Title
Uncertainty in Meta-Analysis: Bridging the Divide Between Ideal and Available Extracted Data

Author
RIZZO VARELA, SHEMRA

Publication Date
2015

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Uncertainty in Meta-Analysis: Bridging the Divide Between Ideal and Available Extracted Data

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Biostatistics

by

Shemra Rizzo Varela

2015

Abstract of the Dissertation

Uncertainty in Meta-Analysis: Bridging the Divide Between Ideal and Available Extracted Data

by

Shemra Rizzo Varela

Doctor of Philosophy in Biostatistics

University of California, Los Angeles, 2015

Professor Robert Erin Weiss, Chair

Meta-analysis in the health sciences combines evidence from multiple studies to derive stronger conclusions about the efficacy of treatments. In the process of data extraction from published papers, it is extremely common for the required data to be ambiguous, incomplete or missing. We consider the case of meta-analysis of odds-ratios with unknown number of events and meta-analysis of mean differences with missing standard errors. Existing approaches consist of computing best-estimates for the missing values then feeding them into the meta-analysis as extracted data without accounting for the uncertainty of the computations. These naive approaches lead to over-certain results and potentially inaccurate conclusions.

Meta-analysis of odds-ratios assumes binomially distributed numbers of events in each treatment group and requires extracted number of events, which are often not available due to loss to follow-up. Common practice consists of inferring the probability of survival from measurements of the Kaplan Meier survival plot and then using it to infer the number of deaths. We propose the Uncertain Reading-Estimated Events model to construct each study's contribution to the meta-analysis separately using the data available for extraction. In our meta-analysis comparing CABG and PCI for ULMCA stenosis, accounting for the uncertainty results in increased standard deviations of the log-odds as compared to a naive meta-analysis that assumes ideal extracted data, equivalent to a reduction of the overall

sample size of 43% in our example. Simulations show that meta-analysis based on the observed number of deaths lead to biased estimates while our model does not.

Meta-analysis of mean differences requires extracted mean differences and their standard errors (SE). However, missing standard errors are pervasive in publications. An algebraic computation to recover the missing SE utilizes the baseline and follow-up standard deviations, and correlations, which are also typically missing. Traditional approaches, that have not been theoretically derived, replace missing SEs with various single-value imputations. We formally derive the Uncertain Standard Error Bayesian model to accommodate multiple patterns of missingness in the standard deviations. In our meta-analysis comparing home monitoring blood pressure to usual care, accounting for the uncertainty results in larger posterior SEs compared to the traditional approaches.

The dissertation of Shemra Rizzo Varela is approved.

Catherine A. Sugar

Michael F. Green

Sudipto Banerjee

Robert Erin Weiss, Committee Chair

University of California, Los Angeles

2015

To my father

Table of Contents

1 Introduction to meta-analysis and data extraction 1

2 Meta-analysis of odds ratios with unknown number of events 7

2.1 Datasets 9

2.1.1 Unprotected left main coronary artery stenosis data 9

2.1.2 Simulated data 11

2.2 Classical and Bayesian random effects meta-analysis models 12

2.2.1 DerSimonian and Laird estimates 13

2.2.2 Maximum Likelihood estimates 13

2.2.3 Bayesian model 14

2.2.4 The naive approach to incomplete extracted data 15

2.3 The uncertain reading-estimated events model 16

2.3.1 The Uncertain Reading density 18

2.3.2 The Estimated Events density 19

2.3.3 Survival probabilities not available for extraction 22

2.4 Results 23

2.4.1 Prior specification 23

2.4.2 Conditional densities 23

2.4.3 Posterior computation 25

2.4.4 Sensitivity to prior specification 26

2.4.5 Results with ULMCA dataset 26

2.4.6 Results with simulated dataset 27

2.5 Discussion 28

3 Meta-analysis of mean differences with missing standard errors 31

3.1 Methods for meta-analysis of mean differences with missing standard errors . 33

3.1.1 Methods for missing mean difference 35

3.1.2 Methods for missing standard errors 35

3.2 Dataset 41

3.3 Uncertain Standard Error model 48

3.4 Results 51

3.4.1 Posterior computation 51

3.4.2 Sensitivity to prior specification 51

3.4.3 Results for he SBP and DBP datasets 53

3.5 Discussion 59

4 Future work in meta-analysis data extraction 61

4.1 Data Extraction Problems in Meta-analysis of Hazard Ratios 61

4.2 Data Extraction Problems in Dose-Response Meta-Analysis 61

Appendices 62

A Ratio of two uniformly distributed random variables 63

B Simulated data for the UR-EE model 64

C USE Bayesian model in JAGS 65

References 70

List of Figures

1.1 Number of articles with *meta-analysis* in their title listed in PubMed. 2

2.1 Naive conditional densities, $f(\pi_{ij}|s^*_{ij})$ in solid lines and UR-EE conditional densities $f(\pi_{ij}|\text{extracted data})$ in dashed lines. Where applicable, the dotted lines (in red) uses the ratio of two uniform random variables while the dashed line (in blue) uses the normal approximation. 24

2.2 Mortality meta-analysis for ULMCA dataset: odds-ratio mean (Bayes and UR-EE) or point estimate (DSL, ML), $\exp(d)$, and 95% confidence interval under various models. 27

2.3 Mortality meta-analysis simulated: odds-ratio mean $\exp(d)$ and 95% confidence interval under various models. 28

3.1 Posterior of μ in the USE model and SBP dataset by prior. Prior 1 is $\tau^2 \sim \text{IG}(7, 60)$, prior 2 is $\tau^2 \sim \text{IG}(3, 20)$, prior 3 is $\tau^2 \sim \text{IG}(2.4, 2.8)$ and prior 4 is $\tau \sim U(0, 10)$. 53

3.2 Posterior of τ in the USE model and SBP dataset by prior. Prior 1 is $\tau^2 \sim \text{IG}(7, 60)$, prior 2 is $\tau^2 \sim \text{IG}(3, 20)$, prior 3 is $\tau^2 \sim \text{IG}(2.4, 2.8)$ and prior 4 is $\tau \sim U(0, 10)$. 53

3.3 Meta-analysis for SBP monitoring vs usual care: Posterior distribution of μ under various models. 55

3.4 Meta-analysis for SBP monitoring vs usual care: Posterior distribution of τ under various models. 56

3.5 Meta-analysis for DBP monitoring vs usual care: Posterior distribution of μ under various models. 56

3.6 Meta-analysis for DBP monitoring vs usual care: Posterior distribution of τ under various models. 57

3.7 Posteriors of missing SE of the difference of mean differences, se_{Di}. Vertical lines are the single value imputations from the four standard algorithms. . . 58

List of Tables

2.1 Components of the extracted data from each study in the ULMCA meta-analysis: n_{ij} is the number of people at baseline, e_{ij} is the number of observed deaths, r_{ij} is the number of people at risk at year 1, x^*_{ij} and y^*_{ij} are the measurements of the KM plot from the x-axis to the curve at baseline and year 1, κ^*_{ij} is the rounded KM survival probability and (a_{ij-}, a_{ij+}) is its confidence interval, m_{ij} and v^2_{ij} are the mean and variance of the follow-up times, $Q1_{ij}$ is the median and $[Q1_{ij}, Q3_{ij}]$ is the interquartile range of the follow-up times for study i and group j. The PCI ($j = 1$) and CABG ($j = 0$) groups have the same type of extractable data but different values. 10

2.2 Extracted data for the ULMCA dataset according to the naive approach. . . 16

2.3 Follow-up times (in days) for studies in the ULMCA dataset by treatment and control group. 20

2.4 Posterior mean and 95% credible intervals for parameters d, σ^2, m and τ^2 in the naive and UR-EE Bayesian models for mortality of the ULMCA dataset. 26

3.1 Extracted data from each study in the SBP meta-analysis includes: sample sizes n_{ij}, mean and SDs at baseline, $(\bar{X}_{i1b}, s_{i1b})$, and at follow-up, $(\bar{X}_{i1f}, s_{i1f})$, mean and SD of the change, $(\bar{X}_{ijf}$ -$\bar{X}_{ijb}), s_{ij}$, and the difference of mean differences and its standard error, (D_i, se_{Di}), for treatment group j and study $i = 1, \ldots, K$. 43

3.2 Extracted data from each study in the DBP meta-analysis includes: sample sizes n_{ij}, mean and SDs at baseline, $(\bar{X}_{i1b}, s_{i1b})$, and at follow-up, $(\bar{X}_{i1f}, s_{i1f})$, mean and SD of the change, $(\bar{X}_{ijf}$ -$\bar{X}_{ijb}), s_{ij}$, and the difference of mean differences and its standard error, (D_i, se_{Di}), for treatment group j and study $i = 1, \ldots, K$. 45

3.3 Patterns of availability of SDs and SE for extraction and corresponding percent of SBP and DBP studies. 48

3.4 Posterior mean, standard deviation and 95% credible interval for parameters μ and τ, for the SBP and DBP dataset, listed by prior in the USE model. . . 52

3.5 Posterior mean, standard deviations and 95% credible intervals for parameters μ and τ for the SBP and DBP meta-analyses under various models. Let Y be all the data input to the model. 54

B.1 Simulated data for the UR-EE model: n_{ij} is the number of people at baseline, e_{ij} is the number of observed deaths, s_{ij} is the number of true deaths and κ_{ij} is the KM survival probability at year 1, for study i and group j. 64

Acknowledgments

I gratefully acknowledge the financial support from the following UCLA Fellowships: the Graduate Summer Research Mentorship, the Graduate Research Mentorship and the Dissertation Year Fellowship.

I want to thank my mentors, friends and family for their guidance and support. Without their help this dissertation could not have been completed.

I would like to express profound gratitude to my advisor Dr. Robert Weiss for his guidance and encouragement at each step of the dissertation, for the unbelievable amount of care in his revisions of my drafts, for allowing me to grow as a statistician, and for his invaluable career and research advice. I am also deeply indebted to my supervisor Dr. Catherine Sugar for financially supporting me, for helping me develop as a collaborator, for her sincere and thoughtful advice, and for being a fantastic female role model. I would like to thank Drs. Michael Green and Sudipto Banerjee for serving as my committee members, for helping to make my defense a pleasant experience, and for their insightful comments and discussion. Special thanks go to Dr. Tom Belin, who was a superb mentor during my first years as a young doctorate student. I want to thank him for the numerous occasions that he was generous with his time and academic advice.

I also want to express my gratitude: to my friends Mienah Sharif, Heidi Fischer, Trina Raval, Sherry Lin, Jacqueline Torres, and many more, for making the journey fun and making Los Angeles feel like home; to my siblings Jeshua, Sheryl, and Ishtar, for inspiring me to pursue excellence in everything I attempt; to my parents for the sacrifices they made that allowed me to cultivate my intellect; to my father Alfredo, for encouraging me to dream bigger and to push farther; to my beloved children, Shazer and Joshua, for their unconditional love; and to my husband and best friend Nick — words cannot describe how grateful I am for his support during the moments of doubt, and for the wonderful life that we share.

Vita

2008	B.S. Engineering Physics with highest honors, Instituto Tecnologico y de Estudios Superiores de Monterrey (ITESM), Nuevo Leon, Mexico.
2008–2009	Teaching Assistant. Courses: Basic Statistics, Introduction to Statistics, and Decision Models for Business. University of North Carolina at Chapel Hill, NC.
2009–2010	Instructor. Courses: Introduction to Statistics, and Decision Models for Business. University of North Carolina at Chapel Hill, NC.
2010	M.S. Statistics and Operations Research, University of North Carolina at Chapel Hill, NC.
2010–2012	Teaching Assistant. Courses: Basic Biostatistics, Introduction to Biostatistics, and Applied Regression. University of California, Los Angeles, CA.
2010–2013	Biostatistics Graduate Student Researcher, Center for Population Health and Health Disparities, University of California, Los Angeles, CA.
2013–2015	Biostatistics Graduate Student Researcher, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA.

Publications

Davis, M.C., Horan, W.P., Numi, E. L., Rizzo, S., Li, W., Sugar, C. A., Green, M. F. (2014) Associations between oxytocin receptor genotypes and social cognitive performance in individuals with schizophrenia. *Schizophrenia Research.* 159 (2); 353–357.

Sharif, M. Z., Rizzo, S., Prelip, M. L., Glik, D. C., Belin, T. R., Garza, J. R., Langellier, B. A., Ortega, A. N. (2014) Nutrition Facts label utilization and comprehension among Latinos in two East Los Angeles neighborhoods. *Journal of the Academy of Nutrition and Dietetics.* 114 (12); 1915–1922.

Reddy, L. F., Green, M.F., Rizzo, S., Sugar, C. A., Blanchard, J. J., Gur, R. E., Kring, A.

M., Horan, W. P. (2014) Behavioral approach and avoidance in schizophrenia: An evaluation of motivational profiles. *Schizophrenia Research*. 159 (1): 164–170.

P P Cook, M Gooch, S Rizzo. (2011) Reduction in fluoroquinolone use following introduction of ertapenem to hospital formulary associated with improvement in susceptibility of Pseudomonas aeruginosa to group 2 carbapenems. *Journal of Antimicrobial Agents and Chemotherapy*. 55 (12): 5597–5601.

P P Cook, S Rizzo, M Gooch, M Jordan, X Fang, S Hudson. (2011) Sustained reduction in antimicrobial use and decrease in methicillin-resistant Staphylcoccus aureus and Clostridium difficile infections following implementation of an electronic medical record at a tertiary-care teaching hospital. *Journal of Antimicrobial Chemotherapy*. 66 (1): 205–209.

CHAPTER 1

Introduction to meta-analysis and data extraction

Meta-analysis combines evidence from multiple studies to derive a stronger conclusion about the research question they have in common. In the health sciences, each study has its own conclusion in the form of a treatment effect and meta-analysis provides a better understanding of the treatment efficacy. Examples of treatment effects for binary outcomes are odds ratios, hazard ratios, and difference of proportions. For continuous outcomes we have the difference in means. Studies in a meta-analysis may be too small or may only focus on only one medical center or sub-population so their results may not have enough power or may not be generalizeable. By aggregating the results of all the studies available in a meta-analysis, the power to detect an effect increases, the precision and accuracy of the estimate of the effect size improves, and the generalizability of results also increases. Two ways to perform a meta-analysis are

(1) Combining the individual patient data (IPD), or

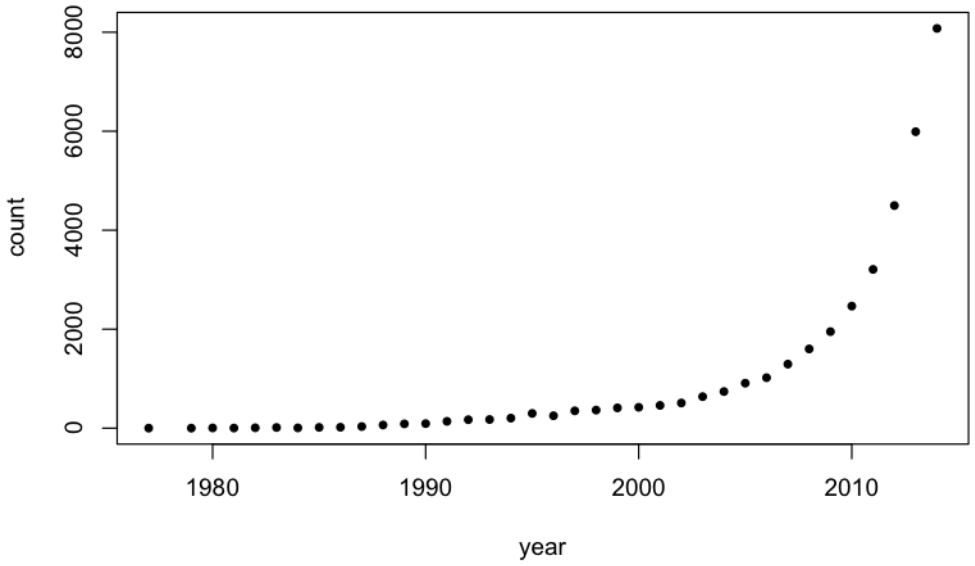
(2) Combining published summary statistics.

Meta-analyses that pool the effect sizes found in the summary statistics of published studies are said to be based on aggregate data. In (1), the meta-analyst collects the dataset from the authors of each participating study. The datasets are merged to create a larger dataset that includes all patients from all studies. The results of each participating study may or may not have been previously published. A hierarchical analysis is performed on the pooled dataset to estimate the treatment effect. In (2), the meta-analyst extracts estimates of effects from published studies. In a random effects meta-analysis, each study's true treatment effect is assumed to be randomly sampled from a distribution of treatment effects centered

around an unknown mean. The meta-analysis estimates the mean of the treatment effect distribution.

Meta-analysis has been heavily criticized since its introduction for its limitations, different sources of bias, the heterogeneity of studies, and for not predicting accurately the results of large studies performed later (Eyseneck, 1994; Walker et al., 2008; LeLorier et al., 1997; Shapiro, 1997). Despite the criticisms, meta-analyses have become extremely popular. A February 2015 PubMed search revealed that the number of articles that included the word "meta-analysis" in the title has increased dramatically in the past decade, from 740 articles in 2004 to 8,077 articles in 2014 (Figure 1). The popularity of meta-analysis is expected to continue rising, because science needs systematic overviews of randomized trials (Peto, 1987) and because rigorous meta-analyses provide definite advantages over the narrative reviews used previously (Thacker, 1988). Additionally, meta-analyses shed light on research questions that need investigation and improve the design of future studies (Berman and Parker, 2002).

Figure 1.1: Number of articles with *meta-analysis* in their title listed in PubMed.



Ideally, a meta-analysis includes individual participant data (IPD) from all relevant studies. However, using IPD in meta-analyses is a rare occurrence. Kovalchik (2012) found that fewer than 5% of published meta-analyses used IPD. Of the meta-analysts surveyed, 71%

did not attempt to collect IPD, and almost half of them never considered using IPD or thought that it did not provide any statistical advantage. While it has been shown that meta-analyses on IPD yield better results (Stewart and Parmar 1993, Lamber et al. 2002, Berlin et al. 2002), meta-analyses are traditionally performed on published summary statistics or on a mixture of IPD and summary statistics (Kovalchik and Cumberland, 2012). For example, in a meta-analysis of aggregate data of odds ratios, the true mean odds ratio is estimated by computing a weighted average of the odds ratios of the studies. Similarly, in meta-analysis of mean differences, the true mean difference is estimated by computing a weighted average of the mean differences of the studies.

Meta-analysis on aggregate data is often considered to be a simple process because the data is already available in published papers. However, a valid meta-analysis is neither quick nor easy (Berman and Parker, 2002). There are five main steps in a meta-analysis of aggregate data:

1. Formulation of the research question. For example: Is treatment A better than treatment B when the outcome is a mortality rate?

2. Definition of the study exclusion/inclusion criteria.

3. Search, identification and evaluation of the research studies.

4. Extraction of data from each study.

5. Data analysis.

Each step presents an opportunity for disaster if not carefully planned and executed. The bulk of time and effort occurs in the first four steps. Of these, extraction of the data from each study is the least studied.

Current guidelines for meta-analysis data extraction suggest that "two individuals should independently abstract the results from every study and differences resolved by consensus" (Berman and Parker, 2002). Consensus can be achieved by discussion among the data extractors. In some cases, "a disagreement may require arbitration by another person"

(Higgins and Deeks, 2011). Furthermore, these individuals should be "blinded to the various treatment groups through a coded photocopying process", and then the inter-observed agreement should be measured (Sacks et al., 1987). However, a study concluded that blinding is extremely time-consuming and that it is not necessary when conducting meta-analyses of randomized clinical trials (Berlin, 1997).

In the process of data extraction, it is extremely common for the required data to be ambiguous, incomplete or missing. For example, Riley et al. (2003) performed a systematic review of prognostic markers for neuroblastoma, the most common extracranial cancerous tumor in childhood. Prognostic markers included DNA or chromosome abnormalities, urinary catecholamines and biological markers. There were 260 papers with relevant information on the relationship between the 13 most commonly studied prognostic markers and survival. Each paper studied a different number of prognostic markers. Riley et al. identified 575 assessments of survival and tumor markers. Hence, to preform 13 meta-analyses, one for each marker, the authors needed to extract 575 hazard ratios and variances. However, only one paper, which studied three markers, reported hazard ratios and variances. For 201 marker-survival associations there were no reported hazard ratios and variances, but Riley et al. were able to infer them from other summary statistics found in the papers. For example, in some cases they extracted the hazard ratio and p-value, or the hazard ratio and confidence interval, or they extracted information from the survival curve and used that information to estimate the hazard ratio and its variance. For the remaining 371 marker-survival associations, there was no useful information in the papers to estimate the hazard ratios and variances.

Thus, in the meta-analyses that Riley et al. (2003) attempted to perform, only half of a percent of the hazard-ratio and variance pairs needed were available, and a further 36% were recovered using other data. Given that 64.5% of the effects were not available or possible to estimate, Riley et al. (2003) concluded that reliable meta-analyses were impossible to perform and that no clinical policy decisions could be made from their evidence-based review. In our datasets, which will be described in detail in the next sections, nine out of ten studies did not provide the data required for a meta-analysis of odds-ratios and 88% did not provide

the data required for a meta-analysis of mean differences.

Poor reporting in the papers that are used as input to meta-analysis has led to the common practice of extracting alternative information and using it to recover or, when recovering is not possible, to "estimate" the needed summary statistics. However, these *best-estimates* have no variance associated, and often no reporting of details of the computation either. These estimates are then inserted in the meta-analysis as extracted data. I argue that treating these insertions as observed summary statistics is a questionable practice because nowhere in the current methodology of meta-analysis is the uncertainty of these estimations accounted for.

For example, in a meta-analysis of odds ratios the necessary data to be extracted includes four numbers per study: the number of people experiencing the outcome and the sample sizes for both the treatment and control groups. The Cochrane Handbook for Systematic Reviews of Interventions recognizes that these data are often not available in published papers (Higgins and Deeks, 2011). Best-estimates of the number of events are computed using other information available in the paper, such as survival probabilities from Kaplan Meier curves. A *best-estimate* of the number of people experiencing the outcome is obtained by multiplying the complement of the survival probability by the sample size. The resulting number is fed into the meta-analysis as if it had been the true number of people experiencing the outcome. However, the survival probability has associated uncertainty that is being ignored in this calculation. Without adequate statistical guidelines, the use of best-estiamtes to replace the necessary data implies unjustified certainty in the resulting inputs to a meta-analysis, which leads to unjustified certainty in the output of the meta-analysis. The resulting over-confident estimates can lead to inaccurate conclusions.

A similar situation occurs in the case of meta-analysis of mean differences. Two numbers are needed per study: the mean difference and the standard error of the mean difference. However, the standard error is often missing, as often as 92% of the time (Philbrook et al., 2007). There are many approaches to estimating the missing standard error, such as inserting a weighted average of the standard errors of the studies that do provide them (Ma et al., 2008), or borrowing a standard error from another study that is considered similar enough or

from a previous meta-analysis. Alternatively, we can look for other information in the paper that may be helpful in constructing a value for the missing standard error. For example, the standard deviations of the treatment and control groups at baseline and follow-up can be used to compute the standard error. Unfortunately, this calculation assumes that the within-subject correlations are known, which are typically missing. None of these strategies have been theoretically justified or empirically tested (Wiebe et al., 2006). As in the case of the meta-analysis of odds ratios, to replace the needed standard errors with *best-estimates* results in uncertainty that is not accounted for in the meta-analysis.

This thesis will address the problems that arise during the data extraction step in meta-analysis of odds-ratios and difference of means. We propose methods to use existing information in published papers to model the unavailable data and to combine that model with a meta-analysis model to incorporate the associated uncertainty in the meta-analysis. This methodology serves as a statistical guideline and improves the validity of results from meta-analysis. Both methodologies are illustrated using real datasets. For the meta-analysis of odds-ratios in chapter 2 we use a dataset that compares two treatments for ULMCA stenosis: CABG and PCI. For the meta-analysis of mean differences in chapter 3 we use a dataset that compares the effectiveness of home blood-pressure monitoring versus usual care. Each chapter is written as a stand alone paper. Hence, they have their own introduction and review of past methodology with some overlap. Chapter 4 lists other cases of unaccounted uncertainty in meta-analysis and serves as a guideline for future research.

CHAPTER 2

Meta-analysis of odds ratios with unknown number of events

Meta-analysis is widely used to quantitatively combine results from multiple studies. In the health sciences, meta-analysis can provide stronger and more broadly-based evidence for treatment efficacy. Ideally, meta-analysis would analyze the combined individual patient data from all studies (Stewart and Parmar, 1993; Lambert et al., 2002; Berlin et al., 2002). In practice, this is rarely done (Kovalchik, 2012). Instead, meta-analyses rely on data extracted from journal articles in the published literature or from presentations at major conferences. This is known as meta-analysis of aggregate data.

A meta-analysis of odds ratios typically requires four quantities to be extracted per study: number of events and non-events in the treatment and control groups, which can easily be summarized in a 2x2 table. The Cochrane Handbook for Systematic Reviews of Interventions recognizes that data required for the meta-analysis are often not available in published papers (Higgins and Deeks, 2011), for example, the true number of events. Software typically requires the data from the 2x2 table from each study regardless of the outcome that was computed in each study (odds, risk and hazard ratio) (Melle et al., 2004). The frequent occurrence of incomplete extracted data has led to the common practice of estimating the missing entries of the 2x2 table using other information available from the study. For example, it is a very common occurrence to have all four entries missing, but the row totals are known which correspond to the numbers of people in each group at baseline. Best-estimates for the missing entries are computed from Kaplan-Meier (KM) survival curves which are often available in the published study. Because the survival curves rarely include survival probability values, meta-analysts take manual measurements from the curve to

estimate them. They then multiply the survival probability and its complement by the row total of each treatment group to fill in the 2x2 table. These estimates are often rounded to the nearest integer. The estimated data is introduced in the meta-analysis as observed data, leading to unjustified certainty in the results and to potentially inaccurate conclusions.

In some cases, the KM survival curve is not available and meta-analysts use the number of observed events reported in the text as the best approximation of the true number of events. However, in the presence of loss to follow-up this is an underestimate of the true number of events and could lead to biased meta-analysis estimates. A typical meta-analysis includes a combination of studies where the KM probabilities are available for extraction and studies where only the observed numbers of events are available.

While there have been efforts to promote better reporting practices (Riley et al., 2003), there is no established protocol for addressing the missing information encountered in the data extraction step of meta-analysis. For time-to-event data, some methods have been proposed to handle the missing extracted data for meta-analyses of hazard ratios but not for odds ratios (Parmar et al., 1998; Tierney et al., 2007). However, the proposed methods involve calculations that are only approximate, and do not account for the uncertainty introduced by the estimation. Methods to account for uncertainty in meta-analysis of odds ratios have been proposed for data from randomized trials that includes number of missing outcomes (non-observed events) (White et al., 2008b,c). However, in time-to-event studies the number of missing outcomes is typically not reported.

We propose the Uncertain Reading-Estimated Events (UR-EE) Bayesian model to account for the uncertainty that arises at the data extraction step of meta-analysis. Our model formally constructs a model to properly account for the contribution of each study to the meta-analysis. Our constructions do not depend on the desired data but depend rather on the actual data available from and extracted from each published study. Data available for extraction includes the number of participants at baseline, and may include one or more of the following: the rounded survival probabilities or measurements taken off the KM plot; confidence intervals for the KM survival probabilities; mean, variance, median, and quartiles of the distributions of follow-up times; the number of observed events, and the number of

people at risk at the time of interest. Because the information available for extraction is different for each study, the extracted data from each study must be modeled individually. The UR-EE model improves the validity of results from meta-analysis by accommodating all the uncertainty in the data input to the meta-analysis.

The chapter is organized as follows. We introduce two datasets in section 2.1. In section 3.1 we briefly review the classical and Bayesian random effects models and describe the *naive* approach to manipulate incomplete extracted data to be able to use the models. In section 2.3 we define the UR-EE Bayesian model for meta-analysis. Results for the two datasets from UR-EE and naive methods are given in section 3.4. The chapter finishes with discussion.

2.1 Datasets

2.1.1 Unprotected left main coronary artery stenosis data

To exemplify the different types of extracted data and proposed methodology for a meta-analysis of odds ratios, we carefully re-evaluate a published meta-analysis that compares two treatments for unprotected left main coronary artery (ULMCA) stenosis (Naik et al., 2009). The current gold standard treatment is coronary artery bypass grafting (CABG), which portends high morbidity. Percutaneous coronary intervention (PCI) has emerged as a plausible alternative. It is desirable to draw a definitive assessment of both treatments. The meta-analysis performed by Naik et al. (2009) included 10 studies with a total of 3,773 patients. The meta-analysis of mortality after 1 year presents multiple challenges in the data extraction step. In all studies the type of extractable data was the same for PCI and CABG. Table 2.1 is a checklist of the components available for extraction from each study.

Table 2.1: Components of the extracted data from each study in the ULMCA meta-analysis: n_{ij} is the number of people at baseline, e_{ij} is the number of observed deaths, r_{ij} is the number of people at risk at year 1, x^*_{ij} and y^*_{ij} are the measurements of the KM plot from the x-axis to the curve at baseline and year 1, κ^*_{ij} is the rounded KM survival probability and (a_{ij-}, a_{ij+}) is its confidence interval, m_{ij} and v^2_{ij} are the mean and variance of the follow-up times, $Q1_{ij}$ is the median and $[Q1_{ij}, Q3_{ij}]$ is the interquartile range of the follow-up times for study i and group j. The PCI ($j = 1$) and CABG ($j = 0$) groups have the same type of extractable data but different values.

Study i	n_{ij}	e_{ij}	r_{ij}	x^*_{ij}	y^*_{ij}	κ^*_{ij}	a^*_{ij-}	a^*_{ij+}	m_{ij}	v^2_{ij}	$Q1_{ij}$	$Q2_{ij}$	$Q3_{ij}$
Brener	✓		✓	✓	✓				✓	✓			
Palmerini	✓		✓	✓	✓						✓	✓	✓
Seung	✓		✓			✓					✓	✓	✓
Wu	✓		✓			✓			✓				
Sanmartin	✓		✓	✓	✓				✓	✓			
Buszman	✓		✓			✓			✓	✓			
Makikallio	✓			✓	✓				✓	✓			
White	✓			✓	✓						✓	✓	✓
Serryus	✓					✓	✓	✓	✓	✓			
Chieffo	✓	✓											

The ten studies have different types of extractable data:

1. All ten papers provide the number of people enrolled at baseline by treatment group.

2. Two papers provide the number of deaths observed after one year. The number of observed deaths is less than or equal to the true number of deaths, which is unknown due to loss to follow-up.

3. Four papers provide the number of people at risk after one year.

4. Eight papers provide a measure of central tendency and spread of the follow-up times. Two papers provide pooled follow-up times only. One paper provides the mean of the follow-up times by group but no variance.

5. Seven papers have a KM survival plot. Three of these plots have numerical values for the survival probabilities at year 1. For the remaining four plots, the values must be manually extracted from the plot using a ruler either in the computer screen or in print. An additional paper has a mortality rate plot with rounded mortality rates.

6. One paper provides the observed number of deaths by treatment group and does not mention follow-up times.

This meta-analysis motivates our development of appropriate methodology to incorporate the uncertainty that arises during data extraction into the meta-analysis model.

2.1.2 Simulated data

To illustrate the dangers of using observed events as a replacement of true number of deaths in a meta-analysis we construct an extreme meta-analysis comprising ten studies, with 100 subjects on average in each arm. The true odds ratio was set to be equal to one. However, loss-to-follow up times were set to be considerably different: 50% in the treatment group and 3% in the control group by year 1. The data can be found in Appendix B. We use this data to compare the results obtained when we assume that a KM survival plot is available for all studies and when it is not available for any.

2.2 Classical and Bayesian random effects meta-analysis models

In this section we briefly review three models for estimating the true population effect size for binomial outcomes, where the estimate of interest is an odds ratio: 1) the classical random effects model using the popular estimates in DerSimonian and Laird (1986) and the maximum likelihood (ML) estimates, and 2) the Bayesian random effects model. Then we describe the *naive* approach to dealing with incomplete data to use these models.

All studies cannot be considered to be equivalent experiments. Between-study variation refers to differences in design, execution and population, and these are reflected in the underlying true odds-ratios found in each study. The random effects model assumes that the odds ratios in each study follow a distribution. Within-study variation is modeled as random sampling error. Thus, the random effects model has two variance components to explain the variation in odds ratios.

Let n_{ij} be the number of subjects at baseline in study i, $i = 1, \ldots, k$ and treatment group j, where $j \in \{1 = \text{treatment}, 0 = \text{control}\}$. The numbers of subjects, s_{ij}, that experience the outcome event (say, death) in each group are independent binomial random variables with π_{ij} probability of dying before year 1, $s_{ij}|n_{ij}, \pi_{ij} \sim \text{Bin}(n_{ij}, \pi_{ij})$ for $i \in \{1, \ldots, k\}$ and $j \in \{0, 1\}$. Study i's odds ratio (OR) is

$$\text{OR}_i = \frac{\pi_{i1}/(1-\pi_{i1})}{\pi_{i0}/(1-\pi_{i0})}. \tag{2.1}$$

Let $\text{O}_i = \log \text{OR}_i$ be the observed log-odds ratio for study i and δ_i be the true log-odds ratio. The random effects model is

$$\text{O}_i|\delta_i \sim \text{N}(\delta_i, \sigma_i^2), \tag{2.2}$$

$$\delta_i \sim \text{N}(d, \tau^2), \tag{2.3}$$

where d is the population mean of true log-odds ratios and the parameter of interest. The between-study variation is τ^2 and the within-study variance is σ_i^2.

2.2.1 DerSimonian and Laird estimates

In the classical random effects model, d can be estimated as a weighted average of the observed log-odds ratios

$$\hat{d} = \frac{\sum_i w_i \mathrm{O}_i}{\sum_i w_i} \tag{2.4}$$

with variance

$$\widehat{\mathrm{Var}}(\hat{d}) = \frac{1}{\sum_i w_i}, \tag{2.5}$$

where the weights are

$$w_i = \frac{1}{\sigma_i^2 + \tau^2}. \tag{2.6}$$

In practice, the variances τ^2 and σ_i^2 are unknown. Estimated variances are used instead and the effect of this practice is generally ignored (Brockwell and Gordon, 2001). The estimate of σ_i^2 is the estimated sampling variance for an odds ratio

$$\hat{\sigma}_i^2 = \widehat{\mathrm{Var}}[\mathrm{O}_i] = \frac{1}{s_{i1}} + \frac{1}{n_{i1} - s_{i1}} + \frac{1}{s_{i2}} + \frac{1}{n_{i2} - s_{i2}}. \tag{2.7}$$

The most widely used estimate of τ^2 is the DerSimonian and Laird (1986) estimator (DSL)

$$\hat{\tau}^2 = \max\left\{0, \frac{Q_a - (k-1)}{\sum_i a_i - \sum_i a_i^2 / \sum_i a_i}\right\}, \tag{2.8}$$

where $a_i = 1/\hat{\sigma}_i^2$, and $Q_a = \sum_i a_i(\mathrm{O}_i - \hat{d})^2$ is the Cochran statistic of heterogeneity.

2.2.2 Maximum Likelihood estimates

Likelihood estimation of τ^2 is an alternative to the DSL estimator in (3.14) (Viechtbauer, 2005). The random effects model can be written as

$$\mathrm{O}_i \sim \mathrm{N}(d, \sigma_i^2 + \tau^2) \tag{2.9}$$

and the log-likelihood function is

$$\log L(d, \tau^2 | \mathrm{O}_1, \ldots, \mathrm{O}_k) = \log\left(\frac{1}{(2\pi)^{k/2}}\right) - \frac{1}{2}\sum_i \log(\sigma_i^2 + \tau^2) - \frac{1}{2}\sum_i \frac{(\mathrm{O}_i - d)^2}{\sigma_i^2 + \tau^2}. \tag{2.10}$$

Because σ_i^2 is unknown, σ_i^2 is replaced by $\hat{\sigma}_i^2$ from (2.7). To obtain the maximum likelihood estimates we take the derivatives of (2.10) with respect to d and τ^2 and set resulting equations to zero. After some manipulation, we obtain

$$\hat{\tau}^2 = \sum_i \frac{(\mathrm{O}_i - \hat{d})^2 - \hat{\sigma}_i^2}{\hat{\sigma}_i^2 + \hat{\tau}^2} \left[\sum_i \frac{1}{\hat{\sigma}_i^2 + \hat{\tau}^2} \right]^{-1}. \tag{2.11}$$

Equations (2.4) and (2.11) can be solved by iterating between $\hat{d}_t = f(\hat{\tau}_{t-1}^2)$ and $\hat{\tau}_t^2 = f(\hat{d}_t)$, with starting value $\hat{\tau}_0^2 = 0$ (Viechtbauer, 2005).

2.2.3 Bayesian model

The Bayesian random effects model allows us to include information or beliefs that may be of importance for the research question being addressed by assigning prior distributions to the parameters in the model. The fully Bayesian model accounts for all parameter uncertainty (Smith et al., 1995; Carlin, 1992).

There are several ways to perform a Bayesian random effects model for meta-analysis. Some prefer the Normal model in (2.2) and (2.3) with priors for d, σ^2 and τ^2. An alternative formulation preferred by some authors (Smith et al., 1995; Sutton and Abrams, 2001) models the outcomes s_{ij} as binomially distributed and relates the observed log-odds ratios to the probability of success using a logit transformation. Let each study's true log-odds ratio be $\delta_i = \mathrm{logit}(\pi_{i1}) - \mathrm{logit}(\pi_{i0})$, where $\mathrm{logit}(\pi_{ij}) = \log(\pi_{ij}/(1-\pi_{ij}))$, and define $u_i = [\mathrm{logit}(\pi_{Ti}) + \mathrm{logit}(\pi_{Ci})]/2$ as the logit scale average death rate for the i^{th} trial. The model is

$$s_{ij}|\pi_{ij} \sim \mathrm{Bin}(\pi_{ij}, n_{ij}) \tag{2.12}$$

$$\mathrm{logit}(\pi_{i0}) = u_i - \delta_i/2 \tag{2.13}$$

$$\mathrm{logit}(\pi_{i1}) = u_i + \delta_i/2 \tag{2.14}$$

$$\delta_i \sim \mathrm{N}(d, \tau^2) \tag{2.15}$$

$$\mu_i \sim \mathrm{N}(m, \sigma^2) \tag{2.16}$$

with priors $p(d)$, $p(\tau^2)$, $p(m)$, and $p(\sigma^2)$. By expanding this model, we are able to account for the uncertainty introduced by the data extraction.

2.2.4 The naive approach to incomplete extracted data

The classical and the Bayesian random effects meta-analysis models require four quantities to be extracted from all studies: n_{i1}, n_{i0}, s_{i1} and s_{i0} for all i. Because in most studies there is loss to follow-up, s_{ij}'s are unknown and *best-estimates* or observed events are used instead. Meta-analysts use the KM survival probability to estimate the number of deaths. Let κ^*_{ij} be the extracted reading of the KM survival probability in study i, treatment group j, which could be a rounded value extracted from the text or the ratio of two measurements off the KM plot. The two measurements off the plot are x^*_{ij}, the distance between the x-axis and the KM survival curve at year 1 and y^*_{ij}, the distance between the x-axis and the KM survival curve at year 0. Then,

$$\kappa^*_{ij} = \frac{x^*_{ij}}{y^*_{ij}}. \tag{2.17}$$

Let κ^+_{ij} be the actual KM survival estimate at year 1 that is available in the computer output, and that κ^*_{ij} is approximating.

The number of deaths calculated based on κ^*_{ij}, and κ^+_{ij} are $s^*_{ij} = n_{ij}(1 - \kappa^*_{ij})$, and $s^+_{ij} = n_{ij}(1 - \kappa^+_{ij})$, respectively. Estimates s^*_{ij} and s^+_{ij} are not necessarily integers. The *naive* approach to meta-analysis is to use s^*_{ij} as s^+_{ij} and, in turn, s^+_{ij} as s_{ij}. Then these values are fed into the classical or Bayesian random effects models. The values for s^*_{ij} for the ULMCA dataset can be found in Table 2.

Let $\mathbf{S}^* = (s^*_{ij})$, $\mathbf{S}^+ = (s^+_{ij})$, $\boldsymbol{\kappa}^* = (\kappa^*_{ij})$ and $\boldsymbol{\kappa}^+ = (\kappa^+_{ij})$, $i = 1, \ldots, k, j = 1, 0$. Let $\boldsymbol{\theta}$ be the vector of parameters $(\pi_{ij}, \delta_i, \mu_i)$ in the model. The naive Bayesian model computes the posterior

$$f(\boldsymbol{\theta}|\mathbf{S}^*) \propto f(\mathbf{S}^*|\boldsymbol{\theta}) f(\boldsymbol{\theta}), \tag{2.18}$$

and incorrectly uses $f(\boldsymbol{\theta}|\mathbf{S}^*)$ as a replacement for $f(\boldsymbol{\theta}|\mathbf{S}^+)$, ignoring that $\boldsymbol{\kappa}^*$ are approximated or rounded values of $\boldsymbol{\kappa}^+$, and in turn uses $f(\boldsymbol{\theta}|\mathbf{S}^+)$ as a substitute for $f(\boldsymbol{\theta}|\mathbf{S})$, ignoring that $\mathbf{S}$ are not observed but estimated from $\boldsymbol{\kappa}^+$, which are estimators themselves with additional associated uncertainty.

Table 2.2: Extracted data for the ULMCA dataset according to the naive approach.

Study	PCI			CABG		
	$\kappa_{i1}^* \times 100\%$	n_{i1}	s_{i1}^*	$\kappa_{i0}^* \times 100\%$	n_{i0}	s_{i0}^*
Brener et al. (2008)	93.3%	97	6.52	94.2%	190	11.08
Palmerini et al. (2006)	89.2%	154	16.68	87.1%	157	20.28
Seung et al. (2008)	96.7%	542	17.89	96.3%	542	20.05
Wu et al. (2008)	83.9%	135	21.74	94.1%	135	7.97
Sanmartin et al. (2007)	88.5%	96	11.04	83.5%	245	40.55
Buszman et al. (2008)	98.1%	52	0.99	92.5%	53	3.98
Makikallio et al. (2008)	94.4%	49	2.74	89.0%	238	26.23
White et al. (2008a)	89.8%	67	6.83	93.2%	67	4.56
Serryus (2008)	95.8%	357	14.99	95.5%	348	15.66
Chieffo et al. (2006)	–	107	3.0^a	–	142	9.0^a

Note: ^{a} Observed number of deaths extracted from the published paper.

2.3 The uncertain reading-estimated events model

We propose the Uncertain Reading-Estimated Events (UR-EE) model, which does not substitute $\mathbf{S}^*$ for $\mathbf{S}^+$ for $\mathbf{S}$. Instead, it incorporates the uncertainty in the estimator $\boldsymbol{\kappa}^+$ by averaging over the possible values of true deaths $\mathbf{S}$ given $\boldsymbol{\kappa}^+$

$$f(\boldsymbol{\theta}|\mathbf{S}^+) = \int f(\boldsymbol{\theta}|\mathbf{S}) f(\mathbf{S}|\mathbf{S}^+) d\mathbf{S}, \tag{2.19}$$

and over the possible values of $\mathbf{S}^+$ given the extracted $\mathbf{S}^*$

$$f(\boldsymbol{\theta}|\mathbf{S}^*) = \int f(\boldsymbol{\theta}|\mathbf{S}^+) f(\mathbf{S}^+|\mathbf{S}^*) d\mathbf{S}^+, \tag{2.20}$$

to obtain a posterior of the parameters given the extracted data $\mathbf{S}^*$

$$f(\boldsymbol{\theta}|\mathbf{S}^*) = \int \int f(\boldsymbol{\theta}|\mathbf{S}) f(\mathbf{S}|\mathbf{S}^+) f(\mathbf{S}^+|\mathbf{S}^*) d\mathbf{S}^+ d\mathbf{S}. \tag{2.21}$$

We call $f(\mathbf{S}^+|\mathbf{S}^*)$ the Uncertain Reading (UR) density and $f(\mathbf{S}|\mathbf{S}^+)$ the Estimated Events (EE) density. The UR density captures the uncertainty due to not having the exact KM survival

probability. The EE density captures the uncertainty of the number of deaths estimated using the KM estimator around the true number of deaths due to censoring.

Both the naive and the UR-EE model compute a $f(\boldsymbol{\theta}|\mathbf{S}^*)$ posterior. The naive model is overly optimistic while the UR-EE model does not make the incorrect assumption that $f(\boldsymbol{\theta}|\mathbf{S}^*) = f(\boldsymbol{\theta}|\mathbf{S}^+) = f(\boldsymbol{\theta}|\mathbf{S})$.

Due to the additional incorporated uncertainty in the UR-EE model, we expect that

$$\text{Var}_{\text{UR-EE}}(d|\mathbf{S}^*) > \text{Var}_{\text{NAIVE}}(d|\mathbf{S}^*). \tag{2.22}$$

The incorporation of additional, previously ignored uncertainty in the model translates to a reduction in the effective sample size of the meta-analysis. Let $n = \sum n_{ij}$, and let $n_{\text{UR-EE}}$ be the effective sample size of the meta-analysis under the UR-EE model. Then

$$n_{\text{UR-EE}} = \frac{\text{Var}_{\text{NAIVE}}(d|\mathbf{S}^*)}{\text{Var}_{\text{UR-EE}}(d|\mathbf{S}^*)} n, \tag{2.23}$$

and we expect that $n_{\text{UR-EE}} < n$.

In any Bayesian model we have observed data and parameters (random variables). In the naive Bayesian model for meta-analysis the "observed" data is not actually observed but a best-estimate with s^*_{ij} substituting for s_{ij}. In contrast, the UR-EE model's observed data is the extracted data, n_{ij}, κ^*_{ij}, and the unknown s^+_{ij} and s_{ij} are treated as random variables.

The fully Bayesian UR-EE model adds models

$$f_{ij}(s^*_{ij}|s^+_{ij}) \tag{2.24}$$

$$f_{ij}(s^+_{ij}|s_{ij}) \tag{2.25}$$

for all i and j to equations (2.12) to (2.15). We call equation (2.24) the Uncertain Reading density and equation (2.25) the Estimated Events density. The choices of densities for (2.24) and (2.25) are different for every i due to the differences in extracted data in each study. In our example, (2.24) and (2.25) have the same form for $j = 0, j = 1$, however that is due to the reporting in the studies and not a requirement of our methodology. We describe the construction of the UR and EE densities next.

2.3.1 The Uncertain Reading density

The UR density $f_{ij}(s_{ij}^+|s_{ij}^*)$ is obtained from $f_{ij}(\kappa_{ij}^+|\kappa_{ij}^*)$, which models the KM survival probability values based on the extracted rounded or manually measured κ_{ij}^*. Because κ_{ij}^* could be extracted in at least two ways, the UR density is different for each study i.

Case 1: Rounded κ_{ij}^*. In studies by Seung, Wu, Buszman and Serryus the rounded value of the survival probability κ_{ij}^* is extracted from the text or a number printed on the KM plot. Assuming that the probability was rounded to three-digit accuracy, we model the actual KM survival probability κ_{ij}^+ as uniformly distributed centered at κ_{ij}^*

$$\kappa_{ij}^+|\kappa_{ij}^* \sim \text{Unif}(\kappa_{ij}^* - 0.0005, \kappa_{ij}^* + 0.0005), \tag{2.26}$$

$$s_{ij}^+|s_{ij}^* \sim \text{Unif}(s_{ij}^* - 0.0005n_{ij}, s_{ij}^* + 0.0005n_{ij}). \tag{2.27}$$

These equations can easily accommodate rounding to different levels of accuracy by changing the minimum and maximum values in (2.26)-(2.27).

Case 2: Measured κ_{ij}^*. Let x_{ij} and y_{ij} be the true unknown distances from the x-axis to the KM survival curve at year 0 and 1, and let x_{ij}^* and y_{ij}^* be the measurements taken off the KM plot from the x-axis to the survival curve at year 0 and year 1. Then $\kappa_{ij}^* = x_{ij}^*/y_{ij}^*$. To account for the measurement error in x_{ij}^* and y_{ij}^* we assume that the uncertainty in x_{ij}^* and y_{ij}^* is similar to rounding error, in that both are measured by a ruler with equally spaced tick marks, and that the maximal error in $x_{ij}^* - x_{ij}$ and $y_{ij}^* - y_{ij}$ is known, typically 1/2 the distance w_{ij} between the tick marks. Then

$$x_{ij} \sim \text{Unif}(x_{ij}^* - (w_{ij}/2), x_{ij}^* + (w_{ij}/2)), \tag{2.28}$$

$$y_{ij} \sim \text{Unif}(y_{ij}^* - (w_{ij}/2), y_{ij}^* + (w_{ij}/2)), \tag{2.29}$$

and set the true unknown KM survival probability to be $\kappa_{ij}^+ = x_{ij}/y_{ij}$. The density $f_{ij}(\kappa_{ij}^+|x_{ij}^*, y_{ij}^*)$ is given by the ratio of two uniform random variables and has positive support on the range of values

$$\frac{x_{ij}^* - \frac{w_{ij}}{2}}{y_{ij}^* + \frac{w_{ij}}{2}} \leq \kappa_{ij}^+ \leq \frac{x_{ij}^* + \frac{w_{ij}}{2}}{y_{ij}^* - \frac{w_{ij}}{2}}. \tag{2.30}$$

The exact form of the piecewise density $f_{ij}(\kappa_{ij}^+|x_{ij}^*, y_{ij}^*)$ is given in Appendix A. Using $f_{ij}(\kappa_{ij}^+|x_{ij}^*, y_{ij}^*)$, an exact density $f_{ij}(s_{ij}^+|x_{ij}^*, y_{ij}^*)$ is immediate, but we derive a distribution to

be a convenient normal approximation to $f_{ij}(s_{ij}^{+}|x_{ij}^{*}, y_{ij}^{*})$ that is

$$s_{ij}^{+}|s_{ij}^{*} \sim \mathrm{N}\left(s_{ij}^{*}, \left(\frac{n_{ij}}{6}\right)^{2}\left[\frac{x_{ij}^{*}+\frac{w_{ij}}{2}}{y_{ij}^{*}-\frac{w_{ij}}{2}} - \frac{x_{ij}^{*}-\frac{w_{ij}}{2}}{y_{ij}^{*}+\frac{w_{ij}}{2}}\right]^{2}\right). \tag{2.31}$$

This has mean and mode at s_{ij}^{*} and standard deviation such that 99.7% of values fall within the support in (2.30). Considering that the exact UR distribution, $f_{ij}(s_{ij}^{+}|s_{ij}^{*})$, is a complicated piecewise function, and that this case holds for several of the studies in the meta-analysis, a normal approximation facilitates implementation of the proposed methodology.

2.3.2 The Estimated Events density

The KM survival estimate κ_{ij}^{+} has two kinds of uncertainty associated with it: binomial sampling and additional uncertainty due to censoring. The binomial sampling is addressed naturally in (2.12). The EE distribution addresses the additional uncertainty due to censoring. The EE distribution, $f_{ij}(s_{ij}|s_{ij}^{+})$, conditions on the estimated number of deaths $s_{ij}^{+} = n_{ij}(1-\kappa_{ij}^{+})$, calculated using κ_{ij}^{+} as input and gives the density of the random variable s_{ij} as output. This distribution is constructed using other information from the i^{th} paper. Since the EE distribution models numbers of deaths, a discrete distribution that assigns a probability to each feasible integer value of s_{ij} is desirable. However, it is not straightforward to determine said probabilities. Thus, we approximate $f_{ij}(s_{ij}|s_{ij}^{+})$ with a truncated normal density, $\mathrm{TN}(s_{ij}^{+}, \mathrm{B}_{ij}, \mathrm{LB}_{ij}, \mathrm{UB}_{ij})$, centered around s_{ij}^{+} with variance B_{ij}, truncated at lower and upper bounds, LB_{ij} and UB_{ij}, where the values of B_{ij}, LB_{ij} and UB_{ij} are dependent on the number of censored people, c_{ij}, in each study. A lower bound, $\mathrm{LB}_{ij} = e_{ij}$, assumes that all censored people survived and an upper bound $\mathrm{UB}_{ij} = e_{ij} + c_{ij}$ assumes that all censored people died. Improved lower and upper bounds can be computed with additional information available for extraction, such as number of people at risk r_{ij}. Simulations (not shown) showed that asymmetry of the density $\mathrm{TN}(s_{ij}^{+}, \mathrm{B}_{ij}, \mathrm{LB}_{ij}, \mathrm{UB}_{ij})$ results in biased estimates of s_{ij}. Thus, we propose a truncated normal with symmetric truncation points

$$s_{ij}|s_{ij}^{+} \sim \mathrm{TN}(s_{ij}^{+}, \mathrm{B}_{ij}, s_{ij}^{+} - \min\{s_{ij}^{+} - \mathrm{LB}_{ij}, \mathrm{UB}_{ij} - s_{ij}^{+}\}, s_{ij}^{+} + \min\{s_{ij}^{+} - \mathrm{LB}_{ij}, \mathrm{UB}_{ij} - s_{ij}^{+}\}). \tag{2.32}$$

To estimate the number of censored people, we use information extracted from the papers about the follow-up distribution times and number of people at risk found in the paper. We model the log follow-up times as $\mathrm{N}(\psi_{ij}, \phi_{ij})$, with mean ψ_{ij} and variance ϕ_{ij}. Let λ_{ij} be the probability of being censored before year 1, then the estimated number of censored subjects is $c_{ij} = n_{ij}\lambda_{ij}$. The information about follow-up times available for extraction in each paper varies; some papers give means and variances, others give quartiles (Table 2.3). We enumerate the following cases of extracted data to calculate ψ_{ij} and ϕ_{ij}.

Table 2.3: Follow-up times (in days) for studies in the ULMCA dataset by treatment and control group.

Study			PCI					CABG		
	m_{i0}	v_{i0}^2	$Q1_{i0}$	$Q2_{i0}$	$Q3_{i0}$	m_{i1}	v_{i1}^2	$Q1_{i1}$	$Q2_{i1}$	$Q3_{i1}$
Brener	1020	840				3660	3180			
Palmerini			2	417	830			105	430	730
Seung			681	1152	1590			688	1017	1451
Wu	732	402.25				753	1060.25			
Sanmartin	474.5	292				1168	584			
Buszman	840	297				840	297			
Makikallio	360	180				360	180			
White			192	362	586			226	600	977
Serryus										
Chieffo										

Case 1: Follow-up time mean and variance by treatment group. Let m_{ij}, v_{ij} be the extracted mean and variance of the follow-up times. Then, $\psi_{ij} = \log\left[m_{ij}^2(v_{ij} + m_{ij}^2)^{-1/2}\right]$, and $\phi_{ij} = \log\left(1 + (v_{ij}/m_{ij}^2)\right)$.

Case 2: Follow-up time median and lower and upper quartiles by treatment group. Let $Q2_{ij}$ and $[Q1_{ij}, Q3_{ij}]$ be the extracted median and upper and lower quartiles. We set

$\psi_{ij} = \log(Q2_{ij})$ and $\phi_{ij} = [\log(Q3_{ij}) - \log(Q1_{ij})] \,/\, [\Phi^{-1}(0.75) - \Phi^{-1}(0.25)]$.

Case 3: Follow-up time not by treatment group. When the follow-up time summary statistics are pooled summaries of the treatment and control groups, set $\psi_{i1} = \psi_{i0}$ and $\phi_{i1} = \phi_{i0}$ and calculate both given the pooled summary statistics.

Case 4: Follow-up time mean but no variance. We use the mean of the standard deviations in the other studies of the meta-analysis as a value for v_{ij}.

To define UB_{ij} and LB_{ij}, we need the number of subjects at risk, r_{ij}, and the number of observed deaths e_{ij} at year 1. The availability of this information varies across studies. We consider the following cases of information available for extraction.

Case 1: Observed deaths and people at risk are both given in the paper. Define $\text{UB}_{ij} = n_{ij} - r_{ij}$ and $\text{LB}_{ij} = e_{ij}$.

Case 2: Observed deaths is given but people at risk is not. Define $\text{UB}_{ij} = e_{ij} + c_{ij}$ and $\text{LB}_{ij} = e_{ij}$.

Case 3: People at risk is given but observed deaths is not. Define $\text{UB}_{ij} = n_{ij} - r_{ij}$ and $\text{LB}_{ij} = \max\{0, n_{ij} - r_{ij} - c_{ij}\}$.

Case 4: Neither people at risk nor observed deaths are given. Due to the lack of information, conservative bounds are $\text{UB}_{ij} = n_{ij}$ and $\text{LB}_{ij} = 0$.

The variance $\text{B}_{ij} = \text{Var}(n_{ij}(1 - \kappa^*_{ij})) = n^2_{ij} b_{ij}$ depends on, b_{ij}, the variance of the KM survival probability. However, studies rarely report KM confidence intervals or KM standard errors so b_{ij} is unknown. To approximate the value of the variance we simulated studies with characteristics similar to the studies contributing to the ULMCA dataset. We found that the preferred formula depends on the amount of censoring.

1. *Greenwood simplified estimate.* When censoring at 12 months is less than 25%, a simplified version of the Greenwood formula was satisfactory $b_{ij} = (\kappa^*_{ij})^2 e_{ij} / [n_{ij}(n_{ij} - e_{ij})]$.

2. *Censoring proportional estimate.* When censoring at 12 months ranged from 25% to 35%, we found the Greenwood simplified estimate and the censoring proportional estimate, $b_{ij} = (c_{ij}/n_{ij})\kappa^*_{ij}(1-\kappa^*_{ij})$ to be very close to each other. So we suggest the average of the two. When censoring ranged from 35% to 50% the censoring proportional estimate was superior while the Greenwood simplified estimate consistently underestimated the KM variance.

3. *Follow-up area under the curve (AUC) proportional estimate.* The censoring proportion c_{ij}/n_{ij} does not take into consideration for how long a censored subject was followed before being lost. Let auc_{ij} be the area under the curve that represents the person-years lost to follow-up, and the total area be $\text{total}_{ij} = n_{ij} * 1$ person-years, then the follow-up AUC proportional estimate is $b_{ij} = (\text{auc}_{ij} \,/\, \text{total}_{ij})\kappa^*_{ij}(1-\kappa^*_{ij})$. For censoring that ranged from 50% to 70%, we found this estimate to be as good as the censoring proportional estimate, so we suggest using the average of the two. For censoring in excess of 70% the AUC proportional estimate was adequate while the censoring proportional estimate consistently overestimated the KM variance.

2.3.3 Survival probabilities not available for extraction

When a paper does not include a KM plot but it includes the number of observed deaths, the naive approach is to use the number of observed deaths as a replacement for the number of true deaths. This underestimates the true value and leads to biased information being input into the meta-analysis. We propose the following estimate of the probability of death at year 1, s_{ij}/n_{ij} when only observed deaths are available

$$k^*_{ij} = \frac{e_{ij}}{n_{ij}}\left(\frac{1}{1-\text{auc}_{ij}}\right), \tag{2.33}$$

for $0 < e_{ij}/n_{ij} < 0.5$ and $0 < \text{auc}_{ij} < 0.5$. Then set $\kappa^+_{ij} = k^*_{ij}$. In our simulations, (2.33) had smaller mean square error than e_{ij}/n_{ij} in estimating s_{ij}/n_{ij} for all ranges of censoring and true values of s_{ij}/n_{ij}.

2.4 Results

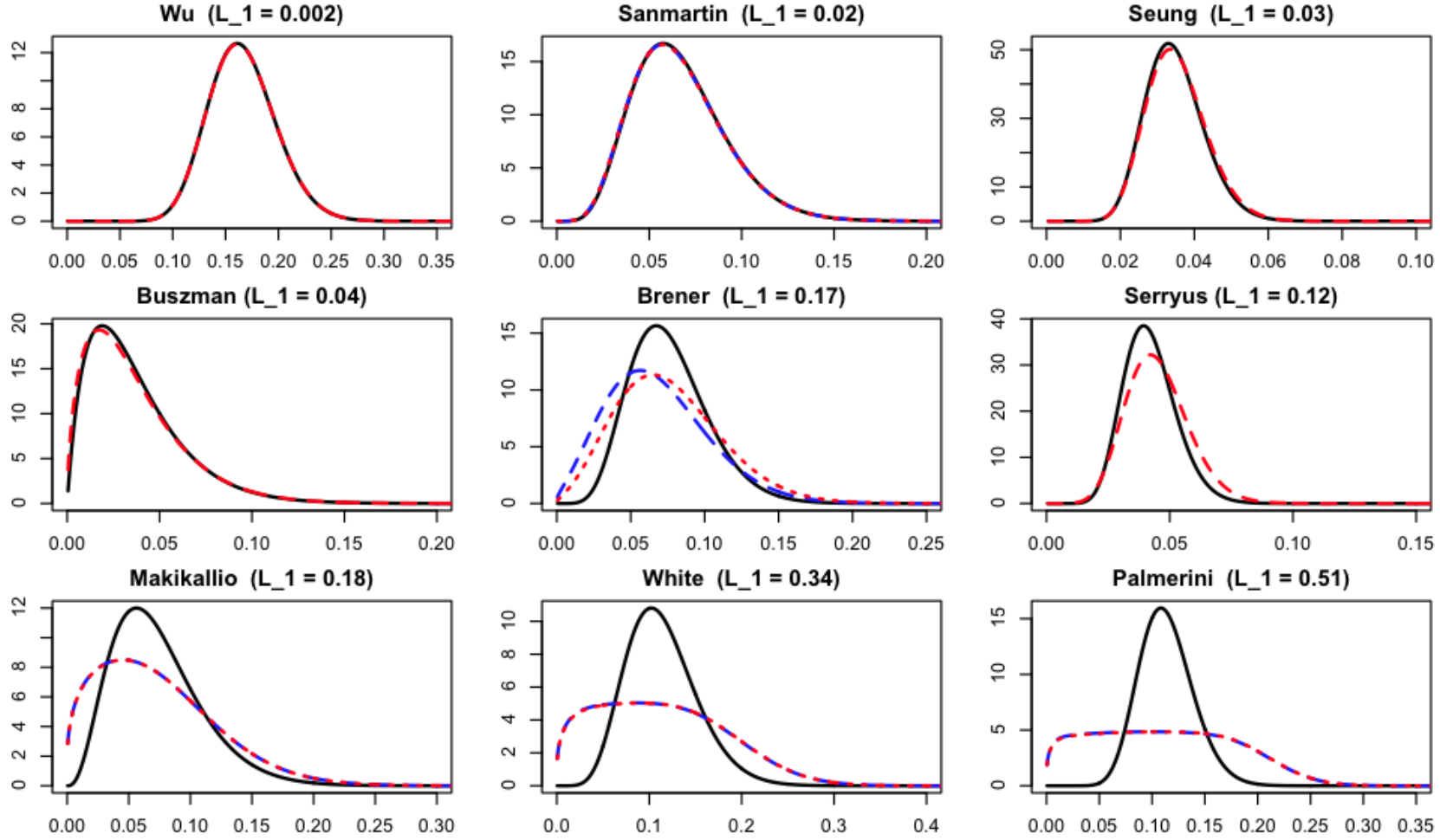
2.4.1 Prior specification

We use proper weakly informative priors. It is unlikely that the underlying odds ratio associated with the model would exceed 100 in favor of either PCI or CABG (Smith et al., 1995). With this constraint, a range for d is $(-4.6, 4.6)$. Thus, a normal $\mathrm{N}(0, 2.35^2)$ prior for d has the mean $\pm$ 1.96 SD covering 95% of the prior range. The event (death) rate logit^{-1} is very likely to lie in the interval $(0.001, 0.999)$ (Larose and Dey, 1997), which corresponds to $(-3.89, 3.89)$ as an interval for m. Assuming a normal distribution, the prior is $\mathrm{N}(0, 1.98^2)$. We use informative Inverse Gamma priors $\mathrm{IG}(3, 2)$ for the variance parameters σ^2 and τ^2 (Gelman et al., 2004).

2.4.2 Conditional densities

Figure 1 compares the conditional densities $f(\pi_{ij}|s^*_{ij})$ using the naive Bayesian model (dashed lines) and $f(\pi_{ij}|\text{extracted data})$ using the UR-EE model (solid lines) in the ULMCA meta-analysis of mortality at year 1. The dotted line corresponds to the use of the ratio of two uniform random variables as described in Appendix A, while the dashed line uses the normal approximation in (2.31). Thus, equation (2.31) serves as an adequate approximation for the ratio of two uniform random variables, in our situation. Curves corresponding to the Chieffo paper are not included as the extracted data did not include follow-up information. Sensitivity analysis (not shown) showed that multiple assumptions about censoring, including no censoring, in Chieffo's data did not change the ULMCA meta-analysis result.

Figure 2.1: Naive conditional densities, $f(\pi_{ij}|s^*_{ij})$ in solid lines and UR-EE conditional densities $f(\pi_{ij}|\text{extracted data})$ in dashed lines. Where applicable, the dotted lines (in red) uses the ratio of two uniform random variables while the dashed line (in blue) uses the normal approximation.



After accounting for the uncertainty that arises from the estimates of the number of deaths, there is an increase in the variance of all conditional densities, confirming the conservativeness of our model. To quantify the effect, we use the L_1 statistic to measure the distance between the naive and proposed densities (Weiss, 1996). The statistic takes values [0,1], where a value of zero indicates no difference and a value of one indicates maximal difference between the densities. The densities in Figure 1 are sorted in increasing order of L_1. The effect of accounting for the different types of extracted data varies greatly among studies. Minimal differences between the naive and proposed densities are found for the studies by Wu, Sanmartin, Seung and Buzman where $L_1 < 0.05$. Large differences of $L_1 > 0.2$ are found in the studies by White and Palmerini. The densities suggest that having a smaller amount of extractable data available from a study results in large increases in the study OR's variance compared to having full information.

2.4.3 Posterior computation

Because the posterior is not available in closed form, we base our inferences on an MCMC simulation from the posterior distribution $f(\boldsymbol{\theta}|\mathbf{S}^*)$. We use a Gibbs step (Gelfand and Smith, 1990) whenever conditional posteriors are available, as for d, σ^2, m, τ^2. When the conditional distribution is of intractable form, such as for m_i and δ_i, we use a Metropolis step (Metropolis et al., 1953), resulting in a hybrid sampler (Tierney, 1994). Gibbs steps are included to sample for s_{ij}^+, and s_{ij}^* using the Uncertain Reading and Estimated Events densities.

The MCMC algorithm to compute the posterior of the UR-EE model was implemented in R (R Core Team, 2012) using the "`ureepkg`." The package was developed specifically for the ULMCA and simulated data meta-analyses, and is available upon request from the authors. After an initial burn-in period of 2,000 iterations, we generated an additional $100,000$ iterations, retaining every tenth iteration. We used three chains with different starting points and assessed convergence by inspecting the densities and time series plots and the convergence diagnostics of Gelman and Rubin (1992).

2.4.4 Sensitivity to prior specification

We examine sensitivity of the inference of d to the prior specification of the variance components. We use $\text{IG}(\epsilon, \epsilon)$ as an attempt at uninformativeness (Gelman, 2006) for the prior distributions of σ^2 and τ^2. The resulting estimates for the log-odds ratio and its standard deviation do not vary substantially for different choices of prior when using $\epsilon \in \{0.1, 1\}$ when compared to those obtained using an $\text{IG}(3, 2)$ as a prior.

2.4.5 Results with ULMCA dataset

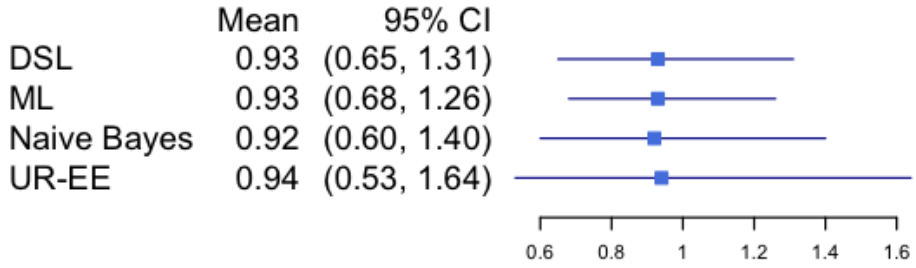
Table 2.4 summarizes the results from the naive and UR-EE Bayesian models or the meta-analyses of mortality of the ULMCA dataset. The results confirm that there is an increase in variance in all parameters as expected from the discussion around (2.22). The increase in the standard deviation of the log-odds ratio is 33%. Across all parameters in the model, there was an average increase of 39% in the standard deviation.

Table 2.4: Posterior mean and 95% credible intervals for parameters d, σ^2, m and τ^2 in the naive and UR-EE Bayesian models for mortality of the ULMCA dataset.

		Mean		SD	
		Naive	UR-EE	Naive	UR-EE
Mortality					
	d	-0.09	-0.07	0.22	0.29
	σ^2	0.25	0.47	0.29	0.48
	m	-2.65	-2.71	0.18	0.20
	τ^2	0.23	0.29	0. 17	0.24

Figure 2 gives OR estimates and 95% intervals for the mortality meta-analysis under the DSL, ML, naive Bayes and UR-EE models. As expected the UR-EE model interval is wider than the naive models. According to (2.23), the increase in the standard deviation over the naive Bayes model is equivalent to a reduction of 43% of the meta-analysis sample size.

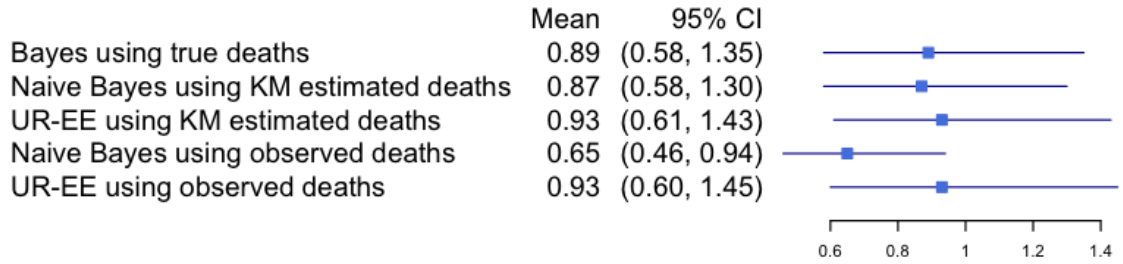
Figure 2.2: Mortality meta-analysis for ULMCA dataset: odds-ratio mean (Bayes and UR-EE) or point estimate (DSL, ML), exp(d), and 95% confidence interval under various models.



2.4.6 Results with simulated dataset

The simulated dataset has large censoring (50%) in the treatment arm and little to no censoring (3%) among controls at year 1 in all ten studies. Log follow-up times $\mathrm{N}(\psi_{ij}, \phi_{ij})$ were set to be $\mathrm{N}(5.89, 0.83)$ and $\mathrm{N}(7.05, 0.63)$, for all i, in the treatment and control groups respectively. The follow-up times are measured in days. The values for the parameters in (2.13) to (2.16) were set to $d = 1.0$, $\tau^2 = 0.4$, $m = -0.8$, and $\sigma^2 = 0.1$. Figure 3 shows the results obtained under different scenarios. Because this is simulated data we are able to compute the odds ratio and confidence interval using the true number of deaths, which is the ideal extracted data that is typically unavailable. If we assume that the KM survival probabilities were available for extraction in both arms in all studies then we can compute naive Bayes and UR-EE models to obtain similar conclusions as in the ULMCA dataset. The UR-EE confidence interval is wider than the one obtained under the naive Bayes model, and both are close to the one obtained from the model based on the true data. The power of the UR-EE model is portrayed in the extreme case where KM survival probabilities are not available in all studies. The meta-analysis that uses observed deaths as a substitute for true deaths results in a biased and inaccurate result: it points to a significant difference in two arms where none exists! Using the UR-EE model on observed deaths results in an adequate odds ratio and confidence interval.

Figure 2.3: Mortality meta-analysis simulated: odds-ratio mean exp(d) and 95% confidence interval under various models.



2.5 Discussion

In the ULMCA meta-analysis we re-evaluated, the estimates from ML were slightly more conservative than those obtained using DSL. The naive Bayesian model was more conservative than both DSL and ML, and the UR-EE model was, as expected, the most conservative of all. The odds ratio for mortality reported in Naik et al. (2009), OR=1.00 [95% CI: 0.70 to 1.41], did not use any of the naive methods. Instead it incorporated uncertainty in the unknown number of deaths by reducing the sample sizes of each study. Naik's interval is wider than the ones from DSL and ML but not as wide as the Bayesian naive and UR-EE models.

The UR-EE model penalizes the lack of important information available for extraction in the papers by increasing the variance in the estimates of the study level and global parameters. Given that each study can present its own challenges, our model potentially requires individual models carefully constructed for each study's contribution to the meta-analysis. While the algorithm increases in complexity the more studies it contains, the results are more conservative and more accurate than traditional meta-analysis methods such as DSL, ML or a naive Bayesian model, which do not account for the uncertainty of their extracted data.

Under the UR-EE model, there is an increase in the standard deviation of all parameters in the model. Thus, non-significant results would be even more non-significant. The conclusion for the ULMCA meta-analysis was the same regardless of the method: there is no

statistical difference in the rates of mortality between PCI and CABG. While the UR-EE model did not change the conclusions here, in other datasets where the odds ratio intervals are different but close to 1.0, it is possible that under the UR-EE model, the wider interval would include 1.0 causing the results to become non-significant. For example, if in our meta-analysis of mortality, the estimated log-odds $\hat{d}$ had been within $(-0.57, -0.43)$ or $(0.43, 0.57)$ instead of zero, with the same standard deviation as in Table 2.4, the resulting confidence intervals for the odds ratio would be significant in the naive model and non-significant in the UR-EE model.

Our model's superiority over the naive models was best exemplified in the extreme case of the simulated data. The data was simulated such that a naive approach would result in biased and inaccurate results, which is the case where observed values are used instead of true values. When feeding that same data into our UR-EE model, we obtained accurate results. While it was also shown that a naive meta-analysis based on best-estimates obtained from KM survival probabilities was adequate in our situation, a typical meta-analysis contains a mix of studies where KM plots are available and where they are not and observed values are used instead, making our UR-EE model a necessary requirement in the computation of meta-analyses of odds-ratio with incomplete extracted data.

We have presented a model that incorporates the uncertainty that arises during data-extraction in the meta-analysis model. Our model would not be necessary if published studies provided better estimates and standard errors and complete information on their follow-up times. Unfortunately, the poor reporting of summary statistics will continue to prevail in the published literature, making the UR-EE model a requirement for any meta-analysis performed that does not include individual patient data. While the overall conclusion may not change from those obtained from naive approaches, our method allows researchers to explore the impact of the uncertainty from the missing extracted data in the final estimates. For example, if a meta-analysis yields a significant result using standard methods, we strongly recommend our method be run in parallel to confirm the significance of the results after accounting for the uncertainty of the missing extracted data to avoid the situation portrayed in our simulated dataset.

To improve estimates from meta-analyses, we recommend that referees and editors of journals require submitted papers to include complete survival plots that contain KM estimates and numbers of people at standardized time points, confidence bands for the KM estimates for both treatment groups, and more complete follow-up time information for both groups, with minimum means, standard deviations and quartiles and maximum times. If these data are not going be part of the published paper, it should be made available in online supplemental materials for future extraction.

CHAPTER 3

Meta-analysis of mean differences with missing standard errors

Meta-analysis is commonly used to quantitatively pool the evidence from multiple studies on a particular research question. In the health sciences, meta-analyses are used to obtain an estimate of a treatment intervention's effectiveness along with a measure of precision. Ideally, meta-analysis would analyze the combined individual patient data (IPD) from all studies (Stewart and Parmar, 1993; Lambert et al., 2002; Berlin et al., 2002). In practice, this is rarely done (Kovalchik, 2012). IPD can be costly, time consuming or not possible to obtain (Riley et al., 2013). Instead, meta-analyses rely on data extracted from journal articles in the published literature or from presentations at major conferences. This is known as meta-analysis of aggregate data. Extracted data from each study typically consists of a treatment effect and its variance.

In a meta-analysis of continuous outcomes the data typically required from each study are the mean difference and standard error of the mean difference. Existing software to perform meta-analysis customarily operate on these data (Viechtbauer, 2010). However, standard errors (SE) for the mean difference are typically missing from published studies (Wiebe et al., 2006; Furukawa et al., 2006). For example, Streiner and Joffe (1998) reported that 87% of the studies in their systematic review of antidepressants did not report SEs, 62% of SEs were missing in the meta-analysis of sodium reduction on blood pressure in Follmann et al. (1992) and 87% to 92% of SEs were missing in a meta-analysis of various renal function outcomes after kidney donation (Philbrook et al., 2007). While contacting the authors of the studies to request the missing data is recommended, in practice the proportion of SEs recovered this way is very low and a large proportion of standard errors remain missing

(Furukawa et al., 2006; Philbrook et al., 2007).

There have been efforts to promote better reporting in clinical trials (Moher et al., 2001). However, the problem of missing SEs has been described as "annoyingly common" (Furukawa et al., 2006) and recovering all missing data by contacting authors as "impossible" (Philbrook et al., 2007). Because missing SEs can occur in the majority of studies in the meta-analysis, omitting studies with missing SEs is not considered a reasonable option. In particular, the standard errors are used in the weighted computation of the overall meta-analysis estimate and omitting studies could lead to bias (Wiebe et al., 2006; Furukawa et al., 2006; Philbrook et al., 2007). Thus, meta-analysts are left to improvise (Wiebe et al., 2006). Sometimes it is possible to compute the missing SE from an available 95% confidence interval or p-value. If these are not available it is common to use other information available in the published paper such as the standard deviations (SD) for the baseline and follow-up mean change for the treatment and control groups to estimate the missing SE. When baseline and follow-up information are not available a common approach is to impute the missing SDs using the average SD from the studies where it is reported. Since studies vary in the amount of information available for extraction, in most cases, a combination of strategies is used to impute all the missing SEs of the meta-analysis. However, introducing imputed SEs in the meta-analysis as observed data is problematic, since it can lead to unjustifiable certainty in the results and to potentially inaccurate conclusions.

We propose the Uncertain Standard Error (USE) Bayesian model to account for the uncertainty that arises at the data extraction step of a meta-analysis of continuous outcomes. We formally construct a model that accounts for the contribution of each study to the meta-analysis using the data available and extracted from it. Data available for extraction includes the number of participants at baseline, and may include one or more of the following: baseline and follow-up means and standard deviations for the treatment and control group, and correlations. The Bayesian model that we propose models the missing SEs using other SDs available in the study, allows for different within-subject correlations in the treatment and control groups, and more importantly, allows for multiple patterns of missingness. The USE model improves the validity of results from meta-analysis by accommodating all the

uncertainty in the data input to the meta-analysis.

In section 3.1 we review existing methods to perform a meta-analysis of mean differences with missing SEs. We introduce the motivating dataset in section 3.2. We describe the USE model in section 3.3. In section 3.4 we apply the existing methods and the USE model to our data. The chapter finishes with discussion.

3.1 Methods for meta-analysis of mean differences with missing standard errors

In this section we briefly review current methods to perform a meta-analysis of mean differences and the strategies taken by meta-analysts to deal with the missing SEs.

Let n_{ij} be the sample size of study i, $i = 1, \ldots, K$ where K is the number of studies, and group j, where $j = 1$ is the treatment group and $j = 0$ is the control group. Let μ_{ijb} be the mean at baseline for study i and treatment group j. Similarly μ_{ijf} is the mean at follow-up. Let $\delta_{ij} = \mu_{ijf} - \mu_{ijb}$ be the mean change in treatment group j and study i. The treatment effect is the difference of mean changes between the two groups

$$\Delta_i = \delta_{i1} - \delta_{i0} = (\mu_{i1f} - \mu_{i1b}) - (\mu_{i0f} - \mu_{i0b}). \tag{3.1}$$

Other measurements of the treatment effect have been used, for example using only the follow-up values, $\Delta_i = \mu_{i1f} - \mu_{i0f}$, and ignoring the baseline values, but this is not recommended (Riley et al., 2013).

The sample estimate of Δ_i is

$$D_i = (\bar{X}_{i1f} - \bar{X}_{i1b}) - (\bar{X}_{i0f} - \bar{X}_{i0b}), \tag{3.2}$$

where $\bar{X}_{ijb}$ and $\bar{X}_{ijf}$ are the sample means at baseline and follow-up for study i and treatment group j. If we assume that the variance of the change for one subject is the same for the two groups, $\mathrm{Var}(\bar{X}_{i1f} - \bar{X}_{i1b}) = \mathrm{Var}(\bar{X}_{i0f} - \bar{X}_{i0b}) = \lambda_i^2$, then the variance of D_i is

$$\mathrm{Var}(D_i) = \left(\frac{1}{n_{i1}} + \frac{1}{n_{i0}}\right)\lambda_i^2 \tag{3.3}$$

and the standard error is

$$\mathrm{se}_{Di} = \sqrt{\mathrm{Var}(D_i)}. \tag{3.4}$$

An estimate of the variance of D_i is

$$\widehat{\mathrm{Var}}(D_i) = \left(\frac{1}{n_{i1}} + \frac{1}{n_{i0}}\right) s_{ip}^2 \tag{3.5}$$

where

$$s_{ip}^2 = \frac{(n_{i1}-1)\mathrm{s}_{i1}^2 + (n_{i0}-1)\mathrm{s}_{i0}^2}{n_{i1}+n_{i0}-2} \tag{3.6}$$

is an unbiased estimator of λ_i^2, and s_{ij} are the sample standard deviations of change in the two groups. If s_{ij} are not available, they can be calculated from the sample standard deviations at baseline and follow-up, s_{ijb} and s_{ijf}, as

$$s_{ij}^2 = s_{ijb}^2 + s_{ijf}^2 - 2r_{ij}s_{ijb}s_{ijf} \tag{3.7}$$

where r_{ij} is the observed within-subject correlation in each group

$$r_{ij} = \frac{\sum_{l=1}^{n_{ij}}(X_{lijb} - \bar{X}_{ijb})(X_{lijf} - \bar{X}_{ijf})}{(n_{ij}-1)s_{ijb}s_{ijf}}, \tag{3.8}$$

where X_{lijb} and X_{lijf} are the measure of the outcome in subject l, study i, treatment group j, at baseline and follow-up respectively.

The necessary data to feed into a meta-analysis are D_i and its variance se_{Di}^2. A random effects model accounts for the heterogeneity of the studies in the meta-analysis. The model is

$$D_i|\theta_i \sim \mathrm{N}(\theta_i, se_{Di}^2), \tag{3.9}$$

$$\theta_i|\mu, \tau^2 \sim \mathrm{N}(\mu, \tau^2) \tag{3.10}$$

where θ_i is the mean effect in study i, μ is the true mean effect of the efficacy of the treatment and τ^2 is the between-study variance.

The random effects model estimate of the mean effect size, $\hat{\mu}$, is a weighted average of the individual studies' effect sizes

$$\hat{\mu}(\tau^2) = \frac{\sum_{i=1}^{K} w_i D_i}{\sum_{i=1}^{k} w_i}, \tag{3.11}$$

where the weights w_i are the inverse of the sum of the within-study and between-study variances

$$w_i = \frac{1}{\text{se}_{Di}^2 + \tau^2}. \tag{3.12}$$

The variance of the mean effect size is

$$\text{Var}(\hat{\mu}|\tau^2) = \frac{1}{\sum_{i=1}^{K} w_i}. \tag{3.13}$$

The most widely used estimate of τ^2 is the DerSimonian and Laird (1986) estimator (DSL)

$$\hat{\tau}^2 = \max\left\{0, \frac{Q_a - (k-1)}{\sum_i a_i - \sum_i a_i^2 / \sum_i a_i}\right\}, \tag{3.14}$$

where $a_i = 1/\widehat{se}_{D_i}^2$, and $Q_a = \sum_i a_i (D_i - \hat{\mu})^2$ is the Cochran statistic of heterogeneity. Then the overall estimate of effect size is $\hat{\mu}(\hat{\tau}^2)$ with variance $\text{Var}(\hat{\mu}|\hat{\tau}^2)$.

3.1.1 Methods for missing mean difference

When D_i is missing, but $\bar{X}_{ijb}$, and $\bar{X}_{ijf}$ are available, D_i is computed using equation (3.2). Sometimes the $\bar{X}_{ijf}$ are not available for extraction but the study evaluated the difference of means at follow-up only, $(\bar{X}_{i1f} - \bar{X}_{i0f})$, and this information is available for extraction as well as baseline information, $\bar{X}_{i1b}, \bar{X}_{i0b}$. In our meta-analysis, this was the case in the study by Bosworth (2011). Equation (3.1) can be rewritten as

$$D_i = (\bar{X}_{i1f} - \bar{X}_{i0f}) - (\bar{X}_{i1b} - \bar{X}_{i0b}), \tag{3.15}$$

and again the value of D_i can be computed. Imputation of the outcome D_i when both $\bar{X}_{ijb}$ and $\bar{X}_{ijf}$ are missing is beyond the scope of this chapter. In general, studies with missing and irrecoverable D_i are excluded from the meta-analysis. In our dataset, this situation arises in two studies: Pierce (1984) and Baque (2005).

3.1.2 Methods for missing standard errors

Missing standard deviations or missing standard errors of the mean difference, se_{Di}, are extremely common. The absence of standard errors in the meta-analysis leads to complications

in the estimation of the pooled treatment effect $\hat{\mu}$. However, excluding trials that do not report standard errors or deviations is not advisable because it could lead to biased estimates of the treatment effect (Fu et al., 2013).

As there is no protocol on how to handle the missing variances, a myriad of methods to impute or calculate the missing variances have been used in the literature. Many of these methods have not been empirically tested and it is not known how they affect the overall estimates of the meta-analysis.

The most commonly used methods to handle missing data can be catalogued as

1. Algebraic calculations
2. Single value imputations of standard deviations
3. Imputation of the correlation

None of these methods account for the uncertainty associated with single imputations. Bayesian strategies have the potential to account for uncertainty and have been recommended but are not commonly used in practice. Other less frequently used methods can be found in the exhaustive search performed by Wiebe et al. (2006).

3.1.2.1 Algebraic calculations

A missing variance can sometimes be recalculated from summary statistics presented in the published study. When standard deviations are reported, standard errors are obtained with the simple computation $\text{SE} = \text{SD}/\sqrt{n}$ and viceversa.

Results of a randomized clinical trial are usually presented in the form of a mean difference and 95% confidence interval. This occurred in four of the studies in our dataset: Soghikian (1992), Rudd (2004), Staessen (2004), and Madsen (2008). The standard error can be calculated as

$$\text{SE} = \frac{\text{UCB} - \text{LCB}}{3.92} \tag{3.16}$$

where UCB and LCB are the upper and lower confidence bounds and $3.92 = 1.96 \times 2$

(Fu et al., 2013). Alternatively, if the results are presented as mean difference and a z- or t-statistic, the standard error can be calculated as

$$\text{SE} = \frac{|D|}{z} \quad \text{or} \quad \text{SE} = \frac{|D|}{t}. \tag{3.17}$$

If the results are presented as mean difference and p-value, the p-value can be converted to a z-statistic. If the sample size is small, then a t-statistic should be used and we recover the SE using (3.17).

Sometimes the p-values provided are not exact. For upper boundary p-values, such as $p < 0.05$, using p-value=0.05 and the calculations described above will result in a conservative estimate (a larger SD) that will down weight the trial that does not provide complete information. For lower boundary p-values, such as $p > 0.05$, variance imputation is not recommended since these p-values contain very little information about the variance (Follmann et al., 1992). Some authors suggest performing a sensitivity analysis using various p-values in the range [0.05, 1] (Wiebe et al., 2006).

If the results of a study are presented in the form of median and interquartile range (IQR), or as median and range = max − min, the mean and missing SD can be calculated only if a bell-curved symmetric distribution is assumed. The distribution of mean differences D is reasonably assumed to have a symmetric sampling distribution. Using the interquartile range (IQR), the SD can be calculated as IQR/1.35 or as range/4 for sample sizes between 15 and 70, and range/6 for sample sizes larger than 70 (Hozo et al., 2005). Alternatively, we can use Pearson's table for a more precise suggestion of which denominator to use depending on sample size (Pearson, 1932). Dividing the range by four is also recommended by the Cochrane Handbook (Higgins and Deeks, 2011). If the results include a p-value from a non-parametric test, the p-value has been used as a substitute for the p-value of a t-test (Wiebe et al., 2006).

3.1.2.2 Single value imputations

The simplest single value imputation method is to borrow an SD or SE from somewhere else. It could be from within the study (using the baseline or follow-up SD, the minimum,

maximum or average of the two) or from an external source (a study that is considered to be similar) (Wiebe et al., 2006; Abrams et al., 2005).

One estimate using complete data from the same meta-analysis is the arithmetic mean of the standard deviations from studies that reported one (Robertson et al., 2004; Follmann et al., 1992). Suppose the first m studies in a meta-analysis have complete data for the standard deviation at baseline while the remaining $K - m$ studies are missing the SD at baseline. Then, the imputed s_{gjb} for study g with missing standard deviation at baseline is the SE

$$s_{gjb} = \frac{\sum_{h=1}^{m} s_{hjb}}{m} \qquad g = m+1, \ldots, K. \tag{3.18}$$

This computation also applies for missing standard deviations at follow-up s_{ijf}. The prognostic method (Ma et al., 2008) imputes the missing standard error of the mean difference by computing the average in the standard deviation scale. In this case, assume the first m studies have standard errors of the mean difference available for extraction. Then, the imputed value for those missing is

$$\text{se}_{Dg} = \frac{\sum_{h=1}^{m} \text{se}_{Dh}\sqrt{n_h}}{m\sqrt{n_g}} \qquad g = m+1, \ldots, K. \tag{3.19}$$

where $n_h = n_{h0}n_{h1}/(n_{h0} + n_{h1})$. Unfortunately, this mean imputation strategy is suboptimal and may underestimate the standard error of the mean difference.

3.1.2.3 Imputation of the correlation

If the standard deviations at baseline and follow-up for the treatment and control groups are available, then the missing standard error of the estimate of the treatment effect is computable using equations (3.7), (3.6) and (3.5) if the correlation r_{ij} is also available. Unfortunately, the correlation r_{ij} is also usually missing. In our dataset, only one study reported correlations: Bosworth (2011). A common strategy is to impute the missing correlations.

Various approaches for imputing the missing correlations have been used: using $\rho = 0.5$ (Follmann et al., 1992), $\rho = 0$ (Abrams et al., 2005), $\rho < 0.5$ (Robertson et al., 2004), performing sensitivity analysis with values that range from 0 to 1 (Abrams et al., 2005),

using the arithmetic mean of the correlations that are available for calculation in the meta-analysis, and using the correlation from an external source such as a previous meta-analysis or a different dataset (Abrams et al., 2005).

Assuming a value of $\rho = 0$ is unrealistic because in clinical outcomes at least some degree of correlation is to be expected (Follmann et al., 1992). Similarly, a value of $\rho = 1$ is unrealistic because outcome values vary to some extent within individuals over time (Philbrook et al., 2007). Robertson et al. (2004) suggests to use low values of the correlation to "err on the safe side" and end up with larger estimates of the SD and wider confidence intervals. Wiebe et al. (2006) found that few authors actually report the value of ρ that they used in their imputations.

3.1.2.4 Bayesian models

A random effects Bayesian model assigns priors to μ and τ^2 in equations (3.9) and (3.10). A Bayesian approach accounts for all the uncertainty of the unknown parameters and the Bayesian framework allows for the modeling of complex data structures. In the systematic review of methods for handling missing variance data performed by Wiebe et al. (2006) it was concluded that Bayesian solutions were rarely used in practice and when used there were not enough details to allow replication of the analysis.

Stevens (2011) presented a Bayesian meta-analysis of mean differences model implemented in WinBUGS that can take missing SEs and deal with them automatically. However, the model does not make use of other existing data in the published papers to aid the imputation of the missing SEs.

Abrams et al. (2005) models the missing SEs based on other SDs available in the published papers and sets a prior on the unknown correlation and handles one pattern of missing data: assuming that s_{ijb} and s_{ijf} are available and only the s_{ij} are missing. They assume that the correlation is the same in the treatment and control group and obtain a distribution for the correlation from an alternative meta-analysis on a similar outcome.

3.1.2.5 Algorithms for imputation

In most meta-analyses each study will contribute different amounts of data presented in different formats such that a single strategy of SD imputation will not suffice. Indeed, a combination of strategies is typically needed and it will not always be clear which method to use (Follmann et al., 1992). Various authors have suggested rules of thumb or paths of imputation in which some methods take precedence over others.

For example, if presented with the choice of imputing SE using a p-value or an imputed correlation, Follmann et al. (1992) suggests to use the minimum of the two resulting SEs. Furukawa et al. (2006) recommends to use the pooled SD from the studies that reported them if the number of missing SDs is small but to borrow an SD from a previous systematic review if the number of missing SDs is large. Robertson et al. (2004) emphasizes the importance of performing a full sensitivity analysis after imputation. Cappuccio et al. (2004), Bray et al. (2010) and Agarwal et al. (2011) followed different algorithms to analyze blood pressure data in their meta-analyses, using combinations of approaches to impute missing SDs:

Cappuccio et al. (2004) algorithm

Step 1 Recover missing SDs using algebraic calculations.

Step 2 Impute missing SD of the mean change using the average of the baseline and follow-up blood pressures:

$$s_{ij} = \frac{s_{ijb} + s_{ijf}}{2} \tag{3.20}$$

for study i with extracted s_{ijb} and s_{ijf}.

Step 3 Impute remaining missing SDs of mean difference, s_{ij}, using the average of all the other studies with SD of mean difference.

Step 4 Compute the missing SEs of the difference of mean differences using equations (3.7) and (3.6).

Bray et al. (2010) algorithm

Step 1 Recover missing SDs using algebraic calculations.

Step 2 Impute correlation using the average of the correlations obtained from studies that provided complete data, where the observed correlations r_{ij} are recovered using

$$r_{ij} = \frac{s_{i1b}^2 + s_{i0f}^2 - s_{ij}^2}{2s_{i1b}s_{i0f}}, \tag{3.21}$$

based on n_{ij} observations, for studies with extracted s_{ijb} and s_{ijf}.

Step 3 If only s_{ijb} or s_{ijf} is missing, but not both, impute it using the average of the corresponding SDs from other studies.

Step 4 Compute missing SDs of mean difference using equation (3.7) with the imputed correlations and imputed SDs from steps 2 and 3.

Step 5 Impute remaining missing SDs of mean difference, s_{ij}, using the average of all other studies.

Step 6 Compute the missing SEs of the difference of mean differences using equation (3.6).

Agarwal et al. (2011) algorithm

Step 1 Recover missing SDs using algebraic calculations.

Step 2 Impute remaining missing SDs of mean difference, s_{ij}, using the average SD of all other studies.

Step 3 Compute the missing SEs of the difference of mean differences using equation (3.6).

3.2 Dataset

To illustrate the proposed methodology for meta-analysis of mean differences with missing SEs we re-evaluate a published meta-analysis that compares the effect of home blood pressure monitoring on systolic and diastolic blood pressure levels when compared to usual care (Cappuccio et al., 2004; Bray et al., 2010; Agarwal et al., 2011). In our analysis, home monitoring of blood pressure is the treatment and usual care is considered the control group.

Hypertension is a preventable cause of stroke and cardiovascular disease. The latter is the leading cause of death worldwide (Bray et al., 2010). However, hypertension is often underdiagnosed and undertreated (Cappuccio et al., 2004). Blood pressure is usually measured by doctors and nurses in the healthcare system, but measuring blood pressure at home has been shown in systematic reviews to be more effective in controlling blood pressure and it is increasingly becoming a part of hypertension management (Agarwal et al., 2011).

We identified a total of 36 randomized controlled trials. Systolic blood pressure (SBP) was evaluated in 31 trials with a total of 8,560 patients. Diastolic blood pressure (DBP) was evaluated in 36 trials with a total of 9,311 patients. The meta-analyses present multiple challenges at the data extraction step. Tables 3.1 and 3.2 display the data available for extraction from each study: The mean and standard deviations at baseline for both treatment groups, followed by the mean and standard deviations at follow-up for both groups, mean difference in each group and its standard deviation and finally, the difference of mean differences and its standard error. A blank spot in the table means that the value was not available for extraction in the publication.

Table 3.1: Extracted data from each study in the SBP meta-analysis includes: sample sizes n_{ij}, mean and SDs at baseline, $(\bar{X}_{i1b}, s_{i1b})$, and at follow-up, $(\bar{X}_{i1f}, s_{i1f})$, mean and SD of the change, $(\bar{X}_{ijf}$ -$\bar{X}_{ijb}), s_{ij}$, and the difference of mean differences and its standard error, (D_i, se_{Di}), for treatment group j and study $i = 1, \ldots, K$.

			Baseline				Follow-up				Treatment		Control		Difference of	
			Treatment		Control		Treatment		Control		difference		difference		differences	
Study			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SE
	n_{i1}	n_{i0}	$\bar{X}_{i1b}$	s_{i1b}	$\bar{X}_{i0b}$	s_{i0b}	$\bar{X}_{i1f}$	s_{i1f}	$\bar{X}_{i0f}$	s_{i0f}	$(\bar{X}_{i1f}$ -$\bar{X}_{i1b})$	s_{i1}	$(\bar{X}_{i0f}$ -$\bar{X}_{i0b})$	s_{i0}	D_i	se_{Di}
Carnahan (1975)	49	48	152.7		156.6		134.7		146.1		-18		-10.5		-7.5	
Pierce (1984)	27	29	184	22	179	26										
Midanik (1991)	74	72	144.4	15.7	144	16.8	142.6	15.6	144.8	18	-1.8	14.5	0.8	14.2	-2.6	
Soghikian (1992)	200	190	137.4	16.9[5]	140.2	17.9[5]	135.9	19.8[5]	142	16.5[5]	-1.5	18.4[5]	1.8	16.5[5]	-3.3	1.8[3]
Friedman (1996)	133	134	169.5		167		158.5		156.4		-11		-10.6		-0.4	
Bailey (1999)	31	29	156	22.3[5]	155	21.5[5]	148	16.7[5]	142	16.2[5]	-8		-13		5	
Vetter (2000)	296	326	166.1	14.5[5]	168.1	14.4[5]	145.1	14.9[5]	147.6	14.6[5]	-21		-20.5		-0.5	
Mehos (2000)	18	18	157.9	16.4	153.9	14.6	140.8		136.8		-17.1	12.9[2]	-7	18.4[2]	-10.1	
Artinian (2001)	6	9	148.8	13.8	142.4	16.5	124.1	13.8	143.3	10.7	-24.7		.95		-25.6	
Broege (2001)	20	18	150	22	144	20	146	11	144	19	-4		0		-4	
Rogers (2001)	99	63									-4.9	24.7[3]	-.13	13.4[3]	-4.7	1.7[3]
Rudd (2004)	69	68	155.9[2]	19.8[2]	154.8[2]	17.1[2]	141.6[2]	19.9[2]	149.4[2]	16.9[2]	-14.2	18.1	-5.7	18.7	-8.5	
Staessen (2004)	203	197	160.8	18.6	159.1	19.3	144.8[2]	36.5[2]	136.8[2]	31.7[2]	-15.9		-22.3		6.3	1.6[3]
Baque (2005)	622	703	161	25.5[3]	162	27.1[3]										
Halme (2005)	113	119	159.5	17.5	159.1	18.9	146.8	17.8	149.5	20.3	-12.7	19.6	-9.6	19.5	-3.1	
McManus (2005)	189	211	157.9	15.7	155	13.6	149		148.4		-8.9		-6.6		-2.3	
Zillich (2005)	64	61	151.5	15.6	151.6	12.9	138.1	15.7[2]	142.6	15.9[2]	-13.4		-9		-4.4	
Marquez (2006)	100	100	159.1	16.6	155.6	14.6	135.6	13.8	136.7	11.2	-23.5	15.9	-18.9	15.9	-4.6	
Kauric (2007)	17	17	161	14	162	12	153	16	161	14	-8.12	9.2[4]	-1	9.8[4]	-7.1	
Verber (2007)	216	214	166.2	19.3	165.1	20.8	143.8	18.4	142.2	20	-22.4		-22.9		0.5	
Green i (2008)	246	247	152.2	10	151.3	10.6	143.8	14.8[3]	146.3	14.8[3]	-8.4		-5		-3.4	

Continued on next page

Table 3.1 – *Continued from previous page*

			Baseline				Follow-up				Treatment		Control		Difference of	
			Treatment		Control		Treatment		Control		difference		difference		differences	
Study			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SE
	n_{i1}	n_{i0}	$\bar{X}_{i1b}$	s_{i1b}	$\bar{X}_{i0b}$	s_{i0b}	$\bar{X}_{i1f}$	s_{i1f}	$\bar{X}_{i0f}$	s_{i0f}	$(\bar{X}_{i1f}$ -$\bar{X}_{i1b})$	s_{i1}	$(\bar{X}_{i0f}$ -$\bar{X}_{i0b})$	s_{i0}	D_i	se_{Di}
Green ii (2008)	237	247	152.2	10.4	151.3	10.6	137.9	15.2[3]	146.3	14.8[3]	-14.3		-5		-9.3	
Madsen (2008)	105	118	153.1	13.2	152.2	13.7	141.1	11.5	142.7	13.3	-12		-9.5		-2.5	1.9[3]
Tobe (2008)	173	97	159	11	160	14	138	13	141	10	-21		-19		-2	
Da Silva (2009)	34	31	157	25	159	21	147	18	154	22	-10		-5		-5	
De Jesus (2009)	7	12	145.4	5.26	149.2	6.98	142.1	20.9	141.4	15.31	-3.29	18.9	-7.8	13.15	4.5	
Marquez (2009)	230	255	152.9	13.8	153.17	12	136.5	9.8	136.7	9.4	-16.4	14.2	-16.5	14.1	0.1	
Parati (2009)	187	111	148.4	12.6	148.7	11.7	137.5	17.8[2]	138	16.3[2]	-10.9		-10.7		-0.2	
Rinfret (2009)	111	112	162.1	16	162	17	143.5	18.5[2]	148.5	18.6[2]	-18.6		-13.5		-5.1	
Godwin (2010)	285	267	149.6	10.5	147.3	9	141.1	18.57[5]	142.8	19.6[5]	-8.5		-4.5		-4	
Bosworth (2011)[1]	131	124	129	19	128	17	129.4[2]	35.4[2]	127.3[2]	34.9[2]	0.36		-0.73		2.09	

[1] Provided value for the correlations r_{ij}.

[2] Extracted from a plot.

[3] Computable from a confidence interval.

[4] Computable from a p-value.

[5] Extracted as a standard error.

Table 3.2: Extracted data from each study in the DBP meta-analysis includes: sample sizes n_{ij}, mean and SDs at baseline, $(\bar{X}_{i1b}, s_{i1b})$, and at follow-up, $(\bar{X}_{i1f}, s_{i1f})$, mean and SD of the change, $(\bar{X}_{ijf}$ -$\bar{X}_{ijb}), s_{ij}$, and the difference of mean differences and its standard error, (D_i, se_{Di}), for treatment group j and study $i = 1, \ldots, K$.

			Baseline				Follow-up				Treatment		Control		Difference of	
			Treatment		Control		Treatment		Control		difference		difference		differences	
Study			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SE
	n_{i1}	n_{i0}	$\bar{X}_{i1b}$	s_{i1b}	$\bar{X}_{i0b}$	s_{i0b}	$\bar{X}_{i1f}$	s_{i1f}	$\bar{X}_{i0f}$	s_{i0f}	$(\bar{X}_{i1f}$ -$\bar{X}_{i1b})$	s_{i1}	$(\bar{X}_{i0f}$ -$\bar{X}_{i0b})$	s_{i0}	D_i	se_{Di}
Carnahan (1975)	50	50	101.7		103.6		91.3		93.2		-10.4		-10.4		0	
Haynes (1976)	20	18	98.5	5.8[5]	98.3	6.4[5]	93.1	5.8[5]	96.4	5.5[5]	-5.4	7.6[5]	-1.9	8.5[5]	-3.5	
Johnson (1978)	35	35	102.6	6.5[5]	103.2	10.1[5]	94.1	9.3[5]	95.7	12.8[5]	-8.5	10.5[5]	-7.5	11.1[5]	-1	
Johnson ii (1978)	35	35	104.2	6.5[5]	103.2	10.1[5]	95.9	9.5[5]	95.7	12.8[5]	-8.3	8.9[5]	-7.5	11.1[5]	-0.8	
Earp (1982)	99	63														
Pierce (1984)	27	29	106	8	103	11										
Stahl (1984)	144	173	109.7		108.6		89.9		88.5		-19.8		-20.1		0.3	
Midanik (1991)	74	72	91.3	9.1	92.7	7.7	92.3	9.5	93.6	9.3	1	10	0.9	8	0.1	
Soghikian (1992)	200	190	86.1	8.5[5]	86.3	11[5]	86.2	9.9[5]	88	9.6[5]	0.1	9.9[5]	1.7	9.6[5]	-1.6	1[3]
Friedman (1996)	133	134	86.1		84						-5.4		-3.3		-2.1	
Bailey (1999)	31	29	93	11.1[5]	95	10.8[5]	89	11.1[5]	89	10.8[5]	-4		-6		2	
Vetter (2000)	296	326	101.9	6.2[5]	102	6.5[5]	88.7	8.3[5]	90.1	7.8[5]	-13.2		-11.9		-1.3	
Mehos (2000)	18	18	91.1	10.8	89.6	9.8	80.6		85.8		-10.5	7.1[2]	-3.8	8.9[2]	-6.7	
Artinian (2001)	6	9	90.2	5.79	91.22	8.66	75.58	11.4	89.05	10.63	-14.62		-2.17		-12.45	
Broege (2001)	20	18	81	12	82	13	80	8	83	12	-1		1		-2	
Rogers (2001)	99	63									-1.95	10.6[3]	2.08	9.3[3]	-4.03	1.6[3]
Rudd (2004)	69	68	86.3[2]	10.3[2]	87.5[2]	10.8[2]	79.7[2]	10.5[2]	83.9[2]	9.3[2]	-6.5	10	-3.4	7.9	-3.1	
Staessen (2004)	203	197	101.8	7.4	101.5	6.5	89.9[2]	14.9[2]	86.7[2]	12.1[2]	-11.81		-14.83		3.02	0.8[3]
Baque (2005)	622	703	94	12.7[3]	94	13.5[3]										
Halme (2005)	113	119	94.1	6.8	94.6	7.5	87	9.3	89.1	8.6	-7.1	10.1	-5.5	8.9	-1.6	
McManus (2005)	189	211	88.7	7.3	88	7.9	83		83.4		-5.7		-4.6		-1.1	

Continued on next page

Table 3.2 – *Continued from previous page*

			Baseline				Follow-up				Treatment		Control		Difference of	
			Treatment		Control		Treatment		Control		difference		difference		differences	
Study			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SE
	n_{i1}	n_{i0}	$\bar{X}_{i1b}$	s_{i1b}	$\bar{X}_{i0b}$	s_{i0b}	$\bar{X}_{i1f}$	s_{i1f}	$\bar{X}_{i0f}$	s_{i0f}	$(\bar{X}_{i1f}$ -$\bar{X}_{i1b})$	s_{i1}	$(\bar{X}_{i0f}$ -$\bar{X}_{i0b})$	s_{i0}	D_i	se_{Di}
Zillich (2005)	64	61	85.3	11.6	85.3	10.7	76.5	11.5[2]	79.7	10.8[2]	-8.8		-5.6		-3.2	
Marquez (2006)	100	100	92.4	10.8	91	9.7	79.5	8.4	81.3	7.6	-12.9	9.9	-9.7	9.8	-3.2	
Kauric-Klein (2007)	17	17	94	7	100	10	90	5	97	10	-4	7.9[4]	-3	12.8[4]	-1	
Verberk (2007)	216	214	97.1	9.9	97.8	10.8	85.4	10.4	84.3	9.6	-11.7		-13.5		1.8	
Green i (2008)	259	258	89	7.9	89.4	8	84.5	9.6[3]	85.7	9.6[3]	-4.5		-3.7		-0.8	
Green ii (2008)	261	258	88.9	8.1	89.4	8	81.6	9.8[3]	85.7	9.6[3]	-7.3		-3.7		-3.6	
Madsen (2008)	113	123	91.2	8.1	90.5	8.9	85	7.1	85.1	8.2	-6.2		-5.4		-0.8	1.1[3]
Tobe (2008)	173	97	91	10	88	10	80	8	78	9	-11		-10		-1	
da Silva (2009)	34	31	89	18	87	16	86	11	89	14	-3		2		-5	
De Jesus (2009)	7	12	68.4	11.6	73.9	13.83	74.1	11.1	71.7	13.9	5.7	10.9	-2.3	12.7	7.9	
Marquez (2009)	230	255	89.7	9.8	91.01	7.9	80.7	7.9	81.49	7.4	-9	9.4	-9.52	8	0.52	
Parati (2009)	187	111	88.7	7.4	88.8	8.6	83.6	7[2]	83.3	10.7[2]	-5.1		-5.5		0.4	
Rinfret (2009)	111	112	91.5	12	90.3	12	82.6	9.1[2]	84.7	11.9[2]	-8.9		-5.6		-3.3	
Godwin (2010)	285	267	81.5	8.4	81.2	8.2	78.7	10.1[5]	79.4	13.1[5]	-2.8		-1.8		-1	
Bosworth (2011)[1]	148	147	77	12	78	14									1.6	

[1] Provided value for the correlations r_{ij}.

[2] Extracted from a plot.

[3] Computable from a confidence interval.

[4] Computable from a p-value.

[5] Extracted as a standard error.

The necessary data from each study to feed into the usual meta-analysis model is the overall difference of mean differences and its standard error. Presence or absence of these data is in the last two columns of tables 3.1 and 3.2. Note that only four studies provided a confidence interval from which to compute the standard error for the meta-analysis. The data for SBP has 88% missing SEs and the data for DBP has 89% missing SEs. Additional data available for extraction to be used to estimate the missing SEs and compute the mean differences varies across studies and the patterns are similar in both datasets. For the SBP dataset:

1. Thirty studies provide a mean blood pressure at baseline for both groups. Of these, twenty-eight studies provide a SD or SE for the mean at baseline.

2. Twenty-eight studies provide mean blood pressure at follow up for both groups. Of these, twenty-four studies provide a SD or SE for the mean at follow-up.

3. Twenty-nine studies provide a value of the mean change in both groups, but only ten studies provide a SE of the mean change in the groups.

4. Twenty-nine studies provide a value for the difference of mean differences, but only four studies provide a SE for the difference of mean differences.

5. Only one paper provides the correlations from each treatment group.

The study by Pierce et al. (1984) did not provide values of blood pressure change, but a table with frequencies and ranges of change. The study by Baqué et al. (2005) only looked at the proportion of people that achieved controlled blood pressure and did not provide the details of the blood pressure changes. These two studies are not included in this meta-analysis.

We identified nine different scenarios of SDs and SEs availability, which we list in table 3.3. Only the first two scenarios are considered complete data while the remaining seven scenarios have missing data. The meta-analyses for SBP and DBP motivate the development of

appropriate methodology to incorporate the uncertainty arising from estimating the missing SEs under all scenarios of missing data.

Table 3.3: Patterns of availability of SDs and SE for extraction and corresponding percent of SBP and DBP studies.

	s_{i1b}	s_{i0b}	s_{i1f}	s_{i0f}	s_{i1}	s_{i0}	se_{Di}	SBP studies	DBP studies
Scenario 1	✓	✓	✓	✓	✓	✓	✓	3%	3%
Scenario 2							✓	0%	0%
Scenario 3	✓	✓	✓	✓			✓	6%	6%
Scenario 4					✓	✓	✓	3%	3%
Scenario 5					✓	✓		0%	0 %
Scenario 6	✓	✓	✓	✓	✓	✓		23%	28%
Scenario 7	✓	✓			✓	✓		3%	3%
Scenario 8	✓	✓	✓	✓				45%	36%
Scenario 9	✓	✓						10%	11%
Scenario 10								6%	11%

3.3 Uncertain Standard Error model

We propose the Uncertain Standard Error (USE) Bayesian model, which incorporates the uncertainty of the variance estimation in the meta-analysis. In the USE Bayesian model, the parameters and unknown data are treated as random variables. The Bayesian framework allows us to borrow strength from other studies in estimating the true effect. More importantly, the Bayesian approach takes account of all parameter uncertainty and allows for a more flexible and robust modeling strategy than the traditional approaches described in sections 3.1.2.1, 3.1.2.2, and 3.1.2.3 (Abrams and Sanso, 1998).

Let Υ be the set of all K studies, Ω be the set of studies that did not report a SD at baseline for the treatment and control groups, Ψ be the set of studies that did not report a

SD at follow-up in the treatment and control groups and Φ be the set of studies with missing SE for the difference of mean differences.

Let SE_i^* be the estimated standard errors using traditional methods and SE_i be actual standard error needed for the meta-analysis for study i with missing SE. The traditional approaches ignore that SE_i^* is only an approximation of SE_i. The USE model does not make that assumption. Instead it incorporates the uncertainty in the estimator. First, it assigns a distribution of plausible values for the within-subject correlations based on data extracted from studies with complete data.

To model the true correlations, ρ_{ij}, we use a truncated normal distribution, $\text{N}(p_j, q_j)$, centered around the mean of the sample correlations, $p_j = \sum_i r_{ij}/L$, from studies with complete data, $i \in \Upsilon \setminus \{\Omega \cup \Psi \cup \Phi\}$, truncated at 0 and 1, where L is the number of studies with complete data. In this dataset the sample correlations have very little variation, $q_j = \sum_i (r_{ij} - p_j)^2/(L-1)$, so the normal distribution is adequate and convenient. Second, the USE model also models the variances at baseline and foliow-up for the treatment and control groups, σ_{ijb}^2 and σ_{ijf}^2, using an Inverse Gamma distribution with known shape and rate parameters a_{jb}, b_{jb} or a_{jf}, b_{jf} that are calculated as functions of the mean and variance of the sample variances, s_{ijb}^2 and s_{ijf}^2, of studies with complete data. Let L_b and L_f be the number of studies with extracted s_{ijb}^2 an s_{ijf}^2 respectively. Then,

$$m_b = \frac{\sum_i (s_{ijb}^2)}{L_b} \tag{3.22}$$

$$v_b = \frac{\sum_i (s_{ijb}^2 - m_b)^2}{L_b - 1} \tag{3.23}$$

$$a_{jb} = \frac{m_b^2}{v_b} + 2 \tag{3.24}$$

$$b_{jb} = m_b(a_{jb} - 1), \tag{3.25}$$

$$\sigma_{ijb}^2 \sim \text{IG}(a_{jb}, b_{jb}) \tag{3.26}$$

and similarly for a_{jf} and b_{jf}.

The USE model computes the variances of mean change by group using $\sigma_{ij}^2 = \sigma_{ijb}^2 + \sigma_{ijf}^2 - 2\rho_{ij}\sigma_{ijb}\sigma_{ijf}$ and estimates the variance of the difference of mean differences with equations (3.5). Typically, se_{Di}^2 are assumed to be known and replaced by the observed within-study

variances, using equation (3.5) as $\widehat{se}^2_{Di}$. A Bayesian model can allow for uncertainty in the within-study variances and model them using the fact that $(n-1)s^2/\sigma^2$ follows a chi-square distribution with $n-1$ degrees of freedom. However, it has been reported that doing so has little impact in the overall uncertainty (Abrams et al., 2005; Hardy and Thompson, 1996). Therefore, we followed the convention to treat $\widehat{se}^2_{Di}$ as equal to se^2_{Di} to facilitate implementation of the model. The full Bayesian model is

$$\rho_{ij} \sim \mathrm{N}(p_j, q_j)\mathrm{I}(0,1), \qquad j \in \{1,0\}, \forall i \in \Phi \cap (\Omega \cup \Psi) \quad (3.27)$$

$$\sigma^2_{ijb} \sim \mathrm{IG}(a_{jb}, b_{jb}), \qquad \forall i \in \Omega, \quad (3.28)$$

$$\sigma^2_{ijf} \sim \mathrm{IG}(a_{jf}, b_{jf}), \qquad \forall i \in \Psi, \quad (3.29)$$

$$\sigma^2_{ij} = \sigma^2_{ijb} + \sigma^2_{ijf} - 2\rho_{ij}\sigma_{ijb}\sigma_{ijf} \qquad \forall i \in \Phi \cap (\Omega \cup \Psi), \quad (3.30)$$

$$se^2_{Di} = \left[\frac{1}{n_{i1}} + \frac{1}{n_{i0}}\right] \frac{(n_{i1}-1)\sigma^2_{i1} + (n_{i0}-1)\sigma^2_{i0}}{n_{i1}+n_{i0}-2} \qquad \forall i \in \Phi \cap (\Omega \cup \Psi), \quad (3.31)$$

$$se^2_{Di} = \left[\frac{1}{n_{i1}} + \frac{1}{n_{i0}}\right] \frac{(n_{i1}-1)s^2_{i1} + (n_{i0}-1)s^2_{i0}}{n_{i1}+n_{i0}-2} \qquad \forall i \in \Phi \cap (\Omega^c \cup \Psi^c), \quad (3.32)$$

$$D_i|\theta_i, se^2_{Di} \sim \mathrm{N}(\theta_i, se^2_{Di}), \qquad i = 1, \ldots, K, \quad (3.33)$$

$$\theta_i|\mu, \tau^2 \sim \mathrm{N}(\mu, \tau^2), \quad (3.34)$$

with priors

$$\mu \sim \mathrm{N}(u, v), \quad (3.35)$$

$$\tau^2 \sim \mathrm{IG}(c, d). \quad (3.36)$$

As a prior for μ we use a non-informative distribution, $\mathrm{N}(u, v)$, with $u = 0$ and $v = 100^2$, to allow Bayesian inference for a parameter about which not much is known beyond the data in the meta-analysis at hand. For τ^2 we use an informative prior, $\mathrm{IG}(c, d)$. We expect considerable between-study variance due to large differences in the study populations, and designs. Thus, we set $c = 7$ and $d = 60$ so that the mean of τ^2 in the prior distribution is 10 and the variance is 20.

3.4 Results

We compare the following models applied to the blood pressure monitoring datasets introduced in section 3.2:

1. The USE Bayesian model with the extracted dataset with missing standard deviations and standard errors.

2. Naive Bayesian models (equations 3.33 to 3.36) with datasets completed using the prognostic algorithm, and the algorithms used in Cappuccio et al. (2004), Bray et al. (2010), and Agarwal et al. (2011) as described in section 3.1.2.5

3. Complete-cases naive Bayesian model, using only on 12 studies in the SBP and 14 in the DBP dataset.

Using complete data from 12 studies in the SBP we obtain the following values to feed into the USE model: $a_{1b} = 5.6$, $b_{1b} = 1212$, $a_{0b} = 6.4$, $b_{0b} = 1322$, $a_{1f} = 3.3$, $b_{1f} = 813$, $a_{0f} = 3.6$, $b_{0f} = 859$, $p_1 = 0.47$, $q_1 = 0.04$, $p_0 = 0.51$, $q_0 = 0.02$.

3.4.1 Posterior computation

The models were implemented using JAGS (Plummer et al., 2003). The USE Bayesian model can be found in the Appendix C. We used a burn-in of 5,000 iterations. Posterior results were generated from 50,000 samples retaining every 10th one. We used three chains with different starting points and assessed convergence by inspecting the posterior densities and the convergence diagnostics of Gelman and Rubin (1992). Autocorrelation plots and time series plots verified that the chains mixed satisfactorily.

3.4.2 Sensitivity to prior specification

We perform sensitivity analysis on the choice of prior for the between-study variance. We compare our chosen prior, $\tau^2 \sim \text{IG}(7, 60)$, to three others: prior 2 is $\tau^2 \sim \text{IG}(3, 20)$ with mean 10 and variance 100, prior 3 is $\tau^2 \sim \text{IG}(2.4, 2.8)$ with mean 2 and variance 10 and

finally prior 4 is the uniform U$(0, A)$ on τ, not τ^2, as recommended by Gelman (2006) with A=10. In the τ^2 scale, prior 4 has mean 33 and variance 871. Table 3.4 displays the results for parameters μ and τ by prior and Figures 3.1 and 3.2 display the posteriors of the parameters μ and τ by prior using the SBP and DBP datasets. All 95% credible intervals are $(2.5\%, 97.5\%)$. For the meta-analysis of SBP, the results for μ are not sensitive to the choice of prior. In contrast, the results for random study effects' standard deviation τ are affected by the choice of prior: we obtain the smallest posterior mean ($\bar{\tau} = 2.6$) using prior 3 and the largest ($\bar{\tau} = 3.2$) using prior 4. Priors 1 and 2 result in very similar posterior mean ($\bar{\tau} = 3$). Priors 2 and 3 result in similar standard deviations for τ. The largest posterior SD occurs when using prior 4 and the smallest when using prior 1. These conclusions apply also to the meta-analysis of DBP.

Table 3.4: Posterior mean, standard deviation and 95% credible interval for parameters μ and τ, for the SBP and DBP dataset, listed by prior in the USE model.

	SBP					
	μ			τ		
Prior	mean	sd	95% CI	mean	sd	95% CI
$\tau^2 \sim$ IG$(7, 60)$	-2.93	0.74	(-4.43, -1.47)	3.03	0.45	(2.30, 4.06)
$\tau^2 \sim$ IG$(3, 20)$	-2.91	0.74	(-4.36, -1.43)	2.94	0.54	(2.05, 4.14)
$\tau^2 \sim$ IG$(2.4, 2.8)$	-2.86	0.68	(-4.23, -1.55)	2.55	0.57	(1.59, 3.79)
$\tau \sim$ U$(0, 10)$	-2.93	0.78	(-4.47, -1.39)	3.24	0.73	(2.06, 4.85)
	DBP					
	μ			τ		
Prior	mean	sd	95% CI	mean	sd	95% CI
$\tau^2 \sim$ IG$(7, 60)$	-1.19	0.52	(-2.22, -0.19)	2.33	0.29	(1.84, 2.99)
$\tau^2 \sim$ IG$(3, 20)$	-1.16	0.46	(-2.06, -0.23)	1.95	0.30	(1.44, 2.63)
$\tau^2 \sim$ IG$(2.4, 2.8)$	-1.06	0.37	(-1.80, -0.35)	1.35	0.29	(0.87, 1.99)
$\tau \sim$ U$(0, 10)$	-1.10	0.40	(-1.93, -0.33)	1.56	0.37	(0.90, 2.36)

Figure 3.1: Posterior of μ in the USE model and SBP dataset by prior. Prior 1 is $\tau^2 \sim \text{IG}(7, 60)$, prior 2 is $\tau^2 \sim \text{IG}(3, 20)$, prior 3 is $\tau^2 \sim \text{IG}(2.4, 2.8)$ and prior 4 is $\tau \sim U(0, 10)$.

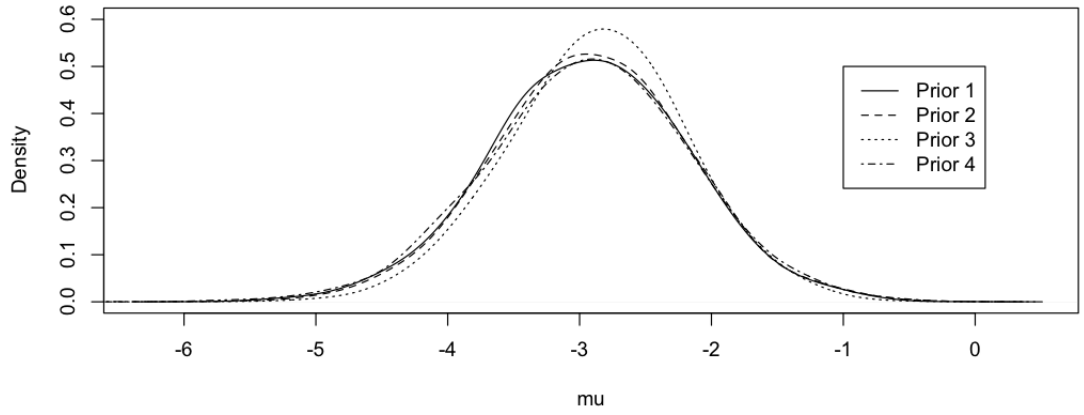
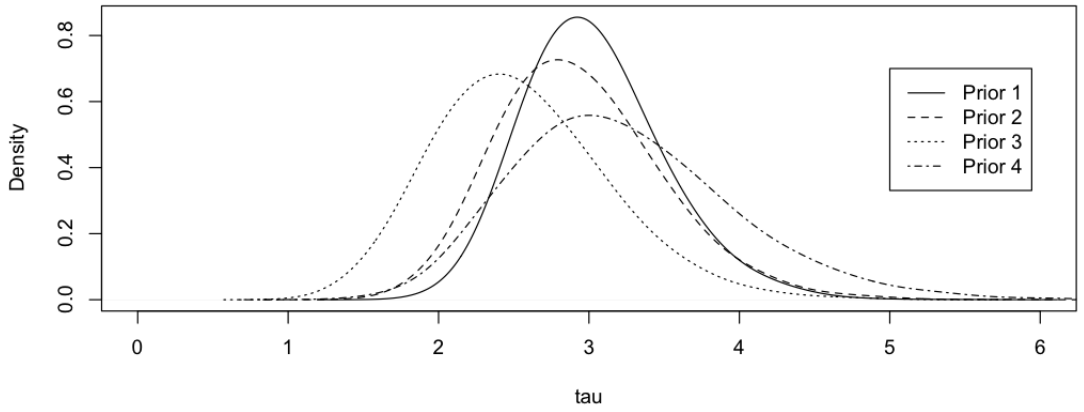


Figure 3.2: Posterior of τ in the USE model and SBP dataset by prior. Prior 1 is $\tau^2 \sim \text{IG}(7, 60)$, prior 2 is $\tau^2 \sim \text{IG}(3, 20)$, prior 3 is $\tau^2 \sim \text{IG}(2.4, 2.8)$ and prior 4 is $\tau \sim U(0, 10)$.



3.4.3 Results for he SBP and DBP datasets

Posterior means, standard deviations and credible intervals for parameters μ and τ under the USE model and naive Bayesian with various algorithms applied to the SBP and DBP datasets can are given in Table 3.5. Figures 3.3 to 3.6 display the corresponding posterior distributions. All models use prior $\tau^2 \sim \text{IG}(7, 60)$.

Table 3.5: Posterior mean, standard deviations and 95% credible intervals for parameters μ and τ for the SBP and DBP meta-analyses under various models. Let Y be all the data input to the model.

SBP

	μ				τ		
Model	mean	sd	95% CI	$P(\mu < 0\|Y)$	mean	sd	95% CI
USE model	-2.92	0.75	(-4.44, -1.46)	1	3.02	0.44	(2.72, 4.00)
Complete cases	-2.71	1.21	(-5.12, -0.33)	0.99	3.20	0.55	(2.30, 4.52)
Algorithm in Cappuccio et al. (2004)	-2.86	0.74	(-4.31, -1.40)	1	3.02	0.44	(2.29, 4.00)
Algorithm in Bray et al. (2010)	-2.84	0.73	(-4.29, -1.43)	1	3.01	0.44	(2.29, 3.99)
Algorithm in Agarwal et al. (2011)	-2.77	0.73	(-4.17, -1.32)	1	3.03	0.44	(2.29, 4.01)
Prognostic algorithm in Ma et al. (2008)	-2.80	0.71	(-4.21, -1.39)	1	3.07	0.45	(2.31, 4.02)

DBP

	μ				τ		
Model	mean	sd	95% CI	$P(\mu < 0\|Y)$	mean	sd	95% CI
USE model	-1.19	0.51	(-2.20, -0.22)	1	2.33	0.30	(1.84, 3.00)
Complete cases	-1.32	0.87	(-3.02, 0.40)	0.94	2.68	0.41	(2.01, 3.65)
Algorithm in Cappuccio et al. (2004)	-1.18	0.51	(-2.19, -0.20)	0.99	2.34	0.29	(1.83, 2.98)
Algorithm in Bray et al. (2010)	-1.14	0.51	(-2.13, -0.14)	0.99	2.32	0.29	(1.82, 2.96)
Algorithm in Agarwal et al. (2011)	-1.21	0.51	(-2.22, -0.24)	0.99	2.34	0.30	(1.84, 3.02)
Prognostic algorithm in Ma et al. (2008)	-1.16	0.50	(-2.16, -0.17)	0.99	2.35	0.30	(1.85, 3.03)

The interval for the complete-cases approach is the largest for all parameters in both meta-analyses. Parameter of interest μ, in the SBP meta-analysis, has a 61% wider credible interval using complete-cases compared to the USE model, and 71% wider in the DBP meta-analysis. This is not surprising due to the substantial loss of information: only 12 out of 29 SBP studies have complete data and only 15 out of 33 SBP studies have complete data. Also, the complete-case analysis operates on a 67% reduced sample size in the SBP meta-analysis and 66% reduced sample size in the DBP meta-analysis. Furthermore, the loss of all the studies without SE leads to non-significants result in the DBP meta-analysis. This strongly supports the existing recommendation to not omit studies in the meta-analysis.

In both meta-analyses, modest differences are observed for the posterior estimates of μ across models, with a slightly larger posterior interval from the USE model. There are no differences in the estimates of τ. All approaches do similarly well for μ possibly because these meta-analyses contain over thirty studies and there is enough information available for extraction to estimate the treatment effect. The conclusions of the meta-analyses about the treatment effect do not change from previously published results and across models: there is a significant reduction of 2.9 (1.46, 4.44) mmHg in systolic blood pressure and 1.2 (0.22, 2.20) mmHg in diastolic blood pressure in the home monitoring groups compared to usual care.

Figure 3.3: Meta-analysis for SBP monitoring vs usual care: Posterior distribution of μ under various models.

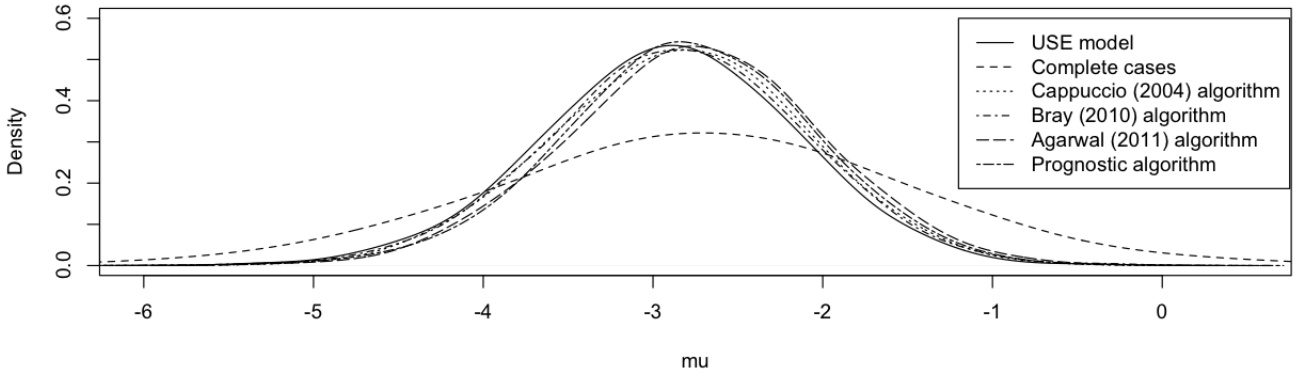


Figure 3.4: Meta-analysis for SBP monitoring vs usual care: Posterior distribution of τ under various models.

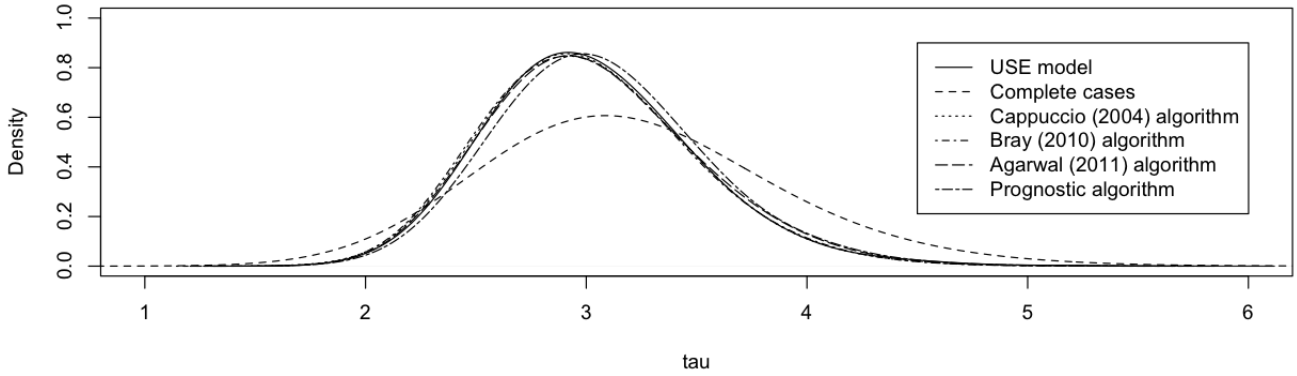


Figure 3.5: Meta-analysis for DBP monitoring vs usual care: Posterior distribution of μ under various models.

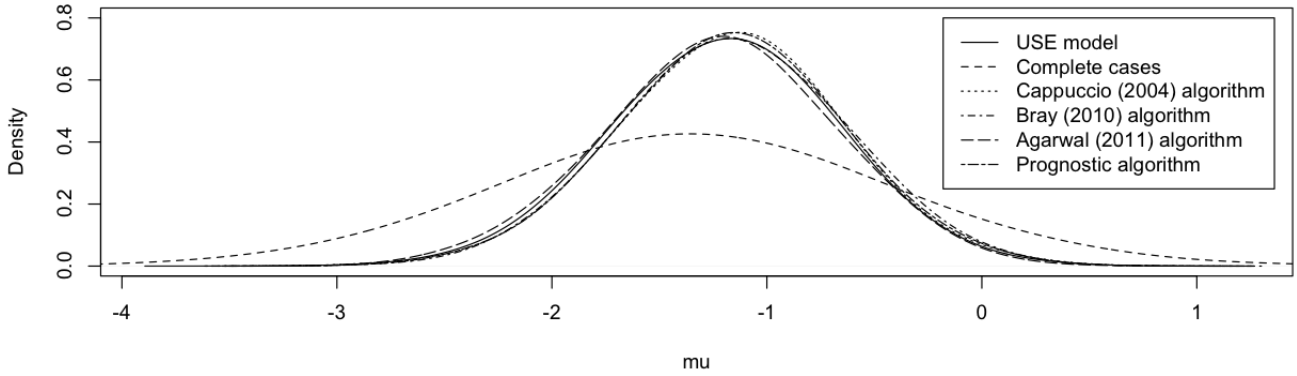
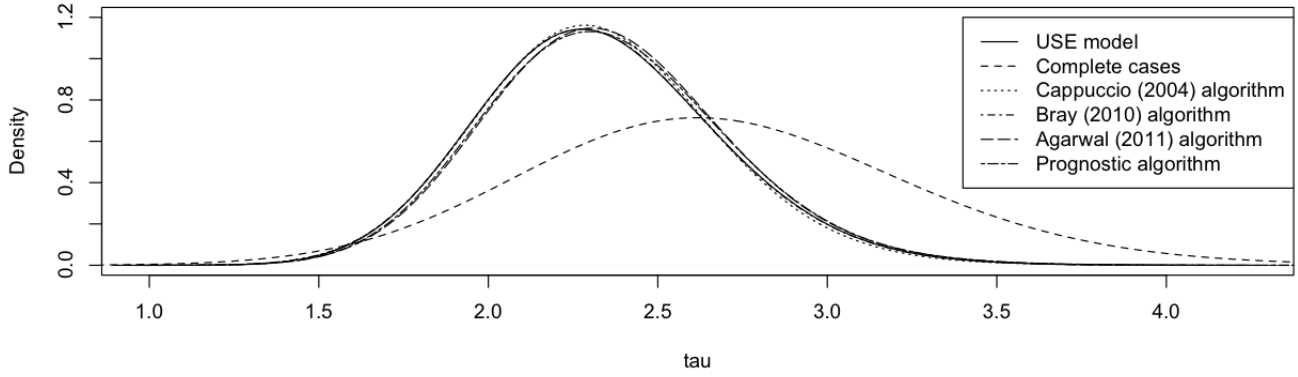
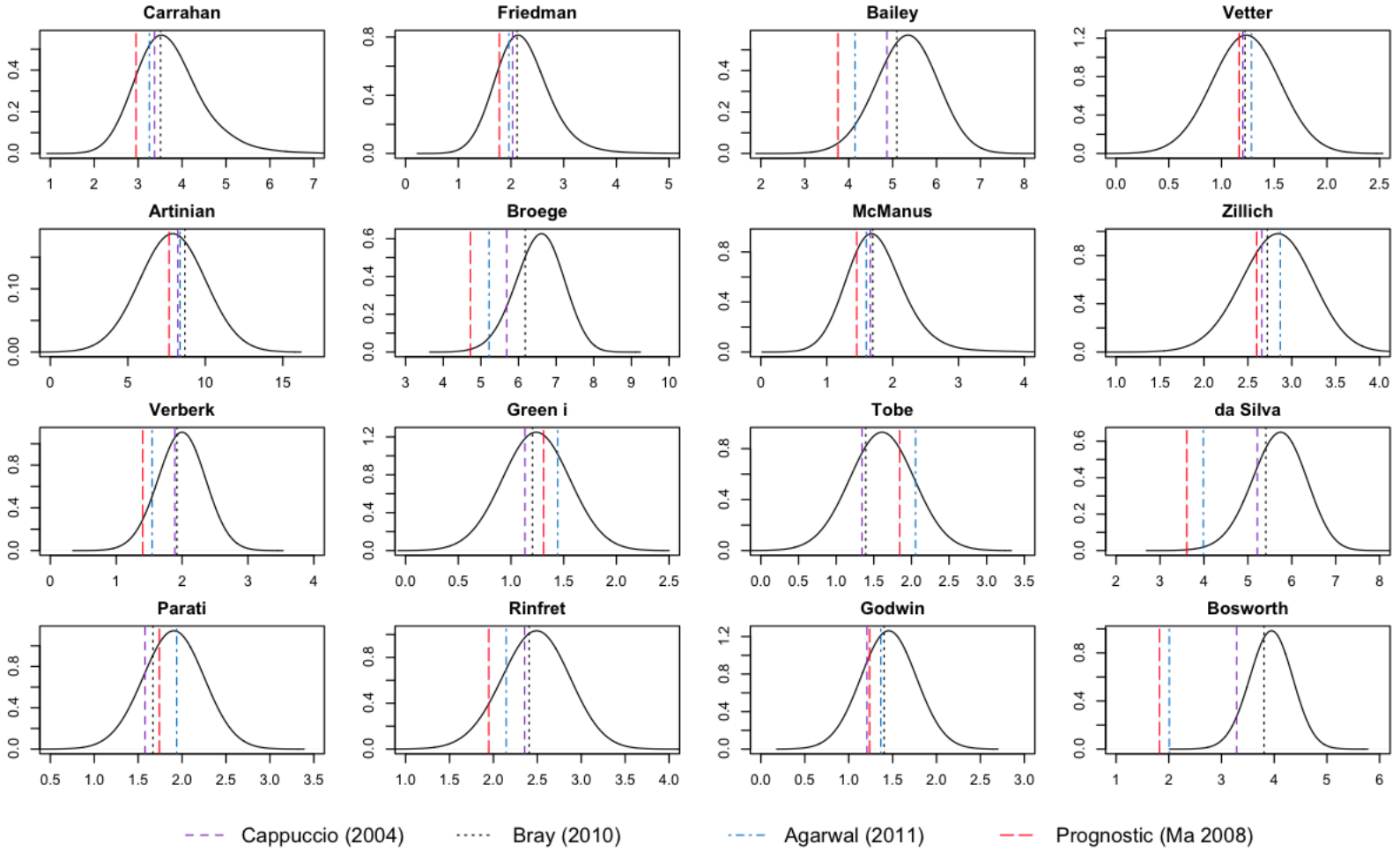


Figure 3.6: Meta-analysis for DBP monitoring vs usual care: Posterior distribution of τ under various models.



The USE model improves the validity of the meta-analysis. It is a formally constructed model, whereas the other approaches are simple algorithms with no theoretical foundation and that have not been empirically tested. Moreover, instead of inserting a single value for the missing SE and hoping for the best, as in the case of the standard algorithms, the USE model computes a posterior distribution for every missing standard deviation, σ_{ijb}, σ_{ijf}, and standard error se_{Di}. Figure 3.4.3 displays the posteriors for the missing standard errors, se_{Di}, and vertical lines which are the single value imputations from various insertion algorithms for the SBP dataset. In all cases the prognostic algorithm yields smaller inserted values for the missing SEs than the Agarwal (2011) algorithm. Also, in all cases the values inserted by the Cappuccio (2004) algorithm are smaller than those from the Bray (2010) algorithm. In general, the inserted values from all algorithms are consistently smaller than the mean of the posterior distributions from the USE model. Thus, accounting for the uncertainty in the data extraction results in somewhat increased SEs.

Figure 3.7: Posteriors of missing SE of the difference of mean differences, se_{Di}. Vertical lines are the single value imputations from the four standard algorithms.



3.5 Discussion

The systematic review of methods for handling missing variance data performed by Wiebe et al. (2006) concluded that Bayesian solutions were rarely used in practice for meta-analyses of continuous outcomes. When Bayesian solutions were used, they were poorly described and were considered "unclear." We have provided a fully Bayesian model to handle missing variance data in meta-analysis of continuous outcomes. Our model is superior to existing approaches to handle the missing data for several reasons: First, existing approaches are only algorithms for inserting values for missing standard deviations and are not theoretically justified. In contrast, we have formally constructed a model that accommodates the complexity of the data. Second, existing approaches consist of the insertion of a single value and the value is then treated as observed extracted data. Instead, our Bayesian model naturally incorporates the uncertainty involved in estimating missing standard deviations and standard errors. Thus, it eliminates the need to perform a sensitivity analysis for every inserted value. Third, our model accommodates all the patterns of missingness observed in real data. However, as with any Bayesian analysis, careful specification of the prior distributions for the variance parameters may be required.

We have applied our model and traditional approaches to meta-analyses of home monitoring of blood pressure compared to usual care with outcomes change in systolic and diastolic blood pressures. Previous meta-analyses have concluded that there exists a significant reduction in systolic and change in diastolic blood pressure in the home monitoring groups. All of the algorithms and the USE model applied to our updated meta-analyses agree with the result, except for the case of a meta-analysis of complete data only. Performing meta-analyses using only studies with complete data is discouraged because, as our data illustrates, much information is lost by a complete-cases analysis. The USE model generates slightly larger posterior SEs than using inserted values from the existing algorithms, as is to be expected because our model accounts for the uncertainty of estimating the missing data, which other approaches ignore.

Since it is unavoidable to encounter missing variances in meta-analyses of mean difference,

further work is needed to determine how much the results are affected by the number of studies, the sample size per study, loss to follow-up, and the percentage of missing data allowed in meta-analysis that undergo variance imputation.

CHAPTER 4

Future work in meta-analysis data extraction

I discuss two additional situations in meta-analysis where the data needed to be extracted from publications is missing or incomplete.

4.1 Data Extraction Problems in Meta-analysis of Hazard Ratios

While the majority of survival meta-analyses analyze odds ratios, time-to-event outcomes are commonly analyzed using a Cox model, which gives a hazard ratio as an output. Tierney et al. (2007) claimed that if there is sufficient data available to estimate an odds ratio, there is usually sufficient data to estimate a hazard ratio. Tierney et al. (2007) proposed methodology to estimate hazard ratios by carefully manipulating summary data found in published papers. However, the uncertainty created with these estimates is not accounted for in the meta-analysis. Appropriate methodology to account for this uncertainty is needed.

4.2 Data Extraction Problems in Dose-Response Meta-Analysis

Dose-response meta-analyses are useful to study how much the risk increases as exposure increases. The meta-analysis is performed on the dose-response patterns reported by each study. However, very few publications provide estimates of a slope with dose as a continuous variable. Instead, publications typically provide estimates of relative risk of the response per unit of exposure for specific dose categories (Ilyasova et al., 2005). For example, in a meta-analysis of the dose-response relationship between coffee intake and risk of heart-failure (Mostofsky et al., 2012) one study contributing to the meta-analysis had the following coffee

consumption categories: 0 to < 1, 1 to < 3, 3 to < 5, 5 to < 7 and ≥ 7 cups a day, while another contributing study had: ≤ 1 cups , 2 cups, 3 cups, 4 cups, and ≥ 5 cups a day. Because the categories are not the same and some are intervals, it is problematic to compute a slope. Current practice assigns a single representative dose value for each category. If the category is an interval, then the midpoint is chosen as a representative dose value. Slopes of the dose-response relationship are calculated for each study using the representative dose values. Using standard methods of meta-analysis, a single slope estimate is obtained from a weighted average of the individual slopes (Shi and Copas, 2004). Appropriate methodology to account for the uncertainty of the slopes is needed.

Appendix A

Ratio of two uniformly distributed random variables

Let $x^{\text{true}} \sim \text{Unif}(x-w, x+w)$ and $y^{\text{true}} \sim \text{Unif}(y-z, y+z)$, where $x > w > 0$ and $y > z > 0$. Then, the density of $p = x^{\text{true}}/y^{\text{true}}$ is

for $\frac{x-w}{y-z} \leq \frac{x+w}{y+z}$

$$g(p|x,y,w,z) = \begin{cases} 0 & \text{for} \quad p \leq \frac{(x-w)}{(y+z)} \\ \frac{(y+z)^2 - \frac{(x-w)^2}{(p)^2}}{8wz} & \text{for} \quad \frac{(x-w)}{(y+z)} \leq p \leq \frac{(x-w)}{(y-z)} \\ \frac{y}{2w} & \text{for} \quad \frac{(x-w)}{(y-z)} \leq p \leq \frac{(x+w)}{(y+z)} \\ \frac{\frac{(x+w)^2}{(p)^2} - (y-z)^2}{8wz} & \text{for} \quad \frac{(x+w)}{(y+z)} \leq p \leq \frac{(x+w)}{(y-z)} \\ 0 & \text{for} \quad \frac{(x+w)}{(y-z)} \leq p, \end{cases}$$

for $\frac{x-w}{y-z} \geq \frac{x+w}{y+z}$

$$g(p|x,y,w,z) = \begin{cases} 0 & \text{for} \quad p \leq \frac{(x-w)}{(y+z)} \\ \frac{\frac{(x-w)^2}{(p)^2} - (y+z)^2}{8wz} & \text{for} \quad \frac{(x-w)}{(y+z)} \leq p \leq \frac{(x+w)}{(y+z)} \\ \frac{x}{2z(v)^2} & \text{for} \quad \frac{(x+w)}{(y+z)} \leq p \leq \frac{(x-w)}{(y-z)} \\ \frac{\frac{(x+w)^2}{(p)^2} - (y-z)^2}{8wz} & \text{for} \quad \frac{(x-w)}{(y-z)} \leq p \leq \frac{(x+w)}{(y-z)} \\ 0 & \text{for} \quad \frac{(x+w)}{(y-z)} \leq p. \end{cases}$$

Appendix B

Simulated data for the UR-EE model

Table B.1: Simulated data for the UR-EE model: n_{ij} is the number of people at baseline, e_{ij} is the number of observed deaths, s_{ij} is the number of true deaths and κ_{ij} is the KM survival probability at year 1, for study i and group j.

Study	n_{i1}	e_{i1}	s_{i1}	κ_{i1}	lost to follow-up $j=1$	n_{i0}	e_{i0}	s_{i0}	κ_{i0}	lost to follow-up $j=0$
1	45	17	23	0.53	16	32	12	12	0.63	0
2	139	21	31	0.82	60	36	6	6	0.83	0
3	75	25	30	0.59	29	15	4	4	0.73	0
4	80	31	35	0.53	22	23	10	10	0.57	0
5	156	43	55	0.66	66	187	72	73	0.61	4
6	87	23	30	0.68	35	90	19	20	0.79	5
7	183	21	24	0.87	72	116	32	32	0.72	3
8	141	26	29	0.78	62	70	22	22	0.69	0
9	158	27	32	0.80	68	168	63	64	0.62	2
10	120	38	45	0.60	47	48	20	20	0.58	1

Appendix C

USE Bayesian model in JAGS

USE model for the SBP dataset:

```
model_USE_SBP <-"model{
phoc ~ dnorm(0.47 , 26.22)I(0,1)
phot ~ dnorm (0.51, 41.85)I(0,1)

# Carrahan
inv.xt.var[1] ~ dgamma(a.xt, b.xt)
inv.xc.var[1] ~ dgamma(a.xc, b.xc)
inv.yt.var[1] ~ dgamma(a.yt, b.yt)
inv.yc.var[1] ~ dgamma(a.yc, b.yc)
xt.var[1]<-1/inv.xt.var[1]
xc.var[1]<-1/inv.xc.var[1]
yt.var[1]<-1/inv.yt.var[1]
yc.var[1]<-1/inv.yc.var[1]
var.t[1]<- xt.var[1] + yt.var[1] - 2*phot*sqrt(xt.var[1])*sqrt(yt.var[1])
var.c[1]<- xc.var[1] + yc.var[1] - 2*phoc*sqrt(xc.var[1])*sqrt(yc.var[1])
pooled.variance[1]<-   ((nt[1]-1) * var.t[1] + (nc[1]-1)*var.c[1])/(nt[1]+nc[1]-2)
sem[1]<- sqrt(pooled.variance[1])*sqrt(1/nt[1]+1/nc[1])

# Friedman
inv.xt.var[4] ~ dgamma(a.xt, b.xt)
inv.xc.var[4] ~ dgamma(a.xc, b.xc)
inv.yt.var[4] ~ dgamma(a.yt, b.yt)
inv.yc.var[4] ~ dgamma(a.yc, b.yc)
xt.var[4]<-1/inv.xt.var[4]
xc.var[4]<-1/inv.xc.var[4]
```

```
yt.var[4]<-1/inv.yt.var[4]
yc.var[4]<-1/inv.yc.var[4]
var.t[4]<- xt.var[4] + yt.var[4] - 2*phot*sqrt(xt.var[4])*sqrt(yt.var[4])
var.c[4]<- xc.var[4] + yc.var[4] - 2*phoc*sqrt(xc.var[4])*sqrt(yc.var[4])
pooled.variance[4]<-   ((nt[4]-1) * var.t[4] + (nc[4]-1)*var.c[4])/(nt[4]+nc[4]-2)
sem[4]<- sqrt(pooled.variance[4])*sqrt(1/nt[4]+1/nc[4])

# Baiely
var.t[5]<- xt.var[5] + yt.var[5] - 2*phot*sqrt(xt.var[5])*sqrt(yt.var[5])
var.c[5]<- xc.var[5] + yc.var[5] - 2*phoc*sqrt(xc.var[5])*sqrt(yc.var[5])
pooled.variance[5]<-   ((nt[5]-1) * var.t[5] + (nc[5]-1)*var.c[5])/(nt[5]+nc[5]-2)
sem[5]<- sqrt(pooled.variance[5])*sqrt(1/nt[5]+1/nc[5])

# Vetter
var.t[6]<- xt.var[6] + yt.var[6] - 2*phot*sqrt(xt.var[6])*sqrt(yt.var[6])
var.c[6]<- xc.var[6] + yc.var[6] - 2*phoc*sqrt(xc.var[6])*sqrt(yc.var[6])
pooled.variance[6]<-   ((nt[6]-1) * var.t[6] + (nc[6]-1)*var.c[6])/(nt[6]+nc[6]-2)
sem[6]<- sqrt(pooled.variance[6])*sqrt(1/nt[6]+1/nc[6])

# Artinian
var.t[8]<- xt.var[8] + yt.var[8] - 2*phot*sqrt(xt.var[8])*sqrt(yt.var[8])
var.c[8]<- xc.var[8] + yc.var[8] - 2*phoc*sqrt(xc.var[8])*sqrt(yc.var[8])
pooled.variance[8]<-   ((nt[8]-1) * var.t[8] + (nc[8]-1)*var.c[8])/(nt[8]+nc[8]-2)
sem[8]<- sqrt(pooled.variance[8])*sqrt(1/nt[8]+1/nc[8])

# Broege
var.t[9]<- xt.var[9] + yt.var[9] - 2*phot*sqrt(xt.var[9])*sqrt(yt.var[9])
var.c[9]<- xc.var[9] + yc.var[9] - 2*phoc*sqrt(xc.var[9])*sqrt(yc.var[9])
pooled.variance[9]<-   ((nt[9]-1) * var.t[9] + (nc[9]-1)*var.c[9])/(nt[9]+nc[9]-2)
sem[9]<- sqrt(pooled.variance[9])*sqrt(1/nt[9]+1/nc[9])

# McManus
inv.yt.var[14] ~ dgamma(a.yt, b.yt)
inv.yc.var[14] ~ dgamma(a.yc, b.yc)
yt.var[14]<-1/inv.yt.var[14]
```

```
yc.var[14]<-1/inv.yc.var[14]
var.t[14]<- xt.var[14] + yt.var[14] - 2*phot*sqrt(xt.var[14])*sqrt(yt.var[14])
var.c[14]<- xc.var[14] + yc.var[14] - 2*phoc*sqrt(xc.var[14])*sqrt(yc.var[14])
pooled.variance[14]<-   ((nt[14]-1) * var.t[14] + (nc[14]-1)*var.c[14])/(nt[14]+nc[14]-2)
sem[14]<- sqrt(pooled.variance[14])*sqrt(1/nt[14]+1/nc[14])

# Zillich
var.t[15]<- xt.var[15] + yt.var[15] - 2*phot*sqrt(xt.var[15])*sqrt(yt.var[15])
var.c[15]<- xc.var[15] + yc.var[15] - 2*phoc*sqrt(xc.var[15])*sqrt(yc.var[15])
pooled.variance[15]<-   ((nt[15]-1) * var.t[15] + (nc[15]-1)*var.c[15])/(nt[15]+nc[15]-2)
sem[15]<- sqrt(pooled.variance[15])*sqrt(1/nt[15]+1/nc[15])

# Verberk
var.t[18]<- xt.var[18] + yt.var[18] - 2*phot*sqrt(xt.var[18])*sqrt(yt.var[18])
var.c[18]<- xc.var[18] + yc.var[18] - 2*phoc*sqrt(xc.var[18])*sqrt(yc.var[18])
pooled.variance[18]<-   ((nt[18]-1) * var.t[18] + (nc[18]-1)*var.c[18])/(nt[18]+nc[18]-2)
sem[18]<- sqrt(pooled.variance[18])*sqrt(1/nt[18]+1/nc[18])

# Green i
var.t[19]<- xt.var[19] + yt.var[19] - 2*phot*sqrt(xt.var[19])*sqrt(yt.var[19])
var.c[19]<- xc.var[19] + yc.var[19] - 2*phoc*sqrt(xc.var[19])*sqrt(yc.var[19])
pooled.variance[19]<-   ((nt[19]-1) * var.t[19] + (nc[19]-1)*var.c[19])/(nt[19]+nc[19]-2)
sem[19]<- sqrt(pooled.variance[19])*sqrt(1/nt[19]+1/nc[19])

# Green ii
var.t[20]<- xt.var[20] + yt.var[20] - 2*phot*sqrt(xt.var[20])*sqrt(yt.var[20])
var.c[20]<- xc.var[20] + yc.var[20] - 2*phoc*sqrt(xc.var[20])*sqrt(yc.var[20])
pooled.variance[20]<-   ((nt[20]-1) * var.t[20] + (nc[20]-1)*var.c[20])/(nt[20]+nc[20]-2)
sem[20]<- sqrt(pooled.variance[20])*sqrt(1/nt[20]+1/nc[20])

# Tobe
var.t[22]<- xt.var[22] + yt.var[22] - 2*phot*sqrt(xt.var[22])*sqrt(yt.var[22])
var.c[22]<- xc.var[22] + yc.var[22] - 2*phoc*sqrt(xc.var[22])*sqrt(yc.var[22])
pooled.variance[22]<-   ((nt[22]-1) * var.t[22] + (nc[22]-1)*var.c[22])/(nt[22]+nc[22]-2)
sem[22]<- sqrt(pooled.variance[22])*sqrt(1/nt[22]+1/nc[22])
```

```
# Da silva
var.t[23]<- xt.var[23] + yt.var[23] - 2*phot*sqrt(xt.var[23])*sqrt(yt.var[23])
var.c[23]<- xc.var[23] + yc.var[23] - 2*phoc*sqrt(xc.var[23])*sqrt(yc.var[23])
pooled.variance[23]<-   ((nt[23]-1) * var.t[23] + (nc[23]-1)*var.c[23])/(nt[23]+nc[23]-2)
sem[23]<- sqrt(pooled.variance[23])*sqrt(1/nt[23]+1/nc[23])

# Parati
var.t[26]<- xt.var[26] + yt.var[26] - 2*phot*sqrt(xt.var[26])*sqrt(yt.var[26])
var.c[26]<- xc.var[26] + yc.var[26] - 2*phoc*sqrt(xc.var[26])*sqrt(yc.var[26])
pooled.variance[26]<-   ((nt[26]-1) * var.t[26] + (nc[26]-1)*var.c[26])/(nt[26]+nc[26]-2)
sem[26]<- sqrt(pooled.variance[26])*sqrt(1/nt[26]+1/nc[26])

# Rinfret
var.t[27]<- xt.var[27] + yt.var[27] - 2*phot*sqrt(xt.var[27])*sqrt(yt.var[27])
var.c[27]<- xc.var[27] + yc.var[27] - 2*phoc*sqrt(xc.var[27])*sqrt(yc.var[27])
pooled.variance[27]<-   ((nt[27]-1) * var.t[27] + (nc[27]-1)*var.c[27])/(nt[27]+nc[27]-2)
sem[27]<- sqrt(pooled.variance[27])*sqrt(1/nt[27]+1/nc[27])

# Godwin
var.t[28]<- xt.var[28] + yt.var[28] - 2*phot*sqrt(xt.var[28])*sqrt(yt.var[28])
var.c[28]<- xc.var[28] + yc.var[28] - 2*phoc*sqrt(xc.var[28])*sqrt(yc.var[28])
pooled.variance[28]<-   ((nt[28]-1) * var.t[28] + (nc[28]-1)*var.c[28])/(nt[28]+nc[28]-2)
sem[28]<- sqrt(pooled.variance[28])*sqrt(1/nt[28]+1/nc[28])

# Bosworth
var.t[29]<- xt.var[29] + yt.var[29] - 2*phot*sqrt(xt.var[29])*sqrt(yt.var[29])
var.c[29]<- xc.var[29] + yc.var[29] - 2*phoc*sqrt(xc.var[29])*sqrt(yc.var[29])
pooled.variance[29]<-   ((nt[29]-1) * var.t[29] + (nc[29]-1)*var.c[29])/(nt[29]+nc[29]-2)
sem[29]<- sqrt(pooled.variance[29])*sqrt(1/nt[29]+1/nc[29])

for (i in 1:length(mean.diff)){
mean.diff[i] ~ dnorm(theta[i], prec[i])
prec[i]<- 1/(sem[i]*sem[i])
theta[i] ~ dnorm(mu, inv.tau2)
```

```
}

mu ~ dnorm(0, 0.001)
inv.tau2 ~ dgamma(7, 60)
tau2<-1/inv.tau2
tau<-sqrt(tau2)
}"
```

Bibliography

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

Abrams, K. R., Gillies, C. L., and Lambert, P. C. (2005). Meta-analysis of heterogeneously reported trials assessing change from baseline. *Statistics in Medicine*, 24(24):3823–3844.

Agarwal, R., Bills, J. E., Hecht, T. J., and Light, R. P. (2011). Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control a systematic review and meta-analysis. *Hypertension*, 57(1):29–38.

Artinian, N. T., Washington, O. G., and Templin, T. N. (2001). Effects of home telemonitoring and community-based monitoring on blood pressure control in urban African Americans: a pilot study. *Heart and Lung: The Journal of Acute and Critical Care*, 30(3):191–199.

Bailey, B., Carney, S., Gillies, A., and Smith, A. (1999). Antihypertensive drug treatment: a comparison of usual care with self blood pressure measurement. *Journal of human hypertension*, 13(2):147–150.

Baqué, A. D., Peréz, R. C., Rocarias, M. G., Sabaterd, M. F., Passola, A. Y.-C., Vilanova, M. B., Taboada, J. V., and Carbó, A. C. (2005). Efectividad de la automedición de la presión arterial en pacientes hipertensos: Estudio DIOAMPA. *Atención Primaria*, 35(5):233–238.

Berlin, J. A. (1997). Does blinding of readers affect the results of meta-analyses? *The Lancet*, 350(9072):185–186.

Berlin, J. A., Santanna, J., Schmid, C. H., Szczech, L. A., and Feldman, H. I. (2002). Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine*, 21(3):371–387.

Berman, N. G. and Parker, R. A. (2002). Meta-analysis: neither quick nor easy. *BMC Medical Research Methodology*, 2(1):10.

Bosworth, H. B., Powers, B. J., Olsen, M. K., McCant, F., Grubber, J., Smith, V., Gentry, P. W., Rose, C., Van Houtven, C., Wang, V., Goldstein, M. K., and Oddone, E. Z. (2011). Home blood pressure management and improved blood pressure control: results from a randomized controlled trial. *Archives of Internal Medicine*, 171(13):1173–1180.

Bray, E. P., Holder, R., Mant, J., and McManus, R. J. (2010). Does self-monitoring reduce blood pressure? Meta-analysis with meta-regression of randomized controlled trials. *Annals of Medicine*, 42(5):371–386.

Brener, S. J., Galla, J. M., Bryant, R., Sabik, J. F., and Ellis, S. G. (2008). Comparison of percutaneous versus surgical revascularization of severe unprotected left main coronary stenosis in matched patients. *American Journal of Cardiology*, 101:169–172.

Brockwell, S. E. and Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, 20:825–840.

Broege, P. A., James, G. D., and Pickering, T. G. (2001). Management of hypertension in the elderly using home blood pressures. *Blood Pressure Monitoring*, 6(3):139–144.

Buszman, P. E., Kiesz, S. R., Bochenek, A., Peszek-Przybyla, E. P., Szkrobka, I., Debinski, M., Bialkowska, B., Dudek, D., Gruska, A., Zurakowski, A., Milewski, K., Wilczynski, M., Rzeszutko, L., Buszman, P., Szymszal, J., Martin, J. L., and Tendera, M. (2008). Acute and late outcomes of unprotected left main stenting in comparison with surgical revascularization. *Journal of the American College of Cardiology*, 51:538–545.

Cappuccio, F. P., Kerry, S. M., Forbes, L., and Donald, A. (2004). Blood pressure control by home monitoring: meta-analysis of randomised trials. *British Medical Journal*, 329(7458):145–151.

Carlin, J. B. (1992). Meta-analysis for 2x2 tables: a Bayesian approach. *Statistics in Medicine*, 11:141–158.

Carnahan, J. E. and Nugent, C. A. (1975). The effects of self-monitoring by patients on the control of hypertension. *The American Journal of the Medical Sciences*, 269(1):69–73.

Chieffo, A., Morici, N., Maisano, F., Bonizzoni, E., Cosgrave, J., Montorfano, M., Airoldi, F., Carlino, M., Michev, I., Melzi, G., Sangiorgi, G., Alfieri, O., and Colombo, A. (2006). Percutaneous treatment with drug-eluting stent implantation versus bypass surgery for unprotected left main stenosis: a single-center experience. *Circulation*, 113:2542–2547.

da Silva, G. V., De Barros, S., Abensur, H., Ortega, K. C., and Mion, D. (2009). Home blood pressure monitoring in blood pressure control among haemodialysis patients: an open randomized clinical trial. *Nephrology Dialysis Transplantation*, 24(12):3805–3811.

DeJesus, R. S., Chaudhry, R., Leutink, D. J., Hinton, M. A., Cha, S. S., and Stroebel, R. J. (2009). Effects of efforts to intensify management on blood pressure control among patients with type 2 diabetes mellitus and hypertension: a pilot study. *Vascular Health and Risk Management*, 5:705.

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

Earp, J., Ory, M. G., and Strogatz, D. S. (1982). The effects of family involvement and practitioner home visits on the control of hypertension. *American Journal of Public Health*, 72(10):1146–1154.

Eyseneck, H. J. (1994). Meta-analysis and its problems. *British Medical Journal*, 309(6957):789–792.

Follmann, D., Elliott, P., Suh, I., and Cutler, J. (1992). Variance imputation for overviews of clinical trials with continuous response. *Journal of Clinical Epidemiology*, 45(7):769–773.

Friedman, R. H., Kazis, L. E., Jette, A., Smith, M. B., Stollerman, J., Torgerson, J., and Carey, K. (1996). A telecommunications system for monitoring and counseling patients with hypertension: impact on medication adherence and blood pressure control. *American Journal of Hypertension*, 9(4):285–292.

Fu, R., Vandermeer, B. W., Shamliyan, T. A., O'Neil, M. E., Yazdi, F., Fox, S. H., and Morton, S. C. (2013). Handling continuous outcomes in quantitative synthesis. *Methods Guide for Comparative Effectiveness Reviews.* AHRQ Publication No. 13-EHC103-EF. Rockville, MD.

Furukawa, T. A., Barbui, C., Cipriani, A., Brambilla, P., and Watanabe, N. (2006). Imputing missing standard deviations in meta-analyses can provide accurate results. *Journal of Clinical Epidemiology*, 59(1):7–10.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

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

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall /CRC, Boca Raton, FL, 2nd edition.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472.

Godwin, M., Lam, M., Birtwhistle, R., Delva, D., Seguin, R., Casson, I., and MacDonald, S. (2010). A primary care pragmatic cluster randomized trial of the use of home blood pressure monitoring on blood pressure levels in hypertensive patients with above target blood pressure. *Family practice*, 27(2):135–142.

Green, B. B., Cook, A. J., Ralston, J. D., Fishman, P. A., Catz, S. L., Carlson, J., Carrell, D., Tyll, L., Larson, E. B., and Thompson, R. S. (2008). Effectiveness of home blood pressure monitoring, web communication, and pharmacist care on hypertension control: a randomized controlled trial. *The Journal of the American Medical Association*, 299(24):2857–2867.

Halme, L., Vesalainen, R., Kaaja, M., and Kantola, I. (2005). Self-monitoring of blood

pressure promotes achievement of blood pressure target in primary health care. *American Journal of Hypertension*, 18(11):1415–1420.

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

Haynes, R. B., Gibson, E., Hackett, B., Sackett, D., Taylor, D. W., Roberts, R., and Johnson, A. (1976). Improvement of medication compliance in uncontrolled hypertension. *The Lancet*, 307(7972):1265–1268.

Higgins, J. and Deeks, J. (2011). In Higgins, J. and Deeks, J., editors, *Cochrane Handbook for Systematic Reviews of Interventions*, chapter 7: Selecting studies and collecting data. The Cochrane Collaboration. Version 5.1.0 [updated March 2011]. Available at: `www.cochrane-handbook.org`.

Hozo, S. P., Djulbegovic, B., and Hozo, I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*, 5(1):13–23.

Ilyasova, D., Hertz-Picciotto, I., Peters, U., Berlin, J. A., and Poole, C. (2005). Choice of exposure scores for categorical regression in meta-analysis: a case study of a common problem. *Cancer Causes and Control*, 16(4):383–388.

Johnson, A. L., Taylor, D. W., Sackett, D. L., Dunnett, C. W., and SHIMIzu, A. G. (1978). Self-recording of blood pressure in the management of hypertension. *Canadian Medical Association Journal*, 119(9):1034.

Kauric-Klein, Z. and Artinian, N. (2007). Improving blood pressure control in hypertensive hemodialysis patients. *Canadian Association of Nephrology Nurses and Technologists Journal*, 17(4):24.

Kovalchik, S. A. (2012). Survey finds that most meta-analysts do not attempt to collect individual patient data. *Journal of Clinical Epidemiology*, 65(12):1296–1299.

Kovalchik, S. A. and Cumberland, W. G. (2012). Using aggregate data to estimate the standard error of a treatment–covariate interaction in an individual patient data meta-analysis. *Biometrical Journal*, 54(3):370–384.

Lambert, P. C., Sutton, A. J., Abrams, K. R., and Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology*, 55:86–94.

Larose, D. R. and Dey, D. K. (1997). Grouped random effects models for Bayesian meta-analysis. *Statistics in Medicine*, 16:1817–1829.

LeLorier, J., Gregoire, G., Benhaddad, A., Lapierre, J., and Derderian, F. (1997). Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine*, 337(8):536–542.

Ma, J., Liu, W., Hunter, A., and Zhang, W. (2008). Performing meta-analysis with incomplete statistical information in clinical trials. *BMC Medical Research Methodology*, 8(1):56–67.

Madsen, L. B., Kirkegaard, P., and Pedersen, E. B. (2008). Blood pressure control during telemonitoring of home blood pressure. A randomized controlled trial during 6 months. *Blood Pressure*, 17(2):78–86.

Makikallio, T. H., Niemela, M., Kervinen, K., Jokinen, V., Laukkanen, J., Yitalo, I., Tulppo, M. P., Juvonen, J., and Huikuri, H. V. (2008). Coronary angioplasty in drug eluting stent era for the treatment of unprotected left main stenosis compared to coronary artery bypass grafting. *Annals of Medicine*, 40:437–443.

Márquez-Contreras, E., Martel-Claros, N., Gil-Guillén, V., de la Figuera-Von Wichmann, M., Casado-Martínez, J. J., Martin-de Pablos, J. L., Figueras, M., Galera, J., and Serra, A. (2006). Efficacy of a home blood pressure monitoring programme on therapeutic compliance in hypertension: the EAPACUM–HTA study. *Journal of Hypertension*, 24(1):169–175.

Márquez-Contreras, E., Martel-Claros, N., Gil-Guillén, V., Martín-de Pablos, J. L., de la Figuera-Von Wichmann, M., Casado-Martínez, J. J., and Espinosa, G. J. (2009). El control de la inertia terapéutica en el tratamiento de la hipertensión arterial mediate differences estrategias [control of therapeutic inertia in the treatment of arterial hypertension by using different strategies]. *Atención Primaria*, 41(6):315–323.

McManus, R., Mant, J., Roalfe, A., Oakes, R., Bryan, S., Pattison, H. M., and Hobbs, F. (2005). Targets and self monitoring in hypertension: randomised controlled trial and cost effectiveness analysis. *British Medical Journal*, 331(7515):493.

Mehos, B. M., Saseen, J. J., and MacLaughlin, E. J. (2000). Effect of pharmacist intervention and initiation of home blood pressure monitoring in patients with uncontrolled hypertension. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 20(11):1384–1389.

Melle, J. P., de Jonge, P., Spijkerman, T. A., Tussen, J. G. P., Ormel, J. van Veldhuisen, D. J., van den Brink, R. H. S., and van den Berg, M. P. (2004). Prognostic association of depression following myocardial infarction with mortality and cardiovascular events: a meta-analysis. *Psychosomatic Medicine*, 66:814–822.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.

Midanik, L. T., Resnick, B., Hurley, L. B., Smith, E. J., and McCarthy, M. (1991). Home blood pressure monitoring for mild hypertensives. *Public Health Reports*, 106(1):85.

Moher, D., Schulz, K. F., and Altman, D. G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet*, 357(9263):1191–1194.

Mostofsky, E., Rice, M. S., Levitan, E. B., and Mittleman, M. A. (2012). Habitual coffee consumption and risk of heart failure a dose-response meta-analysis. *Circulation: Heart Failure*, 5(4):401–405.

Naik, H., White, A. J., Chakravarty, T., J., F., Fontana, G., Kar, S., Shah, P. K., Weiss, R. E., and Makkar, R. (2009). A meta-analysis of 3,773 patients treated with percutaneous coronary intervention or surgery for unprotected left main coronary artery stenosis. *Journal of the American College of Cardiology: Cardiovascular Intervention*, 2:739–747.

Palmerini, T., Marzocchi, A., Marrozzini, C., Ortolani, P., Saia, F., Savini, C., Bacchi-Reggiani, L., Gianstefani, S., Virzi, S., Manara, F., Meron, K. W., Marinelli, G., Bartolomeo, R., and Branzi, A. (2006). Comparison between coronary angioplasty and coronary artery bypass surgery for the treatment of unprotected left main coronary artery stenosis (the Bologna registry). *American Journal of Cardiology*, 98:54–59.

Parati, G., Omboni, S., Albini, F., Piantoni, L., Giuliano, A., Revera, M., Illyes, M., Mancia, G., Group, T. S., et al. (2009). Home blood pressure telemonitoring improves hypertension control in general practice. the teleBPCare study. *Journal of Hypertension*, 27(1):198–203.

Parmar, M. K., Torri, V., and Stewart, L. (1998). Extracting summary statistics to perform meta-analysis of the published literature for survival endpoints. *Statistics in Medicine*, 17:2815–2834.

Pearson, E. (1932). The percentage limits for the distribution of range in samples from a normal population. *Biometrika*, 24:404–417.

Peto, R. (1987). Why do we need systematic overviews of randomized trials? *Statistics in Medicine*, 6:233–140.

Philbrook, H. T., Barrowman, N., and Garg, A. (2007). Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. *Journal of Clinical Epidemiology*, 60(3):228–240.

Pierce, J. P., Watson, D. S., Knights, S., Gliddon, T., Williams, S., and Watson, R. (1984). A controlled trial of health education in the physician's office. *Preventive Medicine*, 13(2):185–194.

Plummer, M. et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, volume 124, pages 125–133. Vienna.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at: `http://www.R-project.org`.

Riley, R., Abrams, K. R., Sutton, A. J., Lambert, P. C., Jones, D. R., Heney, D., and Burchill, S. A. (2003). Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *British Journal of Cancer*, 88:1191–1198.

Riley, R. D., Kauser, I., Bland, M., Thijs, L., Staessen, J. A., Wang, J., Gueyffier, F., and Deeks, J. J. (2013). Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. *Statistics in Medicine*, 32(16):2747–2766.

Rinfret, S., Lussier, M.-T., Peirce, A., Duhamel, F., Cossette, S., Lalonde, L., Tremblay, C., Guertin, M.-C., LeLorier, J., Turgeon, J., et al. (2009). The impact of a multidisciplinary information technology–supported program on blood pressure control in primary care. *Circulation: Cardiovascular Quality and Outcomes*, 2(3):170–177.

Robertson, C., Idris, N. R. N., and Boyle, P. (2004). Beyond classical meta-analysis: can inadequately reported studies be included? *Drug Discovery Today*, 9(21):924–931.

Rogers, M. A., Small, D., Buchan, D. A., Butch, C. A., Stewart, C. M., Krenzer, B. E., and Husovsky, H. L. (2001). Home monitoring service improves mean arterial pressure in patients with essential hypertension: a randomized, controlled trial. *Annals of Internal Medicine*, 134(11):1024–1032.

Rudd, P., Miller, N. H., Kaufman, J., Kraemer, H. C., Bandura, A., Greenwald, G., and Debusk, R. F. (2004). Nurse management for hypertension. A systems approach. *American Journal of Hypertension*, 17(10):921–927.

Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V., and Chalmers, T. C. (1987). Meta-analyses of randomized controlled trials. *The New England Journal of Medicine*, 316(8):450–455.

Sanmartin, M., Baz, J. A., Claro, R., Asorey, V., Duran, D., Pradas, G., and Iniguez, A. (2007). Comparison of drug-eluting stents versus surgery for unprotected left main coronary artery disease. *American Journal of Cardiology*, 100:970–973.

Serryus, P. W. (2008). SYNTAX trial: a randomized trial of CABG vs. PCI for the treatment of 3 vessel disease or unprotected left main disease – end points in the left main subset. New York, NY. Presented at: Transcatheter Therapeutics (TCT). September 10.

Seung, K. B., Park, D., Kim, Y., Lee, S., Lee, C. W., Hong, M., Park, S., Yun, S., Gwon, H., Jeong, M., Jang, Y., Kim, H., Kim, P., Seong, I., Park, H. S., Ahn, T., Chae, I., Tahk, S., Chung, W., and S., P. (2008). Stents versus coronary-artery bypass grafting for left main coronary artery disease. *The New England Journal of Medicine*, 358:1781–1792.

Shapiro, S. (1997). Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *Journal of Clinical Epidemiology*, 50(3):223–229.

Shi, J. Q. and Copas, J. (2004). Meta-analysis for trend estimation. *Statistics in Medicine*, 23(1):3–19.

Smith, T. C., Spiegelhalter, D. J., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine*, 14:2685–2699.

Soghikian, K., Casper, S. M., Fireman, B. H., Hunkeler, E. M., Hurley, L. B., Tekawa, I. S., and Vogt, T. M. (1992). Home blood pressure monitoring: effect on use of medical services and medical care costs. *Medical Care*, 30(9):855–865.

Staessen, J. A., Den Hond, E., Celis, H., Fagard, R., Keary, L., Vandenhoven, G., and O'Brien, E. T. (2004). Antihypertensive treatment based on blood pressure measurement at home or in the physician's office: a randomized controlled trial. *The Journal of the American Medical Association*, 291(8):955–964.

Stahl, S. M., Kelley, C. R., Neill, P. J., Grim, C. E., and Mamlin, J. (1984). Effects of home blood pressure measurement on long-term BP control. *American Journal of Public Health*, 74(7):704–709.

Stevens, J. W. (2011). A note on dealing with missing standard errors in meta-analyses of continuous outcome measures in WinBUGS. *Pharmaceutical Statistics*, 10(4):374–378.

Stewart, L. A. and Parmar, M. K. B. (1993). Meta-analysis of the literature or of individual patient data: is there a difference? *The Lancet*, 341:418–422.

Streiner, D. L. and Joffe, R. (1998). The adequacy of reporting randomized, controlled trials in the evaluation of antidepressants. *Canadian Journal of Psychiatry*, 43:1026–1030.

Sutton, A. J. and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, 10:277–303.

Thacker, S. B. (1988). Meta-analysis: a quantitative approach to research integration. *Journal of the American Medical Association*, 259(11):1685–1689.

Tierney, J. F., Stewart, L. A., Ghersi, D., Burdett, S., and Sydes, M. R. (2007). Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*, 8:1–16.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22:1701–1728.

Tobe, S. W., Hunter, K., Geerts, R., Raymond, N., and Pylypchuk, G. (2008). IMPPACT: investigation of medical professionals and patients achieving control together. *Canadian Journal of Cardiology*, 24(3):205–208.

Verberk, W. J., Kroon, A. A., Lenders, J. W., Kessels, A. G., van Montfrans, G. A., Smit, A. J., van der Kuy, P.-H. M., Nelemans, P. J., Rennenberg, R. J., Grobbee, D. E., Beltman, F. W., Joore, M. A., Brunenberg, D. E. M., Dirksen, C., Thien, T., and de Leeuw, P. W. (2007). Self-measurement of blood pressure at home reduces the need for antihypertensive drugs a randomized, controlled trial. *Hypertension*, 50(6):1019–1025.

Vetter, W., Hess, L., and Brignoli, R. (2000). Influence of self-measurement of blood pressure on the responder rate in hypertensive patients treated with losartan: Results of the SVATCH study. *Journal of Human Hypertension*, 14(4):235–241.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random effects model. *Journal of Educational and Behavioral Statistics*, 30:261–293.

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

Walker, E., Hernandez, A. V., and Kattan, M. W. (2008). Meta-analysis: its strengths and limitations. *Cleveland Clinic Journal of Medicine*, 75(6):431–439.

Weiss, R. E. (1996). An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society. Series B*, 58:739–750.

White, A. J., Kedia, G., Mirocha, J. M., Lee, M. S., Forrester, J. S., Morales, W. C., Dohad, S., Kar, S., Czer, L. S., Fontana, G. P., Trento, A., Shah, P. K., and Makkar, R. R. (2008a). Comparison of coronary artery bypass surgery and percutaneous drug-eluting stent implantation for treatment of left main coronary artery stenosis. *Journal of the American College of Cardiology: Cardiovascular Intervention*, 1:236–245.

White, I. R., Higgins, J. P. T., and Wood, A. M. (2008b). Allowing for uncertainty due to missing data in meta-analysis - part 1: Two-stage methods. *Statistics in Medicine*, 27:711–727.

White, I. R., Welton, N. J., Wood, A. M., Ades, A. E., and Higgins, J. P. T. (2008c). Allowing for uncertainty due to missing data in meta-analysis – part 2: Hierarchical models. *Statistics in Medicine*, 27:728–745.

Wiebe, N., Vandermeer, B., Platt, R. W., Klassen, T. P., Moher, D., and Barrowman, N. J. (2006). A systematic review identifies a lack of standardization in methods for handling missing variance data. *Journal of Clinical Epidemiology*, 59(4):342–353.

Wu, C., Hannan, E. L., Walford, G., and Faxon, D. P. (2008). Utilization and outcomes of unprotected left main coronary artery stunting and coronary artery bypass graft surgery. *The Annals of Thoracic Surgery*, 86:1153–1159.

Zillich, A. J., Sutherland, J. M., Kumbera, P. A., and Carter, B. L. (2005). Hypertension outcomes through blood pressure monitoring and evaluation by pharmacists (home study). *Journal of General Internal Medicine*, 20(12):1091–1096.