A META-ANALYTIC FRAMEWORK FOR COMBINING INCOMPARABLE COX PROPORTIONAL HAZARD MODELS CAUSED BY OMITTING IMPORTANT COVARIATES

by

Xing Yuan

B.S. in Statistics, Beijing Normal University, China, 2004

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Xing Yuan

It was defended on

September 25th 2009

and approved by

**Stewart J. Anderson, PhD, Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh**

Joseph P. Costantino, DrPH, Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Jong-Hyeon Jeong, PhD, Associate Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Garrick L. Wallstrom, PhD, Assistant Professor, Department of Biomedical Informatics,
School of Medicine, University of Pittsburgh

Dissertation Advisor:

**Stewart J. Anderson, PhD, Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh**

Copyright © by Xing Yuan
2009

Stewart J. Anderson, PhD

A META-ANALYTIC FRAMEWORK FOR COMBINING INCOMPARABLE COX PROPORTIONAL HAZARD MODELS CAUSED BY OMITTING IMPORTANT COVARIATES

Xing Yuan, PhD

University of Pittsburgh, 2009

Meta-analysis can be broadly defined as the quantitative review and synthesis of the results of related but independent studies into a single overall result. It is a statistical analysis that combines or integrates the results of several independent clinical trials considered by the analyst to be "combinable". In many biomedical research areas, especially clinical trials in oncology, researchers often use time to some event (or death) as the primary endpoint to assess treatment effects. As the amount of survival analyses continues to increase, there is a greater need to summarize a pool of studies into a coherent overview.

It is well established that in Cox proportional hazard models with censored survival data, estimates of treatment effects with some important covariates omitted will be biased toward zero. This is especially problematic in meta-analyses which combine estimates of parameters from studies where different covariate adjustments are made. Presently, few constructive solutions have been provided to address this issue. We propose a meta-analytic framework for combining incomparable Cox models using both aggregated patient data (APD) and individual patient data (IPD) structures. For APD, two meta-regression models (meta-ANOVA and meta-polynomial models) with indicators of different covariates in Cox models are proposed to adjust for the heterogeneity of treatment effects across studies. Both parametric and nonparametric estimators for the pooled treatment effect and the heterogeneity variance are presented and compared. For IPD, we propose a hierarchical multiple imputation method to handle the unique missing covariates problem when we combine individual data from different studies for a meta-analysis, and results are compared with estimations from the conventional multiple imputation method. We illustrate the advantages of our proposed

analytic procedures over existing methodologies by simulation studies and real data analyses using multiple breast cancer clinical trials.

The public health significance of our work is to provide practical guidance of designing and implementing meta-analyses of incomparable Cox proportional hazard models for researchers in the fields of clinical trials, medical research, and other health care areas. Such guidance is important due to the emerging role of meta-analysis in assessing important public health studies.

Keywords Meta-analysis, Survival analysis, Cox proportional hazard model, Heterogeneity, Random effect model, Missing covariates.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First of all, I would like to express my gratitude to my advisor, Dr. Stewart J. Anderson, for his guidance, encouragement and continuous support throughout the years of my PhD program. He taught me how to ask questions and express my ideas. He showed me different ways to approach scientific problems and demonstrated that persistence is the key for achieving the goals. More important, I have learned not only how to do research but also the philosophy of life from him. I would also like to thank my committee members: Dr. Joseph P. Costantino, Dr. Jong-Hyeon Jeong and Dr. Garrick L. Wallstrom, for giving me invaluable suggestions to improve this dissertation throughout the whole course. Furthermore, I am most thankful to Dr. Lisa A. Weissfeld for her financial support for my PhD program. The experience I gained while working with her and Dr. Lan Kong at the Department of Critical Medicine has been invaluable in the completion of my study, and will continue to benefit me in the future.

I have enjoyed my time in Pittsburgh, and must thank for all the professors who helped me to build a solid foundation for my current and future career through their classes. I also want to thank my classmates, friends and colleagues for the good time we have had, and the support they could provide. My family, especially my wife Dongli, deserves special thanks for their love and unfailing support, which has enabled me to bring this work to fruition. I will be forever grateful.

1.0 INTRODUCTION

Meta-analysis can be broadly defined as the quantitative review and synthesis of the results of related but independent studies into a single overall result [57]. The term meta-analysis was introduced by Gene Glass [20] in his 1976 presidential address to the American Educational Research Association as "the analysis of the results of statistical analyses for the purpose of drawing general conclusions". Since the introduction of meta-analysis, it has been widely used in education, psychology, and biomedical research. The objectives of a meta-analysis include increasing power to detect an overall treatment effect, estimation of degree of benefits associated with a particular study treatment, assessment of the amount of variability between studies, and identification of study characteristics associated with particularly effective treatment [52]. Although Glass coined this new term in 1976, the statistical method of combining results of different studies has a long history. In early times, this often took the form of several p-values being combined into a single value, using procedures such as Fisher's method [16]. Over the decades, there has been enormous number of papers using meta-analyses to combine many types of effect sizes. The commonly used effect sizes fall into one of the two families: the d family (standardized difference between two groups) and the r family (correlation measures of effect size) [7].

1.1 META-ANALYZE SURVIVAL REGRESSION RESULTS

Methodologies of synthesizing the two types of effect sizes in meta-analysis have been intensively studied and clearly documented in several major publications (see Hedges and Olkin (1985) [26]; Rosenthal and Rubin (1986) [62]; Copper and Hedges (1994) [8]; Rosnow and

Rosenthal (1996) [63]; Sutton et al. (2000) [71]; Lipsey and Wilson (2001) [45]; and Hunter and Schmidt (2004) [35].). However, methods of synthesizing evidences from studies using regression, especially in survival regression, have not yet been well studied.

In biomedical research areas, especially clinical trials in oncology, researchers often use time to some event (or death) as the primary endpoint to assess treatment effects. The Cox proportional hazard model [9], introduced by D. R. Cox, has been widely used in survival analysis to accommodate both the effects of treatment and other important covariates. For each subject, the hazard function with a set of covariates, $x_1, x_2, \ldots, x_p$, can be decomposed into two parts: one is the baseline hazard function involves time but not covariates, and one that involves the covariates but not time. Hence,

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p). \tag{1.1}$$

This semi-parametric model makes no assumptions about the nature or shape of the hazard function, which makes Cox method more robust. Another advantage of using Cox proportional hazard model is that it is relatively easy to incorporate different patterns of covariates, such as interaction terms, higher-order variables or time-dependent covariates, in the model to get more precise estimation of treatment effect.

As the amount of survival analyses done by scientists and investigators in universities, government agencies, and other research institutes continues to increase, there is a greater need to summarize a pool of studies into a coherent overview. Meta-analysis enhances the researcher's ability to combine seemingly contradictory results into a summary conclusion of what all data has to say on the subjects. In addition, meta-analysis increases the power of statistical analysis by combining results from separate (but similar) experiments, so that it can provide information that would not otherwise be available. For example, a group of clinical trials may all have positive treatment effect on survival that may be too small to be significant by themselves, but when pooled can give convincing evidence of positive effect.

1.2 POTENTIAL PROBLEMS FROM COMBINING COX MODELS

The most interesting statistics from a Cox model, which can be used as effect sizes in the meta-analysis, are the raw regression coefficients with their variances. When reporting the treatment effect in a randomized controlled trial with a survival outcome, it is recommended that the appropriate summary statistics are the log hazard ratio and its variance [56]. Suppose in a meta-analysis combining k Cox regression models, x_{1i} is designated as the treatment covariate in study i, $i = 1, 2, \ldots, k$. Then the raw regression coefficient of treatment variable β_{1i} is the log hazard ratio while controlling for other covariates in the model. A naïve approach is to take the simple average $\bar{\beta}_1$ but then we do not take into account that in different studies, β_{1i} have different precisions (SE) because of the variation in sample sizes and other characteristics across the clinical trials. A better meta-analysis technique called fixed-effect model can then be used, where a weighted average of all estimated coefficients is viewed as the synthesized log hazard ratio when all trials combined together. Various weights have been proposed but the most commonly used is the reciprocal of variance for each estimated coefficient. However, a risk is that if the variance in one study is close to zero then the combined result will be dominated only by that study. The meta-analysis model based on the random effect approach assumes that studies may be heterogeneous, i.e., variations exist between studies, and this variation is represented by the random effect with an unknown variance.

One of the major problems when combining potential effect sizes from different Cox regression models is the "incomparability" issue, that is, different studies may adjust for different sets of covariates in the Cox models, so the coefficients of interest may have different interpretations across the studies. If the groups are similar, except for the treatment under study, this issue might not be problematic. More often than that, the subjects in the groups have some additional characteristics that may affect their outcomes, such as demographic variables or physiological variables. These variables may be used as covariates in the Cox model in explaining the response variable. However, different trials have different protocols and study designs, some of the significant variables are not observed for certain studies by design or by happenstance. One way to deal with this problem is restricting all Cox models

to have exactly the same set of covariates be included for the meta-analysis. Such a practice may be a waste of useful information and Gail et al. (1984) [19] suggested that biases of estimation are introduced due to omitted covariates. In particular, treatment effect estimates with important covariates omitted will be biased toward zero (hazard ratio toward one) and efficiency is lost for testing treatment effects.

In this dissertation, we will propose a meta-analytic framework for combining incomparable Cox proportional hazard models caused by omitting important covariates. In the next chapter, we give a summary of statistical methodologies commonly used for meta-analysis. Also, three statistical techniques dealing with missing covariate data in the context of modeling time-to-event type data are introduced and compared.

In Chapter 3, we present our proposed methods which account for the incomparability issue of estimated coefficients in the meta-analysis framework. In Chapter 4, simulation studies and data analyses using multiple breast cancer clinical trials provide that proposed methods outperform the commonly used meta-analysis methods. Lastly, in Chapter 5, we make recommendations and outline directions for further research.

2.0 LITERATURE REVIEW

2.1 STATISTICAL METHODOLOGIES FOR META-ANALYSIS

Properly conducted meta-analyses are thought to represent among the highest level of evidence addressing important biomedical issues. There are two types of meta-analysis methodologies based on the data availability. One is called aggregated patient data (APD) meta-analysis. In many meta-analyses an attempt is made to obtain information for all relevant clinical trials by collecting data summaries or APD from previous studies that have been published or presented at professional meetings. A recent search of Medline by Lyman and Kuderer (2005) [48] identified 1,595 reported meta-analyses related to cancer of which 1,519 (95.2%) were based on APD. The other type is called individual patient data (IPD) meta-analysis. IPD meta-analysis has been described as the gold standard and yardstick of systematic review [3]. Although it can be costly and time comsuming, there are many advantages of using IPD meta-analysis: consistent analyses across studies, which is particularly important for survival data; the ability to easily perform subgroup analyses and to use updated follow-up information; the consideration of covariates to examine adjusted differences in treatment effectiveness across patients and trials; and the ability to assess the appropriateness of the original analysis methods. Different statistical methodologies for both APD and IPD meta-analysis are reviewed in the following sections.

2.1.1 APD Meta-Analysis

Although APD meta-analysis is not as desirable as IPD meta-analysis, it conceptually has the same components including estimating the overall summary effect size and allowing one to as-

sess heterogeneity. In certain circumstances, undertaking a large project involving obtaining individual follow-up information may not be feasible owing to time or financial constraints. Hence, using APD may be the only practical alternative. In other cases, researchers may be able to obtain individual data but may wish to perform a APD meta-analysis first in a preliminary analysis. Consequently, they can decide whether it is worthwhile to proceed with the collection of individual data. The methodologies of APD meta-analysis are well developed and documented. Fixed-effect approaches include the use of inverse variance weighted estimator [75], the Mantel-Haenszel (1959) procedures [49] and a method described by Peto (1985) [78]. The most common random effect approach was described by DerSimonian and Laird (1986) [13].

Suppose we have a series of k studies, each comparing an experimental treatment to a control with respect to a time to event outcome. If all of the studies are expected to share a common treatment effect, or if we are interested in the average of the treatment effect in the series of studies, it is natural to pool estimates across studies. In fixed-effect meta-analyses, it is assumed that the true effect of treatment is the same in each study, or "fixed". The differences between study results are considered to be solely due to the play of chance. Hence, the assumption of a fixed effect can be tested using a test of heterogeneity. Among fixed-effect methods, the Mantel-Haenszel procedure is suitable for dichotomous outcomes and Peto's odds ratio method is an alternative to the Mantel-Haenszel method. Details about these two methods can be found elsewhere [49, 78].

2.1.1.1 Inverse Variance Fixed-Effect Method The inverse variance method may be used to pool either binary or continuous data. In general, the true effect size θ_i for study i, $i = 1, 2, \ldots, k$, could be the log odds ratio, log relative risk, risk difference, or difference in means or standardized means. In the setting of APD meta-analysis for survival data, we may only observe the log hazard ratio θ_i and its variance σ_i^2 for study i. The effect sizes are then combined to give a pooled estimate for the true overall treatment effect, θ, by calculating a weighted average of the treatment effects from the individual trials:

$$\hat{\theta}_{\text{IV}} = \frac{\sum_{i=1}^{k} w_i \theta_i}{\sum_{i=1}^{k} w_i}. \tag{2.1}$$

and the weights here are the reciprocals of the variances, that is,

$$w_i = \frac{1}{\sigma_i^2}. \tag{2.2}$$

The variance of $\hat{\theta}_{\text{IV}}$ is given by

$$\text{var}(\hat{\theta}_{\text{IV}}) = \frac{1}{\sum_{i=1}^{k} w_i}. \tag{2.3}$$

When the number of studies is large in the meta-analysis, $\hat{\theta}_{\text{IV}}$ is assumed to be normally distributed with mean θ and variance as in equation (2.3). We can then test the null hypothesis $H_0 : \theta = 0$ by forming $Z = \hat{\theta}_{\text{IV}}/\sqrt{\text{var}(\hat{\theta}_{\text{IV}})}$ and comparing it to a standard normal distribution with significance level α (Wald test). Additionally, an approximate 95% confidence interval for θ is $\left(\hat{\theta}_{\text{IV}} - 1.96\sqrt{\text{var}(\hat{\theta}_{\text{IV}})}, \hat{\theta}_{\text{IV}} + 1.96\sqrt{\text{var}(\hat{\theta}_{\text{IV}})}\right)$.

If the number of studies in an APD meta-analysis is small, the assumption of normality may not be appropriate. Follmann and Proscham (1999) [17] proposed a group permutation method for hypothesis testing, where they permuted the treatment and control group labels *en masse* within each trial. This provides a permutation reference distribution for the estimate of treatment effect, which has better power and preserves the nominal Type I error level better than a Wald test.

The classical measure of heterogeneity is Cochran's Q statistic [6], which is calculated as the weighted sum of squared differences between individual study effects and the pooled estimated effect across all studies, with the weights in (2.2) being used. Hence, following the notation of DerSimonian and Laird (1986) [13],

$$Q = \sum_{i=1}^{k} w_i(\theta_i - \hat{\theta}_{\text{IV}})^2. \tag{2.4}$$

Under the null hypothesis, $H_0 : \theta_1 = \theta_2 = \ldots = \theta_k$, Q is distributed as a χ^2 statistic with $(k-1)$ degrees of freedom. The Q test has low power as a comprehensive test of heterogeneity, especially when the number of studies is small [23]. Conversely, the test arguably has excessive power when there are many studies, especially when those studies are large [31]. Despite these drawbacks, the Q test is still very popular and used in many meta-analyses.

Intuitively, the larger studies which have smaller variances, are given more weights than smaller studies, which have larger variances. Hence, the weights given in equation (2.2) minimize the variability of the pooled treatment effect $\hat{\theta}_{\text{IV}}$. However, a risk is that if one very large study has a variance close to zero (although unlikely in real cases), the pooled estimate will be dominated by that study. The meta-analysis model based on the random effect approach provides something of a compromise assuming that studies are heterogeneous, i.e., there is a random effect.

2.1.1.2 DerSimonian and Laird Random Effect Model In the fixed meta-analysis models, θ_i is treated as an observation and the within-study variance σ_i^2 is treated as a fixed number. In the random effect modeling approach, it is assumed that there exists a between-study variation besides the variation within each study. And this between-study variation is represented by the random effect a_i with mean zero and unknown variance τ^2. So the simple random effect model is written as

$$\theta_i = \theta + a_i + \varepsilon_i, \tag{2.5}$$

where θ_i is the observed study-specific treatment effect, θ is the true overall treatment effect, a_i is the random effect and ε_i is an error term with mean zero and known variance σ_i^2.

The pooled estimate of the common treatment effect provided by DerSimonian and Laird (1986) [13] is

$$\hat{\theta}_{\text{DL}} = \frac{\sum_{i=1}^{k} w_i^* \theta_i}{\sum_{i=1}^{k} w_i^*}, \tag{2.6}$$

with variance

$$\text{var}(\hat{\theta}_{\text{DL}}) = \frac{1}{\sum_{i=1}^{k} w_i^*}, \tag{2.7}$$

where each study's effect size is given weight

$$w_i^* = \frac{1}{\sigma_i^2 + \tau^2}. \tag{2.8}$$

We can also test $H_0 : \theta = 0$ by forming $Z' = \hat{\theta}_{\text{DL}} / \sqrt{\text{var}(\hat{\theta}_{\text{DL}})}$ and comparing to a standard normal distribution. The approximate 95% confidence interval for θ is $\Big(\hat{\theta}_{\text{DL}} - 1.96\sqrt{\text{var}(\hat{\theta}_{\text{DL}})}$, $\hat{\theta}_{\text{DL}} + 1.96\sqrt{\text{var}(\hat{\theta}_{\text{DL}})}\Big)$.

Thus, the heterogeneity variance τ^2 is the key parameter in random effect meta-analysis. If τ^2 were known, the weighted average (2.6) would be unbiased with minimum variance among all linear unbiased estimators following from the Gauss-Markov theorem. There are mainly two methods estimating the variance of the random effect: maximum likelihood estimation and quadratic method of moments estimation [11].

Maximum Likelihood Estimation

If we assume a_i and ε_i follow normal distribution, (2.5) can be written as

$$\theta_i \sim \mathcal{N}(\theta, \sigma_i^2 + \tau^2), \tag{2.9}$$

so that the log-likelihood function is

$$l(\theta, \tau^2) \propto -\frac{1}{2}\sum_{i=1}^{k}\left[\ln(\sigma_i^2 + \tau^2) + \frac{(\theta_i - \theta)^2}{\sigma_i^2 + \tau^2}\right]. \tag{2.10}$$

To use Newton-Raphson algorithm to get the maximum likelihood estimation (MLE), we need the first and second derivatives with respect to θ and τ^2:

$$\begin{aligned}
\frac{\partial l}{\partial \theta} &= \sum_{i=1}^{k}\frac{\theta_i - \theta}{\sigma_i^2 + \tau^2}, \quad \frac{\partial l}{\partial \tau^2} = -\frac{1}{2}\sum_{i=1}^{k}\left[\frac{1}{\sigma_i^2 + \tau^2} - \frac{(\theta_i - \theta)^2}{(\sigma_i^2 + \tau^2)^2}\right], \\
\frac{\partial^2 l}{\partial \theta^2} &= -\sum_{i=1}^{k}\frac{1}{\sigma_i^2 + \tau^2} = H_{11}, \quad \frac{\partial^2 l}{\partial (\tau^2)^2} = \frac{1}{2}\sum_{i=1}^{k}\left[\frac{1}{(\sigma_i^2 + \tau^2)^2} - \frac{2(\theta_i - \theta)^2}{(\sigma_i^2 + \tau^2)^3}\right] = H_{22}, \\
\frac{\partial^2 l}{\partial \theta \partial \tau^2} &= -\sum_{i=1}^{k}\left[\frac{\theta_i - \theta}{(\sigma_i^2 + \tau^2)^2}\right] = H_{12} = H_{21}.
\end{aligned}$$

And the Hessian matrix for $l(\theta, \tau^2)$ is

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}.$$

As we can see, the second derivative of the log-likelihood function with respect to τ^2, i.e. H_{22}, may be positive. This means the Newton-Raphson algorithm may fail if the starting point is far from the maximum.

The information matrix, which is the negative expected Hessian matrix of the log-likelihood function, is given by

$$\boldsymbol{\mathcal{I}} = -E(\mathbf{H}) = \begin{bmatrix} \sum_{i=1}^{k} \frac{1}{\sigma_i^2+\tau^2} & 0 \\ 0 & \frac{1}{2}\sum_{i=1}^{k} \frac{1}{(\sigma_i^2+\tau^2)^2} \end{bmatrix}.$$

Unlike the Hessian matrix, the information matrix, $\mathcal{I}$, is always positive definite and therefore, the Fisher Scoring (FS) algorithm is more reliable than the Newton-Raphson algorithm. Since the information matrix is block-diagonal, the FS algorithm leads to separate maximization over θ and τ^2:

$$\begin{aligned} \hat{\theta}_{s+1} &= \hat{\theta}_s + \left(\sum_{i=1}^{k} \frac{1}{\sigma_i^2+\hat{\tau}_s^2}\right)^{-1} \sum_{i=1}^{k} \frac{\theta_i - \hat{\theta}_s}{\sigma_i^2+\hat{\tau}_s^2} \\ &= \left(\sum_{i=1}^{k} \frac{1}{\sigma_i^2+\hat{\tau}_s^2}\right)^{-1} \sum_{i=1}^{k} \left(\frac{\hat{\theta}_s}{\sigma_i^2+\hat{\tau}_s^2} + \frac{\theta_i - \hat{\theta}_s}{\sigma_i^2+\hat{\tau}_s^2}\right) \\ &= \left(\sum_{i=1}^{k} \frac{1}{\sigma_i^2+\hat{\tau}_s^2}\right)^{-1} \sum_{i=1}^{k} \frac{\theta_i}{\sigma_i^2+\hat{\tau}_s^2}, \end{aligned} \tag{2.11}$$

$$\begin{aligned} \hat{\tau}_{s+1}^2 &= \hat{\tau}_s^2 + \left(\sum_{i=1}^{k} \frac{1}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right)^{-1} \sum_{i=1}^{k} \left(\frac{(\theta_i - \hat{\theta}_s)^2}{(\sigma_i^2+\hat{\tau}_s^2)^2} - \frac{1}{\sigma_i^2+\hat{\tau}_s^2}\right) \\ &= \left(\sum_{i=1}^{k} \frac{1}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right)^{-1} \sum_{i=1}^{k} \left(\frac{\hat{\tau}_s^2}{(\sigma_i^2+\hat{\tau}_s^2)^2} + \frac{(\theta_i - \hat{\theta}_s)^2}{(\sigma_i^2+\hat{\tau}_s^2)^2} - \frac{1}{\sigma_i^2+\hat{\tau}_s^2}\right) \\ &= \left(\sum_{i=1}^{k} \frac{1}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right)^{-1} \sum_{i=1}^{k} \left(\frac{(\theta_i - \hat{\theta}_s)^2 - \sigma_i^2}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right), \end{aligned} \tag{2.12}$$

where s is the iteration index and a good starting point of $\hat{\tau}_0^2$ could be 0. We can see that equation (2.11) takes the form of the weighted average (2.6).

Harville (1977) [24] preferred restricted maximum likelihood (REML) estimation to the maximum likelihood estimation. The log-likelihood function of REML is

$$l_R(\theta, \tau^2) \propto -\frac{1}{2}\left\{\sum_{i=1}^{k}\left[\ln(\sigma_i^2+\tau^2) + \frac{(\theta_i-\theta)^2}{\sigma_i^2+\tau^2}\right] + \ln\sum_{i=1}^{k}(\sigma_i^2+\tau^2)^{-1}\right\}. \tag{2.13}$$

The FS algorithm of τ^2 takes the form

$$\hat{\tau}_{s+1}^2 = \left(\sum_{i=1}^{k} \frac{1}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right)^{-1} \sum_{i=1}^{k} \left(\frac{(\theta_i - \hat{\theta}_s)^2 - \sigma_i^2}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right) + \frac{1}{\sum_{i=1}^{k}(\sigma_i^2+\tau_s^2)^{-1}}, \tag{2.14}$$

with a good starting point, $\hat{\tau}_0^2 = 0$.

Quadratic Method of Moments Estimation

Under the quadratic method of moments estimation, the assumption of normality is relaxed. This distribution-free property is very attractive especially when the number of study k is small. We have already constructed the weighted sum of squares of residuals Q as in (2.4). To calculate the expected value $E(Q)$, Kacker (2004) [38] verified that

$$\begin{aligned} E(\theta_i - \theta_{\text{IV}}) &= 0, \\ E(\theta_i - \theta_{\text{IV}})^2 &= \text{var}(\theta_i - \theta_{\text{IV}}) \\ &= \text{var}(\theta_i) + \text{var}(\theta_{\text{IV}}) - 2\text{cov}(\theta_i, \theta_{\text{IV}}) \\ &= \text{var}(\theta_i) + \frac{\sum_{i=1}^k w_i^2 \text{var}(\theta_i)}{(\sum_{i=1}^k w_i)^2} - \frac{2w_i \text{var}(\theta_i)}{\sum_{i=1}^k w_i}. \end{aligned}$$

So the expected value of Q is

$$\begin{aligned} E(Q) &= E\left(\sum_{i=1}^k w_i(\theta_i - \theta_{\text{IV}})^2\right) \\ &= \sum_{i=1}^k w_i \text{var}(\theta_i) + \frac{\sum_{i=1}^k w_i^2 \text{var}(\theta_i)}{\sum_{i=1}^k w_i} - \frac{2\sum_{i=1}^k w_i^2 \text{var}(\theta_i)}{\sum_{i=1}^k w_i} \\ &= \sum_{i=1}^k w_i \text{var}(\theta_i) - \frac{\sum_{i=1}^k w_i^2 \text{var}(\theta_i)}{\sum_{i=1}^k w_i} \\ &= \sum_{i=1}^k w_i(\sigma_i^2 + \tau^2) - \frac{\sum_{i=1}^k w_i^2(\sigma_i^2 + \tau^2)}{\sum_{i=1}^k w_i} \\ &= \tau^2\left(\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}\right) + \left(\sum_{i=1}^k w_i\sigma_i^2 - \frac{\sum_{i=1}^k w_i^2\sigma_i^2}{\sum_{i=1}^k w_i}\right) \end{aligned} \tag{2.15}$$

If we plug $\sigma_i^2 = w_i^{-1}$ in equation (2.15) and equate the empirical sum Q to its expected value, we have the following weighted method of moments estimate $\hat{\tau}^2_{WMM}$

$$\hat{\tau}^2_{WMM} = \frac{Q - (k-1)}{\left(\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}\right)} \tag{2.16}$$

as derived by DerSimonian and Laird (1986) [13]. Here Q is the heterogeneity statistic and the weights w_i's are calculated as in the inverse variance method (2.2).

However, there is a general problem with estimation of variance components — the estimates may be negative, e.g., when $Q < k - 1$. Accordingly, we truncate the weighted method of moments estimate as

$$\tilde{\tau}^2_{WMM} = \max(0, \hat{\tau}^2_{WMM}). \tag{2.17}$$

The problem with this estimator is that it is slightly positively biased. This bias is often observed for small between-study heterogeneity.

2.1.1.3 Meta-Regression In Section (2.1.1.1), Cochran's Q statistic for testing the heterogeneity among results of k studies was introduced. If there is a substantial heterogeneity between the studies, we need to investigate whether the heterogeneity is related to specific characteristics of the studies. In the context of meta-analysis, that can be done by including covariates on the study level (which could "explain" differences between studies) into the random effect model. Hence equation (2.5) becomes:

$$\theta_i = \mathbf{Z}_i'\boldsymbol{\alpha} + a_i + \varepsilon_i, \tag{2.18}$$

where θ_i is the observed study-specific treatment effect, $\mathbf{Z}_i = (1, z_1, z_2, \ldots, z_m)'$ is the $(1 + m) \times 1$ vector of covariates measured in study i including the constant term, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_m)'$ is the coefficient vector where the intercept term, α_0, is interpreted as the common treatment effect, a_i is the random effect with unknown variance τ^2 and ε_i is an error term with mean zero and study-specific variance σ_i^2.

The relation between θ_i and the "predictor" $\mathbf{Z}_i$ could be modeled by a normal distribution if the number of studies k is large. We get the marginal approximate model

$$\theta_i \sim \mathcal{N}(\mathbf{Z}_i'\boldsymbol{\alpha}, \sigma_i^2 + \tau^2). \tag{2.19}$$

If τ^2 is known, then we apply weighted least squares with

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^{k} w_i^* \mathbf{Z}_i \mathbf{Z}_i' \right)^{-1} \left(\sum_{i=1}^{k} w_i^* \mathbf{Z}_i \theta_i \right), \tag{2.20}$$

where the weight, $w_i^* = (\sigma_i^2 + \tau^2)^{-1}$, is as in the DerSimonian and Laird weighted estimate (2.6).

To get the MLE of τ^2, we can easily generalize the FS algorithm by replacing $\hat{\theta}_s$ with $\mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_s$ in the recursive equation (2.12), where

$$\hat{\boldsymbol{\alpha}}_s = \left(\sum_{i=1}^{k}(\sigma_i^2+\hat{\tau}_s^2)^{-1}\mathbf{Z}_i\mathbf{Z}_i'\right)^{-1}\left(\sum_{i=1}^{k}(\sigma_i^2+\hat{\tau}_s^2)^{-1}\mathbf{Z}_i\theta_i\right). \tag{2.21}$$

The weighted method of moments estimator of τ^2 has the form

$$\hat{\tau}^2_{WMM} = \frac{Q^*-(k-m)}{\sum_{i=1}^{k}\sigma_i^{-2} - tr\left[\left(\sum_{i=1}^{k}\sigma_i^{-2}\mathbf{Z}_i\mathbf{Z}_i'\right)^{-1}\left(\sum_{i=1}^{k}\sigma_i^{-4}\mathbf{Z}_i\mathbf{Z}_i'\right)\right]}, \tag{2.22}$$

where

$$Q^* = \sum_{i=1}^{k}\sigma_i^{-2}(\theta_i-\mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_0)^2, \tag{2.23}$$

$$\hat{\boldsymbol{\alpha}}_0 = \left(\sum_{i=1}^{k}\sigma_i^{-2}\mathbf{Z}_i\mathbf{Z}_i'\right)^{-1}\left(\sum_{i=1}^{k}\sigma_i^{-2}\mathbf{Z}_i\theta_i\right). \tag{2.24}$$

Under the null hypothesis $H_0 : \tau^2 = 0$, we have

$$Q^* \sim \chi^2(k-m-1), \tag{2.25}$$

where m is the number of covariates in the meta-regression model (2.18).

Meta-regression has become a standard model for assessing the potential impact of study-level covariates [71]. Although this method is widely used in systematic review, there are certain problems and limitations associated with this type of analysis: for example, the false-positive rate of the resulting tests for the evidence of effect of covariates can be higher than the nominal level because of the potential presence of heterogeneity [30] and innumerable characteristics of different studies may be identified as potential cause of heterogeneity. It is difficult to identify all of the covariates which, indeed, have influence on the treatment effect. There is a great danger of over-fitting since the numbers of studies in meta-analyses are usually small. Consequently, meta-regression should be carefully designed and analyzed, since the conclusions may be potentially misleading if results from meta-regression are interpreted naïvely.

2.1.2 IPD Meta-Analysis

In Section (2.1.1.3), we considered only analyses of meta-regression at the aggregated meta-analytic level. We did not consider covariates at the individual level. However, if such information is available, the data should be analyzed by IPD meta-analysis methods. IPD meta-analysis has been described as the gold standard and yardstick of systematic review [3]. This approach has several advantages over APD meta-analysis: consistent analyses across studies, which is particularly important for survival data; ability of undertaking subgroup analyses and using updated follow-up information; the consideration of covariates to examine differences in treatment effectiveness across patients and trials; and the possibility of assessing the appropriateness of original analysis methods. However, obtaining and analyzing IPD can be both costly and time consuming.

There are two general approaches to IPD meta-analysis [70]. The first method called a "one-stage analysis", pools IPD from all studies together so that a single analysis can be performed. This could be a "mega-trial" analysis, where distinctions between studies are ignored and the data are analyzed as if they belong to a single trial, or by a stratified analysis where the trial identities are included in the model. The second method called a "two-stage analysis" includes studies in a meta-analysis which are analyzed separately with a consistent modeling approach and then effect sizes and other summarized statistics are combined using APD meta-analysis techniques. In the next section, we will mainly focus on one-stage IPD meta-analysis.

2.1.2.1 One-Stage IPD Meta-Analysis Since we only focus on survival endpoints in this dissertation, the general regression modeling approaches are skipped and we will directly discuss the Cox proportional hazard model under the framework of one-stage IPD meta-analysis. Although we can collect IPD from different trials and combine them to form a "mega-trial", this combined data set should not be analyzed with a simple Cox model without further investigation. For example, factors that vary across studies, such as the inclusion/exclusion criteria, patient characteristics, and therapeutic schemes or practice may substantially influence the results. Thus, outcomes of the patients included in the same trial

are likely to be less heterogeneous than those observed in patients included in different trials. Such trial effects potentially lead to clustering, or dependence of outcomes at each trial. This is sometimes referred as "clustered data", where data from each trial define a cluster, and some general methodologies used for multicenter clinical trials should be transposable to IPD meta-analysis. However, differences between these two frameworks could lead to different recommendations. First, multicenter trials may have more clusters but varying sample sizes across trials whereas in IPD meta-analysis, each trial should have been adequately powered, even if expected benefits may vary from trial to trial. Second, heterogeneity in multicenter trials is often due to heterogeneity in baseline risks only, with fixed treatment effect across trials. On the contrary, in meta-analysis literature, heterogeneity is usually intended as heterogeneity in treatment effects across the trials, and modeling such heterogeneity becomes a key issue [39].

Suppose we observe survival type of IPD from k trials (clusters) and n_i subjects per trial, $i = 1, 2, \ldots, k$. The total sample size of IPD meta-analysis, noted as N, is thus $\sum_{i=1}^{k} n_i$. Let t_{ij}^0 be the latent failure time with c_{ij} denoting the censoring time for subject j in trial i, $j = 1, 2, \ldots, n_i$. So the observed failure time is $t_{ij} = \min(t_{ij}^0, c_{ij})$ and $\delta_{ij} = I\{t_{ij}^0 \leqslant c_{ij}\}$ is the failure indicator. Also assume we observe a $p-$dimensional vector of covariates $\mathbf{X}_{ij}$ for each subject. t_{ij} and c_{ij} are assumed to be independent conditionally on $\mathbf{X}_{ij}$ and identically distributed with each trial. The $(t_{ij}, \delta_{ij}, \mathbf{X}_{ij})$ are the observed variables. In the presence of dependence induced by the trial effect, two distinct approaches are available: marginal (or population-averaged) model and conditional (or trial-specific) model. The two strategies differ in methods for estimation as well as interpretation. The virtues of each approach have been extensively debated [22, 44, 51].

Marginal Cox Model

The fundamental difference between marginal model and conditional model is that instead of modeling trial-specific effects, marginal model "averages out" the trial effects. The marginal hazard of failure for subject j in trial i is

$$h_{ij}(t|\mathbf{X}_{ij}) = h_0(t) \exp(\mathbf{X}_{ij}'\boldsymbol{\beta}). \tag{2.26}$$

where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$. In this model, the baseline hazard $h_0(t)$ is not specific to any trial. The average effect of treatment is between two randomly selected subjects of the population in the whole meta-analysis: one treated and one untreated. Therefore, there is no way to model the heterogeneity in treatment effect across trials. Nevertheless, the "clustered" data structure is taken into account using a robust variance estimator [42, 41].

Conditional Cox Model

Generally, conditional models consist of fixed-effect Cox models, random effect Cox models (frailty models), and stratified Cox models. They share the general form which can be written as

$$h_{ij}(t|\mathbf{X}_{ij}) = h_{0i}(t) \exp(\mathbf{X}'_{ij}\boldsymbol{\beta}), \tag{2.27}$$

where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$. Trial effects are incorporated in equation (2.27) through the trial-specific baseline hazard function $h_{0i}(t)$. The model is conditional in the sense that the treatment effect is conditional on the trial-specific baseline hazard function. Specifically, the hazard ratio for treatment comparing two subjects, e.g., one with treatment and one with placebo, is from the same trial. A key feature of conditional model is the decomposition of the trial-specific baseline hazard $h_{0i}(t)$ into an arbitrary baseline hazard rate $h_0(t)$ and an exponential term which models the multiplicative effect on this baseline hazard.

In the fixed-effect Cox model, we assume trials act proportionally on the risk of failure by adding additional terms to model (2.27). Without loss of generality, we arbitrarily choose trial 1 as the reference trial, so that

$$h_{ij}(t|\mathbf{X}_{ij}) = h_0(t) \exp(\gamma_i + \mathbf{X}'_{ij}\boldsymbol{\beta}) \tag{2.28}$$

for $i = 1, 2, \ldots, k$, where $\gamma_1 = 0$. This model can be estimated by including trial indicator variables in an unstratified Cox model. Model (2.28) assumes common treatment effect adjusted for other covariates and requires estimation of $(k-1) + p$ parameters, where p is the length of vector $\boldsymbol{\beta}$.

When the heterogeneity of treatment effects adjusted by other covariates exists across trials, different trial-specific coefficients $\boldsymbol{\beta}_i$ can be incorporated into conditional model (2.28), leading to

$$h_{ij}(t|\mathbf{X}_{ij}) = h_0(t)\exp(\gamma_i + \mathbf{X}'_{ij}\boldsymbol{\beta}_i), \tag{2.29}$$

which requires the estimation of $(k-1) + k \times p$ parameters.

When the number of trials is small relative to the sample size, the fixed-effect Cox model is a very attractive approach, especially if the trial effect is of special interest. However, when $k \to \infty$, a large number of parameters must be estimated and the asymptotics may break down. This major drawback of fixed-effect model can be very problematic in the extreme cases where very large number of trials are being analyzed.

A better approach is the random effect Cox model or frailty model. Similar to the fixed-effect Cox model, the frailty model assumes trials act proportionally on the baseline hazard rate. However, instead of treating the trial effects as fixed parameters, the frailty model treats them as a sample from a member of a family of probability distributions. In some applications, it is more convenient to write model (2.27) as

$$h_{ij}(t|\mathbf{X}_{ij}) = h_0(t)\xi_i\exp(\mathbf{X}'_{ij}\boldsymbol{\beta}), \tag{2.30}$$

where ξ_i is the frailty for trial i, which represents the unmeasured factors affecting trial-specific baseline hazard functions. Here we assume ξ_i's are an independent sample from some distribution with mean 1 and variance ζ_0^2. When $\zeta_0^2 = 0$, model (2.30) reduces to a simple Cox model, and trial effects are absent so that failures are independent within as well as across trials. As ζ_0^2 increases and nature picks a value of ξ_i greater than 1 in trial i, then subjects in that particular trial tend to fail at a faster rate than under a model where ξ_i is equal or smaller than 1.

The frailty model cannot generally be fitted by an extension of the standard partial likelihood for Cox model. The likelihood may not have a closed form unless the frailty distribution is carefully chosen. The most commonly used distribution in the literature is the one parameter gamma distribution introduced by Clayton (1978) [5]. This distribution yields a closed form of likelihood which can be readily maximized. Other frailty distributions, which may or may not result in closed-form likelihoods, include the positive stable distribution (Hougaard

(1986a) [34]), the inverse Gaussian distribution (Hougaard (1986b) [33]), log-normal distribution (McGilchrist (1993) [50]) and compound Poisson (Henderson (1999) [27]). All of these models can be represented in equation (2.30) and use a single parameter to index the degree of dependence. Thus, in contrast to the fixed-effect Cox model, the number of parameters does not grow with the number of trials. In the gamma frailty distribution, for example, only the variance ζ_0^2 needs to be estimated in addition to the coefficient vector $\boldsymbol{\beta}$.

Again, model (2.30) assumes common treatment effect. It can be modified to accommodate heterogeneity of treatment effects by adding trial-specific parameters $\boldsymbol{\beta}_i$'s into the model, so that

$$h_{ij}(t|\mathbf{X}_{ij}) = h_0(t)\xi_i \exp(\mathbf{X}'_{ij}\boldsymbol{\beta}_i). \tag{2.31}$$

Equation (2.31) implies two model components. The first is the frailty term, ξ_i, modeling the variation in baseline hazard. The second component is a model for the across-trial variation in the treatment effect denoted by $\boldsymbol{\beta}_i$. For the latter component, instead of fitting $\boldsymbol{\beta}_i$ with a fixed-effect strategy, we can denote $\boldsymbol{\beta}_i$ as a sum of common effect $\bar{\boldsymbol{\beta}}$ and a mean zero random effect $\boldsymbol{\omega}_i = \boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}$. Thus, the frailty model with random effects would have the form

$$h_{ij}(t|\mathbf{X}_{ij}) = h_0(t)\xi_i \exp(\mathbf{X}'_{ij}\bar{\boldsymbol{\beta}} + \mathbf{X}'_{ij}\boldsymbol{\omega}_i), \tag{2.32}$$

where the distribution of $\boldsymbol{\omega}_i$, $i = 1, 2, \ldots, k$, could be, for example, a log-gamma or a normal distribution.

However, frailty models with random effects cannot be readily implemented with existing frailty model methodologies. Yamaguchi and Ohashi (1999) [77] presented a REML algorithm to include random treatment effect in a log-normal frailty model as

$$h_{ij}(t|\mathbf{X}_{ij}) = h_0(t) \exp(\gamma_i + \mathbf{X}'_{ij}\bar{\boldsymbol{\beta}} + X'_{i1j}\omega_{i1}), \tag{2.33}$$

where γ_i is the frailty term denoted as in (2.28), X_{i1j} is the first element of vector $\mathbf{X}_{ij}$ indicating the treatment group, which is of primary interest. The random effects γ_i and ω_{i1} are mutually independent and normally distributed with mean 0 and variance ζ_0^2 and ζ_1^2, respectively. In a Bayesian setting, the use of Markov chain Monte Carlo (MCMC) computing allows a more flexible choice of distributions. Sargent (1998) [67] presented a general framework through which random effects can be introduced into the Cox model

using MCMC. This framework is using a Cox partial likelihood as the basis for a hierarchical model, and allows very general random effect structures for the model parameters.

The third type of conditional model is the stratified Cox model. This model takes the form of (2.27) with all k baseline hazard functions h_{0i}, $i = 1, 2, \ldots, k$, completely unspecified. The stratified Cox model assumes trial-specified baseline hazard functions act non-proportionally and it does not provide estimations of trial effects. However, the ease of computing and the applicability across a wide variety of settings make the stratified Cox model an appealing tool, especially when the trial effects are of no special interest. The stratified Cox model can also fit trial-specific parameters into the model as

$$h_{ij}(t|\mathbf{X}_{ij}) = h_{0i}(t)\exp(\mathbf{X}'_{ij}\boldsymbol{\beta}_i). \tag{2.34}$$

However, this formulation is equivalent to fitting a separate model for each trial, and it does not combine information as is done in a meta-analysis. Therefore, this case will not be considered in this dissertation.

So far, under the one-stage IPD meta-analysis framework, we have compared four models presented above, namely, the marginal Cox model, the fixed-effect Cox model, the random effect Cox model or so-called frailty model, and the stratified Cox model. The last three models share the same general form of conditional Cox model.

2.1.2.2 Two-Stage IPD Meta-Analysis The two-stage IPD meta-analysis follows the same statistical methodology of APD meta-analysis, and it is currently the most commonly used method of analyzing IPD. It is conceptually straightforward and is familiar to most researchers. For survival data, a common Cox model with treatment variable and the same set of other adjust covariates is applied for each study and then log-hazard ratios of treatment effect are collected to form a pooled estimate. From a purely statistical perspective, it may not be worthwhile to perform a two-stage IPD meta-analysis unless the appropriate summary results are not available. However, there are many benefits to collecting IPD, including data checking, data updating for longer term follow-up of participants, dealing with missing data, and checking the consistency of original analysis methods. For these

reasons, results from two-stage IPD meta-analyses and equivalent APD meta-analyses can be very different [37, 72].

2.2 STATISTICAL METHODOLOGIES FOR MISSING COVARIATES IN SURVIVAL ANALYSIS

In the previous section, we explained the advantages of IPD meta-analysis as a major improvement over APD meta-analysis. However, problems of missing covariates complicate IPD meta-analysis, and especially for IPD meta-analysis using Cox regression models which require full covariate information. A naïve approach to solve missing covariates problem is to analyze only the subjects with complete observations. This so-called complete case analysis may not only reduce precision because only part of the data is used but also may produce biased results when the missing data are not missing completely at random (MCAR). More sophisticated approaches for handling missing covariates in Cox model have been proposed for a single study. However, not only the frequency of missing data but also the missingness process may vary across studies from which individual data are pooled [40]. Thus different approaches of handling missing covariates for a single study may, therefore, not be directly implemented for IPD meta-analysis.

Statistical methodologies developed to estimate parameters of Cox model with missing covariates have a large bibliography, including several reviews and textbooks. We will review three different types of approaches: multiple imputation, likelihood-based methods and weighted estimating equation methods.

2.2.1 Multiple Imputation Method

Since its introduction 30 years ago by Rubin (1978) [64], multiple imputation (MI) has become an important and influential approach for dealing with the statistical analysis of missing data. The practice of filling in missing data with imputed plausible values has long been reviewed as an attractive approach because it can "lull the user into the pleasure state

of believing that the data are complete after all" [12]. As a result, standard statistical procedures for complete data can be directly implemented.

For survival analysis, Paik and Tsai (1997) [55] proposed a single imputation method for handling missing covariates in using Cox model under missing at random (MAR) mechanism. In their paper, they proposed an estimator that is the solution of the approximated partial likelihood equation obtained by replacing the missing covariates with the expectations of observed covariates from the same risk set. Therefore, information on missing covariates can be recovered from the observed covariates with proper conditioning. This estimator is asymptotically normal and relatively efficient than previous methods under restricted conditions. For one, it works best when the missing covariates are categorical with small number of categories.

However, single imputation cannot reflect the sampling variability under one model for nonresponse about the correct model for nonresponse. MI can be applied to salvage such deficiency. The key idea of the MI procedure is to replace each missing value with a set of M plausible values. Each value is a Bayesian draw from the conditional distribution of the missing observations given the observed data. MI is most straightforward to use under MAR but it is quite possible to apply it under missing not at random (MNAR) settings.

MI involves three distinct tasks [65]:

1. The missing values are filled in M times to generate M complete data sets.

2. The M complete data sets are analyzed by using standard procedures.

3. The results from the M analyses are combined into a single inference.

Paik (1997) [54] implemented a bootstrap-like MI technique proposed by Rubin and Schenker (1986) [66] to estimate Cox model with missing covariates. In her technique, called Approximate Bayesian Bootstrap (ABB), she first sampled with replacement from the non-missing observations, then drew imputes from this bootstrap sample. This method is based on the assumption that non-missing observations are all categorical. If not, we need to employ histogram smoothing technique as suggested by Paik and Tsai (1997) [55], in which each continuous covariate is categorized into appropriately chosen neighborhoods. This smoothing technique, however, may introduce bias.

The limitations of Paik's MI are obvious and it can not impute multiple missing covariates together with a multivariate fashion. Schafer (1997) [68], in his book, presented iterative algorithms for simulating MI of missing categorical, continuous or mixed type of covariates under multivariate models. The multiple imputation algorithm for MIX (estimation/multiple imputation for mixed categorical and continuous data) in his book is based on the general location model under MAR proposed by Olkin and Tate (1961) [53], and multiple imputations are generated based on two concepts: one is the likelihood-based inference with missing data using EM algorithm; the other one is the data augmentation method which uses a Markov Chain Monte Carlo (MCMC) technique. The algorithms from this book are computationally intensive but fortunately they are implemented in the package MIX for software R, which is developed by Schafer (1997) [68] and can be downloaded from CRAN.

2.2.2 Direct Likelihood-Based Method

The Cox model assumes that the hazard function for failure time T, conditional on the covariates $\mathbf{X}_i$, is of the form

$$\lim_{\Delta t \to 0} \frac{P(t \leqslant T \leqslant t + \Delta t)}{\Delta t} = h_i(t|\mathbf{X}_i) = h_0(t) \exp(\mathbf{X}_i' \boldsymbol{\beta}), \quad i = 1, 2, \ldots, n, \tag{2.35}$$

where $\mathbf{X}_i$ is a p-dimensional vector of categorical or continuous covariates, $h_0(t)$ is an unspecified baseline hazard function.

Generally, let T^0 be the latent failure time with C denoting the censoring time. So the observed failure time is $T = \min(T^0, C)$ and $\Delta = I\{T^0 \leqslant C\}$ is the failure indicator. In the case of non-informative censoring, the probability distribution for $(t_i, \delta_i|\mathbf{X}_i)$ conditional on the covariates is given by

$$\begin{aligned} P(t_i, \delta_i|\mathbf{X}_i, \boldsymbol{\beta}) &\propto P(t_i|\mathbf{X}_i, \boldsymbol{\beta})^{\delta_i} S(t_i|\mathbf{X}_i, \boldsymbol{\beta})^{1-\delta_i} \\ &= h_i(t_i|\mathbf{X}_i, \boldsymbol{\beta})^{\delta_i} S(t_i|\mathbf{X}_i, \boldsymbol{\beta}) \\ &= [h_0(t_i) \exp(\mathbf{X}_i' \boldsymbol{\beta})]^{\delta_i} \exp\left(-\exp(\mathbf{X}_i' \boldsymbol{\beta}) H_0(t_i)\right), \end{aligned} \tag{2.36}$$

where $S(t|\mathbf{X}) = P(T > t|\mathbf{X})$ is the survival function for T and $H_0(t) = \int_0^t h_0(u)du$ is the cumulative baseline hazard function.

Thus, the log-likelihood for subject i can be written as

$$l(\boldsymbol{\beta}|t_i, \delta_i, \mathbf{X}_i) = \delta_i \left(\log h_0(t_i) + \mathbf{X}_i'\boldsymbol{\beta}\right) - \exp(\mathbf{X}_i'\boldsymbol{\beta}) H_0(t_i). \tag{2.37}$$

The parameters vector, $\boldsymbol{\beta}$, can be estimated by maximizing function (2.37). Unfortunately, this log-likelihood contains nuisance parameter $h_0(t)$ and $H_0(t)$. To avoid estimating these parameters, a consistent estimator of $\boldsymbol{\beta}$ can be obtained by maximizing the partial likelihood (PL) presented by Cox (1975) [10], which in the absence of ties can be written as

$$L(\boldsymbol{\beta}|t_i, \delta_i, \mathbf{X}_i) = \prod_{i=1}^{n} \left[\frac{\exp(\mathbf{X}_i'\boldsymbol{\beta})}{\sum_{j=1}^{n} I_{t_j \geqslant t_i} \exp(\mathbf{X}_i'\boldsymbol{\beta})} \right]^{\delta_i}. \tag{2.38}$$

Equating the partial likelihood score function to zero as the equation below yields the estimator of $\boldsymbol{\beta}$,

$$\mathbf{u}_{\mathrm{PL}}(\boldsymbol{\beta}) = \frac{\partial \log L(\boldsymbol{\beta}|t_i, \delta_i, \mathbf{X}_i)}{\partial \boldsymbol{\beta}} = 0. \tag{2.39}$$

This estimator has been shown to be consistent and asymptotically normal (Andersen and Gill (1982) [1]; Tsiatis (1981) [73]) and semi-parametrically efficient (Begun et al. (1983) [2]).

When some of the covariates are subject to missingness, we can write $\mathbf{X}_i = (\mathbf{X}_i^{obs}, \mathbf{X}_i^{mis})$. Prentice (1982) [58] suggested using $E(\exp(\mathbf{X}_i'\boldsymbol{\beta})|T \leqslant t, \mathbf{X}_i = \mathbf{X}_i^{obs})$ to replace $\exp(\mathbf{X}_i'\boldsymbol{\beta})$ in the partial likelihood function (2.38). However, this conditional expectation is not independent of the unknown baseline hazard function $h_0(t)$ so that it ruins the advantages of partial likelihood. Zhou and Pepe (1995) [79] overcame this difficulty by using an estimated partial likelihood (EPL), in which an estimated exponential term $\widehat{\exp(\mathbf{X}_i'\boldsymbol{\beta})}$ is expressed as a sum of $\exp(\mathbf{X}_i'\boldsymbol{\beta})$ (when covariates are complete) and a surrogate variable (when covariates are missing). Later $\widehat{\exp(\mathbf{X}_i'\boldsymbol{\beta})}$ is plugged in partial likelihood function (2.38) to get the estimator of $\boldsymbol{\beta}$. The EPL method can yield more efficient estimates than maximum partial likelihood estimates based on complete-case analysis, but additional surrogate variables require strict assumptions to be satisfied.

Alternatively, the information contained in the survival time (t_i, δ_i) can be represented by the counting process $\{N_i(t), Y_i(t) : t \leqslant 0\}$, where $N_i(t) = I_{\{t_i \leqslant t, \delta_i = 1\}}$ and $Y_i(t) = I_{\{t_i \geqslant t\}}$ so that $\sum N_i(t)$ is the number of death at or before time t and $\sum Y_i(t)$ is the risk set at time

t, respectively. Thus, the score function can be written as a stochastic integral in counting process notation as

$$\mathbf{u}_{\mathrm{PL}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\infty} \left[\mathbf{X}_i - \bar{\mathbf{X}}(\boldsymbol{\beta}, u)\right] dN_i(u), \tag{2.40}$$

where

$$\bar{\mathbf{X}}(\boldsymbol{\beta}, t) = \frac{\sum_{j=1}^{n} \mathbf{X}_j Y_j(t) \exp(\mathbf{X}_j' \boldsymbol{\beta})}{\sum_{j=1}^{n} Y_j(t) \exp(\mathbf{X}_j' \boldsymbol{\beta})}. \tag{2.41}$$

That is, $\bar{\mathbf{X}}(\boldsymbol{\beta}, t)$ is the average of $\mathbf{X}_i$ over the risk set $\sum Y_i(t)$, using an "exponentially weighted" form of sampling.

Lin and Ying (1993) [43] introduced an approximate partial likelihood (APL) score function to estimate Cox model with missing covariates. This approach modified the partial likelihood score function (2.40) by replacing the whole term within the integral with a complete-case estimate. This estimator depends critically on the MCAR assumption and sometimes is less efficient than analysis of complete cases.

Lipsitz and Ibrahim (1998) [46] developed a likelihood-based method to estimate the parameters of Cox model when there are only categorical missing covariates. Under MAR, a consistent estimate of $\boldsymbol{\beta}$ can be obtained by letting the conditional expectation of the whole score function (2.40) given observed data equal to 0 and solving for $\hat{\boldsymbol{\beta}}$ as

$$\mathbf{u}^*_{\mathrm{PL}}(\boldsymbol{\beta}) = E[\mathbf{u}_{\mathrm{PL}}(\boldsymbol{\beta}) | \text{observed data}] = 0. \tag{2.42}$$

Herring and Ibrahim (2001) [28] extended the methods proposed by Lipsitz and Ibrahim (1998) [46] by using the Monte Carlo EM algorithm methods to estimate the parameters of Cox model when missing covariates are categorical, continuous or mixed. This proposed methodology is very general, which allows any type of missing covariates and works under MAR, but the computation is quite intensive.

2.2.3 Weighted Estimating Equation Method

The methods that we have reviewed so far for accommodating missing covariates have all involved, directly or indirectly, integration over the distribution of the missing data. For example, this is done directly in constructing the likelihood and indirectly when using multiple imputation. All such methods require assumptions to be made about the distribution, often

in the form of its conditional distribution given the observed data, so that the results can be very sensitive to these assumptions. An alternative method, called a weighted estimating equation (WEE) based on the inverse probability weighting idea of Horvitz and Thompson (1952) [32], is very robust to the distribution of unobserved data. With WEE, the contribution to the estimating equation from a complete observation is weighted by π, the inverse probability that the covariate is observed. However, this method may generate inefficient estimators since in its basic form only the information of complete data is used.

Robins et al. (1994) [61] and Robins and Rotnitzky (1995) [60] developed a class of estimators based on inverse-probability WEEs in a regression setting when data are MAR. Lipsitz et al. (1999) [47] developed WEE methods, which are almost identical to the maximum likelihood estimating equations, for generalized linear models with missing categorical and continuous covariates. As such, they extended Monte Carlo EM algorithm to a Monte Carlo WEE to solve these weighted estimating equations. Following their notation, let y_i be the outcome variable, $\mathbf{X}_i$ be a p-dimensional vector of completely observed covariates, and covariates z_i be missing for some subjects, $i = 1, 2, \ldots, n$. Additionally, an indicator variable o_i is defined, which equals to 1 if z_i is observed and 0 if z_i is missing. Thus, the distribution of o_i given $(y_i, \mathbf{X}_i, z_i)$ is a Bernoulli with probability $\pi_i = P(o_i = 1|y_i, \mathbf{X}_i, z_i)$. If we write

$$\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i, z_i) = \frac{\partial \log P(y_i|\mathbf{X}_i, z_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \tag{2.43}$$

then the score function of WEE is

$$\mathbf{u}_{\text{WEE}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \frac{o_i}{\pi_i} \mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i, z_i) + (1 - \frac{o_i}{\pi_i}) E_{z_i|y_i, \mathbf{X}_i}[\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i, z_i)] \right\}. \tag{2.44}$$

If in equation (2.44), o_i/π_i is replaced by o_i, this estimation equation becomes maximum likelihood score function and thus the solution will be exactly MLE. So one can use a weighted-type EM algorithm to solve $\mathbf{u}_{\text{WEE}} = 0$. The WEEs will produce consistent estimates if π_i is correctly specified or the score vector for the missing data given the observed data is correctly specified, but not necessarily both [36]. This is what Robins and others often called *double robustness*.

In Robins et al. (1994) [61], they discussed the idea of adding an augmentation term to a simple WEE under MAR. Based on this idea, Wang and Chen (2001) [74] proposed an

augmented inverse probability weighted (AIPW) estimator for Cox model with missing categorical covariates, which includes terms that depend on the unknown cumulative baseline hazard function $H_0(t)$ and the conditional distribution of missing covariates given observed data. They imposed a parametric model $P(z_i|\mathbf{X}_i, \boldsymbol{\alpha})$ for this conditional distribution and suggested an EM algorithm to estimate regression coefficients $\boldsymbol{\beta}$, as well as nuisance parameters $H_0(t)$ and $\boldsymbol{\alpha}$. This AIPW estimator holds the *double robustness* property. However, the EM algorithm may not be feasible with a complex conditional distribution $P(z_i|\mathbf{X}_i, \boldsymbol{\alpha})$.

Qi et al. (2005) [59] improved Wang and Chen's AIPW estimator in two ways: they proposed a generalized nonparametric method to estimate selection probability π_i; they also proposed kernel-assisted fully augmented weighted estimators (FAWE) that, unlike Wang and Cheng's methods, do not put a parametric restriction on the association between missing covariates and the observed covariates. In this method, nonparametric kernel smoothing techniques are adopted to estimate conditional expectations that depends on the cumulative baseline hazard function $H_0(t)$ and conditional distribution of the missing covariates given the observed covariates.

The WEE method is more robust than likelihood-based method and it is more flexible in specifying the parametric conditional distribution of missing covariates. However, if multiple covariates are subject to missingness, i.e., $\mathbf{Z}_i$ is a vector instead of an univariate variable z_i, WEE performs best when all elements of $\mathbf{Z}_i$ are observed or unobserved for i, thus the selection probability π_i can be easily and correctly modeled. On the other hand, likelihood-based method can easily handle more than one missing covariates with any pattern of missingness, but it is theoretically complicated.

3.0 PROPOSED METHODOLOGIES

As mentioned in the chapter of introduction, the major problem for combining different Cox proportional hazard models to perform a meta-analysis is that estimates of parameters from different Cox models where adjustment has been made for different subsets of covariates are not comparable. As presented in the chapter of literature review, most of the meta-analysts simply ignore the fact that it is inappropriate to combine effect sizes adjusted by different covariates across studies, or some researchers found this problem (Chastang et al. (1988) [4], Ford et al. (1995) [18], and Hauck et al. (1998) [25]) but did not provide any constructive solution. Thus, in our proposed methods, we want to tackle this problem under both APD and IPD meta-analysis frameworks.

3.1 LINEAR META-REGRESSION MODELS USING APD

The majority of meta-analyses often attempt to obtain data by collecting summaries or APD from published studies, so we would look into the issues under the APD meta-analysis framework first. Since we only observe effect sizes and summarized model information from different studies, it is challenging to combine estimates of treatment effects where adjustment has been made for different subsets of covariates. It appears that many statisticians are aware that estimates of treatment effects may be biased in non-linear models if one fails to adjust on prognostic variables even when the variables are perfectly balanced across treatment groups using randomization. At first glance this statement seems counterintuitive because no such bias would occur from omitting perfectly balanced prognostic variables in the more familiar linear regression model, even thought some efficiency would be lost in estimating

the treatment effect [19]. To better understand this issue, we need to investigate effects of misspecification of Cox model by omitting important covariates in survival analysis.

3.1.1 Effects of Omitting Covariates in Cox Model for Survival Data

Suppose we observe survival time t_j for each subject in a randomized clinical trial, $j = 1, 2, \ldots, n$. Let x_{1j} be the indicator variable (with value 1 or -1) denoting treatment assignment and x_{2j} be a baseline measure of some important covariate which can be assumed related to t_j. The conclusions can be readily generalized to more complex cases. A standard Cox proportional hazard model with both treatment indicator and the baseline covariate, is given by model 1:

$$h_j(t|x_{1j}, x_{2j}) = h_0(t) \exp(\beta_1 x_{1j} + \beta_2 x_{2j}), \tag{3.1}$$

where $h_0(t)$ is an unknown baseline hazard with integral $H_0(t)$.

And we also consider model 2, where covariate x_{2j} has been dropped:

$$h_j(t|x_{1j}) = h_0(t) \exp(\beta_1^* x_{1j}). \tag{3.2}$$

In model 1 (3.1) x_{2j} is a scalar. But multivariate extensions are straightforward by denoting vector $\mathbf{X}_{2j}$, and notation $\mathbf{\Omega}$ is used for the covariance matrix of $\mathbf{X}_{2j}$. Gail et al. (1984) [19] provided an approximated estimation of the bias between β_1^* and β_1 as

$$\delta = \beta_1^* - \beta_1 \approx \frac{1}{4} \boldsymbol{\beta}_2' \mathbf{\Omega} \boldsymbol{\beta}_2 \{R(\beta_1) - R(-\beta_1)\} \tag{3.3}$$

for vector $\mathbf{X}_{2j}$, where

$$R(\eta) = 2\phi'(\eta)/\phi(\eta) + 1, \tag{3.4}$$

$$\phi(\eta) = E[H_0(t)|\eta]. \tag{3.5}$$

The bias function (3.3) illustrates how the bias may be approximated and shows that β_1^* is generally biased unless $\beta_1 = 0$ or $\boldsymbol{\beta}_2 = \mathbf{0}$. Additionally, longer length of the vector $\boldsymbol{\beta}_2$ indicates larger δ, which means omitting more important covariates would increase the bias.

From the suggestion of normal linear model, the motivation for including covariates is twofold. Besides decreasing the bias, we could expect a reduction of variance and hence

increasing the precision of estimation of treatment effect. If we consider a simpler model that the response is now a non-censored survival time from a particular exponential distribution, namely,

$$t_j \sim \text{EXP}(\lambda_j), \tag{3.6}$$

where $\lambda_j = \exp(\beta_1 x_{1j} + \beta_2 x_{2j})$ and $1/\lambda_j$ is the mean of exponential distribution. Ford et al. (1995) [18] showed that the variance of $\hat{\beta}_1$ under model 1 (3.1) can be approximated by

$$\text{var}(\hat{\beta}_1) \approx (1/\sum x_{1j}^2)/\{1 - [\text{corr}(X_1, X_2)]^2\}, \tag{3.7}$$

where $\text{corr}(X_1, X_2)$ is the sample correlation between x_{1j}'s and x_{2j}'s.

Under model 2 (3.2), the variance of $\hat{\beta}_1^*$ can be approximated by

$$\text{var}(\hat{\beta}_1^*) \approx (1/\sum x_{1j}^2). \tag{3.8}$$

It is clear that the inclusion of covariate x_{2j} can increase the variance since $[\text{corr}(X_1, X_2)]^2$ is always non-negative. However, in a randomized clinical trial, $\text{corr}(X_1, X_2)$ will be close to zero with high probability and hence the influence of adding covariates on variance of treatment effect is likely to be trivial.

To demonstrate the bias of estimated treatment effects by omitted covariates with simulation for censored exponential survival data, we generated a simulated data with a sample of 800 patients using a Cox proportional hazard model given by

$$h_j(t) = h_0(t) \exp(\beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3j} + \beta_4 x_{4j} + \beta_5 x_{5j}), \tag{3.9}$$

where the baseline failure time is generated from an exponential distribution with parameter $\lambda_0 = 0.1$. The covariates X_1 and X_4 are categorical variables and X_2, X_3 and X_5 are continuous variables. To simulate data similar to that observed in a breast cancer clinical trial, we generated X_1 from a Bernoulli distribution with $p = 0.5$ to represent a treatment indicator, X_2 as a covariate of centered age from a normal distribution with mean 0 and variance 100 and X_3 as a covariate of tumor size generated from a χ^2 distribution with 3 degrees of freedom. X_4 is a Bernoulli variable with $p = 0.6$, representing an indicator of estrogen receptor positivity and X_5 stands for number of lymph nodes involved with the breast cancer, which was generated using an exponential distribution with rate 1/3 and

rounded up to integers but was modified so that negative nodal status ($X_5 = 0$) would occur in roughly 60% of the patients. To generate the survival times, the coefficients associateed with X_1-X_5 were set to

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)' = (-0.2, -0.1,\ 0.1, -0.15,\ 0.7)'$$

based on values similar to those observed in breast cancer patients. Finally, the survival times were randomly censored with probability 0.1.

We define five different models, each of which includes the treatment indicator X_1. The first model omits all other covariates. The second one omits all but one of the other covariates. In Models 3-5, we continue to randomly add covariates one at a time so that Model 5 is correctly specified. Specifically, covariates included in the Cox models are

Model 1: X_1,

Model 2: $X_1 + X_a$,

Model 3: $X_1 + X_a + X_b$,

Model 4: $X_1 + X_a + X_b + X_c$,

Model 5: $X_1 + X_2 + X_3 + X_4 + X_5$,

where indices a, b and c are randomly chosen without replacement from $\{2, 3, 4, 5\}$ so that $a \neq b \neq c$. Each simulated survival data was fitted with all five models and the estimated treatment effects were recorded.

Figure 3.1 compares the estimated treatment effects $\hat{\beta}_1$'s from different Cox models with all possible model specifications with 200 simulations. When models only have X_1, they yield the largest bias, whereas adding more covariates into the Cox model incrementally decreases the bias, so that the realizations of model with all of the five covariates have a median treatment effect almost identical to the true value -0.2. Hence we illustrate that for Cox models with censored data, estimates of treatment effect with other covariates omitted will be biased toward zero (hazard ratio toward to one), and the more the omitted covariates, the larger the bias.

Figure 3.2 confirms the comparison between equation (3.7) and equation (3.8) that including more covariates into Cox model could slightly increase the variance of treatment

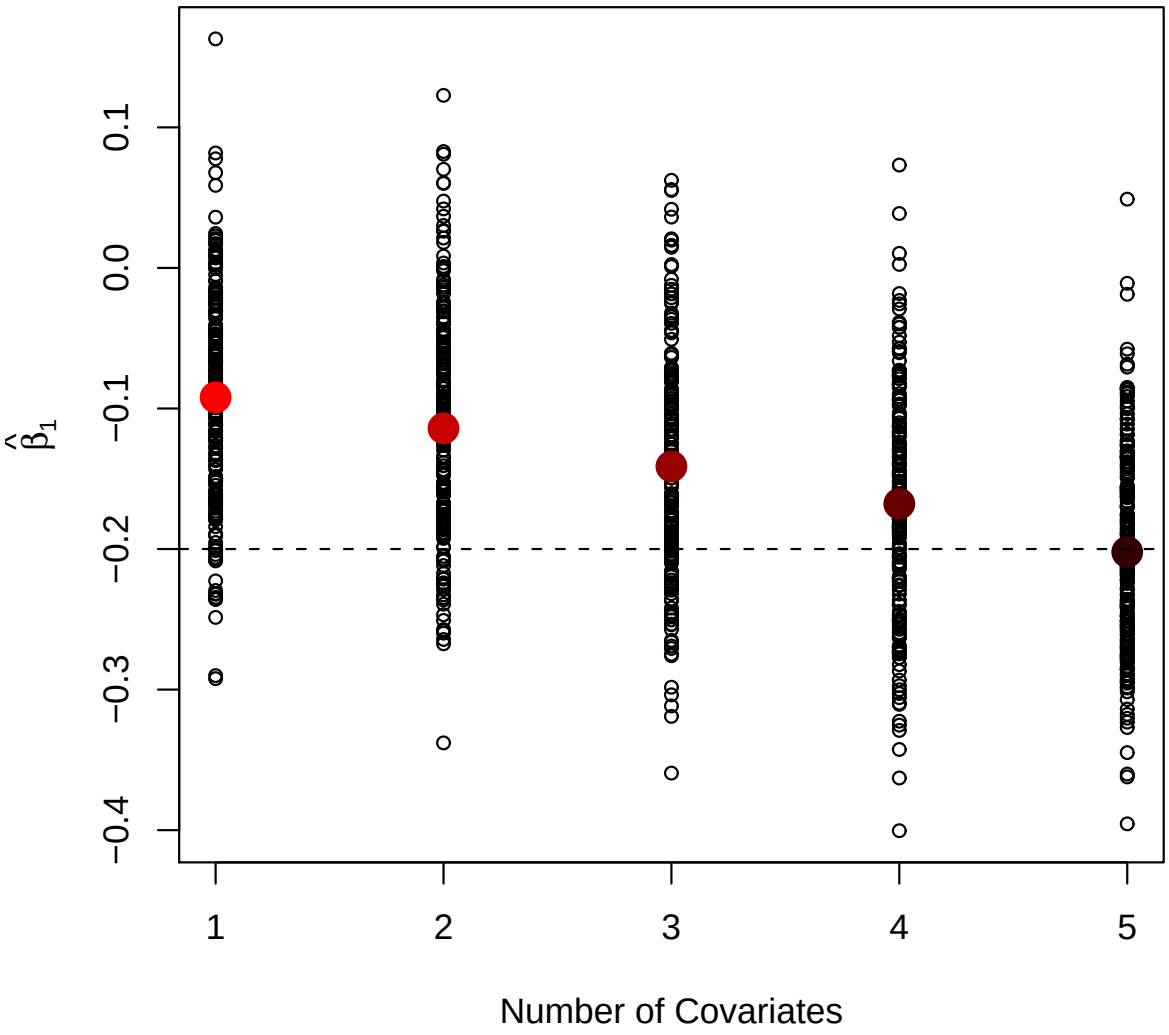

Figure 3.1: Estimated treatment effects from Cox proportional hazard models with different numbers of covariates.

The x-axis indicates the number of covariates in different Cox models with possible misspecification. Treatment indicator X_1 is included in all the models. Other covariates $X_2, \ldots, X_5$ are randomly selected into each model according to the pre-specified number of covariates. For each model, there are 200 estimated $\hat{\beta}_1$s and the dot stands for the median value. -0.2 is the true coefficient of treatment indicator.

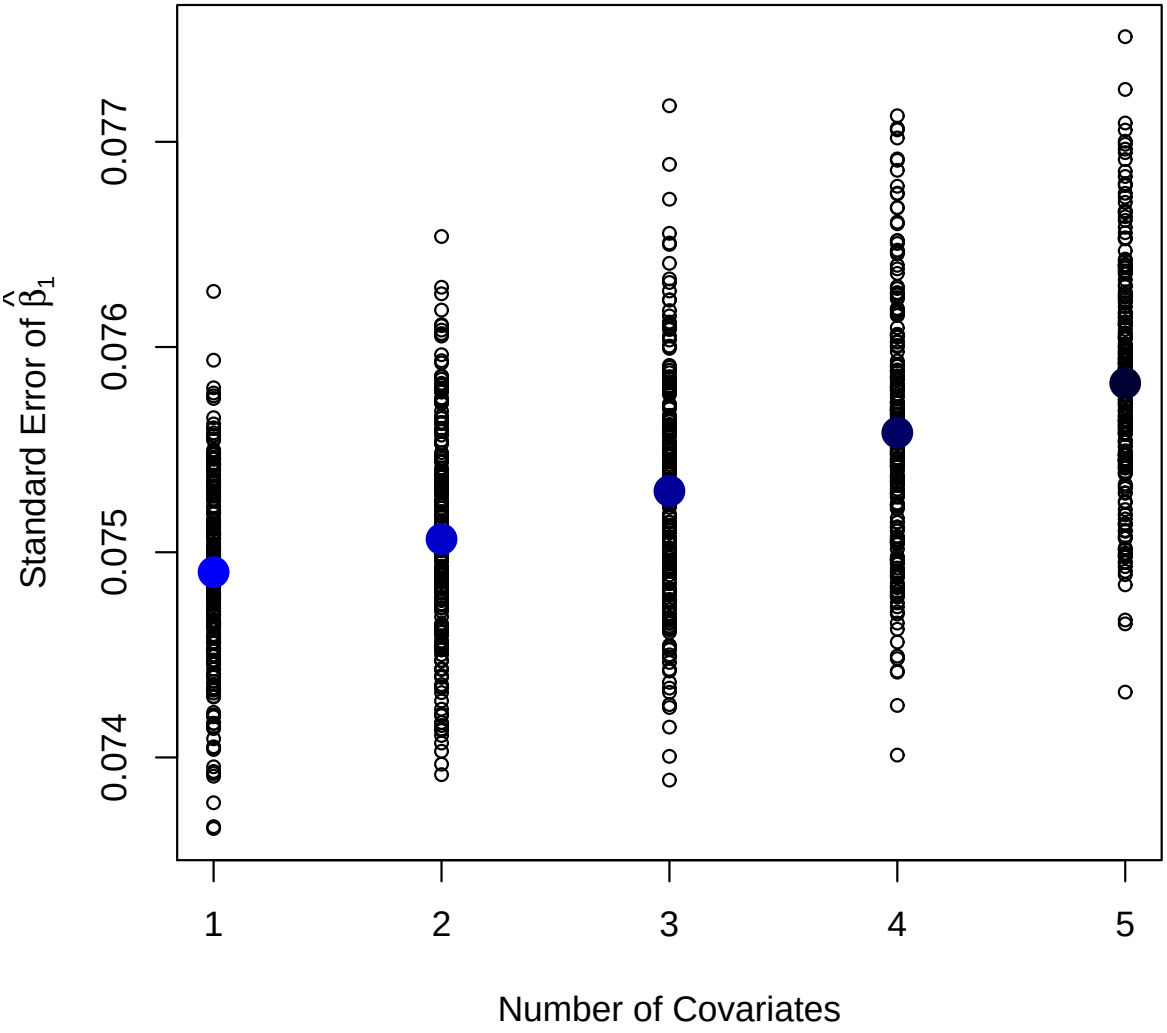

Figure 3.2: Standard errors of estimated treatment effects from Cox proportional hazard models with different numbers of covariates.

The x-axis indicates the number of covariates in different Cox models with possible misspecification. Treatment indicator X_1 is always in all the models. Other covariates $X_2, \ldots, X_5$ are randomly selected into each model according to the pre-specified number of covariates. For each model, there are 200 standard errors $\text{se}(\hat{\beta}_1)$ and the blue dot stands for the median value.

effect estimation. However, the differences are negligibly small. From this simulation, the difference of median standard errors between Model 1 and Model 5 is as small as 0.001.

3.1.2 Proposed Meta-Regression Models

Generally speaking, the basic principle of any research synthesis or meta-analysis is only including related studies with similar study designs and focusing on the same research topic. Otherwise, there exists the "Apples-and-Oranges" problem that Gene Glass talked in his unpublished paper [21], in which he stated that the repeated criticism of meta-analysis was that it was meaningless because it "mixed apples and oranges". Obviously, there are a lot sources of heterogeneity in treatment effects. Different clinical trials may use different treatment protocols and have different study designs. Additionally, patients have baseline covariates such as demographic variables or clinical or pathologic variables which may vary substantially across studies, which may also affect their outcomes. In addition, there is selection bias, publication bias, data irregularities and other debatable fundamental issues with meta-analysis. These problems can generally be avoided if a few basic principles are observed, such as a prior definition of inclusion criteria for studies and a comprehensive trial search strategy. Similar to issues regarding the design of clinical trials, meta-analyses need to be carefully planned with detailed written protocols being prepared in advance. We do not address topics of the selection process or publication bias associated with meta-analyses in this dissertation, but rather assume all studies share certain common characteristics and that a "full" Cox model is globally best for all studies. Here the "full" Cox model is chosen from one or more studies where all important covariates are measured and included, and designate other studies as having some covariates omitted.

We propose two meta-regression models: a meta-ANOVA model and a meta-polynomial model. These two models share the same general form,

$$\theta_i = \mathbf{Z}_i'\boldsymbol{\alpha} + a_i + \varepsilon_i, \quad E(\theta_i) = \mathbf{Z}_i'\boldsymbol{\alpha}, \quad \text{var}(\theta_i) = \sigma_i^2 + \tau^2, \tag{3.10}$$

where θ_i is the observed study-specific treatment effect, $\mathbf{Z}_i = (1, z_1, z_2, \ldots, z_m)'$ is the $(1 + m) \times 1$ vector of covariates measured in study i including the constant term, a_i is the random

effect with mean zero and unknown variance τ^2, and ε_i is an error term with mean zero and study-specific variance σ_i^2.

Suppose we have k studies included in a meta-analysis where we wish to potentially adjust for up to L covariates via Cox regression models. Among the L covariates, X_1 will be designated as the treatment indicator and is included in all studies. However, some studies may not have all of the other $L-1$ covariates. In an APD meta-analysis, each study i provides a value of $\hat{\beta}_{i1}$ (denoted as θ_i) and $\text{var}(\hat{\beta}_{i1})$ (denoted as σ_i^2). Moreover, we can observe other model information such as how many covariates are included in the Cox models and which covariates are included.

Meta-ANOVA Model

Based on the Cox model information for each study i, we can create L indicators $\{z_{il}\} = \{I_{\{X_l \text{ is included in study } i\}}, l = 1, 2, \ldots, L\}$. The indicator matrix $\mathbf{Z} = (z_{il})_{k \times L}$ can be written as

$$\mathbf{Z} = \begin{bmatrix} 1 & z_{12} & z_{13} & \ldots & z_{1L} \\ 1 & z_{22} & z_{23} & \ldots & z_{2L} \\ 1 & z_{32} & z_{33} & \ldots & z_{3L} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & z_{k2} & z_{k3} & \ldots & z_{kL} \end{bmatrix}, \tag{3.11}$$

and each row of $\mathbf{Z}$ is the vector $\mathbf{Z}_i$ for study i.

Thus, the meta-ANOVA model can be written as

$$\theta_i = \alpha_0 + \alpha_1 z_{i2} + \alpha_2 z_{i3} + \ldots + \alpha_{L-1} z_{iL} + a_i + \varepsilon_i, \tag{3.12}$$

where the intercept term α_0 is the common treatment effect since treatment variable X_1 is always in the models, and α_{l-1} is the incremental effect when covariate X_l is included in the Cox model, $l = 2, 3, \ldots, L$.

Assuming that only the θ_is from the models with a full set of covariates are not biased, our proposed estimator for the pooled treatment effect based on the meta-ANOVA model

is the sum of the estimated common treatment effect $\hat{\alpha}_0$ and the offsets due to addition of covariates $X_2, X_3, \ldots, X_L$:

$$\begin{aligned} \hat{\theta}_{\text{M-A}} &= \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \ldots + \hat{\alpha}_{L-1} \\ &= [\ 1\ 1\ 1 \ldots\ 1\ ] \times \hat{\boldsymbol{\alpha}}. \\ &= \mathbf{1}'_L \times \hat{\boldsymbol{\alpha}}. \end{aligned} \tag{3.13}$$

The variance of $\hat{\theta}_{\text{M-A}}$ is given by

$$\text{var}(\hat{\theta}_{\text{M-A}}) = \mathbf{1}'_L \text{var}(\hat{\boldsymbol{\alpha}}) \mathbf{1}_L. \tag{3.14}$$

Meta-polynomial Model

It is well established that for Cox proportional hazard models, the more important covariates omitted, the larger the bias in treatment effect estimation [25]. In Figure 3.1, we observe a clear linear relationship between the estimated treatment effects θ_is and the number of covariates included in the Cox models. Accordingly, we may assign the number of covariates in the Cox model as a score for each study and treat the score as a study-level covariate z_i in the meta-regression model (3.10), $z_i = 1, 2, \ldots, L$. However, different covariates, when omitted from the Cox model, will have different effects on the treatment effect estimation. If two studies have different subsets but the same number of covariates included in the Cox models, it is problematic to assign the same score to both studies because they are still incomparable. We wish to define a quantitative variable which could reflect the effects of different subsets of covariates on the estimation of treatment effect. It is possible that we could estimate these different effects and put that information into the meta-regression models to adjust the estimation bias. However, as suggested by Gail et al. (1984) [19], it requires individual patient data information such as the variance of the omitted covariates to estimate the bias of treatment effect, which is not available in an APD meta-analysis setting. Thus, we propose a nonparametric method to rank the covariates and assign scores to them. Our proposed scoring system is constructed as follows:

1. Assign score s_l to covariate X_l, $l = 1, 2, \ldots, L$ so that $s_1 = 1$ and $\sum_{l=2}^{L} s_l = (L-1)$;

2. If the rank of $(X_2, X_3, \ldots, X_L)$ is $(r_2, r_3, \ldots, r_L)$, where higher rank indicates larger potential bias of estimated treatment effect when that covariate is omitted from the Cox model, then $s_2 : s_3 : \ldots : s_L = r_2 : r_3 : \ldots : r_L$.

We can rank $(X_2, X_3, \ldots, X_L)$ based on the absolute values of estimated coefficients $(|\hat{\alpha}_1|, |\hat{\alpha}_2|, \ldots, |\hat{\alpha}_{L-1}|)$ from the meta-ANOVA model (3.12). Here, $|\hat{\alpha}_{l-1}|$ is the estimated incremental effect when covariate X_l is included in the Cox model, $l = 2, 3, \ldots, L$. Additionally, the ranking should be carefully chosen to be biologically or clinically meaningful in case where two coefficients have very similar values. Furthermore, we use $z_i = \sum_{l=1}^{L} (I_{\{X_l \text{ is included in study } i\}} \times s_l)$ to denote the score for each study. Hence, z_i is the "study-level" covariate which has lowest value 1 if only treatment indicator X_1 is in the Cox model and has highest value L if the Cox model includes all L covariates. Additionally, studies with the same subsets of covariates share the same value of score z_i, and the higher the score the smaller the bias in treatment effect.

Similar to the conditional linear model proposed by Wu and Bailey (1989) [76], assuming treatment effect θ_i given the score z_i is a polynomial function of z_i, we propose a meta-polynomial model as

$$\begin{aligned} (\theta_i | z_i) &= \mathbf{Z}_i' \boldsymbol{\alpha} + a_i + \varepsilon_i \\ &= \sum_{j=0}^{n} \alpha_j (z_i)^j + a_i + \varepsilon_i, \end{aligned} \tag{3.15}$$

where $\mathbf{Z}_i = (1, z_i, z_i^2, \ldots, z_i^n)'$ for each study.

Explicitly, this conditional polynomial model (3.15) can be written as

$$\begin{bmatrix} \theta_1 \mid z_1 \\ \theta_2 \mid z_2 \\ \theta_3 \mid z_3 \\ \vdots \\ \theta_k \mid z_k \end{bmatrix} = \begin{bmatrix} 1 & z_1 & z_1^2 & \ldots & z_1^n \\ 1 & z_2 & z_2^2 & \ldots & z_2^n \\ 1 & z_3 & z_3^2 & \ldots & z_3^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & z_k & z_k^2 & \ldots & z_k^n \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_k \end{bmatrix}. \tag{3.16}$$

If $z_{\max} = \max\{z_i\}$ is the score of the Cox model with a full set of covariates, our proposed pooled treatment effect estimator based on the meta-polynomial model is

$$\begin{aligned}\hat{\theta}_{\text{M-P}} &= \sum_{j=0}^{n} \hat{\alpha}_j (z_{\max})^j \\ &= [\ 1\ z_{\max}\ z_{\max}^2\ \ldots\ z_{\max}^n\ ] \times \hat{\boldsymbol{\alpha}} \\ &= \mathbf{Z}'_{\max} \times \hat{\boldsymbol{\alpha}}. \end{aligned} \tag{3.17}$$

The variance of $\hat{\theta}_{\text{M-P}}$ is given by

$$\text{var}(\hat{\theta}_{\text{M-P}}) = \mathbf{Z}'_{\max} \text{var}(\hat{\boldsymbol{\alpha}}) \mathbf{Z}_{\max}. \tag{3.18}$$

To better demonstrate our proposed models, suppose that we pooled five studies, generated from the same data simulation scheme as what we did in Section (3.1.1), to form a meta-analysis. We arbitrarily removed different covariates from different studies: Study 1 contains only covariate X_1; Study 2 contains both X_1 and X_2; Study 3 contains X_1, X_2 and X_3; Study 4 contains X_1, X_4 and X_5; and Study 5 contains all five covariates. Thus, Study 5 has the "full" Cox model and Study 1-4 have some covariates omitted. Among the five covariates, X_1 is the treatment indicator of primary interest and it is always included in all models. The specifications of the Cox models for the five studies are:

Study 1: $h_{1j}(t) = h_0(t) \exp(\beta_{11} x_{11j})$, $j = 1, 2, \ldots, n_1$,

Study 2: $h_{2j}(t) = h_0(t) \exp(\beta_{21} x_{21j} + \beta_{22} x_{22j})$, $j = 1, 2, \ldots, n_2$,

Study 3: $h_{3j}(t) = h_0(t) \exp(\beta_{31} x_{31j} + \beta_{32} x_{32j} + \beta_{33} x_{33j})$, $j = 1, 2, \ldots, n_3$,

Study 4: $h_{4j}(t) = h_0(t) \exp(\beta_{41} x_{41j} + \beta_{44} x_{44j} + \beta_{45} x_{45j})$, $j = 1, 2, \ldots, n_4$,

Study 5: $h_{5j}(t) = h_0(t) \exp(\beta_{51} x_{51j} + \beta_{52} x_{52j} + \beta_{53} x_{53j} + \beta_{54} x_{54j} + \beta_{55} x_{55j})$, $j = 1, 2, \ldots, n_5$,

where x_{ilj} is the value of variable X_l for subject j in study i.

In this example, each study i provides a value of $\hat{\beta}_{i1}$ (denoted as θ_i) and $\text{var}(\hat{\beta}_{i1})$ (denoted as σ_i^2). To combine the k studies to come up with an overall treatment effect estimation, the meta-ANOVA model can be written as

$$\theta_i = \alpha_0 + \alpha_1 \times z_{i2} + \alpha_2 \times z_{i3} + \alpha_3 \times z_{i4} + \alpha_4 \times z_{i5} + a_i + \varepsilon_i. \tag{3.19}$$

Explicitly, equation (3.19) has the form in matrix notation as following

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}, \tag{3.20}$$

where the intercept term α_0 is the common treatment effect since treatment variable X_1 is always in the models, and α_{l-1} is the incremental effect when covariate X_l is included in the Cox model, $l = 2, 3, 4, 5$.

Since each z_{il} in the four-way meta-ANOVA model (3.19) is a two-level predictor, $l = 2, 3, 4, 5$, there is a total of $2^4 = 16$ categories of estimated treatment effects θ_is. We re-do Figure 3.1 to plot estimated treatment effects over the 16 categories in Figure 3.3.

From our assumption, only θ_is from the "full" models are not biased. Thus our proposed estimator for the pooled treatment effect bases on the meta-ANOVA model is

$$\begin{aligned} \theta_{\text{M-A}} &= \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 + \hat{\alpha}_4 \\ &= [\ 1\ 1\ 1\ 1\ 1\ ] \times \hat{\boldsymbol{\alpha}}, \\ &= \mathbf{1}'_{(5)} \times \hat{\boldsymbol{\alpha}}. \end{aligned} \tag{3.21}$$

The variance of $\theta_{\text{M-A}}$ is given by

$$\text{var}(\theta_{\text{M-A}}) = \mathbf{1}'_{(5)} \text{var}(\hat{\boldsymbol{\alpha}}) \mathbf{1}_{(5)}. \tag{3.22}$$

We can rank (X_2, X_3, X_4, X_5) based on the absolute values of estimated coefficients $(|\hat{\alpha}_1|, |\hat{\alpha}_2|, |\hat{\alpha}_3|, |\hat{\alpha}_4|)$ from the four-way meta-ANOVA model (3.19). In our example, we rank (X_2(Age), X_3(Tumor Size), X_4(ER-Positive), X_5(Number of Nodes)) as $(3, 1, 2, 4)$, then the scores for the five covariates are $(1, 1.2, 0.4, 0.8, 1.6)$. Furthermore, we use $z_i = \sum_{l=1}^{5} (I_{\{X_l \text{ is included in study } i\}} \times s_l)$ to denote the score for each study. Figure 3.4 illustrates that there is a strong linear relationship between the estimated treatment effects and scores.

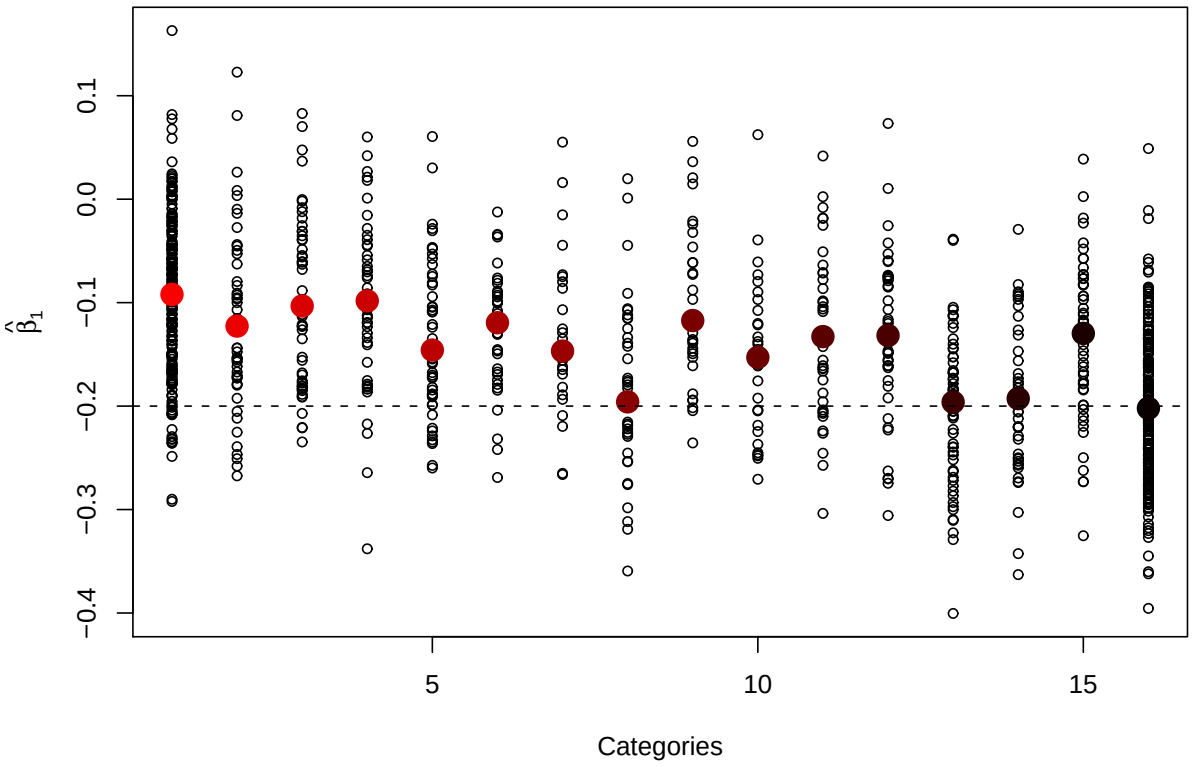

Figure 3.3: Estimated treatment effects from Cox proportional hazard models with different categories.

The x-axis indicates the categories of the four-way meta-ANOVA model. Category 1 only includes treatment indicator X_1 and Category 16 includes all five covariates. Other categories indicate different combinations of covariates $X_1, \ldots, X_5$ but X_1 is always included. The red dot stands for the median value for each category. -0.2 is the true coefficient of treatment indicator.

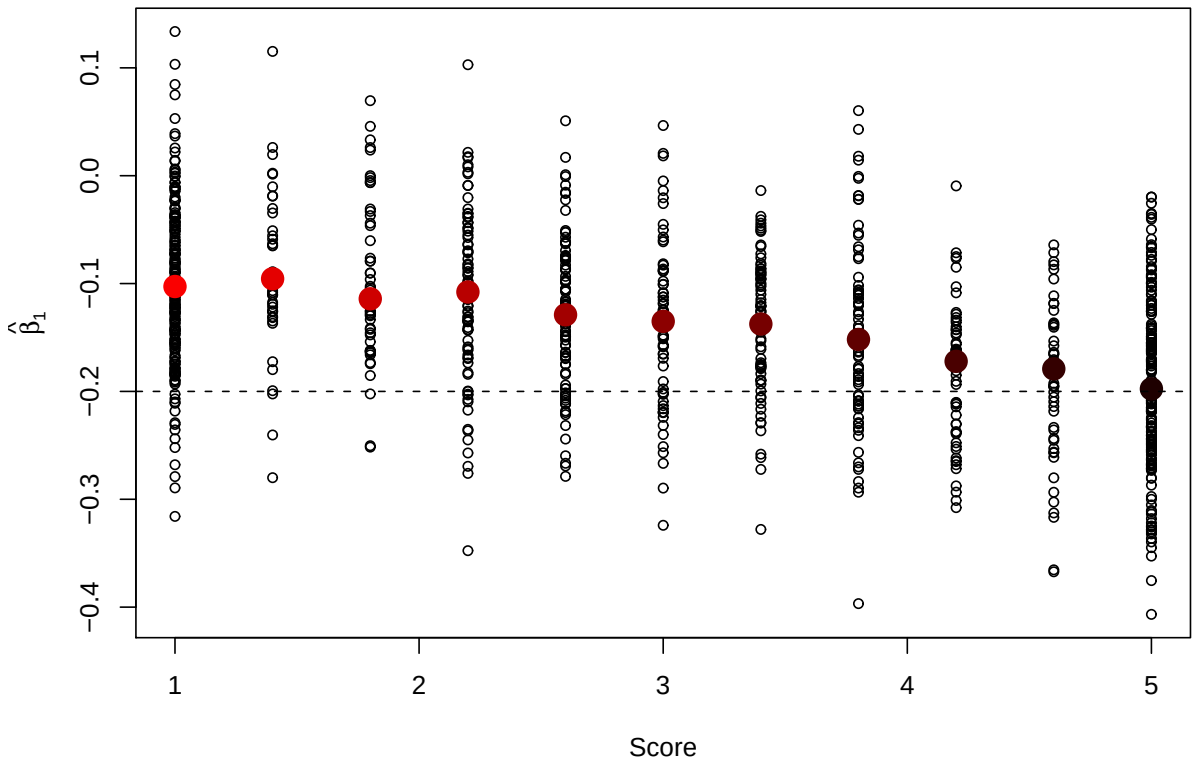

Figure 3.4: Estimated treatment effects from Cox proportional hazard models with different scores.

The x-axis indicates the score for each study of the proposed meta-polynomial model. Score = 1 indicates only treatment indicator X_1 is included. Score = 5 indicates all five covariates are included. The red dot stands for the median value for each score. -0.2 is the true coefficient of treatment indicator.

Thus, the conditional polynomial model (3.15) for our example can be written as

$$\begin{bmatrix} \theta_1 \mid 1.0 \\ \theta_2 \mid 2.2 \\ \theta_3 \mid 2.6 \\ \theta_4 \mid 3.4 \\ \theta_5 \mid 5.0 \end{bmatrix} = \begin{bmatrix} 1 & 1.0 & 1.0^2 & \dots & 1.0^n \\ 1 & 2.2 & 2.2^2 & \dots & 2.2^n \\ 1 & 2.6 & 2.6^2 & \dots & 2.6^n \\ 1 & 3.4 & 3.4^2 & \dots & 3.4^n \\ 1 & 5.0 & 5.0^2 & \dots & 5.0^n \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}. \tag{3.23}$$

If $z_{\max} = 5.0$ is the score of the "full" Cox model, our proposed pooled treatment effect estimator based on the meta-polynomial model is

$$\begin{aligned} \theta_{\text{M-P}} &= \sum_{j=0}^{n} \hat{\alpha}_j (5.0)^j \\ &= [\ 1\ 5.0\ 5.0^2\ \dots\ 5.0^n\ ] \times \hat{\boldsymbol{\alpha}}. \end{aligned} \tag{3.24}$$

The variance of $\theta_{\text{M-P}}$ is given by

$$\text{var}(\theta_{\text{M-P}}) = [\ 1\ 5.0\ 5.0^2\ \dots\ 5.0^n\ ] \times \text{var}(\hat{\boldsymbol{\alpha}}) \times [\ 1\ 5.0\ 5.0^2\ \dots\ 5.0^n\ ]'. \tag{3.25}$$

From the previous Section (3.1.1), we demonstrated that the estimated treatment effect θ_i is highly dependent on other covariates included in the Cox model. Then for both meta-ANOVA and meta-polynomial models, $\mathbf{Z}_i$ is the study level characteristic vector which helps to explain the heterogeneity of treatment effects across studies. Since these two models share the same general form as equation (3.10), we demonstrate how both parametric maximum likelihood and nonparametric quadratic method of moments estimations are calculated for $\boldsymbol{\alpha}$ and the heterogeneity variance τ^2 under the general form of notation.

Maximum Likelihood Estimation

If the assumption of normality of model (3.10) is valid, the log-likelihood is written as

$$l(\boldsymbol{\alpha}, \tau^2) \propto -\frac{1}{2} \sum_{i=1}^{k} \left[\ln(\sigma_i^2 + \tau^2) + \frac{(\theta_i - \mathbf{Z}_i' \boldsymbol{\alpha})^2}{\sigma_i^2 + \tau^2} \right]. \tag{3.26}$$

To use iterative algorithm to get the maximum likelihood estimation (MLE), we need the first and second derivatives with respect to $\boldsymbol{\alpha}$ and τ^2:

$$\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\alpha}} &= \sum_{i=1}^{k} \frac{(\theta_i - \mathbf{Z}_i'\boldsymbol{\alpha})\mathbf{Z}_i}{\sigma_i^2 + \tau^2}, \quad \frac{\partial l}{\partial \tau^2} = -\frac{1}{2}\sum_{i=1}^{k}\left[\frac{1}{\sigma_i^2+\tau^2} - \frac{(\theta_i - \mathbf{Z}_i'\boldsymbol{\alpha})^2}{(\sigma_i^2+\tau^2)^2}\right], \\
\frac{\partial^2 l}{\partial \boldsymbol{\alpha}^2} &= -\sum_{i=1}^{k} \frac{\mathbf{Z}_i\mathbf{Z}_i'}{\sigma_i^2 + \tau^2} = \mathbf{H}_{11}, \quad \frac{\partial^2 l}{\partial (\tau^2)^2} = \frac{1}{2}\sum_{i=1}^{k}\left[\frac{1}{(\sigma_i^2+\tau^2)^2} - \frac{2(\theta_i - \mathbf{Z}_i'\boldsymbol{\alpha})^2}{(\sigma_i^2+\tau^2)^3}\right] = H_{22}, \\
\frac{\partial^2 l}{\partial \boldsymbol{\alpha}\partial \tau^2} &= -\sum_{i=1}^{k} \frac{(\theta_i - \mathbf{Z}_i'\boldsymbol{\alpha})\mathbf{Z}_i}{(\sigma_i^2 + \tau^2)^2} = \mathbf{H}_{12}, \\
\frac{\partial^2 l}{\partial \tau^2\partial \boldsymbol{\alpha}} &= -\sum_{i=1}^{k} \frac{\mathbf{Z}_i'(\theta_i - \mathbf{Z}_i'\boldsymbol{\alpha})}{(\sigma_i^2 + \tau^2)^2} = \mathbf{H}_{21}.
\end{aligned}$$

And the Hessian matrix for $l(\boldsymbol{\alpha}, \tau^2)$ is

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & H_{22} \end{bmatrix}.$$

The information matrix is the negative of the expected Hessian matrix

$$\boldsymbol{\mathcal{I}} = -E(\mathbf{H}) = \begin{bmatrix} \sum_{i=1}^{k} \frac{\mathbf{Z}_i\mathbf{Z}_i'}{\sigma_i^2+\tau^2} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\sum_{i=1}^{k} \frac{1}{(\sigma_i^2+\tau^2)^2} \end{bmatrix}.$$

The Fisher Scoring (FS) algorithm leads to the MLE of $\boldsymbol{\alpha}$ as

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_{s+1} &= \hat{\boldsymbol{\alpha}}_s + \left(\sum_{i=1}^{k} \frac{\mathbf{Z}_i\mathbf{Z}_i'}{\sigma_i^2 + \hat{\tau}_s^2}\right)^{-1} \sum_{i=1}^{k} \frac{(\theta_i - \mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_s)\mathbf{Z}_i}{\sigma_i^2 + \hat{\tau}_s^2} \\
&= \left(\sum_{i=1}^{k} \frac{\mathbf{Z}_i\mathbf{Z}_i'}{\sigma_i^2 + \hat{\tau}_s^2}\right)^{-1} \sum_{i=1}^{k} \left(\frac{\mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_s\mathbf{Z}_i}{\sigma_i^2 + \hat{\tau}_s^2} + \frac{(\theta_i - \mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_s)\mathbf{Z}_i}{\sigma_i^2 + \hat{\tau}_s^2}\right) \\
&= \left(\sum_{i=1}^{k} \frac{\mathbf{Z}_i\mathbf{Z}_i'}{\sigma_i^2 + \hat{\tau}_s^2}\right)^{-1} \sum_{i=1}^{k} \frac{\mathbf{Z}_i\theta_i}{\sigma_i^2 + \hat{\tau}_s^2} \qquad (3.27)
\end{aligned}$$

with a good starting point $\hat{\boldsymbol{\alpha}}_0 = \left(\sum_{i=1}^{k} \sigma_i^{-2}\mathbf{Z}_i\mathbf{Z}_i'\right)^{-1}\left(\sum_{i=1}^{k} \sigma_i^{-2}\mathbf{Z}_i\theta_i\right)$, where s is the iteration index.

Then the estimation of heterogeneity variance τ^2 is the key in our random effect meta-regression models. It is well known that MLE of variances is biased for finite samples. Thus

we prefer restricted MLE for our analysis. The restricted log-likelihood of model (3.10) is written as

$$l_R(\boldsymbol{\alpha}, \tau^2) \propto -\frac{1}{2}\left\{\sum_{i=1}^{k}\left[\ln(\sigma_i^2+\tau^2)+\frac{(\theta_i-\mathbf{Z}_i'\boldsymbol{\alpha})^2}{\sigma_i^2+\tau^2}\right]+\ln\left|\sum_{i=1}^{k}\frac{\mathbf{Z}_i\mathbf{Z}_i'}{\sigma_i^2+\tau^2}\right|\right\}. \tag{3.28}$$

From matrix calculus we know that $\partial(\ln(\det(\mathbf{M}))) = tr(\mathbf{M}^{-1}\partial\mathbf{M})$. Then we have the first derivative with respect to τ^2 as

$$\frac{\partial l_R}{\partial \tau^2} = -\frac{1}{2}\left\{\sum_{i=1}^{k}\left[\frac{1}{\sigma_i^2+\tau^2}-\frac{(\theta_i-\mathbf{Z}_i'\boldsymbol{\alpha})^2}{(\sigma_i^2+\tau^2)^2}\right]-D(\tau^2)\right\}, \tag{3.29}$$

where

$$D(\tau^2) = tr\left[\left(\sum_{i=1}^{k}\frac{\mathbf{Z}_i\mathbf{Z}_i'}{\sigma_i^2+\tau^2}\right)^{-1}\left(\sum_{i=1}^{k}\frac{\mathbf{Z}_i\mathbf{Z}_i'}{(\sigma_i^2+\tau^2)^2}\right)\right]. \tag{3.30}$$

As Demidenko (2004) [11] pointed out that the information matrices of ML and REML asymptotically coincide, the FS algorithm for the REML estimation of τ^2 is

$$\begin{aligned}
\hat{\tau}_{s+1}^2 &= \hat{\tau}_s^2 + \left(\sum_{i=1}^{k}\frac{1}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right)^{-1}\sum_{i=1}^{k}\left(\frac{(\theta_i-\mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_s)^2}{(\sigma_i^2+\hat{\tau}_s^2)^2}-\frac{1}{\sigma_i^2+\hat{\tau}_s^2}+D(\hat{\tau}_s)^2\right)\\
&= \left(\sum_{i=1}^{k}\frac{1}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right)^{-1}\sum_{i=1}^{k}\left(\frac{\hat{\tau}_s^2}{(\sigma_i^2+\hat{\tau}_s^2)^2}+\frac{(\theta_i-\mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_s)^2}{(\sigma_i^2+\hat{\tau}_s^2)^2}-\frac{1}{\sigma_i^2+\hat{\tau}_s^2}+D(\hat{\tau}_s)^2\right)\\
&= \left(\sum_{i=1}^{k}\frac{1}{(\sigma_i^2+\hat{\tau}_s^2)^2}\right)^{-1}\sum_{i=1}^{k}\left(\frac{(\theta_i-\mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_s)^2-\sigma_i^2}{(\sigma_i^2+\hat{\tau}_s^2)^2}+D(\hat{\tau}_s)^2\right),
\end{aligned} \tag{3.31}$$

where s is the iteration index and a good starting point $\hat{\tau}_0^2 = 0$.

Quadratic Method of Moments Estimation

The use of likelihood methods assumes model (3.10) following a normal distribution. However, a quadratic method of moments estimation relaxes this assumption and therefore is free of distribution. The weighted sum of squares Q-statistic for heterogeneity test is

$$Q = \sum_{i=1}^{k}\sigma_i^{-2}(\theta_i-\mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_0)^2, \tag{3.32}$$

where $\hat{\boldsymbol{\alpha}}_0$ is estimated as the starting point in (3.27).

To calculate the expectation of Q, we write the general meta-regression model (3.10) in matrix notation as

$$\boldsymbol{\theta} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{a} + \boldsymbol{\varepsilon}, \quad E(\mathbf{a}+\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \text{var}(\mathbf{a}+\boldsymbol{\varepsilon}) = \mathbf{V} = \tau^2\mathbf{I} + \boldsymbol{\Sigma}, \tag{3.33}$$

where $\mathbf{Z}$ is a $k \times (m+1)$ design matrix, k is the number of studies and m is the number of covariates included in the meta-regression model, and $\boldsymbol{\Sigma} = diag(\sigma_1^2, \sigma_1^2, \ldots, \sigma_k^2)$.

From Seber and Lee (2003) [69], if assume vector $\mathbf{x}$ with $E(\mathbf{x}) = \mathbf{c}$ and $\text{var}(\mathbf{x}) = \boldsymbol{\Sigma}$, $\mathbf{A}$ is a symmetric matrix, then

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \mathbf{c}'\mathbf{A}\mathbf{c}. \tag{3.34}$$

Thus, the expectation of Q is

$$\begin{aligned}
E(Q) &= E\left(\sum_{i=1}^{k} \sigma_i^{-2}(\theta_i - \mathbf{Z}_i'\hat{\boldsymbol{\alpha}}_0)^2\right) \\
&= E\left[\boldsymbol{\theta}'\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\theta}\right] \\
&= tr\left[\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\right)(\tau^2\mathbf{I} + \boldsymbol{\Sigma})\right] \\
&\quad +(\mathbf{Z}\boldsymbol{\alpha})'\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\right)\mathbf{Z}\boldsymbol{\alpha} \\
&= tr\left[\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\right)(\tau^2\mathbf{I} + \boldsymbol{\Sigma})\right] \\
&= \tau^2 tr\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\right) + k - m \\
&= \tau^2\left(\sum_{i=1}^{k}\sigma_i^{-2} - D(0)\right) + k - m.
\end{aligned} \tag{3.35}$$

If we equate the empirical sum Q to its expected value, the weighted method of moments estimator of τ^2 has the form

$$\hat{\tau}^2_{WMM} = \frac{Q - (k-m)}{\sum_{i=1}^{k}\sigma_i^{-2} - D(0)}, \tag{3.36}$$

where $D(0)$ is defined as in (3.30) when $\tau^2 = 0$.

However, the estimates may be negative when $Q < k - m$. Accordingly, we truncate the weighted method of moments estimate as

$$\tilde{\tau}^2_{WMM} = \max(0, \hat{\tau}^2_{WMM}). \tag{3.37}$$

The problem with this estimator is that it is slightly positively biased. This bias is often observed for small between-study heterogeneity.

To test the heterogeneity of treatment effects between studies, under the null hypothesis $H_0 : \tau^2 = 0$ we have

$$Q \sim \chi^2(k - m - 1), \tag{3.38}$$

where k is the number of studies and m is the number of covariates in the meta-regression model.

3.2 HIERARCHICAL MULTIPLE IMPUTATION

As mentioned in the chapter of literature review, missing covariates is problematic for IPD meta-analysis using Cox model. However, not only the frequency of missing data but also the missingness process may vary across studies from which individual data are pooled [40]. Thus current approaches of handling missing covariates for a single study may, therefore, not be directly implemented for IPD meta-analysis. Furthermore, we mainly focus on another type of "missing" covariates may occur in the pooled data set: some covariates might not be observed at all in a specific study included in the IPD meta-analysis. That is to say, a set of covariates is always missing for some studies in the pooled data set by design. And within each study, we assume there is no missing data problem. To address this specific type of missing covariates problem in IPD meta-analysis, we propose a hierarchical multiple imputation (HMI) method.

As we reviewed at Section (2.1.2), we can pool IPD from k studies and n_i subjects per study together to form a "mega-trial" so that a single analysis can be perform, $i = 1, 2, \ldots, k$. The total sample size of IPD meta-analysis is thus $N = \sum_{i=1}^{k} n_i$. Also assume we observe a p-dimensional vector of covariates $\mathbf{X}_{ij}$ for subject j in study i, $j = 1, 2, \ldots, n_i$. Then $(t_{ij}, \delta_{ij}, \mathbf{X}_{ij})$ are the observed variables, where t_{ij} is the observed failure time and δ_{ij} is the censoring indicator which equals to 1 if the observed event is a failure and 0 otherwise. When some of the covariates are subject to missingness, we can write $\mathbf{X}_{ij} = (\mathbf{X}_{ij}^{obs}, \mathbf{X}_{ij}^{mis}) = ((x_{i1j}^{obs}, x_{i2j}^{obs}, \ldots, x_{iqj}^{obs}), (x_{i(q+1)j}^{mis}, x_{i(q+2)j}^{mis}, \ldots, x_{ipj}^{mis}))$ which means for the vector $\mathbf{X}_{ij}$, the first

q covariates are always completely observed and some of the last $(p-q)$ covariates are sometimes omitted from study i by design or by happenstance. Additionally, an indicator vector $\mathbf{O}_{ij} = \{o_{ilj}\}_{1 \leqslant l \leqslant p}$ is defined, in which o_{ilj} equals to 1 if x_{ilj} is observed and 0 if x_{ilj} is missing. It is highly likely that $\mathbf{X}_{ij}$ is a mixture of both categorical and continuous covariates. So we will use the MIX package developed by Schafer (1997) [68] to impute the missing data. The procedure of our proposed HMI has several steps:

Step 1. Re-arrange the columns of the covariate matrix $\mathbf{X}^{mis}$ so that $\sum_{i=1}^{k}\sum_{j=1}^{n_i} o_{i(q+1)j} \geqslant \sum_{i=1}^{k}\sum_{j=1}^{n_i} o_{i(q+2)j} \geqslant \cdots \geqslant \sum_{i=1}^{k}\sum_{j=1}^{n_i} o_{ipj}$;

Step 2. Impute $x^{mis}_{i(q+1)j}$ with $\mathbf{X}^{obs}_{ij}$ to get $\tilde{x}^{mis}_{i(q+1)j}$;

Step 3. Impute $x^{mis}_{i(q+2)j}$ with $(\mathbf{X}^{obs}_{ij}, \tilde{x}^{mis}_{i(q+1)j})$ to get $\tilde{x}^{mis}_{i(q+2)j}$. Carry the imputation forward until x^{mis}_{ipj} is imputed, then the data set is complete;

Step 4. Analyze this complete data set with frailty model to get the treatment effect estimate $\hat{\theta}_1$ and its variance $\hat{\sigma}_1^2$ for the first imputation;

Step 5. Repeat Step 1-4 M times.

Then, our proposed HMI estimate for the pooled treatment effect θ is given by

$$\hat{\theta}_{\text{HMI}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m. \tag{3.39}$$

The variance of $\hat{\theta}_{\text{HMI}}$ has two components of variability: one is the average within-imputation variance

$$\bar{\sigma}^2 = \frac{1}{M}\sum_{m=1}^{M}\hat{\sigma}_m^2; \tag{3.40}$$

and the other is the between-imputation variance

$$\tau^2 = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\theta}_m - \hat{\theta}_{\text{HMI}})^2. \tag{3.41}$$

Thus the total variance is the sum

$$\text{var}(\theta_{\text{HMI}}) = \bar{\sigma}^2 + \left(1 + \frac{1}{M}\right)\tau^2, \tag{3.42}$$

where $1/M$ is the adjustment for finite M.

The overall standard error is the square root of $\text{var}(\theta_{\text{HMI}})$. Confidence intervals are obtained by taking the overall estimate θ_{HMI} plus or minus the standard error multiplied by some number, where that number is a quantile of Student's $t-$distribution with degree of freedom

$$\nu = (M-1)\left(1 + \frac{\bar{\sigma}^2}{(1+1/M)\tau^2}\right)^2. \tag{3.43}$$

4.0 SIMULATION AND DATA EXAMPLE

In this chapter, we will use data simulation to investigate the performance of different methodologies we proposed and compare them to the the standard approaches people use in their meta-analyses. These simulations are implemented to mimic real clinical data in breast cancer research. The simulations were programed with the R language and executed on a PC with Intel®Core^{TM}2 Duo CPU and 3.0 GB memory(RAM). Later in this chapter, the proposed meta-analysis methodologies are implemented to a data example.

4.1 SIMULATION

4.1.1 Comparison of Methodologies for APD Meta-Analysis

To begin our simulation study, we assumed that we had a total of k studies included in the APD meta-analysis, and thus generated the survival data for each study i using a Cox proportional hazard model given by

$$h_{ij}(t) = h_0(t) \exp(\beta_{i1}x_{i1j} + \beta_{i2}x_{i2j} + \beta_{i3}x_{i3j} + \beta_{i4}x_{i4j} + \beta_{i5}x_{i5j}). \tag{4.1}$$

This is a similar data generating scheme as what we did in Section (3.1.1), where the baseline failure time is generated from an exponential distribution with parameter $\lambda_0 = 0.1$ and $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, n_i$.

The covariates, X_1 and X_4, are categorical variables and X_2, X_3 and X_5 are continuous variables. To simulate data similar to that observed in a breast cancer clinical trial, we generated X_1 from a Bernoulli distribution with $p = 0.5$ to represent a treatment indicator,

X_2 as a covariate of centered age from a normal distribution with mean 0 and variance 100 and X_3 as a covariate of tumor size generated from a χ^2 distribution with 3 degrees of freedom. X_4 is a Bernoulli variable with $p = 0.6$, representing an indicator of estrogen receptor positivity and X_5 stands for number of lymph nodes involved with the breast cancer, which was generated using an exponential distribution with rate 1/3 and rounded up to integers but was modified so that negative nodal status ($X_5 = 0$) would occur in roughly 60% of the patients. To generate the survival times, the coefficients associated with X_1-X_5 were set to

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)' = (-0.2, -0.1,\ 0.1, -0.15,\ 0.7)'.$$

However, the number of patients, n_i, for each study was randomly picked from a uniform distribution with endpoints 100 and 500, and the survival times for each study were censored with probabilities uniformly randomly chosen between 0.1 and 0.3, inclusively. After generating the data, each study was randomly fitted with either a correctly specified Cox model with all five covariates or a misspecified Cox model with some covariates but X_1 omitted. Next, we obtained the treatment effect $\theta_i (= \hat{\beta}_1)$ and its variance $\sigma_i^2 (= \text{var}(\hat{\beta}_1))$ from each study. Additionally, if we denote the model information, such as which covariates are included in the Cox model, as $\mathcal{F}_i$, then $(\theta_i, \sigma_i^2, \mathcal{F}_i)$ represents the observed APD information for study i, $i = 1, 2, \ldots, k$.

Our proposed meta-ANOVA (M-A) model and meta-polynomial (M-P) model are implemented to analyze the observed APD, which are compared with commonly used DerSimonian and Laird (DL) random effect model. There are two objectives for this simulation study:

1. Compare the the accuracy of the three methodologies using estimation biases;

2. Compare the efficiency of the three methodologies using estimated standard errors.

The data generating and simulation program for APD meta-analysis is listed in Appendix A.

The results of comparison are shown in the following two figures for 200 simulation:

In Figure 4.1, the number of studies included in the APD meta-analysis varies from 10 to 30. There is little fluctuation of the estimation bias for all three models. Among them, DL constantly yielded biases > 0.05 while our two proposed methods had small biases around 0.

When the number of studies is small, M-A had larger biases than M-P, which results from the M-A model requiring more covariates, thus, it has a great danger of over-fitting when k is small. As k increases, the biases of M-A decrease and gradually overlap that in the plot of M-P biases.

In Figure 4.2, the standard error of DL model is much smaller than our proposed methods due to the fact that it includes no covariates. The standard error of M-P estimator is consistently smaller than the M-A estimator, especially when the number of studies k is small.

Another issue is how our proposed models perform compared with DL, if we fix k but vary β_1 when generating the data sets. Thus, in the second simulation study, we fixed $k = 20$ and varied β_1 from -1.0 to -0.1 while other coefficients were left unchanged. In Figure 4.3, the bias of the DL estimator becomes considerably large when β_1 reaches -1.0 and goes down linearly as β_1 increases. However, our proposed methods have consistently small biases around 0 and the M-P estimator yields smaller biases than the M-A estimator. Figure 4.4 illustrates the standard error for each method in the second simulation study. We can see that the DL estimator has relatively smaller errors than the other two. The M-A and M-P estimators have similar efficiency but the standard error for M-P is a little smaller.

4.1.2 Comparison of Methodologies for IPD Meta-Analysis

Following the same data generating scheme as what we did in Section (4.1.1), we generated survival data for k studies and pooled them together study by study. Similarly, the sample size of each study was uniformly randomly picked from 100 to 500, and the survival times for each study were censored with probabilities randomly chosen between 0.1 and 0.3, inclusively. In a randomized clinical trial, treatment indicator (X_1) is always recorded, and we believe age (X_2) is always (and easy to be) observed, too. So to create the pooled IPD with missingness, we arbitrarily and randomly pick one of the k studies with no observations of ER status (X_4), two studies with no observations of tumor size (X_3) and three studies with no observations of nodal status (X_5). With the help of MIX package developed by Schafer (1997) [68], we used the information from fully observed covariates (X_1, X_2) to impute X_4 and denoted the

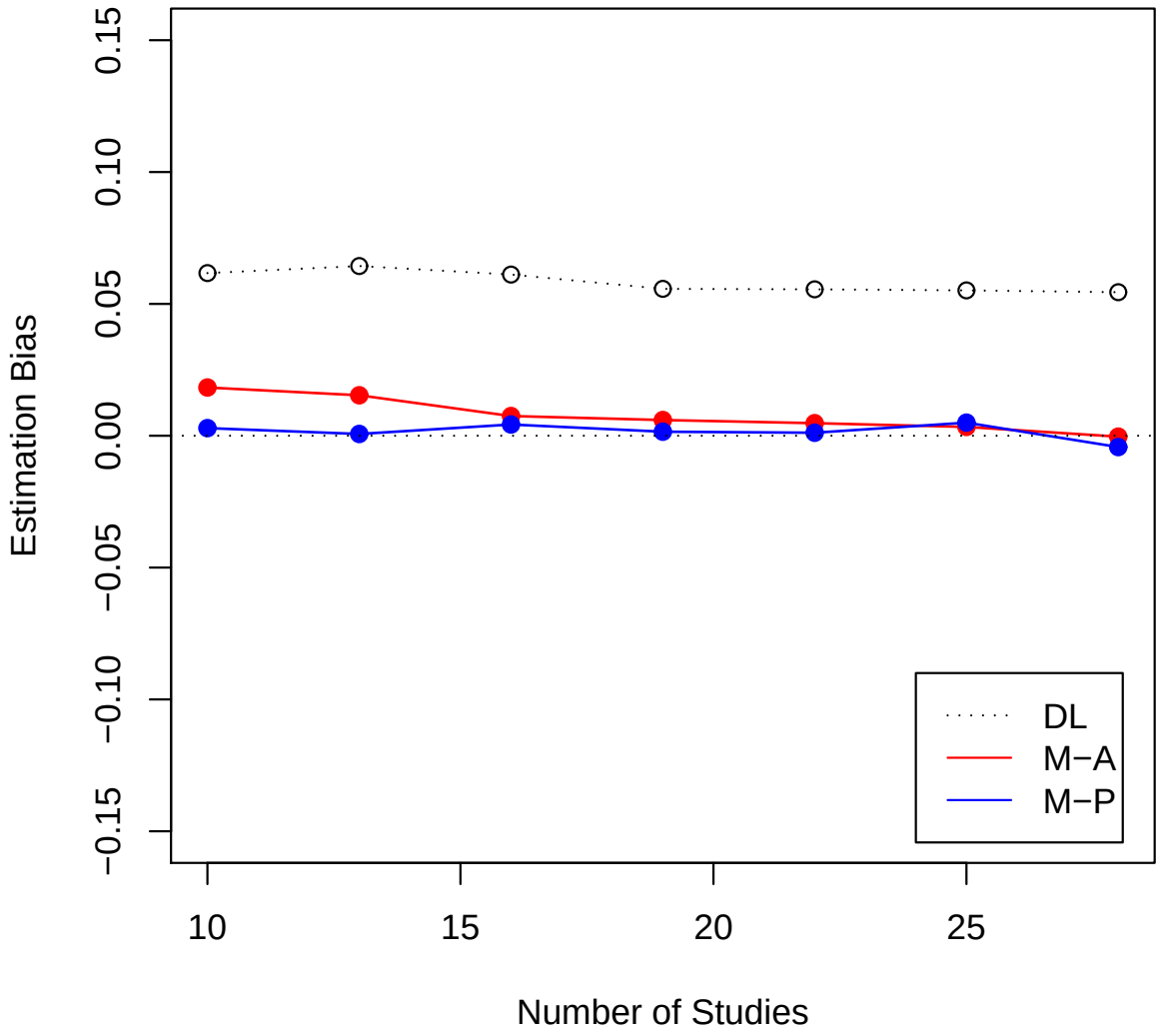

Figure 4.1: Comparison of estimation bias with different numbers of studies

The x-axis indicates the number of studies included in the APD meta-analysis. DL: DerSimonian and Laird random effect model; M-A: meta-ANOVA model; M-P: meta-polynomial model. Each point on the graphs is averaged over 200 simulation realizations.

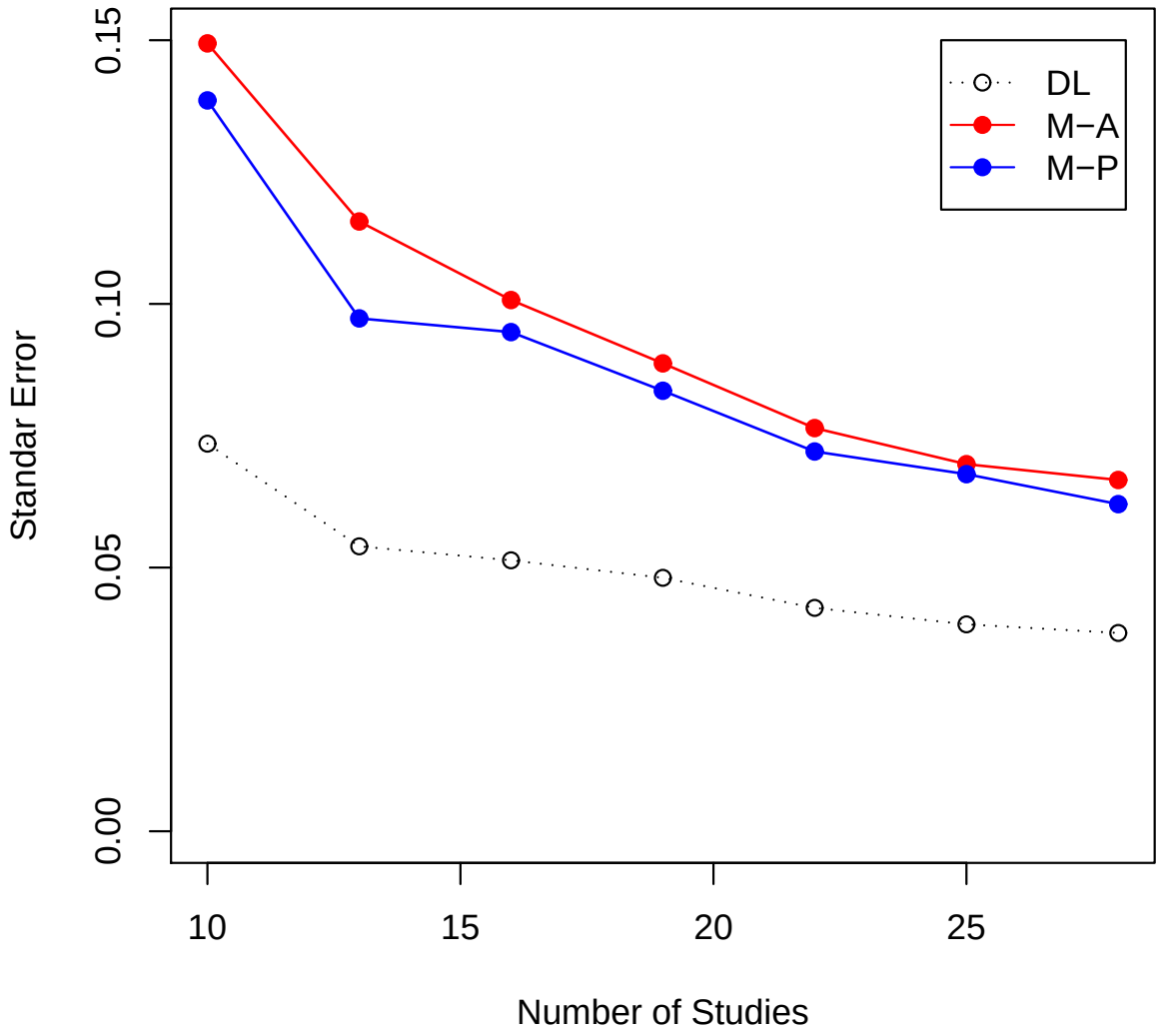

Figure 4.2: Comparison of standard error estimate with different numbers of studies

The x-axis indicates the number of studies included in the APD meta-analysis. DL: DerSimonian and Laird random effect model; M-A: meta-ANOVA model; M-P: meta-polynomial model. Each point on the graphs is averaged over 200 simulation realizations.

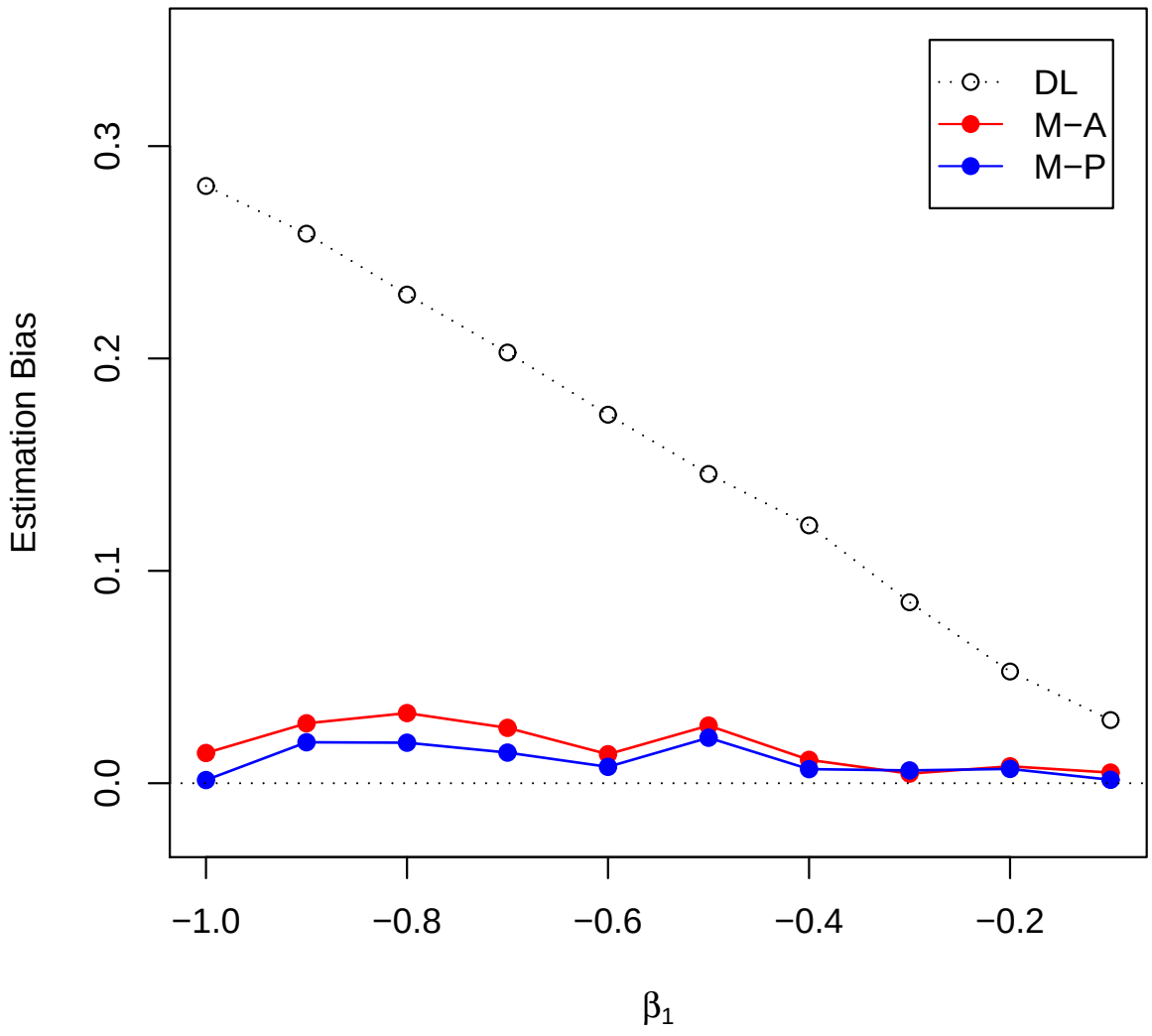

Figure 4.3: Comparison of estimation bias with different values of coefficient

The x-axis is the true value of β_1 from a Cox proportional hazard model ranging in values from -1.0 to -0.1. DL: DerSimonian and Laird random effect model; M-A: meta-ANOVA model; M-P: meta-polynomial model. Each point on the graphs is averaged over 200 simulation realizations.

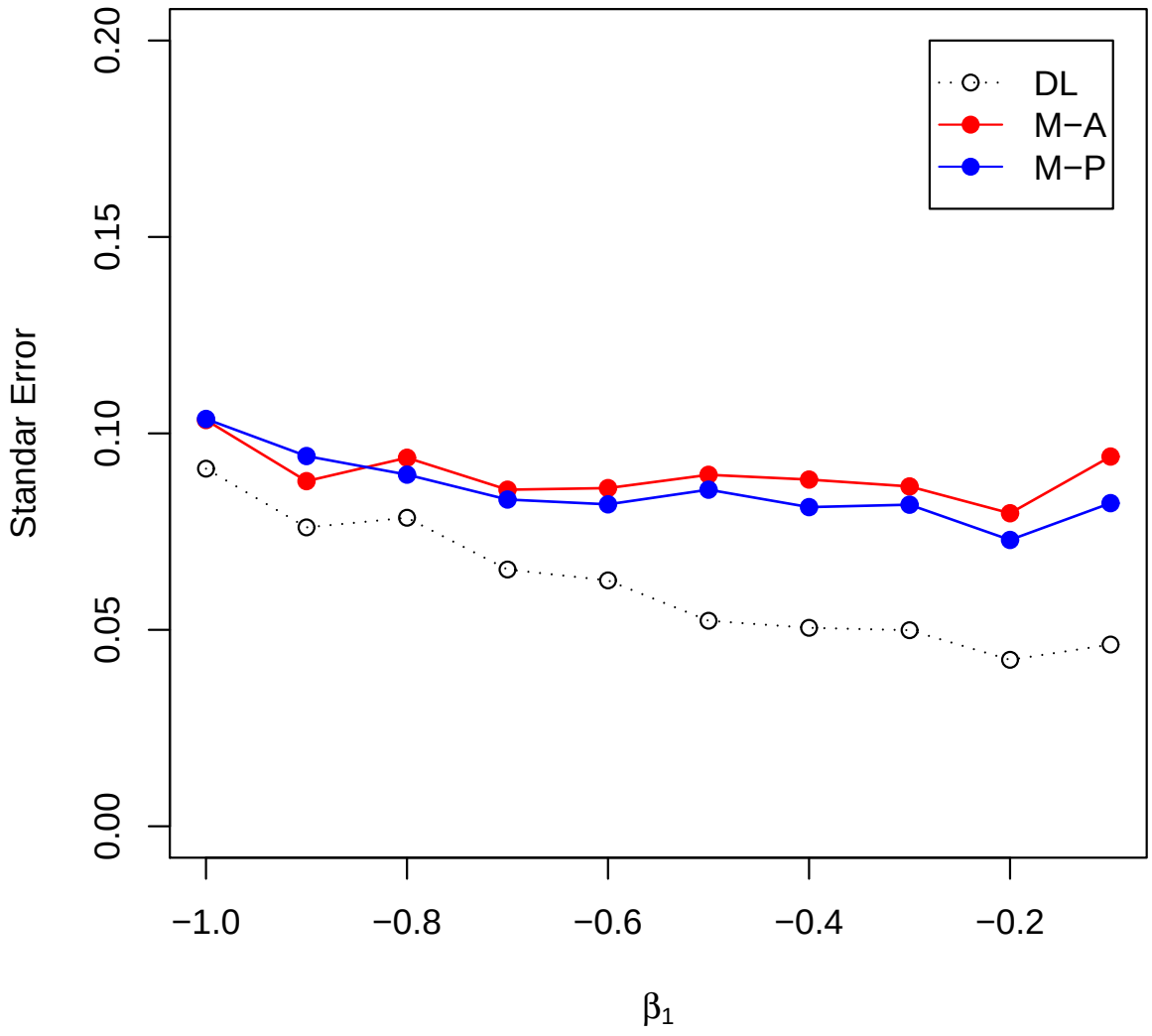

Figure 4.4: Comparison of standard error estimate with different values of coefficient

The x-axis is the value of β_1 from -1.0 to -0.1. DL: DerSimonian and Laird random effect model; M-A: meta-ANOVA model; M-P: meta-polynomial model. Each point on the graphs is averaged over 200 simulation realizations.

imputed covariate as $\tilde{X}_4$. Then X_3 was imputed based on $(X_1, X_2, \tilde{X}_4)$. Lastly, X_5 was imputed based on $(X_1, X_2, \tilde{X}_3, \tilde{X}_4)$. Later, a gamma frailty model was used to analyze the imputed dataset and coefficients and their variances were recorded for comparison under different data generating scenarios. The above procedure was repeated M times for the hierarchical multiple imputation. The data generating and simulation program for IPD meta-analysis is listed in Appendix B.

The true coefficients of (X_1, X_2, X_3, X_4, X_5) used to generate the survival data were set as (-0.2,-0.1, 0.1,-0.15, 0.7) as before. In Table 4.1, we compared the mean estimated coefficients and their standard errors with 10 imputations ($M = 10$). "Full" represents the situation that there is no missing covariates in the pooled data set. "MIX (Total)" means that we use the MIX package to impute the missing covariates (X_3, X_4, X_5) all at once, which is the most conventional method to implement data imputation. "MIX (HMI)" means that we use the MIX package to impute the missingness using the hierarchical method as described above. When only 10 studies were pooled ($k = 10$), all the estimated coefficients are underestimated (absolute-value-wise) with noticeable bias, especially for X_5 which has the most frequent missingness (roughly 30%). When more studies were pooled together ($k = 20$), all the estimated coefficients moved toward the true values but still were underestimated (absolute-value-wise). When we pooled 30 studies, the estimated coefficients had very good estimation accuracy except for $\hat{\beta}_5$, which mainly because of the substantial missingness in X_5. Through out all the scenarios, the proposed HMI method consistently had better estimation accuracy than the convetional method. Both imputation methods did not increase the standard errors and thus had similar effciency as when the full data were analyzed.

As can be seen from Figure 4.5, when we increased the number of imputation gradually from 10 to 50, the estimation accuracy of the treatment effect did not improve as M increased, and the confidence intervals remained the same width. Thus, we believe for multiple imputation with the MIX package, $M = 10$ is adequate enough to reach the estimation accuracy and reflect the sampling variability.

Later, we want to investigate the influence of missingness frequency on the estimation accuracy of all the coefficients. Suppose there are 20 studies pooled together. If we define

Table 4.1: Mean estimated coefficients (and standard errors) with different number of studies from 10 imputations

		$k = 10$	$k = 20$	$k = 30$
X_1	True β_1	-0.20	-0.20	-0.20
	Full	-0.21 (0.058)	-0.22 (0.040)	-0.20 (0.033)
	MIX (Total)	-0.14 (0.058)	-0.16 (0.040)	-0.18 (0.035)
	MIX (HMI)	-0.15 (0.058)	-0.18 (0.041)	-0.19 (0.034)
X_2	True β_2	-0.10	-0.10	-0.10
	Full	-0.10 (0.001)	-0.10 (0.001)	-0.10 (0.001)
	MIX (Total)	-0.07 (0.001)	-0.08 (0.001)	-0.08 (0.001)
	MIX (HMI)	-0.08 (0.001)	-0.08 (0.001)	-0.09 (0.001)
X_3	True β_3	0.10	0.10	0.10
	Full	0.10 (0.010)	0.10 (0.010)	0.10 (0.001)
	MIX (Total)	0.06 (0.014)	0.07 (0.010)	0.08 (0.001)
	MIX (HMI)	0.06 (0.014)	0.08 (0.010)	0.08 (0.001)
X_4	True β_4	-0.15	-0.15	-0.15
	Full	-0.16 (0.059)	-0.15 (0.041)	-0.15 (0.035)
	MIX (Total)	-0.10 (0.060)	-0.11 (0.042)	-0.10 (0.035)
	MIX (HMI)	-0.12 (0.060)	-0.12 (0.041)	-0.12 (0.034)
X_5	True β_5	0.70	0.70	0.70
	Full	0.69 (0.020)	0.71 (0.014)	0.70 (0.010)
	MIX (Total)	0.25 (0.014)	0.43 (0.010)	0.52 (0.010)
	MIX (HMI)	0.29 (0.014)	0.49 (0.010)	0.54 (0.010)

the missing covariates scenario described above as "Severe", in which one study has X_4 missing, two studies have X_3 missing and three studies have X_5 missing, then we can define a "Moderate" missing covariates scenario as one study has X_3 missing and two studies have X_5 missing. Subsequently, we can define a "Mild" missing covariates scenario as only one study has X_5 missing. In Table 4.2, we compared the mean estimated coefficients and their standard errors with 10 imputations. In the "Mild" scenario, HMI method reduces to the conventional method since there is only one covariate subject to missingness, so MIX (Total) and MIX (HMI) have the same values of estimated coefficients and the accuracy is very good. In the "Moderate" scenario, the estimated coefficients are slightly biased toward zero and in the "Severe" scenario, the bias is even more obvious. Based on these results, we can conclude that both the conventional method and our proposed HMI method are sensitive to the frequency of missingness. In our simulation studies, if more covariates are subject to more frequent missingness, the estimated coefficients would have larger bias toward zero. However, the frequency of missingness does not influence the efficiency of estimation quite well.

In conclusion, our proposed HMI method has somewhat better estimation accuracy than the conventional imputation method, using the MIX package from R. However, this accuracy is highly dependent on the frequency of missingness. When there is substantial proportion of patients with covariates subject to missingness, the estimated coefficients from the gamma frailty model are tend to be biased toward zero. The estimation from multiple imputation using MIX package is as efficient as using the full data set without any missing covariates.

4.2 DATA EXAMPLE

4.2.1 Comparison of Methodologies for APD Meta-Analysis

To assess the average effect on overall survival of positive-node breast cancer patients receiving "more" chemotherapy (hypothesized by some investigators to improve outcome) versus those receiving "less" chemotherapy (control group: including no chemotherapy), we com-

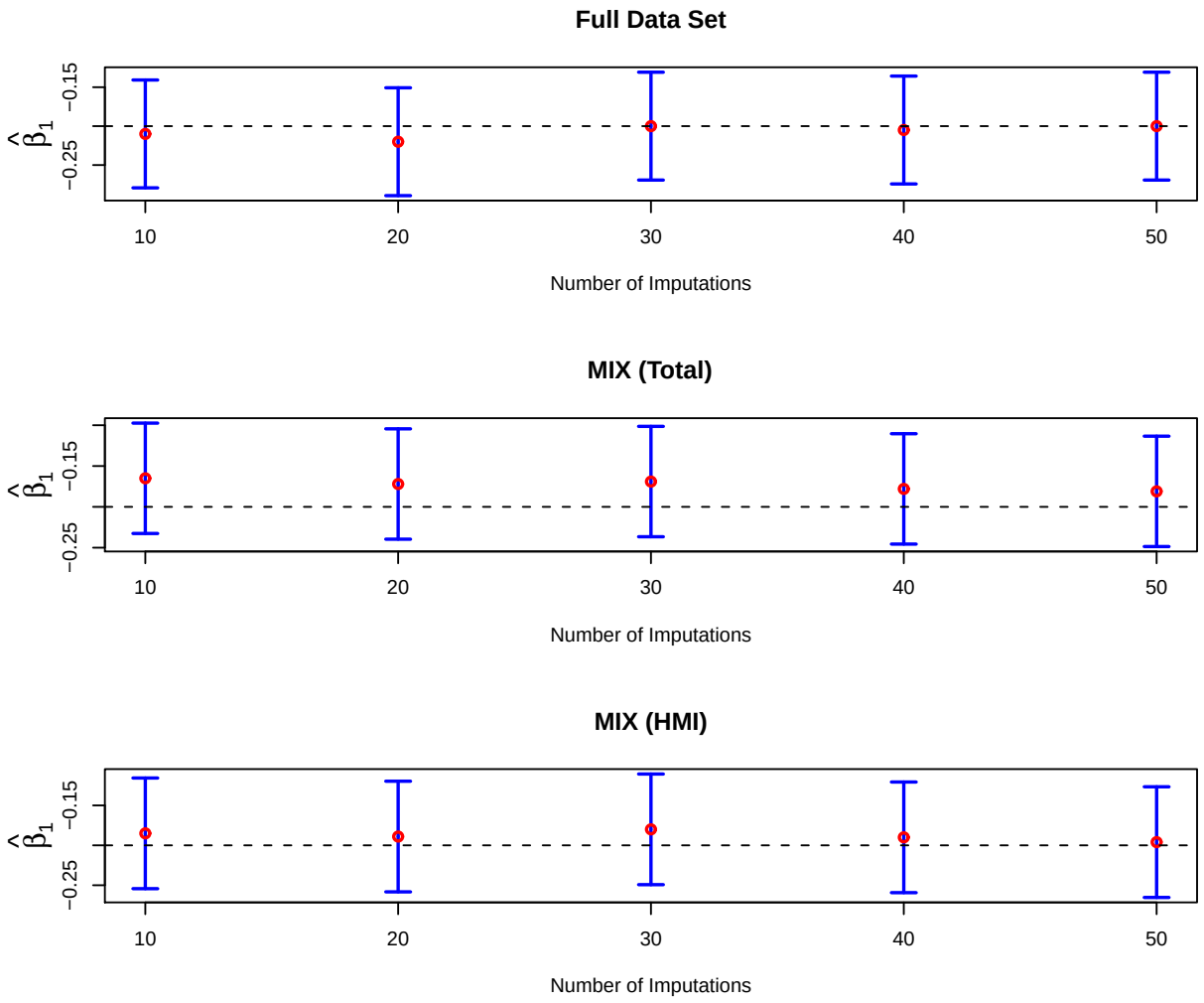

Figure 4.5: Comparison of estimated treatment effect with 95% CI for different numbers of imputations

The x-axis is the numbers of imputations from 10 to 50. Dashed line represents the true value -0.2 of β_1.

Table 4.2: Mean estimated coefficients (and standard errors) with 10 imputations in different missing covariates scenarios

		Mild	Moderate	Severe
X_1	True β_1	-0.20	-0.20	-0.20
	Full	-0.20 (0.041)	-0.19 (0.041)	-0.22 (0.040)
	MIX (Total)	-0.19 (0.042)	-0.16 (0.041)	-0.16 (0.040)
	MIX (HMI)	-0.19 (0.042)	-0.18 (0.041)	-0.18 (0.041)
X_2	True β_2	-0.10	-0.10	-0.10
	Full	-0.10 (0.001)	-0.10 (0.001)	-0.10 (0.001)
	MIX (Total)	-0.09 (0.001)	-0.08 (0.001)	-0.08 (0.001)
	MIX (HMI)	-0.09 (0.001)	-0.08 (0.001)	-0.08 (0.001)
X_3	True β_3	0.10	0.10	0.10
	Full	0.10 (0.010)	0.10 (0.010)	0.10 (0.010)
	MIX (Total)	0.09 (0.010)	0.08 (0.010)	0.07 (0.010)
	MIX (HMI)	0.09 (0.010)	0.09 (0.010)	0.08 (0.010)
X_4	True β_4	-0.15	-0.15	-0.15
	Full	-0.15 (0.042)	-0.15 (0.042)	-0.15 (0.041)
	MIX (Total)	-0.14 (0.042)	-0.11 (0.042)	-0.11 (0.042)
	MIX (HMI)	-0.14 (0.042)	-0.13 (0.042)	-0.12 (0.041)
X_5	True β_5	0.70	0.70	0.70
	Full	0.70 (0.014)	0.71 (0.014)	0.71 (0.014)
	MIX (Total)	0.63 (0.014)	0.51 (0.010)	0.43 (0.010)
	MIX (HMI)	0.63 (0.014)	0.53 (0.010)	0.49 (0.010)

bined four NSABP clinical trials B-15, B-16, B-22 and B-25 to carry out a meta-analysis. The individual-level covariates considered are treatment ($X_1 = 0$ if less chemotherapy; $= 1$ if more chemotherapy), age (X_2: age at entry) and number of positive nodes ($X_3 = 1$ if 1-4 nodes; $= 2$ if 4+ nodes). If the Cox proportional hazard models for these four clinical trials include all three covariates, the estimated log hazard ratios $\hat{\beta}_{i1}$ are listed in the second column of the first part of Table 4.3. The Q test for heterogeneity is not significant ($p = 0.61$). Thus the DL estimator degenerates to the inverse variance (IV) fixed effect estimator and hence the pooled treatment effect is -0.057. When only X_3 is omitted from all four Cox models, the pooled treatment effect decreases to -0.048 and when only X_2 is omitted, the pooled treatment effect decreases to -0.053. If all four models include only X_1, the pooled estimate becomes -0.045. This supports our findings that the treatment effect in Cox model will be biased toward zero if important covariates are omitted from the model, and the more covariates omitted, the larger the bias. Additionally, since X_3 (number of positive nodes) is a stronger predictor than X_2 (age), the bias is larger when X_3 is omitted than when X_2 is omitted.

To further demonstrate our proposed methods, we arbitrarily removed different covariates from only two of the studies, B-15 and B-25. The specifications of the Cox models are list below:

Study 1: B-15 $h_1(t) = h_0(t)\exp(\beta_{11}x_{11})$,

Study 2: B-16 $h_2(t) = h_0(t)\exp(\beta_{21}x_{21} + \beta_{22}x_{22} + \beta_{23}x_{23})$,

Study 3: B-22 $h_3(t) = h_0(t)\exp(\beta_{31}x_{31} + \beta_{32}x_{32} + \beta_{33}x_{33})$,

Study 4: B-25 $h_4(t) = h_0(t)\exp(\beta_{41}x_{41} + \beta_{43}x_{43})$.

Thus, only the Cox models for B-16 and B-22 are "correctly specified". Suppose also that we have only the estimated treatment coefficients $\hat{\beta}_{i1}$ (denoted as θ_i) and its variance $\text{var}(\hat{\beta}_{i1})$ (denoted as σ_i^2) available from each of the studies, which we index with $i = 1, 2, 3, 4$. Then

Table 4.3: Estimated log hazard ratios from Cox proportional hazard models with different model specifications and corresponding meta-analysis results

Protocol	$\hat{\beta}_{i1}$	$\text{var}(\hat{\beta}_{i1})$	Covariates in Cox Models	Q Test[a]	$\hat{\beta}_{1,\text{IV}}$[b]	$\text{var}(\hat{\beta}_{1,\text{IV}})$
B-15	-0.061	0.0038	X_1, X_2, X_3	$p = 0.61$	-0.057	0.0012
B-16	-0.142	0.0061				
B-22	-0.007	0.0044				
B-25	-0.039	0.0047				
B-15	-0.051	0.0038	X_1, X_2	$p = 0.58$	-0.048	0.0012
B-16	-0.140	0.0061				
B-22	-0.004	0.0044				
B-25	-0.020	0.0047				
B-15	-0.060	0.0038	X_1, X_3	$p = 0.70$	-0.053	0.0012
B-16	-0.126	0.0061				
B-22	-0.006	0.0044				
B-25	-0.039	0.0047				
B-15	-0.049	0.0038	X_1	$p = 0.65$	-0.045	0.0012
B-16	-0.126	0.0061				
B-22	-0.005	0.0044				
B-25	-0.020	0.0047				

[a] χ^2 test with 3 degrees of freedom.

[b] IV: inverse variance fixed effect estimator.

the meta-ANOVA model in matrix form can be written as

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}. \tag{4.2}$$

Our proposed estimator for the pooled treatment effect based on the meta-ANOVA model is

$$\begin{aligned} \theta_{\text{M-A}} &= \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 \\ &= -0.063 \end{aligned} \tag{4.3}$$

with variance 0.0025.

In this example, we rank $[X_2, X_3]$ as $(1, 2)$ since we believe number of positive nodes is a stronger predictor than age at entry. Consequently, following the scoring system described earlier, the scores associated with $[X_1, X_2, X_3]$ are $(1, 0.67, 1.33)$. Explicitly, the conditional polynomial model (3.15) can then be written as

$$\begin{bmatrix} \theta_1 \mid 1.00 \\ \theta_2 \mid 3.00 \\ \theta_3 \mid 3.00 \\ \theta_4 \mid 2.33 \end{bmatrix} = \begin{bmatrix} 1 & 1.00 & 1.00^2 & \dots & 1.00^n \\ 1 & 3.00 & 3.00^2 & \dots & 3.00^n \\ 1 & 3.00 & 3.00^2 & \dots & 3.00^n \\ 1 & 2.33 & 2.33^2 & \dots & 2.33^n \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}. \tag{4.4}$$

If we only use the first polynomial term and the score of the "full" Cox model is 3, our proposed pooled treatment effect estimator based on the meta-polynomial model is

$$\begin{aligned} \theta_{\text{M-P}} &= [\ 1\ 3\ ] \times \hat{\boldsymbol{\alpha}} \\ &= -0.058 \end{aligned} \tag{4.5}$$

with variance 0.0021.

If we apply the IV method for this example, the pooled treatment effect is -0.053 (with variance 0.0011) and is biased toward zero as compared to the true value of -0.057 from Table 4.3. However, the M-P estimator has good accuracy but with a relatively larger

variance compared to the IV estimator. The M-A estimator is somewhat biased low largely because of the small number of studies included in the meta-regression.

There are 256 different ways to choose one set of B-15, B-16, B-22 and B-25 from the first column of Table 4.3. Each box plot in Figure 4.6 shows the pooled treatment effects from all the 256 meta-analyses for one of the three estimators. The results concur with our simulation analyses that the simple meta-analysis method (IV) has pooled treatment effects biased toward zero. Our proposed APD meta-regression models have better estimation accuracy but larger variance.

4.2.2 Comparison of Methodologies for IPD Meta-Analysis

To better demonstrate our proposed methods for IPD meta-analysis, we again pooled four NSABP clinical trials B-15, B-16, B-22 and B-25 with individual patient data and this time, we included another covariate to represent ER status ($X_4 = 1$ if ER-positive; $= 0$ if ER-negative), besides the already included covariates treatment (X_1), age (X_2), and number of positive nodes (X_3). Further, we assume that one out the four trials has X_4 missing and two out the four have X_3 missing. Then, we implement the MIX package to do the multiple imputation 10 times with both the conventional way and our proposed HMI method. A gamma frailty model was later used to analyzed the imputed data set. The results are compared with the full data set and listed in Table 4.4. We could see that the estimated coefficients with imputed data set all have very good estimation accuracy, except for X_3 since it has almost 50% of its observations missing. Again, across all covariates, the proposed HMI method has better estimation accuracy than the conventional method, and both methods have similar efficiency when compared with the results from a full data set.

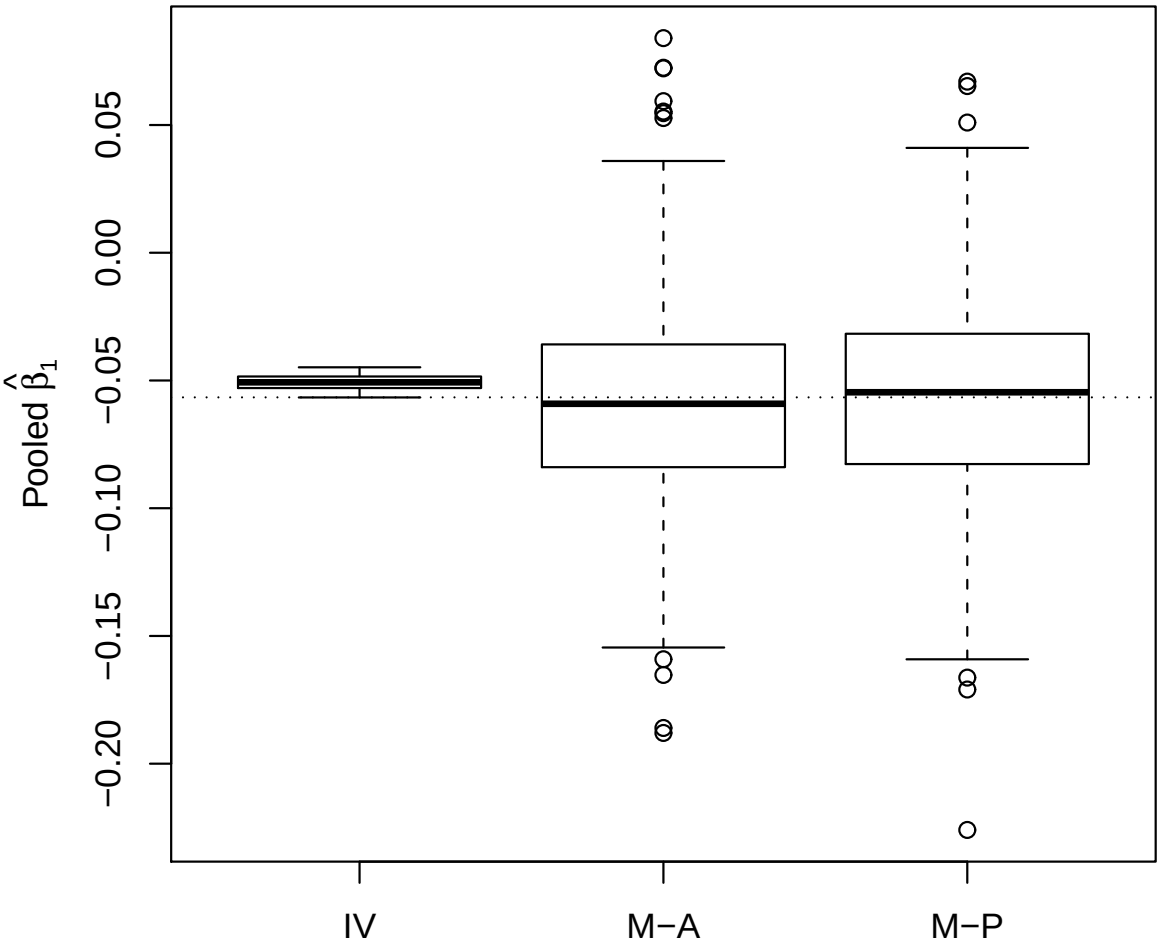

Figure 4.6: Comparison of estimation of the pooled treatment effect from three estimators

IV: inverse variance fixed effect model; M-A: meta-ANOVA model; M-P: meta-polynomial model.

Table 4.4: Mean estimated coefficients (and standard errors) with 10 imputations for NSABP data

	X_1	X_2	X_3	X_4
Full	−0.050 (0.035)	0.003 (0.001)	0.790 (0.033)	−0.242 (0.035)
MIX (Total)	−0.044 (0.035)	0.004 (0.001)	0.396 (0.033)	−0.221 (0.035)
MIX (HMI)	−0.047 (0.035)	0.004 (0.001)	0.458 (0.033)	−0.236 (0.035)

5.0 DISCUSSION AND CONCLUSIONS

5.1 APD META-ANALYSIS

Modeling heterogeneity across studies is a key issue in meta-analysis, especially in APD settings where we have limited information regarding treatment effect from each trial. Since, in APD settings, we only observe effect sizes and summarized model information from different studies, it is challenging to combine estimates of the treatment effects from Cox proportional hazard models where adjustments have been made for different subsets of covariates. It has been established that estimation biases are likely to be high in cases where non-randomized studies are combined. However, estimates of treatment effects may also be biased in Cox proportional hazard models in randomized prospective trials if one fails to adjust for prognostic variables even when the covariates are perfectly balanced across treatment groups using randomization [4]. At first glance, this statement seems counterintuitive because no such bias would occur from omitting perfectly balanced prognostic variables in the more familiar linear regression model, even though some efficiency would be lost in estimating the treatment effect [19]. Hauck et al. (1998) [25] provided an intuitive explanation regarding this unique problem in non-linear regression models such as Cox proportional hazard model: models with a "full" set of covariates measure the best reflection of the risks or benefits for patients with specific covariate values, which they called "subject-specific" models. In contrast, the treatment effects from models with some covariates omitted is a comparison of averaged outcomes, where the averaging is over all omitted covariates. In the extreme case where a Cox model has only treatment variable, the treatment effect is a population average comparing all different outcomes in groups of patients. Thus, the latter are called "population-averaged" models. Further, adding more covariates moves the population-averaged models closer to the

subject-specific models and subsequently, changes the interpretation of treatment effects.

Obviously, there are other sources of heterogeneity associated with treatment effects. Different clinical trials may use different treatment protocols and have different study designs. Additionally, the sets of baseline covariates collected such as demographic variables or clinical or pathologic variables may vary substantially across studies, and such variables are likely to be related to outcomes. In addition, there is selection bias, publication bias, data irregularities and other debatable fundamental issues with meta-analysis [15]. These problems can generally be avoided if a few basic principles are observed, such as a prior definition of inclusion criteria for studies and a comprehensive trial search strategy [14]. Similar to the need to carefully design clinical trials, meta-analyses also need to be carefully planned with detailed written protocols being prepared in advance. We do not address topics of the selection process or publication bias associated with meta-analyses in this dissertation, but rather assume all studies share certain common characteristics and that a "full" Cox model is globally best for all studies. Here, the "full" Cox model is chosen from one or more studies where the largest number of covariates are measured and included, and other studies are designated as having some covariates omitted.

Similar to the regular linear regression models, our proposed meta-regression models work better with larger sample sizes, which, in the particular case of APD meta-analysis, means more clinical trials should be included. This is especially true for a meta-ANOVA model which usually requires more covariates and one is in danger of over fitting when the number of studies is small. As of now, there is no globally accepted guideline of how many studies should be pooled to design a meta-analysis. With the booming development of the internet and information searching technology, identifying and locating relevant studies may be easily achieved. For published studies, electronic databases are useful. The Cochrane Central Register of Controlled Trials [29] has provided extensive searches of medical journals in English and many other languages and is probably the best single electronic source of clinical trials for meta-analyses. Thus, including as many studies as possible which meet the prespecified selection criteria in a meta-analysis will provide better results when implementing the meta-regression models.

Another issue with the meta-polynomial model is how to score covariates. The analytic

procedure proposed in this paper scores the covariates included in the Cox model based on their ranks. However, the method to rank covariates is somewhat arbitrary and thus, different researchers many have different options to assign scores. One needs to gather as much information as possible from literature reviews and expert advice to explore different ways of ranking the covariates. In some cases, sensitivity analyses should be performed to explore the robustness of scoring procedures. Additionally, a plot of the observed treatment effects over the scores of studies could be very helpful. The usual stepwise regression technique can be used to determine the lowest-degree polynomial that fits the APD best.

The random effect method provided by DerSimonian and Laird (1986) [13] allows for treatment effects to vary across studies and uses a simple non-iterative method to estimate the between-study treatment effect variance. Because it incorporates differences into the analysis of overall treatment efficacy, and because of its simplicity, this estimator has been widely used. However, this method does not address problems of combining incomparable Cox models caused by omitting covariates. Our proposed meta-ANOVA and meta-polynomial models outperform the existing method in the presence of the "incomparability" issue. Both estimators produce very accurate estimation with price of losing a small amount of efficiency. Further, the meta-polynomial method performs better than the meta-ANOVA with a smaller bias and a smaller standard error, especially when the number of studies included in the APD meta-analysis is small.

5.2 IPD META-ANALYSIS

Although due to the time or financial constrains, IPD meta-analysis is not as prevalent as APD meta-analysis, it is however viewed as the gold standard of doing meta-analysis, and researchers should try their best to obtain individual data, which has enormous advantages where we stated at the beginning of Section (2.1.2) in details. However, problems of missing covariates complicate IPD meta-analysis, and especially for IPD meta-analysis using Cox regression model which requires full covariate information. We proposed the solution for the missing covariates problems via a hierarchical multiple imputation (HMI) method.

From the simulation study (in Section (4.1.2)) and real data analysis (in Section (4.2.2)), the HMI method is consistently superior than the conventional imputation method which imputes all missing covariates at the same time. The reason behind this phenomenon is that, for the conventional way, it only utilizes the information from all the covariates without missing value across all studies, which could be only a small portion of the whole information the full data set holds if the frequency of missingness is severe. In a hypothetic example of an IPD meta-analysis combining six studies, as depicted in Figure 5.1, we only use the information from study 4, 5 and 6 to do the imputation in the conventional way. However, for the HMI method, we utilize the information of fully observed information from the covariates without missing value to impute the first covariate with the smallest percentage of missingness, which means we use all the information from X_1 and X_2 across all six studies to impute X_3. Then we combined the imputed X_3 with X_1 and X_2 to impute X_4. In this way, we utilize the maximum amount of information we could get out from this data set. Thus, HMI could improve the accuracy of the estimation based on the imputed data set.

We also observed that the estimated coefficients for the covariates with considerably large amount of missingness could be highly biased, no matter which imputation methods were used or how many times of imputations were implemented. Missing data problem should be thoroughly investigated when conducting a IPD meta-analysis. We should not assume the imputed data set as the true "full" data set and take whatever conclusion we could get for granted. Besides, in this dissertation, we only consider the situation where some covariates are omitted as a whole chunk from certain studies so that there is not a single observation. Other than that, we assume there is no missing data problem within each observed covariate. If the second type of missing data problem exists, further analysis should be implemented to tackle this problem.

However, the fundamental assumption we made for the validity of the hierarchical multiple imputation, or other likelihood or weighted estimating equation methods, is that the distributions of covariates are the same across all studies, which means all the studies included in an IPD meta-analysis should share the same study design, same protocol, and same patient baseline characteristics. This could be somehow achieved by planning a carefully designed selection criteria and a comprehensive study searching strategy. But, still, this is a

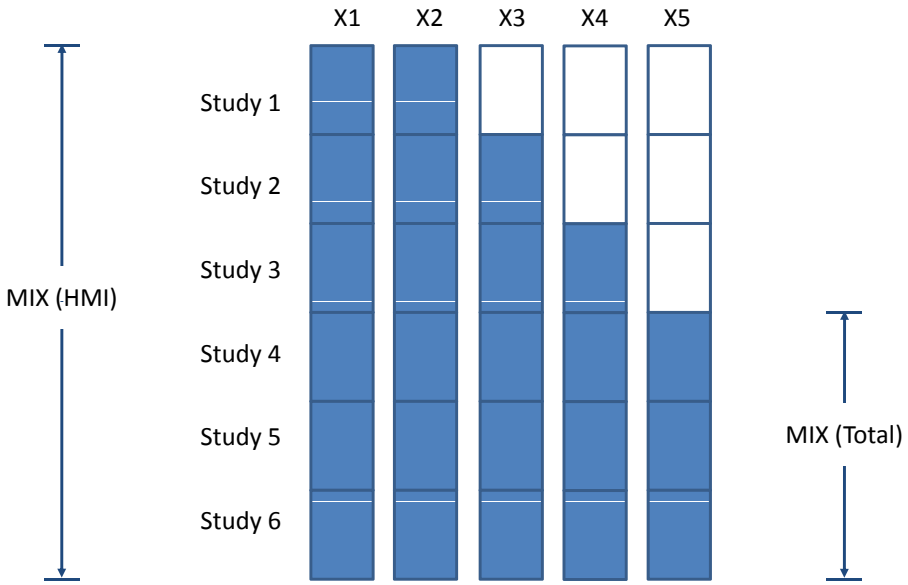

Figure 5.1: Comparison of MIX (HMI) and MIX (Total) in a hypothetic IPD meta-analysis

MIX (HMI): the proposed hierarchical multiple imputation; MIX (Total): the conventional way to impute all missingness together at once.

very aggressive assumption and often it is open to questions when doing a meta-analysis in a real world. The proposed HMI method utilizes the maximum amount of information from the data set as a whole and could best reflect the variability and heterogeneity of covariates across studies, that is the reason why it is more appropriate than the conventional imputation method in the IPD meta-analysis. However, caution has to be taken when interpreting the results from HMI method and the descriptive statistics of the baseline characteristics could help us to make the right decision.

Thus, for IPD meta-analysis, our recommendation is to take the two-stage approach as we discussed in Section (2.1.2.2). The purpose of doing a meta-analysis is twofold: one is to get the combined treatment effect estimation; the other is to assess the heterogeneity of treatment effects across studies and explore why. The IPD meta-analysis will simply not provide the latter information. By doing a two-stage IPD meta-analysis, we can get estimated treatment effect from each study via a standard Cox proportional hazard model and then combine them with the general APD meta-analysis methods. Or if the missing covariates problem persists, the proposed meta-ANOVA model or meta-polynomial model could be used. However, future works which might have better results are list in the next section.

5.3 FUTURE WORKS

In this section, we present two directions for future work. One is an attempt to extend the existing weighted estimating equation for IPD meta-analysis with missing covariates problem. The other one is a proposal of integrating the information provided by IPD meta-analysis into APD meta-analysis based on the findings of Gail et al. (1984) [19].

5.3.1 Multiple Missingness Weighted Estimating Equation (MMWEE)

As we asserted in Section (2.2.3), the weighted estimating equation (WEE) method works best when there is only one covariate subject to missingness, or when all elements of $\mathbf{X}_{ij}^{mis}$ are

always observed or unobserved. However, either scenario is unlikely to occur in a IPD meta-analysis. Thus, we may extend the usual WEE to accommodate multi-missingness patterns of $\mathbf{X}_{ij}^{mis}$ and we call it a "multiple missingness weighted estimating equation" (MMWEE).

Without loss of generality, we will demonstrate multi-missingness with a simple example. More complicated cases are easy to envision. Suppose we have five covariates $X_1 - X_5$ but only X_4 and X_5 are subject to missingness. If we have o_{ilj} defined as before, a hypothetic IPD of o_{ilj} are shown in Table (5.1(a)). Previous WEE methods can only handle missing patterns such as in Study 1 and Study 4, but not in Study 2 and Study 3. Assuming that there is no within-study missing data problem, it implies that o_{ilj} have the same value for all j within one study. Thus, we can remove the subscript j and have o_{il} to denote the indicator of observed covariate l in study i, as in Table (5.1(b)).

Because only two covariates might be missing for some studies, there are $2^2 = 4$ types of missing patterns. Therefore, we create four new indicator variables for each study i as

$$\mathbf{R}_i = (R_{i1}, R_{i2}, R_{i3}, R_{i4}) \begin{cases} R_{i1} = I_{\{o_{i4}=1 \text{ and } o_{i5}=1\}} \\ R_{i2} = I_{\{o_{i4}=1 \text{ and } o_{i5}=0\}} \\ R_{i3} = I_{\{o_{i4}=0 \text{ and } o_{i5}=1\}} \\ R_{i4} = I_{\{o_{i4}=0 \text{ and } o_{i5}=0\}} \end{cases}.$$

$\mathbf{R}$ is thus a Multi-Bernoulli variable with probabilities $(\pi_1, \pi_2, \pi_3, \pi_4)$, $\sum_{a=1}^{4} \pi_a = 1$. It is safe to assume MACR in our IPD meta-analysis, since $\mathbf{R}$ does not depend on either y_i, $\mathbf{X}_i^{obs}$ or $\mathbf{X}_i^{mis}$. Now we can extend the score function (2.44) to be

$$\begin{aligned} \mathbf{u}_{\text{MMWEE}}(\boldsymbol{\beta}) \quad = \quad & \sum_{i=1}^{k} \Bigg\{ \frac{R_{i1}}{\pi_1} \mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, \mathbf{X}_i^{mis}) + (1 - \frac{R_{i2}}{\pi_2}) E_{X_5|y_i, \mathbf{X}_i^{obs}, X_4}[\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, X_4)] \\ & + (1 - \frac{R_{i3}}{\pi_3}) E_{X_4|y_i, \mathbf{X}_i^{obs}, X_5}[\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, X_5)] \\ & + (1 - \frac{R_{i4}}{\pi_4}) E_{\mathbf{X}_i^{mis}|y_i, \mathbf{X}_i^{obs}}[\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs})] \Bigg\}. \end{aligned} \tag{5.1}$$

Table 5.1: Multi-missingness patterns

(a)

	X^{obs}			X^{mis}	
	X_1	X_2	X_3	X_4	X_5
Study 1	1	1	1	1	1
	⋮	⋮	⋮	⋮	⋮
	1	1	1	1	1
Study 2	1	1	1	1	0
	⋮	⋮	⋮	⋮	⋮
	1	1	1	1	0
Study 3	1	1	1	0	1
	⋮	⋮	⋮	⋮	⋮
	1	1	1	0	1
Study 4	1	1	1	0	0
	⋮	⋮	⋮	⋮	⋮
	1	1	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮

$\Rightarrow$

(b)

	X^{obs}			X^{mis}	
	X_1	X_2	X_3	X_4	X_5
Study 1	1	1	1	1	1
Study 2	1	1	1	1	0
Study 3	1	1	1	0	1
Study 4	1	1	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮

(a). Table of indicators o_{ilj} for observed covariate l of subject j in Study i. (b). Table of indicators o_{il} for observed covariate l in Study i.

Suppose the probabilities $(\pi_1, \ldots, \pi_4)$ are correctly specified, the first term of $\mathbf{u}_{\text{MMWEE}}(\boldsymbol{\beta})$ in (5.1) has expectation equal to 0; that is,

$$\begin{aligned} & E\left[\frac{R_{i1}}{\pi_1}\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, \mathbf{X}_i^{mis})\right] \\ = \; & E\left[E_{R_{i1}}\left(\frac{R_{i1}}{\pi_1}\right)\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, \mathbf{X}_i^{mis})\right] \\ = \; & E\left[1 \times \mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, \mathbf{X}_i^{mis})\right] \\ = \; & 0, \end{aligned} \tag{5.2}$$

because the score function has expectation 0.

Now the second term of $\mathbf{u}_{\text{MMWEE}}(\boldsymbol{\beta})$ in (5.1) has expectation equal to

$$\begin{aligned} & E\left[(1 - \frac{R_{i2}}{\pi_2})E_{X_5|y_i, \mathbf{X}_i^{obs}, X_4}[\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, X_4)]\right] \\ = \; & E\left[E_{R_{i2}}\left(\frac{1 - R_{i2}}{\pi_2}\right)E_{X_5|y_i, \mathbf{X}_i^{obs}, X_4}[\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, X_4)]\right] \\ = \; & E\left[0 \times E_{X_5|y_i, \mathbf{X}_i^{obs}, X_4}[\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, X_4)]\right] \\ = \; & 0, \end{aligned} \tag{5.3}$$

regardless of whether $E_{X_5|y_i, \mathbf{X}_i^{obs}, X_4}[\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, X_4)]$ can be correctly calculated or not.

Similarly the third and the fourth term of $\mathbf{u}_{\text{MMWEE}}(\boldsymbol{\beta})$ equal to 0 too. Thus $\mathbf{u}_{\text{MMWEE}}(\boldsymbol{\beta})$ has expectation equal to 0 so $\hat{\boldsymbol{\beta}}_{\text{MMWEE}}$ is asymptotically unbiased. Note that if $(\pi_1, \ldots, \pi_4)$ are correctly specified, $\mathbf{u}_{\text{MMWEE}}(\hat{\boldsymbol{\beta}}) = 0$ will give a consistent estimate of $\boldsymbol{\beta}$.

In our IPD meta-analysis, an intuitive way to model probabilities $(\pi_1, \ldots, \pi_4)$ is

$$\pi_a = \frac{\sum_{i=1}^{k} R_{ia} \times n_i}{N}, \qquad a = 1, \ldots, 4, \tag{5.4}$$

where $N = \sum_{i=1}^{k} n_i$.

To make the MMWEE method work, two conditions have to be met (double-robustness): one is to correctly specify the conditional distribution of missing covariates give observed ones. This can hardly be done because when different studies are pooled, it is not viable to define a universally applicable conditional distribution which could work for every study. The other condition is that the missing probabilities $(\pi_1, \ldots, \pi_4)$ has to be correctly modeled.

Since we can assume MCAR in IPD meta-analysis, model (5.4) is a very constructive attempt, but it is rather intuitive and is in many ways open to questions.

An other problem is that, the score function $\mathbf{u}_i(\boldsymbol{\beta}|y_i, \mathbf{X}_i^{obs}, \mathbf{X}_i^{mis})$ in Equation (5.1) is easy to define for simple Cox proportional hazard model but it is extremely difficult to get for the frailty model. We need to put the frailty parameters into the equation and the score function may not have a closed form. This adds another huge challenge for the future work.

5.3.2 Integrated IPD and APD meta-analysis

As we discussed in Section (3.1.1), suppose we have a Cox model with treatment variable x_{1j} and a baseline measure of some important covariate x_{2j} like this:

$$h_j(t|x_{1j}, x_{2j}) = h_0(t)\exp(\beta_1 x_{1j} + \beta_2 x_{2j}), \tag{5.5}$$

and we also consider model 2, where covariate x_{2j} has been dropped:

$$h_j(t|x_{1j}) = h_0(t)\exp(\beta_1^* x_{1j}). \tag{5.6}$$

Gail et al. (1984) [19] provided an approximated estimation of the bias between β_1^* and β_1 as

$$\delta = \beta_1^* - \beta_1 \approx \frac{1}{4}\boldsymbol{\beta}_2'\boldsymbol{\Omega}\boldsymbol{\beta}_2 \left\{R(\beta_1) - R(-\beta_1)\right\} \tag{5.7}$$

for vector $\mathbf{X}_{2j}$, where

$$R(\eta) = 2\phi'(\eta)/\phi(\eta) + 1, \tag{5.8}$$

$$\phi(\eta) = E[H_0(t)|\eta]. \tag{5.9}$$

Thus, when some studies have certain covariates omitted in an IPD meta-analysis, another way to deal with this problem is to get the estimated bias (5.7) from IPD and integrate this information into an APD meta-analysis. Specifically, treatment effect from study i is firstly estimated from the Cox proportional hazard model as $\hat{\beta}_{1i}$ with some covariates omitted. Then $\tilde{\beta}_{1i} = \hat{\beta}_{1i} + \delta_{1i}$ could be viewed as the adjusted treatment effect from study i and later are combined using APD meta-analysis techniques to get the pooled treatment effect estimation.

APPENDIX A

R PROGRAM FOR APD META-ANALYSIS

```
library(survival)
library(MASS)

################### Data Generation ######################

Gen.Data = function(size, beta, base.hz, prob.cens, cov.ind)
{
x1 = rbinom(size,1,0.5) # Treatment
x2 = rnorm(size, 0, 10) # Age
x3 = rchisq(size, 3) # Tumor size
x4 = rbinom(size,1,0.6) # ER-positive
x5 = ceiling(rexp(size,1/3))
x5[rbinom(size,1,0.6)==0] = 0 # Nodal Status
x = as.matrix(cbind(x1,x2,x3,x4,x5))
beta = as.matrix(beta)
hr = exp(x%*%beta)
time = rexp(size,hr*base.hz)
status = rbinom(size,1,1-prob.cens)

form.full = as.formula(paste("Surv(time, status)~", paste("x",
   c(1:5), sep = "", collapse = "+")))
fit.full = summary(coxph(form.full))

form.redu = as.formula(paste("Surv(time, status)~", paste("x",
   cov.ind, sep = "", collapse = "+")))
fit.redu = summary(coxph(form.redu))

Z = rep(0,6)
```

```
Z[cov.ind] = 1
Z[6] = length(cov.ind)

output = c(fit.full$coef[1,1], fit.full$coef[1,3], fit.
   redu$coef[1,1], fit.redu$coef[1,3], Z)
return(output)
}

################### Estimation Functions ####################

theta.IV = function(theta, sigma2)
{
output=sum(ginv(sigma2)%*%theta)/sum(diag(ginv(sigma2)))
return(output)
}

var.IV = function(sigma2)
{
output=1/sum(diag(ginv(sigma2)))
return(output)
}

Q.DL = function(theta, sigma2)
{
output=as.numeric(t(theta-theta.IV(theta,sigma2))%*%ginv(sigma2
   )%*%(theta-theta.IV(theta,sigma2)))
return(output)
}

tau2.wmm = function(theta, sigma2)
{
weight=1/diag(sigma2)
output=max(0,(Q.DL(theta, sigma2)-(K-1))/(sum(weight)-sum(
   weight^2)/sum(weight)))
return(output)
}

alpha = function(theta, sigma2, Z)
{
output=ginv(t(Z)%*%ginv(sigma2)%*%Z)%*%(t(Z)%*%ginv(sigma2)%*%
   theta)
return(output)
}

var.alpha = function(sigma2, Z)
```

```
{
output=ginv(t(Z)%*%ginv(sigma2)%*%Z)
return(output)
}

D = function(tau2, Z)
{
part1 = ginv(t(Z)%*%ginv(sigma2+tau2)%*%Z)
part2 = t(Z)%*%ginv((sigma2+tau2)^2)%*%Z
return(sum(diag(part1%*%part2)))
}

Q.REG = function(theta, sigma2, Z)
{
output=as.numeric(t(theta)%*%(ginv(sigma2)-ginv(sigma2)%*%Z%*%
   ginv(t(Z)%*%ginv(sigma2)%*%Z)%*%t(Z)%*%ginv(sigma2))%*%theta
   )
return(output)
}

tau2.wmm.REG = function(theta, sigma2, Z)
{
M=diag(x=rep(0,dim(theta)[1]))
output=max(0,(Q.REG(theta,sigma2,Z)-(K-4))/(sum(1/diag(sigma2))
   -D(M,Z)))
return(output)
}

#####################  Size varies  ##########################

meta.size = seq(10, 30, by=3) # Pool of number of studies
size.pool = seq(100,500,by=10) # Pool of size for each study
beta=c(-0.2,-0.1,0.1,-0.15,0.7) # Coefficients
iter = length(meta.size)

est.naive = rep(NA,iter)
est.adv1 = rep(NA,iter)
est.adv2 = rep(NA,iter)
est.adv3 = rep(NA,iter)

se.naive = rep(NA,iter)
se.adv1 = rep(NA,iter)
se.adv2 = rep(NA,iter)
se.adv3 = rep(NA,iter)
```

```
for (k in 1:iter)
{
K = meta.size[k]

ns = 60

est1 = rep(NA,ns)
est2 = rep(NA,ns)
est3 = rep(NA,ns)
est4 = rep(NA,ns)

var1 = rep(NA,ns)
var2 = rep(NA,ns)
var3 = rep(NA,ns)
var4 = rep(NA,ns)

for (j in 1:ns)
{

Data = matrix(NA, nrow=K, ncol=10)

for (i in 1:K)
{
size=sample(size.pool,1)
base.hz=sample(seq(0.1,0.3,length=100),1)
prob.cens=sample(seq(0.1,0.3,length=100),1)
cov.ind = c(1, sample(c(2:5), sample(c(0:4),1),replace=F))
Data[i,] = Gen.Data(size,beta,base.hz,prob.cens,cov.ind)
}

theta.full = as.matrix(Data[,1])
sigma2.full = diag(x=Data[,2]^2)
theta = as.matrix(Data[,3])
sigma2 = diag(x=Data[,4]^2)
tau2 = diag(x=rep(0,length(Data[,2])))
V = sigma2+tau2
Z1 = Data[,5:9]
Z2 = Data[,c(5,10)]
score = Z1%*%as.matrix(c(1,1.2,0.4,0.8,1.6))
score2 = score^2
Z3 = cbind(Z2[,1],score)
I5 = as.matrix(c(1,1,1,1,1))
S5 = as.matrix(c(1,5))

V.new = sigma2+diag(x=tau2.wmm(theta,sigma2),nr=K,nc=K)
```

```
V1.new=sigma2+diag(x=tau2.wmm.REG(theta,sigma2,Z1),nr=K,nc=K)
V2.new=sigma2+diag(x=tau2.wmm.REG(theta,sigma2,Z2),nr=K,nc=K)
V3.new=sigma2+diag(x=tau2.wmm.REG(theta,sigma2,Z3),nr=K,nc=K)

est1[j] = theta.IV(theta,V.new)
est2[j] = sum(alpha(theta,V1.new,Z1))
est3[j] = sum(c(1,5)*alpha(theta, V2.new, Z2))
est4[j] = sum(c(1,5)*alpha(theta, V3.new, Z3))

var1[j] = var.IV(V.new)
var2[j] = t(I5)%*%var.alpha(V1.new,Z1)%*%I5
var3[j] = t(S5)%*%var.alpha(V2.new,Z2)%*%S5
var4[j] = t(S5)%*%var.alpha(V3.new,Z3)%*%S5

}

est.naive[k] = mean(est1)
est.adv1[k] = mean(est2)
est.adv2[k] = mean(est3)
est.adv3[k] = mean(est4)

se.naive[k] = sqrt(mean(var1)+(1+1/ns)*var(est1))
se.adv1[k] = sqrt(mean(var2)+(1+1/ns)*var(est2))
se.adv2[k] = sqrt(mean(var3)+(1+1/ns)*var(est3))
se.adv3[k] = sqrt(mean(var4)+(1+1/ns)*var(est4))

}

################### Beta Varies ########################

size.pool = seq(100,500,by=10) # Pool of size for each study
meta.size = 20
iter = length(seq(-1,-0.1,by=0.1))
beta.pool=cbind(seq(-1,-0.1,by=0.1),rep(-0.1,iter),rep(0.1,iter
   ),rep(-0.15,iter),rep(0.7,iter)) # Coefficients

est.naive = rep(NA,iter)
est.adv1 = rep(NA,iter)
est.adv2 = rep(NA,iter)
est.adv3 = rep(NA,iter)

se.naive = rep(NA,iter)
se.adv1 = rep(NA,iter)
se.adv2 = rep(NA,iter)
se.adv3 = rep(NA,iter)
```

```
for (k in 1:iter)
{
K = meta.size

ns = 50

est1 = rep(NA,ns)
est2 = rep(NA,ns)
est3 = rep(NA,ns)
est4 = rep(NA,ns)

var1 = rep(NA,ns)
var2 = rep(NA,ns)
var3 = rep(NA,ns)
var4 = rep(NA,ns)

for (j in 1:ns)
{

Data = matrix(NA, nrow=K, ncol=10)

for (i in 1:K)
{
size=sample(size.pool,1)
beta = beta.pool[k,]
base.hz=sample(seq(0.1,0.3,length=100),1)
prob.cens=sample(seq(0.1,0.3,length=100),1)
cov.ind = c(1, sample(c(2:5), sample(c(0:4),1),replace=F))
Data[i,] = Gen.Data(size,beta,base.hz,prob.cens,cov.ind)
}

theta = as.matrix(Data[,3])
sigma2 = diag(x=Data[,4]^2)
tau2 = diag(x=rep(0,length(Data[,2])))
V = sigma2+tau2
Z1 = Data[,5:9]
Z2 = Data[,c(5,10)]
score = Z1%*%as.matrix(c(1,1.2,0.4,0.8,1.6))
Z3 = cbind(Z2[,1],score)
I5 = as.matrix(c(1,1,1,1,1))
S5 = as.matrix(c(1,5))

V.new = sigma2+diag(x=tau2.wmm(theta,sigma2),nr=K,nc=K)
V1.new=sigma2+diag(x=tau2.wmm.REG(theta,sigma2,Z1),nr=K,nc=K)
```

```
V2.new=sigma2+diag(x=tau2.wmm.REG(theta,sigma2,Z2),nr=K,nc=K)
V3.new=sigma2+diag(x=tau2.wmm.REG(theta,sigma2,Z3),nr=K,nc=K)

est1[j] = theta.IV(theta,V.new)
est2[j] = sum(alpha(theta,V1.new,Z1))
est3[j] = sum(c(1,5)*alpha(theta, V2.new, Z2))
est4[j] = sum(c(1,5)*alpha(theta, V3.new, Z3))

var1[j] = var.IV(V.new)
var2[j] = t(I5)%*%var.alpha(V1.new,Z1)%*%I5
var3[j] = t(S5)%*%var.alpha(V2.new,Z2)%*%S5
var4[j] = t(S5)%*%var.alpha(V3.new,Z3)%*%S5

}

est.naive[k] = mean(est1)
est.adv1[k] = mean(est2)
est.adv2[k] = mean(est3)
est.adv3[k] = mean(est4)

se.naive[k] = sqrt(mean(var1)+(1+1/ns)*var(est1))
se.adv1[k] = sqrt(mean(var2)+(1+1/ns)*var(est2))
se.adv2[k] = sqrt(mean(var3)+(1+1/ns)*var(est3))
se.adv3[k] = sqrt(mean(var4)+(1+1/ns)*var(est4))

}
```

APPENDIX B

R PROGRAM FOR IPD META-ANALYSIS

```
library(mix)
library(survival)

MI.times = 10 # Number of multiple imputations

F.coef = matrix(NA,nc=MI.times,nr=5)
F.var = matrix(NA,nc=MI.times,nr=5)

T.coef = matrix(NA,nc=MI.times,nr=5)
T.var = matrix(NA,nc=MI.times,nr=5)

H.coef = matrix(NA,nc=MI.times,nr=5)
H.var = matrix(NA,nc=MI.times,nr=5)

for (m in 1:MI.times)

{

k = 20 # Number of studies
index = seq(1:k)
size = sample(seq(100,500,by=1),k)
id = rep(NA,sum(size))

for (i in 1:k)
{
if (i==1) {id[i:size[i]] = rep(index[i],size[i])}
else {id[(sum(size[1:(i-1)])+1):(sum(size[1:(i-1)])+size[i])] =
    rep(index[i],size[i])}
```

```
}

x1 = rbinom(sum(size),1,0.5) # Treatment
x1[x1==0] = 2
x2 = rnorm(sum(size), 0, 10) # Age
x3 = rchisq(sum(size), 3) # Tumor size

x4 = rbinom(sum(size),1,0.6) # ER-positive
x4[x4==0] = 2

x5 = ceiling(rexp(sum(size),1/3))
x5[rbinom(sum(size),1,0.6)==0] = 0 # Nodal Status

x.raw = as.matrix(cbind(x1,x4,x2,x3,x5))
beta=as.matrix(c(-0.2,-0.15,-0.1,0.1,0.7))
hr = exp(x.raw%*%beta)
time = rexp(sum(size),hr*0.1)
status = rbinom(sum(size),1,1-0.4)

x4.mis = x4
m4 = sample(index,1) # 1 study has X4 missing
x4.mis[id==m4] = NA

x3.mis = x3
m3 = sample(index,2) # 2 studies have X3 missing
x3.mis[id==m3[1]|id==m3[2]] = NA

x5.mis = x5
m5 = sample(index,3) # 3 studies have X5 missing
x5.mis[id==m5[1]|id==m5[2]|id==m5[3]] = NA

x.mis = as.matrix(cbind(x1,x4.mis,x2,x3.mis,x5.mis))

rngseed(123456)

D1 = x.mis[,c(1,2,3)]
s1 = prelim.mix(D1,2)
t1 = em.mix(s1,showits=F)
nt1 = da.mix(s1,t1,steps=500,showits=F)
D1.imp = imp.mix(s1,nt1)

D2 = cbind(D1.imp,x3.mis)
s2 = prelim.mix(D2,2)
t2 = em.mix(s2,showits=F)
nt2 = da.mix(s2,t2,steps=500,showits=F)
```

```
D2.imp = imp.mix(s2,nt2)
D2.imp[,4][D2.imp[,4]<0] = 0

D3 = cbind(D2.imp,x5.mis)
s3 = prelim.mix(D3,2)
t3 = em.mix(s3,showits=F)
nt3 = da.mix(s3,t3,steps=500,showits=F)
D3.imp = imp.mix(s3,nt3)
D3.imp[,5][D3.imp[,5]<0] = 0

s = prelim.mix(x.mis,2)
t = em.mix(s,showits=F)
nt = da.mix(s,t,steps=1000,showits=F)
D.imp = imp.mix(s,nt)

M.F = coxph(Surv(time,status)~x1+x4+x2+x3+x5+frailty(id,
   distribution="gamma"))
M.T = coxph(Surv(time,status)~D.imp[,1]+D.imp[,2]+D.imp[,3]+D.
   imp[,4]+D.imp[,5]+frailty(id, distribution="gamma"))
M.H = coxph(Surv(time,status)~D3.imp[,1]+D3.imp[,2]+D3.imp[,3]+
   D3.imp[,4]+D3.imp[,5]+frailty(id, distribution="gamma"))

F.coef[,m] = M.F$coef
F.var[,m] = diag(M.F$var)

T.coef[,m] = M.T$coef
T.var[,m] = diag(M.T$var)

H.coef[,m] = M.H$coef
H.var[,m] = diag(M.H$var)

}
```

BIBLIOGRAPHY

[1] Andersen, P. K., and Gill, R. D. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics 10*, 4 (Dec. 1982), 1100–1120.

[2] Begun, J. M., Hall, W. J., Huang, W., and Wellner, J. A. Information and asymptotic efficiency in Parametric-Nonparametric models. *The Annals of Statistics 11*, 2 (June 1983), 432–452.

[3] Chalmers, I. The cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Annals of the New York Academy of Sciences 703* (Dec. 1993), 156–63; discussion 163–5.

[4] Chastang, C., Byar, D., and Piantadosi, S. A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models. *Statistics in Medicine 7*, 12 (1988), 1243–1255.

[5] Clayton, D. G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65*, 1 (Apr. 1978), 141–151.

[6] Cochran, W. G. The combination of estimates from different experiments. *Biometrics 10*, 1 (1954), 101–129.

[7] Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.

[8] Cooper, H., and Hedges, L. V. *The Handbook of research synthesis*. New York: Russell Sage Foundation, 1994.

[9] Cox, D. R. Regression models and life tables. *Journal of the Royal Statistical Society 34*, 2 (1972), 187–220.

[10] Cox, D. R. Partial likelihood. *Biometrika 62*, 2 (1975), 269–276.

[11] Demidenko, E. *Mixed Models: Theory and Applications*, 1 ed. Wiley-Interscience, Hoboken, NJ, Aug. 2004.

[12] Dempster, A. P., and Rubin, D. B. *Overview. Incomplete Data in Sample Surveys, Vol. II: Theory and Annotated Bibliography.* New York: Academic Press, 1983.

[13] DerSimonian, R., and Laird, N. Meta-analysis in clinical trials. *Controlled Clinical Trials 7*, 3 (1986), 177–88.

[14] Egger, M., Smith, G. D., and Phillips, A. N. Meta-analysis: Principles and procedures. *BMJ 315*, 7121 (Dec. 1997), 1533–1537.

[15] Egger, M., Smith, G. D., Schneider, M., and Minder, C. Bias in meta-analysis detected by a simple, graphical test. *BMJ 315*, 7109 (Sept. 1997), 629–634.

[16] Fisher, R. A. Statistical methods for research workers. *Edinburgh and London* (1925).

[17] Follmann, D. A., and Proschan, M. A. Valid inference in random effects meta-analysis. *Biometrics 55*, 3 (Sept. 1999), 732–7.

[18] Ford, I., Norrie, J., and Ahmadi, S. Model inconsistency, illustrated by the cox proportional hazards model. *Statistics in Medicine 14*, 8 (1995), 735–746.

[19] Gail, M. H., Wieand, S., and Piantadosi, S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika 71*, 3 (Dec. 1984), 431–444.

[20] Glass, G. V. Primary, secondary, and meta-analysis of research. *Educational Researcher 5*, 10 (1976), 3–8.

[21] Glass, G. V. Meta-Analysis at 25. *Unpublished paper* (2000).

[22] Glidden, D. V., and Vittinghoff, E. Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine 23*, 3 (2004), 369–388.

[23] Hardy, R. J., and Thompson, S. G. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine 17*, 8 (1998), 841–856.

[24] Harville, D. A. Maximum likelihood approaches to variance component estimation and to related problems: Rejoinder. *Journal of the American Statistical Association 72*, 358 (June 1977), 339–340.

[25] Hauck, W. W., Anderson, S., and Marcus, S. M. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials 19*, 3 (1998), 249–256.

[26] Hedges, L. V., and Olkin, I. *Statistical methods for meta-analysis.* Academic Press Orlando, 1985.

[27] Henderson, R., and Oman, P. Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society: Series B: Statistical Methodology 61*, 2 (1999), 367–379.

[28] Herring, A. H., and Ibrahim, J. G. Likelihood-Based methods for missing covariates in the cox proportional hazards model. *Journal of the American Statistical Association 96*, 453 (2001), 292–302.

[29] Higgins, J., and Green, S., Eds. *Cochrane Handbook for Systematic Reviews of Interventions*, version 5.0.1 ed. The Cochrane Collaboration, 2008.

[30] Higgins, J. P. T., and Thompson, S. G. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine 23*, 11 (2004), 1663–1682.

[31] Higgins, J. P. T., Thompson, S. G., Deeks, J. J., and Altman, D. G. Measuring inconsistency in meta-analyses. *British Medical Journal 327*, 7414 (Sept. 2003), 557–560.

[32] Horvitz, D. G., and Thompson, D. J. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*, 260 (1952), 663–685.

[33] Hougaard, P. A class of multivanate failure time distributions. *Biometrika 73*, 3 (Dec. 1986), 671–678.

[34] Hougaard, P. Survival models for heterogeneous populations derived from stable distributions. *Biometrika 73*, 2 (Aug. 1986), 387–396.

[35] Hunter, J. E., and Schmidt, F. L. *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*. Sage, 2004.

[36] Ibrahim, J. G., Chen, M. H., Lipsitz, S. R., and Herring, A. H. Missing-Data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association 100*, 469 (2005), 332–346.

[37] Jeng, G. T., Scott, J. R., and Burmeister, L. F. A comparison of meta-analytic results using literature vs individual patient data. paternal cell immunization for recurrent miscarriage. *Journal of the American Medical Association 274*, 10 (1995), 830–836.

[38] Kacker, R. N. Combining information from interlaboratory evaluations using a random effects model. *Metrologia 41*, 3 (2004), 132–136.

[39] Katsahian, S., Latouche, A., Mary, J., Chevret, S., and Porcher, R. Practical methodology of meta-analysis of individual patient data using a survival outcome. *Contemporary Clinical Trials 29*, 2 (Mar. 2008), 220–230.

[40] Koopman, L., van der Heijden, G. J. M. G., Grobbee, D. E., and Rovers, M. M. Comparison of methods of handling missing data in individual patient data meta-analyses: An empirical example on antibiotics in children with acute otitis media. *American Journal of Epidemiology 167*, 5 (Mar. 2008), 540–545.

[41] Lee, E. W., Wei, L. J., and Amato, D. A. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. *Survival Analysis: State of the Art 211* (1992), 237–247.

[42] Lin, D. Y., and Wei, L. J. The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association 84*, 408 (Dec. 1989), 1074–1078.

[43] Lin, D. Y., and Ying, Z. Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association 88*, 424 (1993), 1341–1349.

[44] Lindsey, J. K., and Lambert, P. On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine 17*, 4 (1998), 447–469.

[45] Lipsey, M. W., and Wilson, D. B. *Practical Meta-Analysis*. Sage Publications Inc, 2001.

[46] Lipsitz, S. R., and Ibrahim, J. G. Estimating equations with incomplete categorical covariates in the cox model. *Biometrics 54*, 3 (1998), 1002–13.

[47] Lipsitz, S. R., Ibrahim, J. G., and Zhao, L. P. A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistical Association 94*, 448 (1999), 1147–1160.

[48] Lyman, G. H., and Kuderer, N. M. The strengths and limitations of meta-analyses based on aggregate data. *BMC Medical Research Methodology 5* (2005), 14.

[49] Mantel, N., and Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute 22*, 4 (Apr. 1959), 719–48.

[50] McGilchrist, C. A. REML estimation for survival models with frailty. *Biometrics 49*, 1 (Mar. 1993), 221–5.

[51] Neuhaus, J. M., Kalbfleisch, J. D., and Hauck, W. W. A comparison of Cluster-Specific and Population-Averaged approaches for analyzing correlated binary data. *International Statistical Review 59*, 1 (Apr. 1991), 25–35.

[52] Normand, S. L. T. Tutorial in biostatistics meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine 18* (1999), 321–359.

[53] Olkin, I., and Tate, R. F. Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics 32*, 2 (1961), 448465.

[54] Paik, M. C. Multiple imputation for the cox proportional hazards model with missing covariates. *Lifetime Data Analysis 3*, 3 (1997), 289–298.

[55] Paik, M. C., and Tsai, W. On using the cox proportional hazards model with missing covariates. *Biometrika 84*, 3 (1997), 579–593.

[56] Parmar, M., and Machin, D. *Survival Analysis: A Practical Approach*, 1 ed. John Wiley & Sons, Nov. 1995.

[57] Parmar, M. K. B., Torri, V., and Stewart, L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine 17*, 24 (1998), 2815–2834.

[58] Prentice, R. L. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika 69*, 2 (1982), 331–342.

[59] Qi, L., Wang, C. Y., and Prentice, R. L. Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association 100*, 472 (2005), 1250–1263.

[60] Robins, J. M., and Rotnitzky, A. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association 90*, 429 (1995), 122–129.

[61] Robins, J. M., Rotnitzky, A., and Zhao, L. P. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association 89* (1994), 846–866.

[62] Rosenthal, R., and Rubin, D. B. Meta-analytic procedures for combining studies with multiple effects sizes. *Psychological bulletin 99*, 3 (1986), 400–406.

[63] Rosnow, R. L., and Rosenthal, R. Computing contrasts, effect sizes, and counternulls on other peoples published data: General procedures for research consumers. *Psychological Methods 1*, 4 (1996), 331–40.

[64] Rubin, D. B. Multiple imputations in sample surveys - a phenomenological bayesian approach to nonresponse. In *Proceedings of the Survey Research Methods Section, American Statistical Association* (1978), pp. 20–34.

[65] Rubin, D. B. Multiple imputation for nonresponse in surveys. new york: J. *New York: John Wiley & Sons, Inc.* (1987).

[66] Rubin, D. B., and Schenker, N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association 81*, 394 (1986), 366–374.

[67] Sargent, D. J. A general framework for random effects survival analysis in the cox proportional hazards setting. *Biometrics 54*, 4 (Dec. 1998), 1486–1497.

[68] Schafer, J. L. *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC: New York, 1997.

[69] Seber, G. A. F., and Lee, A. J. *Linear Regression Analysis*, 2 ed. Wiley-Interscience, Hoboken, NJ, 2003.

[70] Simmonds, M. C., Higginsa, J. P. T., Stewartb, L. A., Tierneyb, J. F., Clarke, M. J., and Thompson, S. G. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials 2*, 3 (June 2005), 209–217.

[71] Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., and Song, F. *Methods for Meta-Analysis in Medical Research*, 1 ed. Wiley, Dec. 2000.

[72] Tierney, J. F., Clarke, M., and Stewart, L. A. Is there bias in the publication of individual patient data Meta-Analyses? *International Journal of Technology Assessment in Health Care 16*, 02 (2000), 657–667.

[73] Tsiatis, A. A. A large sample study of cox's regression model. *The Annals of Statistics 9*, 93 (1981), 108.

[74] Wang, C. Y., and Chen, H. Y. Augmented inverse probability weighted estimator for cox missing covariate regression. *Biometrics 57*, 2 (2001), 414–419.

[75] Whitehead, A., and Whitehead, J. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine 10*, 11 (1991), 1665–1677.

[76] Wu, M. C., and Bailey, K. R. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics 45*, 3 (Sept. 1989), 939–955.

[77] Yamaguchi, T., and Ohashi, Y. Investigating centre effects in a multi-centre clinical trial of superficial bladder cancer. *Statistics in Medicine 18*, 15 (1999), 1961–1971.

[78] Yusuf, S., Peto, R., Lewis, J., Collins, R., and Sleight, P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases 27*, 5 (1985), 335–71.

[79] Zhou, H., and Pepe, M. S. Auxiliary covariate data in failure time regression. *Biometrika 82*, 1 (1995), 139–149.