_A theta_B; /* includes study-specific random effects
*/
array beta[2] beta_A beta_B;
array omega[2,2] tau_A cov_AB cov_AB2 tau_B; /* Between-study covariance
matrix */
array Sigma[2,2] s_A wscov_AB wscov_AB2 s_B; /* Within-study covariance
matrix */
array y[2] y_A y_B;
array thetanew[2] thetanew_A thetanew_B;

tau_A=sdtau_A*sdtau_A;
tau_B=sdtau_B*sdtau_B;
cov_AB=rhoB*sdtau_A*sdtau_B;
cov_AB2=cov_AB;
y_A=yA;
y_B=yB;
s_A=sA*sA;
s_B=sB*sB;
wscov_AB=rhoW*sA*sB;
wscov_AB2=wscov_AB;

parms rhoB ;
parms beta: ;
parms sdtau_A sdtau_B;

prior beta: ~ normal(0,var=1000);
prior sdtau: ~ normal(0,var=2,lower=0); /*truncated normal distribution*/
prior rhoB ~ uniform(0,1);

random theta ~ mvn(beta, omega) subject=id;
model y ~ mvn(theta, Sigma);
random thetanew ~ mvn(beta, omega) subject=id;
```

```

prob_thetanewA = (thetanew_A>0);
prob_thetanewB = (thetanew_B>2);
prob_thetanewjoint = (thetanew_A>0 & thetanew_B>2);

run;

/* create different data set to run analyses from */
data bivsim.buse_ws08bs08tN02rh01_pool;
    set bivsim.b_ws08bs08tN02rh01_pool;
run;

/* sort dataset by parameter */
proc sort data=bivsim.buse_ws08bs08tN02rh01_pool;
    by parameter;
run;

/* estimate mean pooled estimate, stderr pooled estimate, average variance
*/
proc means data=bivsim.buse_ws08bs08tN02rh01_pool n mean std min max
median stderr;
    ods listing;
    var mean StdDev P50;
    by parameter;
run;

/* compute bias-squared in each simulated dataset */
data bivsim.buse_ws08bs08tN02rh01_pool;
    set bivsim.buse_ws08bs08tN02rh01_pool;
    MSE_A=(mean-0)**2; /* bias squared for outcome 1 */
    MSE_B=(mean-2)**2; /* bias squared for outcome 2 */
run;

/* estimate MSE of pooled estimate for outcome 1 */
proc means data=bivsim.buse_ws08bs08tN02rh01_pool n mean std min max
median stderr ;
    ods listing;
    var MSE_A;
    by parameter;
    where parameter='beta_A';
run;

/* estimate MSE of pooled estimate for outcome 2 */
proc means data=bivsim.buse_ws08bs08tN02rh01_pool n mean std min max
median stderr ;
    ods listing;
    var MSE_B;
    by parameter;
    where parameter='beta_B';
run;

data bivsim.buse_ws08bs08tN02rh01_cred;
    set bivsim.b_ws08bs08tN02rh01_cred;
run;

proc sort data=bivsim.buse_ws08bs08tN02rh01_cred;
    by parameter;
run;

```

```

/* Performance of Credible intervals */
data bivsim.cri_ws08bs08tN02rhou01;
    set bivsim.buse_ws08bs08tN02rhou01_cred;
    if CredibleLower >0 then blhigh=1;
    else blhigh=0;
    if CredibleUpper <0 then bllow=1;
    else bllow=0;
    if CredibleLower >2 then b2high=1;
    else b2high=0;
    if CredibleUpper <2 then b2low=1;
    else b2low=0;
run;

/* Performance of Credible intervals for beta1 */
proc freq data=bivsim.cri_ws08bs08tN02rhou01;
    by parameter;
    where parameter='beta_A';
run;

/* Performance of Credible intervals for beta2 */
proc freq data=bivsim.cri_ws08bs08tN02rhou01;
    by parameter;
    where parameter='beta_B';
run;

```

Appendix D2: Results of simulation study with various prior distributions for between-study variance in univariate meta-analysis model

Table D1: Univariate simulation study results for various prior distributions for between-study variance, τ_j .

Prior for τ_j	Mean of $\hat{\beta}_1$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	No. of 95% CrIs for $\hat{\beta}_1$ including β_1 (%)	Mean prob ($\hat{\theta}_{i_{new}} > 0$)	Mean of $\hat{\beta}_2$ (s.e. of mean)	Mean/median s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	No. of 95% CrIs for $\hat{\beta}_2$ including β_2 (%)	Mean prob ($\hat{\theta}_{i_{new}} > 2$)	Mean/median $\hat{\tau}_1$	Mean/median $\hat{\tau}_2$
1a. $\tau_j \sim N(0,1), \tau_j > 0$	-0.0018 (0.1954)	0.2106/ 0.2094	0.0382	951 (95.1)	0.4955	2.0003 (0.2184)	0.2502/ 0.2492	0.0476	962 (96.2)	0.4988	0.4812/ 0.4839	0.5070/ 0.5081
1b. $\tau_j \sim N(0,2), \tau_j > 0$	-0.0019 (0.1953)	0.2156/ 0.2133	0.0381	954 (95.4)	0.4962	2.0011 (0.2189)	0.2594/ 0.2575	0.0479	965 (96.5)	0.4989	0.4934/ 0.4952	0.5249/ 0.5225
2a. $1/\tau_j^2 \sim \text{Gamma}(0.1,0.1)$	-0.0017 (0.1979)	0.5909/ 0.5888	0.0391	1000 (100.0)	0.4996	1.9982 (0.2585)	0.6820/ 0.6765	0.0668	1000 (100.0)	0.4993	1.6820/ 1.6740	1.8346/ 1.8189
2b. $1/\tau_j^2 \sim \text{Gamma}(1,1)$	-0.0018 (0.1962)	0.2611/ 0.2567	0.0384	988 (98.8)	0.4984	1.9996 (0.2206)	0.3106/ 0.3049	0.0486	989 (98.9)	0.4993	0.6835/ 0.6703	0.7315/ 0.7148
3. $\tau_j \sim \text{Unif}(0,2)$	-0.0021 (0.1953)	0.2210/ 0.2174	0.0381	958 (95.8)	0.4962	2.0007 (0.2190)	0.2675/ 0.2655	0.0479	968 (96.8)	0.4991	0.5061/ 0.5032	0.5460/ 0.5446
4. $\text{Log}(\tau_j^2) \sim \text{Unif}(-10,1.386)$	-0.0023 (0.1955)	0.1925/ 0.1878	0.0382	919 (91.9)	0.4955	2.0008 (0.2191)	0.2254/ 0.2203	0.0480	923 (92.3)	0.4990	0.3870/ 0.4002	0.3934/ 0.3935

Additional detail for the findings in the univariate simulation study

The mean pooled treatment effect estimates across the 1000 simulated datasets are very similar for each prior distribution, and they are close to the true values of the pooled treatment effects. For example, for between-study variance prior distribution 1a, the mean $\hat{\beta}_1$ is -0.002 compared to $\beta_1=0$ (s.e of the mean 0.195), and the mean $\hat{\beta}_2$ is 2.000 compared to $\beta_2=2$ (s.e of mean 0.218). However, as expected, the average between-study standard deviations, $\hat{\tau}_1$ and $\hat{\tau}_2$, vary depending on the prior distribution for these parameters. Since we know that the true values of τ_1 and τ_2 are 0.5, the most appropriate prior distribution for τ_j in this example is $N(0,2)/(0,.)$. For this prior, the posterior mean and median τ_1 are 0.493 and 0.495, respectively, and the posterior mean and median τ_2 are 0.525 and 0.523, respectively. Therefore, this prior distribution is chosen for the bivariate meta-analysis simulation study. For this prior distribution the number of 95% Crls that contain the true treatment effects is good with 95.4% for β_1 and 96.5% for β_2 . The number of credible intervals that contain the true parameter value is slightly higher for β_2 due to a slightly larger estimate of τ_2 . The average probability that $\hat{\theta}_1$ is greater than zero is 0.497, which is close to the true value of 0.5. The same finding is true for θ_2 .

The results from this simulation study reveal that some of the prior distributions for τ_j estimate the variances particularly poorly. For example, the $\text{gamma}(0.1,0.1)$ prior distribution for $1/\tau_j^2$ hugely overestimates the true between-study variances. The posterior mean and median τ_1 are 1.682 and 1.674, respectively, compared to $\tau_1=0.5$. Similarly, the posterior mean and median τ_2 are 1.835 and 1.819, respectively. This highlights the potential dangers of selecting a prior distribution for the between-study variances without a sensitivity analysis.

LIST OF REFERENCES

1. ClinicalTrials.gov. Learn about clinical studies [updated August 2012; cited 2014 07-11]. Available from: <https://clinicaltrials.gov/ct2/about-studies/learn#ClinicalTrials>.
2. Ivy SP, Siu LL, Garrett-Mayer E, Rubinstein L. Approaches to Phase 1 Clinical Trial Design Focused on Safety, Efficiency, and Selected Patient Populations: A Report from the Clinical Trial Design Task Force of the National Cancer Institute Investigational Drug Steering Committee. *Clinical Cancer Research*. 2010;16(6):1726-36.
3. Desfrere L, Zohar S, Morville P, Brunhes A, Chevret S, Pons G, et al. Dose-finding study of ibuprofen in patent ductus arteriosus using the continual reassessment method. *Journal of Clinical Pharmacy and Therapeutics*. 2005;30(2):121-32.
4. Foss FM, Sjak-Shie N, Goy A, Jacobsen E, Advani R, Smith MR, et al. A multicenter phase II trial to determine the safety and efficacy of combination therapy with denileukin diftitox and cyclophosphamide, doxorubicin, vincristine and prednisone in untreated peripheral T-cell lymphoma: the CONCEPT study. *Leuk Lymphoma*. 2013;54(7):1373-9.
5. Wieand HS. Randomized phase II trials: what does randomization gain? *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2005;23(9):1794-5.
6. Estey EH, Thall PF. New designs for phase 2 clinical trials. *Blood*. 2003;102(2):442-8.
7. Byrd JC, Peterson BL, Morrison VA, Park K, Jacobson R, Hoke E, et al. Randomized phase 2 study of fludarabine with concurrent versus sequential treatment with rituximab in symptomatic, untreated patients with B-cell chronic lymphocytic leukemia: results from Cancer and Leukemia Group B 9712 (CALGB 9712). *Blood*. 2003;101(1):6-14.
8. Baigent C, Collins R, Appleby P, Parish S, Sleight P, Peto R. ISIS-2: 10 year survival among patients with suspected acute myocardial infarction in randomised comparison of intravenous streptokinase, oral aspirin, both, or neither. The ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *BMJ*. 1998;316(7141):1337-43.
9. ABPI. Guidelines for Phase 1 Clinical Trials. 2012.
10. Schmartz D, Tabardel Y, Preiser JC, Barvais L, d'Hollander A, Duchateau J, et al. Does aprotinin influence the inflammatory response to cardiopulmonary bypass in patients? *The Journal of thoracic and cardiovascular surgery*. 2003;125(1):184-90.
11. Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? *BMJ*. 1998;316(7126):201.

12. Ashby D. Bayesian statistics in medicine: A 25 year review. *Statistics in Medicine*. 2006;25(21):3589-631.
13. OQuigley J, Pepe M, Fisher L. Continual Reassessment Method - A Practical Design for Phase-1 Clinical-Trials in Cancer. *Biometrics*. 1990;46(1):33-48.
14. Chevret S. The Continual Reassessment Method in Cancer Phase-I Clinical-Trials - A Simulation Study. *Statistics in Medicine*. 1993;12(12):1093-108.
15. Heitjan DF. Bayesian interim analysis of phase II cancer clinical trials. *Stat Med*. 1997;16(16):1791-802.
16. Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat Med*. 1985;4(4):521-6.
17. Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DPM, et al. Multivariate meta-analysis of mixed outcomes: a Bayesian approach. *Statistics in Medicine*. 2013;32:3926-43.
18. Wei Y, Higgins JPT. Bayesian multivariate meta-analysis with multiple outcomes. *Statistics in Medicine*. 2013;32:1191-205.
19. Sutton AJ, Abram KR, Jones DR, Sheldon TA, Song F. Bayesian methods in meta-analysis. *Methods for Meta-Analysis in Medical Research*. Chichester: John Wiley & Sons,Ltd; 2000. p. 163-89.
20. Sutton AJ, Abram KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*. 2001;10:277-303.
21. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: John Wiley & Sons; 2004.
22. Spiegelhalter DJ, Abrams KR, Myles JP. Prior Distributions. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: John Wiley & Sons Ltd; 2004. p. 139-80.
23. Bayes T. An essay towards solving a problem in the doctrine of chances. *Philosophical Translations*. 1763;53:370-418.
24. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall; 1996.
25. Spiegelhalter DJ, Abrams KR, Myles JP. An Overview of the Bayesian Approach. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: John Wiley & Sons Ltd; 2004. p. 49-117.
26. O'Hagan AB, C.E; Daneshkhah, A; Eiser, R; Garthwaite, P.H; Jenkinson, D.J; Oakley, J.E; Rakow, T. *Uncertain Judgements: Eliciting Experts' Probabilities*. 1 ed: Wiley-Blackwell; 2006.

27. Burke DL, Billingham LJ, Girling AJ, Riley RD. Meta-analysis of randomized phase II trials to inform subsequent phase III decisions. *Trials*. 2014;15:346.
28. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian Approaches to Randomized Trials. *Journal of the Royal Statistical Society Series A-Statistics in Society*. 1994;157:357-87.
29. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of prior distributions in MCMC using WinBUGS. *Statistics in Medicine*. 2005;24:2401-28.
30. Casella G, George EI. Explaining the Gibbs Sampler. *The American Statistician*. 1992;46(3):167-74.
31. Gilks WR, Thomas A, Spiegelhalter DJ. A Language and Program for Complex Bayesian Modelling. *Journal of the Royal Statistical Society Series D (The Statistician)*. 1994;43(1):169-77.
32. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS Version 1.4. User Manual. Cambridge: MRC Biostatistics Unit [Internet]. 2003. Available from: <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>.
33. Rawlins M. De Testimonio: on the evidence for decisions about the use of therapeutic interventions. *ClinMed*. 2008;8(6):579-88.
34. Higgins JPT, Green S, (editors). *Cochrane Handbook for Systematic Reviews of Interventions 2008* [updated 2/2008]. Available from: www.cochrane-handbook.org.
35. Alderson P, Green S, (editors). *Cochrane Collaboration open learning material for reviewers 2002* [updated 11/2002]. 1.1:[Available from: <http://www.cochrane-net.org/openlearning/index.htm>].
36. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *IntJEpidemiol*. 2012;41(3):818-27.
37. Pullenayegum EM. An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. *Stat Med*. 2011;30(26):3082-94.
38. Rhodes KM, Turner RM, Higgins JP. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of clinical epidemiology*. 2014;68(1):52-60.
39. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *British Medical Journal*. 2011;342:964-7.

40. Bradburn MJ, Deeks JJ, Berlin JA, Russell LA. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *StatMed*. 2007;26(1):53-77.
41. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986;7(3):177-88.
42. Morris RK, Malin GL, Quinlan-Jones E, Middleton LJ, Hemming K, Burke D, et al. Percutaneous vesicoamniotic shunting versus conservative management for fetal lower urinary tract obstruction (PLUTO): a randomised trial. *Lancet*. 2013;382(9903):1496-506.
43. Spiegelhalter DJ, Abrams KR, Myles JP. Randomised controlled trials. Bayesian approached to clinical trials and health-care evaluation. Chichester: John Wiley & Sons Ltd; 2004. p. 181-250.
44. Parmar MKB, Griffiths GO, Spiegelhalter D, Souhami RL, Altman DG, van der Scheuren E, et al. Monitoring of large randomised clinical trials: a new approach with Bayesian methods. *Lancet*. 2001;358(9279):375-81.
45. Ashby D. Bayesian statistics in medicine: a 25 year review. *StatMed*. 2006;25(21):3589-631.
46. Rohrig B, du Prel JB, Wachtlin D, Kwiecien R, Blettner M. Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *DtschArzteblInt*. 2010;107(31-32):552-6.
47. Web Of Science 2014 [updated 2014; cited 2013 11/4/2013]. Available from: http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=Q1iKiyO3oJ3a1WRK3d&preferencesSaved=.
48. Lucas BP, Trick WE, Evans AT, Mba B, Smith J, Das K, et al. Effects of 2-vs 4-Week Attending Physician Inpatient Rotations on Unplanned Patient Revisits, Evaluations by Trainees, and Attending Physician Burnout A Randomized Trial. *Jama-Journal of the American Medical Association*. 2012;308(21):2199-207.
49. Newman MF, Ferguson T, White JA, Ambrosio G, Koglin J, Nussmeier NA, et al. Effect of Adenosine-Regulating Agent Acadesine on Morbidity and Mortality Associated With Coronary Artery Bypass Grafting The RED-CABG Randomized Controlled Trial. *Jama-Journal of the American Medical Association*. 2012;308(2):157-64.
50. Hashkes PJ, Spalding SJ, Giannini EH, Huang B, Johnson A, Park G, et al. Riloncept for Colchicine-Resistant or -Intolerant Familial Mediterranean Fever A Randomized Trial. *Annals of Internal Medicine*. 2012;157(8):533-41.

51. Raudenbush S. A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics*. 1993;18(4):321-49.
52. Leckie G, Charlton C. runmlwin: A program to run the MLwiN multilevel modelling software from within Stata. *Journal of Statistical Software*. 2012;52(11):1-40.
53. Link W, Cam E, Nichols J, Cooch E. Of *Bugs* and Birds: Markov Chain Monte Carlo for hierarchical modeling in wildlife research. *Journal of Wildlife Management*. 2002;66(2):277-91.
54. Koren MJ, Scott R, Kim JB, Knusel B, Liu T, Lei L, et al. Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 as monotherapy in patients with hypercholesterolaemia (MENDEL): a randomised, double-blind, placebo-controlled, phase 2 study. *Lancet*. 2012;380(9858):1995-2006.
55. Giugliano RP, Desai NR, Kohli P, Rogers WJ, Somaratne R, Huang F, et al. Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 in combination with a statin in patients with hypercholesterolaemia (LAPLACE-TIMI 57): a randomised, placebo-controlled, dose-ranging, phase 2 study. *Lancet*. 2012;380(9858):2007-17.
56. Sabchareon A, Wallace D, Sirivichayakul C, Limkittikul K, Chanthavanich P, Suvannadabba S, et al. Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. *Lancet*. 2012;380(9853):1559-67.
57. Solomon SD, Zile M, Pieske B, Voors AA, Shah A, Kraigher-Krainer E, et al. The angiotensin receptor neprilysin inhibitor LCZ696 in heart failure with preserved ejection fraction: a phase 2 double-blind randomised controlled trial. *Lancet*. 2012;380(9851):1387-95.
58. Kirsner RS, Marston WA, Snyder RJ, Lee TD, Cargill D, I, Slade HB. Spray-applied cell therapy with human allogeneic fibroblasts and keratinocytes for the treatment of chronic venous leg ulcers: a phase 2, multicentre, double-blind, randomised, placebo-controlled trial. *Lancet*. 2012;380(9846):977-85.
59. Diacon AH, Dawson R, von Groote-Bidlingmaier F, Symons G, Venter A, Donald PR, et al. 14-day bactericidal activity of PA-824, bedaquiline, pyrazinamide, and moxifloxacin combinations: a randomised trial. *Lancet*. 2012;380(9846):986-93.
60. Papp K, Cather JC, Rosoph L, Sofen H, Langley RG, Matheson RT, et al. Efficacy of apremilast in the treatment of moderate to severe psoriasis: a randomised controlled trial. *Lancet*. 2012;380(9843):738-46.

61. Pavord ID, Korn S, Howarth P, Bleecker ER, Buhl R, Keene ON, et al. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. *Lancet*. 2012;380(9842):651-9.
62. Stein EA, Gipe D, Bergeron J, Gaudet D, Weiss R, Dufour R, et al. Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a phase 2 randomised controlled trial. *Lancet*. 2012;380(9836):29-36.
63. Roberts JD, Wells GA, Le May MR, Labinaz M, Glover C, Froeschl M, et al. Point-of-care genetic testing for personalisation of antiplatelet treatment (RAPID GENE): a prospective, randomised, proof-of-concept trial. *Lancet*. 2012;379(9827):1705-11.
64. Burant CF, Viswanathan P, Marcinak J, Cao C, Vakilynejad M, Xie B, et al. TAK-875 versus placebo or glimepiride in type 2 diabetes mellitus: a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet*. 2012;379(9824):1403-11.
65. Makkar RR, Smith RR, Cheng K, Malliaras K, Thomson LE, Berman D, et al. Intracoronary cardiosphere-derived cells for heart regeneration after myocardial infarction (CADUCEUS): a prospective, randomised phase 1 trial. *Lancet*. 2012;379(9819):895-904.
66. Budde K, Becker T, Arns W. Everolimus-based, calcineurin-inhibitor-free regimen in recipients of de-novo kidney transplants: an open-label, randomised, controlled trial. *Lancet*. 2011;377(9768):837-47.
67. Bolla M, van Poppel H, Tombal B, Vekemans K, Da Pozzo L, de Reijke TM, et al. Postoperative radiotherapy after radical prostatectomy for high-risk prostate cancer: long-term results of a randomised controlled trial (EORTC trial 22911). *Lancet*. 2012;380(9858):2018-27.
68. Laharie D, Bourreille A, Branche J, Allez M, Bouhnik Y, Filippi J, et al. Ciclosporin versus infliximab in patients with severe ulcerative colitis refractory to intravenous steroids: a parallel, open-label randomised controlled trial. *Lancet*. 2012;380(9857):1909-15.
69. Pickard R, Lam T, MacLennan G, Starr K, Kilonzo M, McPherson G, et al. Antimicrobial catheters for reduction of symptomatic urinary tract infection in adults requiring short-term catheterisation in hospital: a multicentre randomised controlled trial. *Lancet*. 2012;380(9857):1927-35.
70. Rini B, Escudier B, Tomczak P. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet*. 2011;378(9807):1931-9.

71. Cohen JA, Coles AJ, Arnold DL, Confavreux C, Fox EJ, Hartung HP, et al. Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: a randomised controlled phase 3 trial. *Lancet*. 2012;380(9856):1819-28.
72. Coles AJ, Twyman CL, Arnold DL, Cohen JA, Confavreux C, Fox EJ, et al. Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. *Lancet*. 2012;380(9856):1829-39.
73. Simmons RK, Echouffo-Tcheugui JB, Sharp SJ, Sargeant LA, Williams KM, Prevost A, et al. Screening for type 2 diabetes and population mortality over 10 years (ADDITION-Cambridge): a cluster-randomised controlled trial. *Lancet*. 2012;380(9855):1741-8.
74. Lopes RD, Al-Khatib SM, Wallentin L, Yang H, Ansell J, Cecilia Bahit M, et al. Efficacy and safety of apixaban compared with warfarin according to patient risk of stroke and of bleeding in atrial fibrillation: a secondary analysis of a randomised controlled trial. *Lancet*. 2012;380(9855):1749-58.
75. Ryan NM, Birring SS, Gibson PG. Gabapentin for refractory chronic cough: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2012;380(9853):1583-9.
76. Sabate M, Cequier A, Iniguez A, Serra A, Hernandez-Antolin R, Mainar V, et al. Everolimus-eluting stent versus bare-metal stent in ST-segment elevation myocardial infarction (EXAMINATION): 1 year results of a randomised controlled trial. *Lancet*. 2012;380(9852):1482-90.
77. Camenzind E, Wijns W, Mauri L, Kurowski V, Parikh K, Gao R, et al. Stent thrombosis and major clinical events at 3 years after zotarolimus-eluting or sirolimus-eluting coronary stent implantation: a randomised, multicentre, open-label, controlled trial. *Lancet*. 2012;380(9851):1396-405.
78. Wandt H, Schaefer-Eckart K, Wendelin K, Pilz B, Wilhelm M, Thalheimer M, et al. Therapeutic platelet transfusion versus routine prophylactic transfusion in patients with haematological malignancies: an open-label, multicentre, randomised study. *Lancet*. 2012;380(9850):1309-16.
79. Nogueira RG, Lutsep HL, Gupta R, Jovin TG, Albers GW, Walker GA, et al. Trevo versus Merci retrievers for thrombectomy revascularisation of large vessel occlusions in acute ischaemic stroke (TREVO 2): a randomised trial. *Lancet*. 2012;380(9849):1231-40.
80. Scirica BM, Bonaca MP, Braunwald E, De Ferrari GM, Isaza D, Lewis BS, et al. Vorapaxar for secondary prevention of thrombotic events for patients with previous myocardial infarction: a prespecified subgroup analysis of the TRA 2 degrees P-TIMI 50 trial. *Lancet*. 2012;380(9850):1317-24.

81. Saver JL, Jahan R, Levy EI, Jovin TG, Baxter B, Nogueira RG, et al. Solitaire flow restoration device versus the Merci Retriever in patients with acute ischaemic stroke (SWIFT): a randomised, parallel-group, non-inferiority trial. *Lancet*. 2012;380(9849):1241-9.
82. Rickard CM, Webster J, Wallis MC, Marsh N, McGrail MR, French V, et al. Routine versus clinically indicated replacement of peripheral intravenous catheters: a randomised controlled equivalence trial. *Lancet*. 2012;380(9847):1066-74.
83. Crawley J, Sismanidis C, Goodman T, Milligan P. Effect of intermittent preventive treatment for malaria during infancy on serological responses to measles and other vaccines used in the Expanded Programme on Immunization: results from five randomised controlled trials. *Lancet*. 2012;380(9846):1001-10.
84. Fairall L, Bachmann MO, Lombard C, Timmerman V, Uebel K, Zwarenstein M, et al. Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): a pragmatic, parallel, cluster-randomised trial. *Lancet*. 2012;380(9845):889-98.
85. Zinkstok SM, Roos YB. Early administration of aspirin in patients treated with alteplase for acute ischaemic stroke: a randomised controlled trial. *Lancet*. 2012;380(9843):731-7.
86. Wong C, Jayaram L, Karalus N, Eaton T, Tong C, Hockey H, et al. Azithromycin for prevention of exacerbations in non-cystic fibrosis bronchiectasis (EMBRACE): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2012;380(9842):660-7.
87. Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the JUPITER trial. *Lancet*. 2012;380(9841):565-71.
88. Gallwitz B, Rosenstock J, Rauch T, Bhattacharya S, Patel S, von Eynatten M, et al. 2-year efficacy and safety of linagliptin compared with glimepiride in patients with type 2 diabetes inadequately controlled on metformin: a randomised, double-blind, non-inferiority trial. *Lancet*. 2012;380(9840):475-83.
89. Sandberg T, Skoog G, Hermansson AB, Kahlmeter G, Kuylenstierna N, Lannergard A, et al. Ciprofloxacin for 7 days versus 14 days in women with acute pyelonephritis: a randomised, open-label and double-blind, placebo-controlled, non-inferiority trial. *Lancet*. 2012;380(9840):484-90.
90. Davalos A, Alvarez-Sabin J, Castillo J, Diez-Tejedor E, Ferro J, Martinez-Vila E, et al. Citicoline in the treatment of acute ischaemic stroke: an international, randomised, multicentre, placebo-controlled study (ICTUS trial). *Lancet*. 2012;380(9839):349-57.

91. Hauschild A, Grob JJ, Demidov LV, Jouary T, Gutzmer R, Millward M, et al. Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *Lancet*. 2012;380(9839):358-65.
92. Kirchhof P, Andresen D, Bosch R, Borggrefe M, Meinertz T, Parade U, et al. Short-term versus long-term antiarrhythmic drug treatment after cardioversion of atrial fibrillation (Flec-SL): a prospective, randomised, open-label, blinded endpoint assessment trial. *Lancet*. 2012;380(9838):238-46.
93. Mees SM, Algra A, Vandertop W, van Kooten F, Kuijsten HA, Boiten J, et al. Magnesium for aneurysmal subarachnoid haemorrhage (MASH-2): a randomised placebo-controlled trial. *Lancet*. 2012;380(9836):44-9.
94. DeJesus E, Rockstroh JK, Henry K, Molina JM, Gathe J, Ramanathan S, et al. Co-formulated elvitegravir, cobicistat, emtricitabine, and tenofovir disoproxil fumarate versus ritonavir-boosted atazanavir plus co-formulated emtricitabine and tenofovir disoproxil fumarate for initial treatment of HIV-1 infection: a randomised, double-blind, phase 3, non-inferiority trial. *Lancet*. 2012;379(9835):2429-38.
95. Sax PE, DeJesus E, Mills A, Zolopa A, Cohen C, Wohl D, et al. Co-formulated elvitegravir, cobicistat, emtricitabine, and tenofovir versus co-formulated efavirenz, emtricitabine, and tenofovir for initial treatment of HIV-1 infection: a randomised, double-blind, phase 3 trial, analysis of results after 48 weeks. *Lancet*. 2012;379(9835):2439-48.
96. Jamieson DJ, Chasela CS, Hudgens MG, King CC, Kourtis AP, Kayira D, et al. Maternal and infant antiretroviral regimens to prevent postnatal HIV-1 transmission: 48-week follow-up of the BAN randomised controlled trial. *Lancet*. 2012;379(9835):2449-58.
97. Sandercock P, Wardlaw JM, Lindley RI, Dennis M, Cohen G, Murray G, et al. The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial. *Lancet*. 2012;379(9834):2352-63.
98. Aschner P, Chan J, Owens DR, Picard S, Wang E, Dain MP, et al. Insulin glargine versus sitagliptin in insulin-naive patients with type 2 diabetes mellitus uncontrolled on metformin (EASIE): a multicentre, randomised open-label trial. *Lancet*. 2012;379(9833):2262-9.
99. Gallwitz B, Guzman J, Dotta F, Guerci B, Simo R, Basson BR, et al. Exenatide twice daily versus glimepiride for prevention of glycaemic deterioration in patients with type 2 diabetes with metformin failure (EUREXA): an open-label, randomised controlled trial. *Lancet*. 2012;379(9833):2270-8.

100. Bhatnagar S, Wadhwa N, Aneja S, Lodha R, Kabra SK, Natchu UCM, et al. Zinc as adjunct treatment in infants aged between 7 and 120 days with probable serious bacterial infection: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2012;379(9831):2072-8.
101. Cima RF, Maes IH, Joore MA, Scheyen DJ, El Refaie A, Baguley DM, et al. Specialised treatment based on cognitive behaviour therapy versus usual care for tinnitus: a randomised controlled trial. *Lancet*. 2012;379(9830):1951-9.
102. van der Graaf WT, Blay JY, Chawla SP, Kim DW, Binh B, Casali PG, et al. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet*. 2012;379(9829):1879-86.
103. Biere SS, Henegouwen MI, Maas KW, Bonavina L, Rosman C, Roig Garcia J, et al. Minimally invasive versus open oesophagectomy for patients with oesophageal cancer: a multicentre, open-label, randomised controlled trial. *Lancet*. 2012;379(9829):1887-92.
104. Engert A. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. *Lancet*. 2012;379(9828):1791-9.
105. Goya M, Pratcorona L, Merced C, Rodo C, Valle L, Romero A, et al. Cervical pessary in pregnant women with a short cervix (PECEP): an open-label randomised controlled trial. *Lancet*. 2012;379(9828):1800-6.
106. Meirik O, Nguyen Thi MH, Piaggio G, Bergel E, von Hertzen H. Complications of first-trimester abortion by vacuum aspiration after cervical preparation with and without misoprostol: a multicentre randomised trial. *Lancet*. 2012;379(9828):1817-24.
107. van Vollenhoven RF, Geborek P, Forslind K, Albertsson K, Ernestam S, Petersson IF, et al. Conventional combination treatment versus biological treatment in methotrexate-refractory early rheumatoid arthritis: 2 year follow-up of the randomised, non-blinded, parallel-group Swefot trial. *Lancet*. 2012;379(9827):1712-20.
108. Guelmezoglu A, Lumbiganon P, Landoulsi S, Widmer M, Abdel-Aleem H, Festin M, et al. Active management of the third stage of labour with and without controlled cord traction: a randomised, controlled, non-inferiority trial. *Lancet*. 2012;379(9827):1721-7.
109. Kobayashi T, Saji T, Otani T, Takeuchi K, Nakamura T, Arakawa H, et al. Efficacy of immunoglobulin plus prednisolone for prevention of coronary artery abnormalities in severe Kawasaki disease (RAISE study): a randomised, open-label, blinded-endpoints trial. *Lancet*. 2012;379(9826):1613-20.
110. Heller S, Buse J, Fisher M, Garg S, Marre M, Merker L, et al. Insulin degludec, an ultra-longacting basal insulin, versus insulin glargine in basal-bolus treatment with mealtime

insulin aspart in type 1 diabetes (BEGIN Basal-Bolus Type 1): a phase 3, randomised, open-label, treat-to-target non-inferiority trial. *Lancet*. 2012;379(9825):1489-97.

111. Garber AJ, King AB, Del Prato S, Sreenan S, Balci MK, Munoz-Torres M, et al. Insulin degludec, an ultra-longacting basal insulin, versus insulin glargine in basal-bolus treatment with mealtime insulin aspart in type 2 diabetes (BEGIN Basal-Bolus Type 2): a phase 3, randomised, open-label, treat-to-target non-inferiority trial. *Lancet*. 2012;379(9825):1498-507.

112. Castaigne S, Pautas C, Terre C, Raffoux E, Bordessoule D, Bastie JN, et al. Effect of gemtuzumab ozogamicin on survival of adult patients with de-novo acute myeloid leukaemia (ALFA-0701): a randomised, open-label, phase 3 study. *Lancet*. 2012;379(9825):1508-16.

113. Nijhof SL, Bleijenberg G, Uiterwaal CS, Kimpen JL, van de Putte EM. Effectiveness of internet-based cognitive behavioural treatment for adolescents with chronic fatigue syndrome (FITNET): a randomised controlled trial. *Lancet*. 2012;379(9824):1412-8.

114. Manaseki-Holland S, Maroof Z, Bruce J, Mughal M, Masher MI, Bhutta ZA, et al. Effect on the incidence of pneumonia of vitamin D supplementation by quarterly bolus dose to infants in Kabul: a randomised controlled superiority trial. *Lancet*. 2012;379(9824):1419-27.

115. Corbacioglu S, Cesaro S, Faraci M, Valteau-Couanet D, Gruhn B, Rovelli A, et al. Defibrotide for prophylaxis of hepatic veno-occlusive disease in paediatric haemopoietic stem-cell transplantation: an open-label, phase 3, randomised controlled trial. *Lancet*. 2012;379(9823):1301-9.

116. Avery AJ, Rodgers S, Cantrill JA, Armstrong S, Cresswell K, Eden M, et al. A pharmacist-led information technology intervention for medication errors (PINCER): a multicentre, cluster randomised, controlled trial and cost-effectiveness analysis. *Lancet*. 2012;379(9823):1310-9.

117. Baird SJ, Garfein RS, McIntosh CT, Oezler B. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *Lancet*. 2012;379(9823):1320-9.

118. Fleshner NE, Lucia M, Egerdie B, Aaron L, Eure G, Nandy I, et al. Dutasteride in localised prostate cancer management: the REDEEM randomised, double-blind, placebo-controlled trial. *Lancet*. 2012;379(9821):1103-11.

119. Garcia PJ, Holmes KK, Carcamo CP, Garnett GP, Hughes JP, Campos PE, et al. Prevention of sexually transmitted infections in urban communities (Peru PREVEN): a multicomponent community-randomised controlled trial. *Lancet*. 2012;379(9821):1120-8.

120. Arifeen SE, Mullany LC, Shah R, Mannan I, Rahman SM, Talukder MR, et al. The effect of cord cleansing with chlorhexidine on neonatal mortality in rural Bangladesh: a community-based, cluster-randomised trial. *Lancet*. 2012;379(9820):1022-8.
121. Soofi S, Cousens S, Imdad A, Bhutto N, Ali N, Bhutta ZA. Topical application of chlorhexidine to neonatal umbilical cords for prevention of omphalitis and neonatal mortality in a rural district of Pakistan: a community-based, cluster-randomised trial. *Lancet*. 2012;379(9820):1029-36.
122. Thiele H, Woehrle J, Hambrecht R, Rittger H, Birkemeyer R, Lauer B, et al. Intracoronary versus intravenous bolus abciximab during primary percutaneous coronary intervention in patients with acute ST-elevation myocardial infarction: a randomised trial. *Lancet*. 2012;379(9819):923-31.
123. Griffin S, Borch-Johnsen K, Davies M. Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): a cluster-randomised trial. *Lancet*. 2011;378(9786):156-67.
124. Soofi S, Ahmed S, Fox MP, MacLeod WB, Thea DM, Qazi SA, et al. Effectiveness of community case management of severe pneumonia with oral amoxicillin in children aged 2-59 months in Matiari district, rural Pakistan: a cluster-randomised controlled trial. *Lancet*. 2012;379(9817):729-37.
125. Elena Santolaya M, O'Ryan ML, Teresa Valenzuela M, Prado V, Vergara R, Munoz A, et al. Immunogenicity and tolerability of a multicomponent meningococcal serogroup B (4CMenB) vaccine in healthy adolescents in Chile: a phase 2b/3 randomised, observer-blind, placebo-controlled study. *Lancet*. 2012;379(9816):617-24.
126. Baselga J, Bradbury I, Eidtmann H, Di Cosimo S, de Azambuja E, Aura C, et al. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *Lancet*. 2012;379(9816):633-40.
127. Johnston C, Saracino M, Kuntz S, Magaret A, Selke S, Huang MI, et al. Standard-dose and high-dose daily antiviral therapy for short episodes of genital HSV-2 reactivation: three randomised, open-label, cross-over trials. *Lancet*. 2012;379(9816):641-7.
128. Hanney M, Prasher V, Williams N, Jones EL, Aarsland D, Corbett A, et al. Memantine for dementia in adults older than 40 years with Down's syndrome (MEADOWS): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2012;379(9815):528-36.
129. Abraham W, Adamson P, Bourge R, Aaron M, Costanzo M, Stevenson L, et al. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. *Lancet*. 2011;377(9766):658-66.

130. Glazener C, Boachie C, Buckley B, Cochran C, Dorey G, Grant A, et al. Urinary incontinence in men after formal one-to-one pelvic-floor muscle training following radical prostatectomy or transurethral resection of the prostate (MAPS): two parallel randomised controlled trials. *Lancet*. 2011;378(9788):328-37.
131. Bang YJ, Kim YW, Yang HK, Chung HC, Park YK, Lee KH, et al. Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. *Lancet*. 2012;379(9813):315-21.
132. Mitja O, Hays R, Ipai A, Penias M, Paru R, Fagaho D, et al. Single-dose azithromycin versus benzathine benzylpenicillin for treatment of yaws in children in Papua New Guinea: an open-label, non-inferiority, randomised trial. *Lancet*. 2012;379(9813):342-7.
133. Coovadia HM, Brown ER, Fowler MG, Chipato T, Moodley D, Manji K, et al. Efficacy and safety of an extended nevirapine regimen in infant children of breastfeeding mothers with HIV-1 infection for prevention of postnatal HIV-1 transmission (HPTN 046): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2012;379(9812):221-8.
134. Gao Smith F, Perkins GD, Gates S, Young D, McAuley DF, Tunnicliffe W, et al. Effect of intravenous beta-2 agonist treatment on clinical outcomes in acute respiratory distress syndrome (BALTI-2): a multicentre, randomised controlled trial. *Lancet*. 2012;379(9812):229-35.
135. Mehilli J, Pache J, Abdel-Wahab M. Drug-eluting versus bare-metal stents in saphenous vein graft lesions (ISAR-CABG): a randomised controlled superiority trial. *Lancet*. 2011;378(9796):1071-8.
136. Buller HR, Gallus AS, Pillion G, Prins MH, Raskob GE. Enoxaparin followed by once-weekly idrabiotaparinux versus enoxaparin plus warfarin for patients with acute symptomatic pulmonary embolism: a randomised, double-blind, double-dummy, non-inferiority trial. *Lancet*. 2012;379(9811):123-9.
137. Gebre T, Ayele B, Zerihun M, Genet A, Stoller NE, Zhou Z, et al. Comparison of annual versus twice-yearly mass azithromycin treatment for hyperendemic trachoma in Ethiopia: a cluster-randomised trial. *Lancet*. 2012;379(9811):143-51.
138. Enden T, Haig Y, Klow NE, Slagsvold CE, Sandvik L, Ghanima W, et al. Long-term outcome after additional catheter-directed thrombolysis versus standard treatment for acute iliofemoral deep vein thrombosis (the CaVenT study): a randomised controlled trial. *Lancet*. 2012;379(9810):31-8.
139. Smith MR, Saad F, Coleman R, Shore N, Fizazi K, Tombal B, et al. Denosumab and bone-metastasis-free survival in men with castration-resistant prostate cancer: results of a phase 3, randomised, placebo-controlled trial. *Lancet*. 2012;379(9810):39-46.

140. Smith FG, Perkins GD, Gates S, Young D, McAuley DF, Tunnicliffe W, et al. Effect of intravenous beta-2 agonist treatment on clinical outcomes in acute respiratory distress syndrome (BALTI-2): a multicentre, randomised controlled trial. *Lancet*. 2012;379(9812):229-35.
141. El Arifeen S, Mullany LC, Shah R, Mannan I, Rahman SM, Talukder M, et al. The effect of cord cleansing with chlorhexidine on neonatal mortality in rural Bangladesh: a community-based, cluster-randomised trial. *Lancet*. 2012;379(9820):1022-8.
142. Chevret S. Bayesian adaptive clinical trials: a dream for statisticians only? *StatMed*. 2012;31(11-12):1002-13.
143. Lee JJ, Chu CT. Bayesian clinical trials in action. *StatMed*. 2012;31(25):2955-72.
144. Biswas S, Liu DD, Lee JJ, Berry DA. Bayesian clinical trials at the University of Texas M. D. Anderson Cancer Center. *ClinTrials*. 2009;6(3):205-16.
145. Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Statistics in Medicine*. 2014;33(24):4186-201.
146. Goodman S, Sladky J. A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barre syndrome. *Clinical trials*. 2005;2:305-10.
147. Baiardi P, Giaquinto C, Giroto S, Manfredi C, Ceci A. Innovative study design for paediatric clinical trials. *European Journal of Clinical Pharmacology*. 2011;67(Supplement 1):109-15.
148. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. *British Medical Journal*. 1995;311(7020):1621-5.
149. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340:c723.
150. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
151. Mills EJ, Wu P, Gagnier J, Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. *ContempClinTrials*. 2005;26(4):480-7.
152. Birmingham Uo. PLUTO. University of Birmingham [Internet]. 2012. Available from: <http://www.pluto.bham.ac.uk/index.shtml>.
153. Morris RK, Malin GL, Quinlan-Jones E, Middleton LJ, Diwakar L, Hemming K, et al. The Percutaneous shunting in Lower Urinary Tract Obstruction (PLUTO) study and

randomised controlled trial: evaluation of the effectiveness, cost-effectiveness and acceptability of percutaneous vesicoamniotic shunting for lower urinary tract obstruction. *Health Technol Assess*. 2013;17(59):1-232.

154. Lissauer D, Morris RK, Kilby MD. Fetal lower urinary tract obstruction. *Seminars in Fetal and Neonatal Medicine Fetal Intervention* 2007. p. 464-70.

155. Philadelphia TCsHo. Lower Urinary Tract Obstruction (LUTO) - CHOP. CHOP [Internet]. 2012. Available from: <http://www.chop.edu/service/fetal-diagnosis-and-treatment/fetal-diagnoses/lower-urinary-tract-obstruction-luto.html>.

156. Clark T, Martin W, Divakaran T, Whittle M, Kilby M, Khan K. Prenatal bladder drainage in the management of fetal lower urinary tract obstruction: a systematic review and meta-analysis. *The American College of Obstetricians and Gynecologists*. 2003;102(2):367-82.

157. Magann EF, Sanderson M, Martin JN, Chauhan S. The amniotic fluid index, single deepest pocket, and two-diameter pocket in normal human pregnancy. *American journal of obstetrics and gynecology*. 2000;182(6):1581-8.

158. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)*. 1972;34(2):187-220.

159. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*. 1982;10(4):1100-20.

160. Vittinghoff E, McCulloch CE. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *American Journal of Epidemiology*. 2007;165(6):710-8.

161. Kleinbaum DG, Klein M. The Cox Proportional Hazards Model and Its Characteristics. *Survival Analysis: A Self-Learning Text*. 2nd ed. New York: Springer; 2005. p. 83-130.

162. Klein JPvH, H.C.; Ibrahim, J.G.; Scheike, T.H. Bayesian Analysis of the Cox Model. *Handbook of Survival Analysis*. Florida: Chapman and Hall/CRC; 2014.

163. Long AE. Leuk: survival analysis using Cox regression. MRC-BSU [Internet]. 1999 5/22/2012. Available from: <http://www.mrc-bsu.cam.ac.uk/bugs/documentation/exampVol1/node29.html#SECTION00019000000000000000>.

164. Brown C, Morris RK, Daniels J, Khan KS, Lilford RJ, Kilby MD. Effectiveness of percutaneous vesico-amniotic shunting in congenital lower urinary tract obstruction: divergence in prior beliefs among specialist groups. *European journal of obstetrics, gynecology, and reproductive biology*. 2010;152(1):25-9.

165. Spiegelhalter D. Incorporating Bayesian ideas into health-care evaluation. *Statistical Science*. 2004;19(1):156-74.

166. Lunn D, Thomas A, Best N, Spiegelhalter D. A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*. 2002;10(4):325-37.
167. StataCorp. *Stata: Release 12*. College Station, TX: StataCorp LP; 2011 2011.
168. Excellence NfHaC. Fetal vesico-amniotic shunt for lower urinary tract outflow obstruction: NICE intervention procedure guidance 202. 2006.
169. Eikelboom JW, Mehta SR, Pogue J, Yusuf S. Safety outcomes in meta-analyses of phase 2 vs phase 3 randomized trials: Intracranial haemorrhage in trials of bolus thrombolytic therapy. *The Journal of the American Medical Association*. 2001;285(4):444-50.
170. Smolina K, Wright L, Rayner M, Goldacre M. Determinants of the decline in mortality from acute myocardial infarction in England between 2002 and 2010: linked national database study. *British Medical Journal*. 2012;344.
171. Smalling RW, Bode C, Kalbfleisch J, Sen S, Limbourg P, Forycki F, et al. More rapid, complete, and stable coronary thrombolysis with bolus administration of Reteplase compared with Alteplase infusion in acute myocardial infarction. *Circulation*. 1995;91:2725-32.
172. Bode C, Smalling RW, Berg G, Burnett C, Lorch G, Kalbfleisch J, et al. 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*. 1996;94:891-8.
173. Kawai C, Yui Y, Hosoda S, Nobuyoshi M, Suzuki S, Sato H, et al. 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*. 1997;29(7):1447-53.
174. Vanderschueren S, Dens J, Kerdsinchai P, Desmet W, Vrolix M, Man F, et al. Randomized coronary patency trial of double-bolus recombinant Staphylokinase versus front-loaded Alteplase in acute myocardial infarction. *American Heart Journal*. 1997;134(2):213-9.
175. Bar F, Meyer J, Boland J, Betriu A, Artmeyer B, Charbonnier B, et al. Bolus administration of Saruplase in Europe (BASE), a pilot study in patients with acute myocardial infarction. *Journal of Thrombosis and Thrombolysis*. 1998;6:147-53.
176. Bleich S, Adgey A, McMechan S, Love T, Investigators DS. An angiographic assessment of Alteplase: double-bolus and front-loaded infusion regimens in myocardial infarction. *American Heart Journal*. 1998;136(4):741-8.

177. den Heijer P, Vermeer F, Ambrosioni E, Sadowski Z, Lopez-Sendon J, von Essen R, et al. Evaluation of a weight-adjusted single-bolus plasminogen activator in patients with myocardial infarction. *Circulation*. 1998;98:2117-25.
178. Cannon C, Gibson M, McCabe C, Adgey A, Schweiger M, Sequeira R, et al. TNK-tissue plasminogen activator compared with front-loaded Alteplase in acute myocardial infarction: results of the TIMI 10B trial. *Circulation*. 1998;98:2805-14.
179. Park S, Group TS. Comparison of double bolus Urokinase versus front-loaded Alteplase regimen for acute myocardial infarction. *The American Journal of Cardiology*. 1998;82(6):811-3.
180. TheGUSTOinvestigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *New England Journal of Medicine*. 1993;329(10):673-82.
181. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*. 2004;23:1351-75.
182. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases*. 1985;27(5):335-71.
183. Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Stat Med*. 1990;9(3):247-52.
184. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. *JClinEpidemiol*. 2008;61(1):41-51.
185. Abo-Zaid G, Guo B, Deeks JJ, Debray TP, Steyerberg EW, Moons KG, et al. Individual participant data meta-analyses should not ignore clustering. *JClinEpidemiol*. 2013;66(8):865-73.
186. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*. 2003;22(17):2693-710.
187. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. 2002;21:1539-58.
188. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *British Medical Journal*. 2003;327:557-60.
189. Smith AFM, Roberts GO. Bayesian computation via the Gibbs Sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*. 1993;55(1):3-23.

190. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2009;172(Part 1):137-59.
191. Broglio KR, Stivers DN, Berry DA. Predicting clinical trial results based on announcements of interim analyses. *Trials*. 2014;15(1):73.
192. Bland JM, Altman DG. Statistics notes. The odds ratio. *BMJ*. 2000;320(7247):1468.
193. Hampton JR, Schroder R, Wilcox RG, Skene AM, Meyersabellek W, Heikkila J, et al. Randomized, Double-Blind Comparison of Reteplase Double-Bolus Administration with Streptokinase in Acute Myocardial-Infarction (Inject) - Trial to Investigate Equivalence. *Lancet*. 1995;346(8971):329-36.
194. Van de Werf F, Adgey A, Agnelli G, Aylward P, Binbrek A, Col J, et al. A comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction. *New England Journal of Medicine*. 1997;337(16):1124-30.
195. Topol E, Califf R, Ohman E, Wilcox R, Grinfeld L, Aylward P, et al. A comparison of reteplase with alteplase for acute myocardial infarction. *New England Journal of Medicine*. 1997;337(16):1118-23.
196. Van de Werf F, Adgey J, Ardissino D, Armstrong PW, Aylward P, Barbash G, et al. Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial. *Lancet*. 1999;354(9180):716-22.
197. Antman EM, Wilcox RG, Giugliano RP, al e. Long-term comparison of lanoteplase and alteplase in ST elevation myocardial infarction: 6 month follow-up in InTIME II Trial [abstract]. *Circulation*. 1999;100(Supplement):I-498.
198. Bar F, Hopkins G, Dickhoet S. The bolus versus infusion in Rescueplase (Saruplase) development (BIRD) study in 2410 patients with acute myocardial infarction [abstract]. *Circulation*. 1998;98(Supplement):I-505.
199. Rucker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. *BMC MedResMethodol*. 2008;8:79.
200. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *StatMed*. 2011;30(20):2481-98.
201. Gurwitz JH, Gore JM, Goldberg RJ, Barron HV, Breen T, Rundle AC, et al. Risk for intracranial hemorrhage after tissue plasminogen activator treatment for acute myocardial infarction. Participants in the National Registry of Myocardial Infarction 2. *Ann Intern Med*. 1998;129(8):597-604.
202. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. *JClinEpidemiol*. 2000;53(2):207-16.

203. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002.
204. Higgins JPT, Spiegelhalter DJ. Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. *International Journal of Epidemiology*. 2002;31:96-104.
205. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med*. 2000;19(24):3417-32.
206. Simmonds MC, Higgins JP. A general framework for the use of logistic regression models in meta-analysis. *Stat Methods Med Res*. 2014.
207. Li Y, Li S, Zhu Y, Liang X, Meng H, Chen J, et al. Incidence and risk of sorafenib-induced hypertension: a systematic review and meta-analysis. *JClinHypertens(Greenwich)*. 2014;16(3):177-85.
208. Zhan P, Wang Q, Qian Q, Li-Ke Y. Megestrol acetate in cancer patients with anorexia-cachexia syndrome: a meta-analysis. *Translational Cancer Research*. 2013;2(2):74-9.
209. Hu J, Zhao G, Wang HX, Tang L, Xu YC, Ma Y, et al. A meta-analysis of gemcitabine containing chemotherapy for locally advanced and metastatic pancreatic adenocarcinoma. *JHematolOncol*. 2011;4:11.
210. Qi WX, Tang LN, Shen Z, Yao Y. Treatment-related mortality with aflibercept in cancer patients: a meta-analysis. *EurJClinPharmacol*. 2014;70(4):461-7.
211. Al-Shahi SR. Haemostatic drug therapies for acute spontaneous intracerebral haemorrhage. *Cochrane DatabaseSystRev*. 2009(4):CD005951.
212. Sitole M, Silva M, Spooner L, Comee MK, Malloy M. Telaprevir versus boceprevir in chronic hepatitis C: a meta-analysis of data from phase II and III trials. *ClinTher*. 2013;35(2):190-7.
213. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2009;172(4):789-811.
214. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *StatMethods MedRes*. 2013;22(2):133-58.
215. Riley RD, Thompson JR, Abram KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*. 2008;9(1):172-86.

216. Stegmayr B, Asplund K, Kuulasmaa K, Rajakangas AM, Thorvaldsen P, Tuomilehto J. Stroke incidence and mortality correlated to stroke risk factors in the WHO MONICA Project. An ecological study of 18 populations. *Stroke*. 1997;28(7):1367-74.
217. Trikalinos TA, Hoaglin DC, Schmid CH. An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes. *StatMed*. 2014;33(9):1441-59.
218. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC MedResMethodol*. 2007;7:3.
219. Ishak KJ, Platt RW, Joseph L, Hanley JA. Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *StatMed*. 2008;27(5):670-86.
220. Bagos PG. On the covariance of two correlated log-odds ratios. *StatMed*. 2012;31(14):1418-31.
221. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *StatMed*. 2012;31(20):2179-95.
222. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *EvalHealth Prof*. 2002;25(1):76-97.
223. Berkey CS, Antczak-Bouckoms A, Hoaglin DC, Mosteller F, Pihlstrom BL. Multiple-outcomes meta-analysis of treatments for periodontal disease. *JDentRes*. 1995;74(4):1030-9.
224. Wang JG, Staessen JA, Franklin SS, Fagard R, Gueyffier F. Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome. *Hypertension*. 2005;45(5):907-13.
225. Riley RP, MJ.; Jackson, D.; Wardle, M.; Gueyffier, F.; Wang, J.; Staessen, JA.; White, IR. Multivariate meta-analysis using individual participant data. *Research Synthesis Methods*. 2014.
226. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *StatMed*. 2003;22(14):2309-33.
227. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. *Stat Med*. 2005;24(24):3823-44.
228. Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med*. 2008;27(29):6111-36.
229. Zwinderman AH, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic reviews. *Stat Med*. 2008;27(5):687-97.

230. Daniels MJK, R.E. Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*. 1999;94(448):1254-63.
231. Leonard TH, J.S. Bayesian inference for a covariance matrix. *Annals of Statistics*. 1992;20(4):1669-96.
232. Riley RDA, K.R.; Lambert, P.C.; Sutton,A.J.; Thompson,J.R. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine*. 2007;26:78-97.
233. Symons MJ, Moore DT. Hazard rate ratio and prospective epidemiological studies. *Journal of clinical epidemiology*. 2002;55(9):893-9.
234. Perneger TV. Estimating the relative hazard by the ratio of logarithms of event-free proportions. *Contemp Clin Trials*. 2008;29(5):762-6.
235. Fibrinogen Studies C, Jackson D, White I, Kostis JB, Wilson AC, Folsom AR, et al. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Stat Med*. 2009;28(8):1218-37.
236. Inc SI. SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc; 2011.
237. Wei Y, Higgins JPT. Estimating the within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in Medicine*. 2013;32:1191-205.
238. Davey J, Turner RM, Clarke MJ, Higgins JP. Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. *BMC Med Res Methodol*. 2011;11:160.
239. Senn S. Trying to be precise about vagueness. *Stat Med*. 2007;26(7):1417-30.
240. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. Comments on 'trying to be precise about vagueness' by Stephen Senn, *Statistics in Medicine* 2007; 26:1417-1430. *Stat Med*. 2008;27(4):619-22, author reply 22-4.
241. Berry DA. A case for Bayesianism in clinical trials. *StatMed*. 1993;12(15-16):1377-93.
242. Berry DA. Bayesian clinical trials. *Nature Reviews*. 2006;5:27-36.
243. Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *J Natl Cancer Inst*. 2012;104(8):590-8.
244. Amiri-Kordestani L, Fojo T. Why do phase III clinical trials in oncology fail so often? *J Natl Cancer Inst*. 2012;104(8):568-9.
245. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JP. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Stat Med*. 2014.

246. Claxton K, Neumann PJ, Araki S, Weinstein MC. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. *International journal of technology assessment in health care*. 2001;17(1):38-55.
247. O'Hagan A. Eliciting expert beliefs in substantial practical applications. *The Statistician*. 1998;47:21-35.
248. Menke J. Bayesian bivariate meta-analysis of sensitivity and specificity: summary of quantitative findings in 50 meta-analyses. *Journal of evaluation in clinical practice*. 2014.
249. Rossi PA, G; McCulloch, R. *Bayesian Statistics and Marketing*. Chichester: John Wiley & Sons Ltd; 2005.
250. Breslow N. Biostatistics and Bayes. *Statistical Science*. 1990;5(3):269-98.
251. Wijeyesundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. *Journal of clinical epidemiology*. 2009;62(1):13-21 e5.