ABSTRACT

Title of thesis: Investigating center effects in a multi-center clinical trial study using a parametric proportional hazards meta-analysis model

Mathewos S. Demissie
Master of Science, 2009

Thesis directed by: Professor Eric V. Slud
Mathematical Statistics Program
Department of Mathematics

In this paper, we investigate meta-analysis of the overall treatment effect in the setting of a multi-center clinical trial study in which patient level data are available. We estimate the overall treatment effect using two methods: meta-analysis, which uses the summary statistics from each center and a unified combined analysis of patient level data. In the meta-analysis we use a random effects meta-analysis model and in both analyses we use a parametric proportional hazards model.

In a randomized clinical trial study, subjects are recruited at multiple centers to accrue large enough samples within an acceptable period of time and to enhance the generalizability of study results. Heterogeneity between trials may arise from the center effects or treatment effect itself. To take into account the heterogeneities, random effects models are used. We performed a data analysis

based on a multi-center clinical trial study in small-cell lung cancer conducted by the Eastern Cooperative Oncology Group and then parallel data analysis within a simulation study.

In the simulation study we vary the magnitude of the center and the treatment-by-center heterogeneity in the data generation and estimated the over all treatment effect using the two methods. We compared the two methods in terms of bias, mean square error and percentage of significant treatment effect. The simulation study shows that meta-analysis treatment effects estimate are slightly biased when covariates are included in the analysis.

Investigating center effects in a multi-center clinical trial study using a parametric proportional hazards meta-analysis model

by

Mathewos S. Demissie

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2009

Advisory Committee:
Professor Eric V. Slud, Chair/Advisor
Professor Abram M. Kagan
Professor Paul J. Smith

Acknowledgments

I owe my sincere gratitude to all the people who have made this thesis possible.

First and foremost I would like to thank my advisor, Professor Eric V. Slud for giving me the opportunity to work on such an interesting project, and for his advice and help through the thesis process.

I would like to acknowledge the Eastern Cooperative Oncology Group for the permission to use the EST 1582 data.

I owe my deepest thanks to my manager James Shulgold and my team colleagues at the Department of Risk Adjustment, MedAssurant for their help in balancing my time for work and school.

Contents

Chapter 1

Introduction

1.1 Background

In a randomized clinical trial, it is often necessary to recruit subjects at multiple study centers to accrue enough sample size within an acceptable period of time and to enhance the generalizability of study results. However, there might be factors that vary by center, which exert influence on the study's outcomes. These factors include patient characteristics and medical practice patterns. Such center effects potentially lead to dependence between outcomes at each center. In addition to factors that vary by center, heterogeneity between trials may arise from the treatment effect itself. That is, the treatment may have worked better in some centers than others. Such treatment-by-trial interaction heterogeneity

can be accounted for by using random interaction effects between the treatment and trial in a random effect model. If these effects are sufficiently large, a model which ignores the center effect or the treatment-by-center effect may lead to incorrect inference.

When the primary endpoint of interest in the study is the time it takes for a certain event to occur (time-to-event), a proportional hazards model with random effects which includes the treatment-by-center interaction term as well as the baseline risk term can be used [1]. Another alternative to analyze this kind of data is meta-analysis in terms survival-analysis models fitted to data for the individual centers.

Meta-analysis is defined as the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings [12]. The objectives of a meta-analysis include increasing power to detect an overall treatment effect, estimation of the degree of benefit associated with a particular study treatment, assessment of the amount of variability between studies, or identification of study characteristics associated with particularly effective treatments. When several studies have conflicting conclusions, a meta-analysis can be used to estimate an average effect or to identify a subset of studies associated with a beneficial effect. In meta-analysis, variation between centers can be captured using a random effects model [3, 4, 13]. As indicated in [4], meta-analysis includes the following basic steps:

1. Identification of literature
2. Study selection
3. Data extraction
4. Statistical analysis

Identification of literature: Meta-analysis needs to find all of the relevant articles on the topic. Sources to be searched include the published literature, unpublished literature, dissertations and drug company studies. Including all trials which could potentially contribute to the meta-analysis minimizes problems associated with selection or publication bias [4, 13, 23]. Although, in general it is true that meta-analysis needs to find all relevant articles on the topic, a multi-center clinical trial will have been designed prospectively with a combined analysis of the data from all centers as its main objective and a meta-analysis can be used [13]. Individual centers are expected to follow a common protocol.

Study selection: Once the author of a meta-analysis has assembled a large number of studies, it is important to select studies based on eligibility criteria used in accepting or rejecting a study. The criteria include whether the study includes enough information for analysis (a point estimate and a standard deviation or standard error), year and demographic features of the study, study design (observational or randomized), treatment dose, and sample size.

Data extraction: After identifying an appropriate group of studies, the researcher has to extract the relevant data from each study. There are many sources

of potential error in data extraction, such as misinterpreting tables and data entry.

Statistical analysis: In meta-analysis the type of model to be used should be specified. The type of model (fixed effects versus random effects) to be used is specified based on the degree of heterogeneity of between-study variability of the collected studies. To determine the degree of heterogeneity, a chi-square test is often used [4]. A fixed effect model assumes that the parameter measuring treatment difference is the same across all trials. A random effects model assumes this parameter acts as a random variable taking different values from one trial to the next. More discussion about meta-analysis models including parameter estimation methods is given in Chapter 3.

1.2 Motivation

In this thesis research we use a multi-center study and investigate the overall treatment effect using two approaches, the first a unified parametric proportional hazards model and the second a meta-analysis model. We use a dataset from a multi-center clinical trial EST 1582 in small-cell lung cancer conducted by the Eastern Cooperative Oncology Group (ECOG). Previous studies on this dataset have shown that there is significant variation in treatment effect by center [9, 10]. Thus, in the unified parametric proportional hazards model we include

random effects for both the treatment-by-center interaction and for baseline risk. In the meta-analysis model, we assume the studies in separate centers are independent, and we take from each center the minimum information that might be published in a journal article (center treatment effect estimate and standard error). We use a meta-analysis random effects model to account for the heterogeneity of treatment effects among centers, beyond the variation accounted for by fixed effects. In this thesis we are motivated to see how the overall treatment effect estimate varies from the two models: the meta-analysis random effects model and the unified parametric proportional hazards model with both center and treatment-by-center random effects. We have been unable to find any previous study that used patient level data and analyzed using both unified and meta-analysis to investigate what could go wrong in the meta-analysis.

1.3 Multi-center data: example

The ECOG EST 1582 study compares two different chemotherapy regimens: a standard therapy consisting of cyclophosphamide, adriamycin and vincristine (CAV) and an alternating regimen (CAV-HEM) where cycles of CAV were alternated with HEM (hexamethylmelamine, etoposide and methotrexate). Gray ([9] and [10]) has made a detailed analysis of institutional variation in this dataset and has shown the variation of treatment effect by center. In Chapter

4, we use these data to investigate the center effects using a unified parametric proportional hazards model and a random effects meta-analysis model.

1.4 Survival proportional hazards model with random effects

Let i be a cluster, $i = 1, \ldots, g$, and let j index subjects in cluster i, $j = 1, \ldots, n_i$. Let Z_{ij} be a random death time and let C_{ij} be the corresponding right-censoring time for subject j in cluster i . We assume that the censoring times are independent of the survival times and that $T_{ij} = \min(Z_{ij}, C_{ij})$, and $\delta_{ij} = I_{[Z_{ij} \leq C_{ij}]}$ are observed rather than Z_{ij}, C_{ij}. Each patient will have a binary variable X_{ij1} representing the treatment group to which the patient is randomized ($X_{ij1} = 0$ if a patient is in the control group and $X_{ij1} = 1$ if the patient is in the treatment group). The proportional hazards model with random center effects u_i and random treatment-by-center interaction v_i is given as

$$h_{ij}(t|u_i, v_i) = h_0(t) \exp\left(\sum_{k=1}^{p} X_{ijk}\beta_k + u_i + v_i X_{ij1}\right) \tag{1.1}$$

$$u_i \sim N(0, \sigma^2), v_i \sim N(0, \tau^2), cov(u_i, v_i) = 0$$

where $h_0(t)$ is the baseline hazard function, β is the fixed effects vector of dimension p, and X_{ij} is the vector of covariates. The variance σ^2 of the u_i represents the heterogeneity between centers of the overall underling baseline risk and the variance τ^2 of v_i represents the heterogeneity between centers of the overall treatment effect β_1. More discussion about this model and other survival proportional hazards models, in which $h_0(t)$ is known except for a finite-dimensional parameter, is given in Chapter 2.

1.5 Meta-analysis random effects model

Let θ be the central parameter of interest and assume there are $i = 1, 2, \ldots, g$ independent studies. Assume that Y_i is such that $E(Y_i) = \theta$ and let $s_i^2 = var(Y_i)$ be the variance of the summary statistic in the ith study. The random effects model assumes each study is associated with a different but related parameter. It postulates that each study summary statistic, Y_i, is drawn from a distribution with a study specific mean, θ_i and variance, s_i^2

$$Y_i|\theta_i, s_i^2 \overset{indep.}{\sim} N(\theta_i, s_i^2) \tag{1.2}$$

Furthermore, each study-specific mean, θ_i, is assumed to be drawn from some superpopulation of effects with θ and variance τ, with $\theta_i|\theta, \tau^2 \overset{indep.}{\sim} N(\theta, \tau^2)$. θ and

τ represent, respectively, the average treatment effect and inter-study variation. Although s_i^2 are assumed known, in reality the the estimated variance of θ_i, $\text{var}(\theta_i)$, is treated as if it were the true variance s_i^2. Details of this model and the fixed effects meta-analysis model are given in Chapter 3.

1.6 Objective of the study

The objective of this study is to investigate the validity of overall treatment effects estimates from meta-analysis. We do this by comparing treatment effects estimate from a meta-analysis model and a unified parametric proportional hazards model. In the meta-analysis, we use a random effect model to capture center variation and treatment-by-center variation. In the unified parametric proportional hazards model, we include a center random effect and treatment-by-center random effect. We study the bias and the mean square errors of the estimates from the two models. The thesis includes results of data analysis from a multi-center clinical trial and from a simulation study. In the simulation study, we vary the design matrix and the degree of center-by-treatment random effect and compare the treatment effects estimate from the two models.

1.7 Organization of the thesis

This thesis is organized as follows: Chapter 2 gives a brief review of survival proportional hazards models. It includes data classifications, basic quantities, development of a parametric proportional hazards model, and estimation methods. Chapter 3 discusses meta-analysis models and estimation methods. Chapter 4 reviews the ECOG data from Gray's paper and presents the results from the survival analysis and meta-analysis models discussed in Chapter 2 and 3. Chapter 5 presents simulation results. Chapter 6 presents a discussion and some problems for future research.

Chapter 2

Survival proportional hazards models

In Section 2.1 we explain survival data structure and the basic parameters that are used in modeling survival data. In Section 2.2 we define the proportional hazards model. First, we define the general model; then we show the presence of random effects in the model. Finally, we show how we use the NLMIXED procedure in SAS to get parameter estimates.

2.1 Survival data and basic quantities

Many types of non-life data, as well as some life data, are complete or fully observed. In many cases, life data contains uncertainty as to when exactly an event happened (i.e., when the unit failed). Data containing such uncertainty as to exactly when the event happened are termed censored data. There are at least three types of possible censoring schemes, right, left and interval censoring. Right censoring is the most common type of censoring. For right censored data all that is known for some individuals is a time beyond which the subject is still alive. In the second type of censoring, left censoring, a failure time is only known to be before a certain time. Interval censoring data reflects uncertainty as to the exact time the units failed within an interval [16, 17, 21].

The two basic quantities used in modeling survival data are the survival function and the hazard rate function. Let T denote a continuous non-negative random survival time, with probability density function $f(t)$ and cumulative distribution function $F(t) = Pr(T \leq t)$. The survival function is the probability of an individual surviving beyond time t. It is defined as

$$S(t) = Pr(T > t). \tag{2.1}$$

The hazard function, which is also called the hazard rate or conditional failure rate in reliability, is defined as

$$h(t) = \frac{f(t)}{S(t)}. \tag{2.2}$$

The cumulative hazard function, $H(t)$, is the accumulation of the hazard up to time t; that is, $H(t) = \int_0^t h(u)du$. The survival function and the cumulative hazard function are related through the identity $S(t) = \exp\{-H(t)\}$.

2.2 Parametric proportional hazards models

The proportional hazards model is the most popular model in survival data analysis. It relates the underlying hazard function, describing hazard changes over time, to the effect parameters, describing how hazard relates to other factors. The proportional hazards assumption is the assumption that non-time-dependent effect parameters multiply the time-dependent hazard. The effect parameters specified by the proportional hazards model can be reported as log hazard ratios for population members with specified covariate values differing by a unit amount. For the simplest case of treated and control, the proportional hazards model states that the hazard of treated subject over the hazard of control subject does not change over time [6, 16]. In general the proportional hazards

model can be written as

$$h_i(t) = h_0(t) \exp(X_i^{tr} \beta) \tag{2.3}$$

where $h_0(t)$ is the baseline hazard function corresponding to the hazard function of a subject with covariate information X_i equal to $\mathbf{0}$ and X_i^{tr} is the transpose of the vector X_i.

The baseline hazard function $h_0(t)$ either can be assumed to have a particular parametric form or can be left unspecified. In parametric proportional hazards models, we assume a particular parametric function for the baseline hazard $h_0(t)$. One of the most important models is the Weibull baseline hazard, with hazard function given by

$$h_0(t) = \lambda \rho t^{\rho - 1} \tag{2.4}$$

with $\lambda > 0$, $\rho > 0$. The scale parameter[1] λ provides information on the way the hazard (or density) is stretched out over time, and the shape parameter ρ parameterizes a variety of shapes for the density.

When ρ is smaller than 1, the hazard decreases monotonically with time. However, when ρ is larger than 1, the hazard increases monotonically with time. When ρ is equal to 1, this hazard (2.4) is the exponential hazard and is constant

[1]In this thesis we use λ as the scale parameter, but some people use $\lambda^{-1/\rho}$ as the scale parameter.

over time. For a fixed value of λ, ρ determines how fast the hazard function increases ($\rho > 1$) or decreases ($\rho < 1$)

Substituting $h_0(t)$ from (2.4) into (2.3), we get

$$h_i(t) = \lambda\rho t^{\rho-1} \exp(X_i^{tr}\beta) \tag{2.5}$$

The corresponding survival function and density function for subject i are given as

$$S_i(t) = \exp(-\int_0^t \lambda\rho s^{\rho-1} \exp(X_i^{tr}\beta)ds) = \exp(-\lambda t^{\rho} \exp(X_i^{tr}\beta)) \tag{2.6}$$

and

$$f_i(t) = h_i(t)S_i(t) = \lambda\rho t^{\rho-1} \exp(X_i^{tr}\beta) \exp(-\lambda t^{\rho} \exp(X_i^{tr}\beta)) \tag{2.7}$$

The random survival time T_i with this density has a Weibull distribution, denoted as $T_i \sim Weib(\lambda \exp(X_i^{tr}\beta), \rho)$. Thus all subjects following model (2.5) are Weibull distributed with the same shape parameter ρ but differ with respect to the scale parameter.

Other kinds of distributions that can be used instead of the Weibull include Gamma, lognormal, loglogistic, normal, and Gompertz. Details of these choices can be seen in survival analysis books such as the one by Klein and Moeschberger [16]. When the baseline hazard $h_0(t)$ in (2.3) is left unspecified, it has one parametric factor, $X_i^{tr}\beta$, and one factor which is not specified in a parametric way, $h_0(t)$, and we call the model semiparametric [6].

In survival data analysis, situations where survival times are not independent are frequently encountered. Such data tends to arise when different individuals have some features in common. An example of such data arises in a multi-center study, in which the survival experience of individuals from the same center may be more similar than that for individuals from different centers. In this kind of situation, we might represent the effect of the center by introducing a random effect in the model. In survival data analysis, a random effect is often referred to as a frailty [7, 17]. A frailty model is a random effect model for time-to-event data, where the random effect (the frailty) has a multiplicative effect on the baseline hazard function. The most common model for frailty is the shared frailty model where subjects in the same cluster all share the same frailty factor [8, 10, 17,22]. The shared frailty model is defined as

$$h_{ij}(t) = h_0(t) \exp(X_{ij}^{tr}\beta + u_i) \tag{2.8}$$

where $h_{ij}(t)$ is the conditional hazard function for the j^{th} subject from the i^{th} cluster, $h_0(t)$ is the baseline hazard, β is the fixed effects vector of dimension p, X_{ij} is the vector of covariates, and u_i is the logarithm of the shared multiplicative frailty parameter for the i^{th} cluster. The model can be written as

$$h_{ij}(t) = h_0(t) w_i \exp(X_{ij}^{tr}\beta) \tag{2.9}$$

where $w_i = \exp(u_i)$ is the frailty for the i^{th} cluster. We see that individuals in a group i with $w_i > 1$ are frail (high risk) and individuals in a group i with $w_i < 1$ are strong (lower risk). The w_i are an independent and identically distributed sample from a distribution with mean 1 and some unknown variance. A mathematically convenient choice for the distribution of the w_i is the one-parameter gamma distribution written as

$$f_W(w) = \frac{w^{1/\theta - 1} \exp(-w/\theta)}{\theta^{1/\theta}\Gamma(1/\theta)}$$

where Γ denotes the gamma function. With this frailty distribution, the mean of W is 1 and the variance of W is θ, so that large values of θ reflect a greater degree of heterogeneity among groups and a stronger association within groups.

Model (2.8) is a single frailty model, where there is only one random term. Next we discuss a model where there are two random terms within the same cluster. A typical example of such a model is a multi-center clinical trial with a frailty term describing the heterogeneity between centers and a second random term modeling the center-by-treatment interaction. The second random effect term describes the treatment heterogeneity between centers [1, 2, 17]. In this situation a model with two random effects shown in (1.1) can be used. That is

$$h_{ij}(t|u_i, v_i) = h_0(t) \exp(\sum_{k=1}^{p} X_{ijk}\beta_k + u_i + v_i X_{ij1}).$$

One factor limiting the practical use of models with random terms is the lack of a good estimation method. The two approaches used in practice are the EM algorithm, which treats the frailties as missing values, and the partial penalized likelihood (PPL) approach, which considers the density of the frailty as a penalty term in the likelihood.

Another technique of likelihood based analysis of random effect survival data is to use numerical integration of the random effects based on Gaussian quadrature [14]. This technique can be implemented in SAS using the NLMIXED procedure [15]. For the review of the first two methods (EM algorithm approach and PPL approach) and detailed discussion about the numerical integration of the random effects based on Gaussian quadrature, see Liu and Huang [14]. In this paper we use numerical integration, allowing multiple random effects in the random statement in the NLMIXED procedure. The complete code to analyze the ECOG lung cancer data using model (1.1) is given in the data analysis section.

PROC NLMIXED fits by maximizing an approximation to the likelihood integrated over the random effects using adaptive Gaussian quadrature integral approximation. The Dual quasi-Newton algorithm is the default optimization technique. Other integral approximation methods and optimization techniques can be specified. PROC NLMIXED gives parameter estimates along with their approximate standard errors based on the second derivative matrix of the likelihood function. Approximate standard errors are computed using the delta

method. NLMIXED can be used to analyze data that are normal, binomial, or Poisson or that have any likelihood programmable with SAS statements. In the model statement, the conditional distribution of the data given the random effects should be specified. One possibility is to use *'general(ll)'* which specifies a general log likelihood function constructed using SAS programming statements. The random statement defines the random effects and their distribution. The only available distribution for the random effects is *normal(m,v)* with mean m and variance v. The syntax can be written as:

```
random u ~ normal(0,s2u) subject=study;
```

where subject specifies the clusters. When two effects are present as in (1.1), they can be specified as follows

```
random b1 b2 ~ normal([0,0],[g11,g21,g22]) subject=study;
```

where [g11,g21,g22] is the lower triangle of the random-effects variance matrix listed in row order.

Chapter 3

Meta-analysis models

As mentioned in the Introduction, meta-analysis involves combining summary information from related but independent studies. In combining information from different trials, one should consider two possible types of models, fixed effects and random effects models. In a fixed effects model, the true treatment difference is considered to be the same for all trials. The standard error of each trial estimate is based on the variation of the sample within the trial. In a random effects model, the true treatment difference in each trial is itself assumed to be a realization of a random variable, which is usually assumed to be normally distributed [3, 4, 5].

In Section 3.1, estimation of the treatment difference in an individual trial is discussed. In Section 3.2, the meta-analysis fixed effects model and the random effects model are reviewed. A likelihood approach to the estimation of parameters

and its implementation in PROC MIXED in SAS is shown.

3.1 Estimating the treatment difference in an individual trial

Often the meta-analyst has little control over the choice of the summary measure because most of the decision is dictated by what was employed in the primary studies [4]. Meta-analysis may be performed on studies for which the available data are in the form of summary information from trial reports or publications or on studies for which individual patient data are available. The form of the data available from each study has implications for the meta-analysis. Three forms are commonly encountered [4, 13]. The first consists of an estimate of the treatment difference and its variance or standard error, which is the minimum amount of information needed. The second form of data is slightly more detailed, consisting of summary statistics for each treatment group, enabling a choice to be made between several different parameterizations of the treatment difference. The third form, individual patient data, allows the most flexibility. In this case it is possible to choose any sensible parametrization of the treatment difference and the method of estimation. In addition, if all the studies provide individual patient data, a more thorough analysis can be undertaken by employing a statistical

modeling approach. In this section, we describe the estimation of the treatment difference from an individual study for survival data. Estimation of the treatment difference for other kinds of data in an individual study, including binary data and normally distributed data, is described in [4, 13].

A model in survival analysis is expressed in terms of the hazard function or the survivor function. The hazard function is the limiting probability per unit time in which the event occurs, conditional on survival until time t. The survivor function is the probability that the event occurs after time t. Let $h_T(t|1)$ and $h_T(t|0)$ represent the hazard functions for the treated and control groups and $S_T(t|1)$ and $S_T(t|0)$ their respective survivor functions. Assume the proportional hazards model under which $h_T(t|1) = \exp(\theta X)h_T(t|0)$ for all t. The treatment difference can be measured using the log-hazard ratio

$$\theta = \log\left\{\frac{h_T(t|1)}{h_T(t|0)}\right\}. \tag{3.1}$$

3.2 Fixed and random effects models

3.2.1 Fixed study effects model

The fixed-effects model assumes each study measures the same underlying parameter and that there is no inter-study variation. Let θ be the central parameter

of interest and assume there are $i = 1, 2, \ldots, g$ independent studies. Let $\hat{\theta}_i$ be an estimate of θ from the ith study. The fixed effects model is given by

$$\hat{\theta}_i = \theta + \varepsilon_i, \tag{3.2}$$

where the ε_i are the error terms and are realizations of normally distributed random variables with expected value 0 and variance ξ_i^2 assumed known. That is

$$\hat{\theta}_i \sim N(\theta, \xi_i^2).$$

Although ξ_i^2 are assumed known, in reality the estimated variance of θ_i, $\text{var}(\theta_i)$, is treated as if it were the true variance ξ_i^2, that is, no allowance is made for the error in the calculated term $\text{var}(\theta_i)$. Let $w_i = 1/\xi_i^2$. When ξ_i^2 is assumed known, the maximum likelihood estimator (MLE) of θ is

$$\hat{\theta}_{MLE} = \frac{\Sigma_{i=1}^{g} w_i \hat{\theta}_i}{\Sigma_{i=1}^{g} w_i}. \tag{3.3}$$

3.2.2 Random study effects model

In a random effects model, it is assumed that the treatment difference parameters in the g studies are a sample of independent observations from an underlying distribution. In real situations, the underlying distribution is not known. If the

underlying distribution is normal, then it is completely described by its mean and its variance. In this paper, we restrict the underlying distribution to normally distributed random effects with mean θ and unknown variance τ^2 (i.e., $\hat{\theta}_i \overset{indep.}{\sim} N(\theta, \tau^2)$). It is important to note that, when the log-hazard ratio (3.1) is used for center treatment difference, the center-to-center variation is not due to the baseline risk difference between the centers but is due to treatment-by-center variation. This is because the baseline risk affects the hazard rates of both the treated group and control group in the same way, in particular in the same direction (decrease or increase); the log-hazard ratio remains unaffected. In general the random study effects model is given by

$$\hat{\theta}_i = \theta + v_i + \varepsilon_i, \tag{3.4}$$

for $i = 1, 2, ..., g$, where the v_i are normally distributed random effects with mean 0 and variance τ^2. The terms v_i and ε_i are assumed to be independently distributed. It follows that

$$\hat{\theta}_i \sim N(\theta, \xi_i^2 + \tau^2).$$

If both ξ_i^2 and τ^2 are assumed known, the maximum likelihood (MLE) of θ, based on data $\hat{\theta}_i$, is given by

$$\hat{\theta}_{(\tau)MLE} = \frac{\Sigma_{i=1}^{g} w_{i(\tau)} \hat{\theta}_i}{\Sigma_{i=1}^{g} w_{i(\tau)}} \tag{3.5}$$

where $w_{i(\tau)} = 1/\xi_i^2 + \tau^2$.

In reality, τ^2 is unknown and statistical methods are used to estimate it from data, either by a likelihood or a Bayesian approach. Here we discuss the likelihood approach and its implementation in SAS PROC MIXED. A Bayesian method can be seen in [4].

Assuming $w_i^{-1} = \xi_i^2$ is known, the contribution to the likelihood function from study i is

$$L(\theta, \tau^2; \hat{\theta}_i) = \frac{1}{\sqrt{2\pi(w_i^{-1} + \tau^2)}} \exp\left\{ \frac{-(\hat{\theta}_i - \theta)^2}{2(w_i^{-1} + \tau^2)^2} \right\}.$$

The likelihood function is given by the product of the individual study likelihood functions, and the log-likelihood function is given by

$$l(\theta, \tau^2; \hat{\theta}_i) = \text{constant} - \frac{1}{2} \sum_{i=1}^{g} \log(w_i^{-1} + \tau^2) - \frac{1}{2} \sum_{i=1}^{g} \frac{(\hat{\theta}_i - \theta)^2}{(w_i^{-1} + \tau^2)}.$$

Maximum likelihood estimates of τ^2 and θ can be found through an iterative scheme. Each iteration involves two steps: first τ^2 is treated as fixed and the value of θ, which maximizes the log-likelihood, is calculated. Then θ is treated as fixed and the value of τ^2, which maximizes the log-likelihood, is calculated.

Thus, the estimate of θ at the $(t+1)$th cycle of the iteration is given by

$$\hat{\theta}^*_{t+1} = \frac{\sum_{i=1}^{g} \hat{\theta}_i w^*_{it}}{\sum_{it}^{g} w^*_{it}}, \tag{3.6}$$

for $t = 0, 1, 2, \ldots$, where $w^*_{it} = (w_i^{-1} + \hat{\tau}^2_{M,t})^2$ and $\hat{\tau}^2_{M,t}$ is the ML estimate of τ^2 at the tth cycle of the iteration.

The ML estimate of τ^2 is given by

$$\hat{\tau}^2_{M,t+1} = \frac{\sum_{i=1}^{g} (w^*_{it})^2 \{(\hat{\theta}_i - \hat{\theta}^*_{t+1})^2 - w_i^{-1}\}}{\sum_{i=1}^{g} w^{*2}_{it}}. \tag{3.7}$$

An initial estimate of τ^2, $\hat{\tau}^2_{M,0}$, can be obtained using the method of moments to start the iteration process [13].

The maximum likelihood estimator of τ^2 does not take into account the information used in estimating θ, and, thus, will usually underestimate [13]. We follow [13] (page 94 to 96) in describing the alternative estimation steps leading to REML estimates in the model (3.4), which takes account of this loss of information. The REML log-likelihood function is based on the residual term $(\hat{\theta}_i - \hat{\theta}^*_{t+1})$, instead of the observation $\hat{\theta}_i$, and given by

$$l_R(\tau^2; (\hat{\theta}_i - \hat{\theta}^*_{t+1})_{i=1}^{g}) = \text{constant} - \frac{1}{2}\sum_{i=1}^{g} \log(w_i^{-1} + \tau^2) - \frac{1}{2}\sum_{i=1}^{g} \frac{(\hat{\theta}_i - \hat{\theta}^*_{t+1})^2}{(w_i^{-1} + \tau^2)}$$

$$-\frac{1}{2}\log(\sum_{i=1}^{g} \frac{1}{(w_i^{-1} + \tau^2)}).$$

REML estimates are found via a similar iterative scheme to that described before, where now $w_{it}^* = (w_i^{-1} + \hat{\tau}_{R,t}^2)^{-1}$. At the $(t+1)$th cycle of the iteration, (3.6) is used to calculate an updated estimate of θ. The REML estimate of the τ^2 at the $(t+1)$th cycle of the iteration is given by

$$\hat{\tau}_{R,t+1}^2 = \frac{\sum_{i=1}^{g}(w_{it}^*)^2\{g(\hat{\theta}_i - \hat{\theta}_{t+1}^*)^2/(g-1) - w_i^{-1}\}}{\sum_{i=1}^{g}(w_{it}^*)^2}. \tag{3.8}$$

3.3 SAS statements for meta-analysis

The ML and REML estimates can be found using PROC MIXED procedure in SAS. As indicated in [13], the following steps are needed to get REML estimates; first we need to create a diagonal variance matrix with the estimated within-study variance components as the diagonal elements. To do this, suppose that the values of i, θ_i, w_i have been entered into the dataset 'temp' under the variable names 'study', 'survtime', and 'w' respectively. Then the following code can be used to create the diagonal matrix [13].

```
data lung;
set temp;
var=1/w;
col=_n_;
row=_n_;
value=var;
```

Then we use the following PROC MIXED program.

```
proc mixed data=lung method=reml order=data;
class study;
model survtime=/solution;
random study/gdata=lung;
```

```
repeated diag;
```

ML estimates can be obtained by replacing 'method=reml' with 'method=ml'.

Chapter 4

Data analysis

4.1 Review of the ECOG EST 1582 study

In this Chapter, we analyze the treatment effect of the ECOG EST 1582 data, using the unified parametric proportional hazards model and meta-analysis model. The ECOG EST 1582 study, a multi-center clinical trial in small-cell lung cancer, compares two different chemotherapy regimens: a standard therapy consisting of cyclophosphamide, adriamycin and vincristine (CAV) and an alternating regimen (CAV-HEM), where cycles of CAV were alternated with HEM (hexamethylmelamine, etoposide and methotrexate). The primary end point is patient death. In addition to the two treatment arms, there are four important covariates that affected patient survival: presence or absence of bone metastases,

presence or absence of liver metastases, performance status at study entry, and weight loss prior to entry.

Gray [9] has made a detailed analysis of center variation in this data. He excluded centers contributing three or fewer patients, which left 570 patients from 26 centers, with the number of patients per center varying from 5 to 56 (median=18.5). In his analysis, he showed the presence of fixed treatment difference between the CAV-HEM and the CAV arms across the participating centers and discussed the possible causes for center variations in multi-center clinical trials. Despite the tightly structured protocols, in multi-center clinical trials, center-to-center variations can be caused by different standards of practice, types of supportive care, interpretation of dose modifications, patient populations, and so forth. He used a proportional hazards model shown in (4.1). let X_{ijk} be covariate k for subject j from center i, and in general assume g centers with n_i cases from center i, and $p-1$ covariates, with X_{ij1} the treatment variable. A piecewise constant model is used for the underlying hazard. Let $0 = t_0 < t_1 < \ldots < t_m$ be the fixed boundaries of time intervals, and set $I_l(t) = I(t_{l-1} < t \leq t_l)$. Then, the full model for the hazard for subject ij is

$$h(t|X_{ij}, \alpha, \beta, u_i, v_i) = \exp\left\{\sum_{l=1}^{m} \alpha_l I_l(t) + u_i + \sum_{k=1}^{p} X_{ijk}\beta_k + v_i X_{ij1}\right\} \qquad (4.1)$$

where the $\alpha = (\alpha_1, ..., \alpha_m)', \beta = (\beta_1, ...\beta_p)'$, u_i and v_i are unknown parameters, and $X_{ij} = (X_{ijl}, ...X_{ijp})'$. He used $m = 30$ intervals for the underlying hazard.

Gray [10] developed tests for group variation and showed that there is significant center variation among the treatment effects, but the variation among the baseline hazards is not significant. This means that there is significant institutional variation in patients' survival in the CAV-HEM arm, but not in the CAV arm.

4.2 Data analysis results

We fit unified parametric proportional hazards models discussed in Chapter 2 and the meta-analysis models discussed in Chapter 3 to the 570 patient dataset. For the unified proportional hazards model, we first fit three basic models and then tried other models with more parameters. The first three models are a model with no random effect (model (2.3)), a model with center random effect (model (2.8)), and a model with center and center-by-treatment random effects (model (1.1)). To perform meta-analysis in this dataset, the following procedures were used. First, to imitate the usual meta-analysis methods, we fit a parametric proportional hazards model (model (2.3)) for each center and collected the log-hazard ratio and the standard error. Then, using the center estimates and standard errors, we fit the random effect meta-analysis model (3.4).

Parameter estimates and standard errors of the three basic unified proportional hazards model are shown in Table 4.1. Model 0 has the five covariate

fixed effects but has no random effect. Model 1 has the five covariate effects and a random center effect. The third model, Model 2, has the five covariate fixed effects and two random effects, center and treatment-by-center interaction. The log-hazard ratio of the CAV-HEM treatment relative to the HEM treatment is -0.339 in model 0, -0.318 in model 1 and -0.332 in Model 2. This shows that the treatment estimate of the fixed regression effect did not change substantially after including the random effects. All four covariates in all three models are statistically significant at the 0.05 level. Although standard errors for the variance components are given, they should not be used directly to test against zero by assuming normality, since the null hypothesis lies on the boundary of the parameter space. The asymptotic null distribution of the change in deviance[1] is a mixture of χ^2_1 and χ^2_0 with equal weights 0.5 [18, 19]. The χ^2_0 distribution is the distribution which gives probability mass 1 to the value 0. If normality is assumed for the change in deviance, the P-values would be overestimated and the null hypothesis of no variance would be accepted too often. Here we look at the magnitude of the change in deviance when we add the random effects in the model. When center random effect is included (Model 1), the change in deviance is only 3.2 $(1105.0 - 1101.8)$ compared to Model 0. But when we include both the center and the treatment-by-center random effects (Model 2), there is a substantial amount of change in the deviance $(1101.8 - 1081.1 = 20.7)$ compared to

[1]Deviance is defined as the log-liklihood multiplied by (-2). Note that, some authors use "deviance" to refer to the difference between deviance of the original model and a reference model.

Table 4.1: Parameter estimates of the unified parametric proportional hazards regression models for the ECOG EST 1582 data.

	Model 0[a]		Model 1[b]		Model 2[c]	
	Estimate		Estimate		Estimate	
Parameter	(se)	P-value	(se)	P-value	(se)	P-value
Scale	1.133	<.0001	1.271	<.0001	1.337	<.0001
	(0.129)		(0.040)		(0.046)	
Shape	1.244	<.0001	1.141	<.0001	1.129	<.0001
	(0.037)		(0.142)		(0.144)	
Treatment	-0.339	0.0001	-0.318	0.0014	-0.332	0.0207
	(0.087)		(0.089)		(0.134)	
Bone	0.227	0.0155	0.241	0.0198	0.229	0.0300
	(0.094)		(0.097)		(0.099)	
Liver	0.293	0.0012	0.283	0.0050	0.293	0.0047
	(0.090)		(0.092)		(0.094)	
Perform	-0.523	<.0001	-0.550	<.0001	-0.577	<.0001
	(0.104)		(0.107)		(0.113)	
Weight	0.230	0.0097	0.223	0.021	0.270	0.0075
	(0.088)		(0.091)		(0.093)	
σ_u^2	-	-	0.032	-	0.011	-
			(0.026)		(0.023)	
σ_v^2	-	-	-	-	0.234	-
					(0.109)	
	-2LogLik=1105.0		-2LogLik=1101.8		-2LogLik=1081.1	

[a]Model 0 is a labeling for model (2.3).
[b]Model 1 is a labeling for model (2.8).
[c]Model 2 is a labeling for model (1.1).

Model 1, which indicates the importance of treatment-by-center random effect in the model. Next, we fit other models with more parameters and look at the change in the deviance. We consider Model 2 as our reference and compare the change in deviance. The results are shown in Table 4.2. Each model in the table has the treatment and the four fixed covariate effects. Model 3b, which is fitted with the treatment and the four fixed covariate effects, center random effect, treatment-by-center random effect and random bone effect has a deviance value

of 1073.1, which is a change of 8 from Model 2. Model 5, which is fitted with the treatment and the four covarite fixed effects, bone-by-treatment fixed effect, center random effect, treatment-by-center random effect, and bone random effect has a substantially smaller deviance from Model 2 ($1081.1 - 1069.6 = 11.5$). As shown in the table, the other models did not result in substantial change.

For the meta-analysis to have more data in each center, we decided to merge

Table 4.2: Proportional hazards regression and -2logLik values for ECOG data. (Fixed=Treatment and the four covariate fixed effects, u_i=center random effect, v_i=treatment random effect, b_i=bone random effect, l_i=liver random effect, p_i=performance random effect, w_i=weight random effect)

Model		-2logLik	# of parameters
Model 0:	Fixed	1105.0	7
Model 1:	Fixed+u_i	1101.8	8
Model 2:	Fixed+u_i+v_i	1081.1	9
Model 3b:	**Fixed+u_i+v_i+b_i**	**1073.1**	**10**
Model 3l:	Fixed+u_i+v_i+l_i	1080.2	10
Model 3p:	Fixed+u_i+v_i+p_i	1081.1	10
Model 3w:	Fixed+u_i+v_i+w_i	1080.5	10
Model 4:	Fixed+Bone*Treat+u_i+v_i	1077.0	10
Model 5:	**Fixed+Bone*Treat+u_i+v_i+b_i**	**1069.6**	**11**

some centers based on similarity of the unweighed center average of the covariates (bone, liver, performance, weight) using a method of divisive hierarchical clustering. Divisive hierarchical clustering is a top-down clustering method which starts with a single cluster containing all objects, and then successively splits resulting clusters until only clusters of individual objects remain [20]. At each stage of the divisive algorithm the cluster with the largest diameter that is available after the previous step is selected for the next split. The diameter of a cluster is the largest

dissimilarity between any two of its observations. Dissimilarities are calculated using the Euclidean distances, which are the root sum-of-squares of differences. To divide the selected cluster, the algorithm first looks for its most disparate observation, that is, the observation which has the largest average dissimilarity to the other observations of the selected cluster. This observation initiates a new group (splinter group). In subsequent steps within the same splitting stage, for each object of the old group we compute the average dissimilarity with the remaining objects, and compare it to the average dissimilarity with the object of the splinter group. The average dissimilarity of an object is the average dissimilarity to all other objects in the group. The algorithm reassigns observations that are closer to the splinter group than to the old group. The algorithm results in a division of the selected cluster into two new clusters. We used the `diana` function in the R `cluster` library, based on the center average vectors of the four covariates from the 26 centers. That is, the data matrix used in `diana` is organized in such a way that each row corresponds to a center, and each column corresponds to a covariate average, where there are 26 rows and 4 columns. We required each center to have at least 15 patients as a stopping criterion of the splitting process.

We were left with 18 centers where the sample size varied from 17 to 56. A proportional hazards model (2.3) on each center was fitted. The log-hazard ratio and standard error estimates for a CAV-HEM treatment relative to the CAV

treatment for each center were collected. Table 4.3 shows the treatment effect estimates for each center, and Figure 4.2 shows the corresponding forest plot. A forest plot is a graphical display designed to illustrate the relative strength of treatment effects in multiple studies addressing the same question. In meta-analysis forest plot is used as a means of graphically representing the results of each trials (e.g. [4] and [13]). The left-hand column lists the names of the studies and the right-hand column is a plot of the measure of effect (e.g. log hazard ratio) for each of these studies incorporating confidence intervals represented by horizontal lines. The measure of effects is often represented by square, and the area of each square is proportional to the study's weight in the meta-analysis. A vertical line representing no effect is also plotted. If the confidence intervals for individual studies overlap with this line, it indicates that at the given level of confidence, their effect size do not differ significantly from no effect for the individual study.

A negative estimate indicates that CAV-HEM treatment has a beneficial effect in longer survival. Thirteen centers have negative estimates and the remaining five centers have positive estimates. Four of the centers (7, 22, 25+41, and 26) showed a statistically significant difference of the CAV-HEM treatments. In the other fourteen centers there was no significant difference between the CAV-HEM and HEM treatments. A random effect meta-analysis model (3.4) was fitted. A restricted maximum likelihood (REML) estimate of -0.339 with standard error

of 0.126 was obtained. This shows that the unified proportional hazards model and the meta-analysis model give very similar estimates. The normality assumption in the meta-analysis model was checked graphically. Histogram, Q-Q plot and Empirical cumulative density function (CDF) plot, for the error terms based on 18 error points, are shown in Figure 4.2. None of these plots show violation of the normality assumption.

Table 4.3: Estimates of the log hazard ratio for treatment in each center for ECOG EST 1582 study

Study center	# of patients	$\hat{\theta}_i$	$se(\hat{\theta}_i)$	95% CI
1	21	-1.018	0.569	(-2.133, 0.098)
5+59	17	-0.782	0.616	(-1.989, 0.425)
7	18	-1.717	0.615	(-2.921, -0.512)
10	27	-0.545	0.480	(-1.485, 0.395)
13	46	0.303	0.352	(-0.387, 0.993)
18+61	31	-0.529	0.383	(-1.280, 0.222)
19+21	28	0.022	0.448	(-0.856, 0.901)
20	48	-0.073	0.298	(-0.656, 0.510)
22	56	-0.696	0.318	(-1.319, -0.073)
25+41	31	-0.974	0.475	(-1.904, -0.044)
26	22	-1.085	0.493	(-2.050, -0.119)
28+52+60	39	-0.432	0.354	(-1.126, 0.262)
33	27	-0.394	0.450	(-1.276, 0.488)
36+51	53	0.405	0.323	(-0.227, 1.037)
38	17	-0.435	0.809	(-2.021, 1.151)
39+49	42	-0.056	0.352	(-0.747, 0.635)
40	23	0.144	0.540	(-0.914, 1.202)
42	24	0.410	0.578	(-0.724, 1.543)

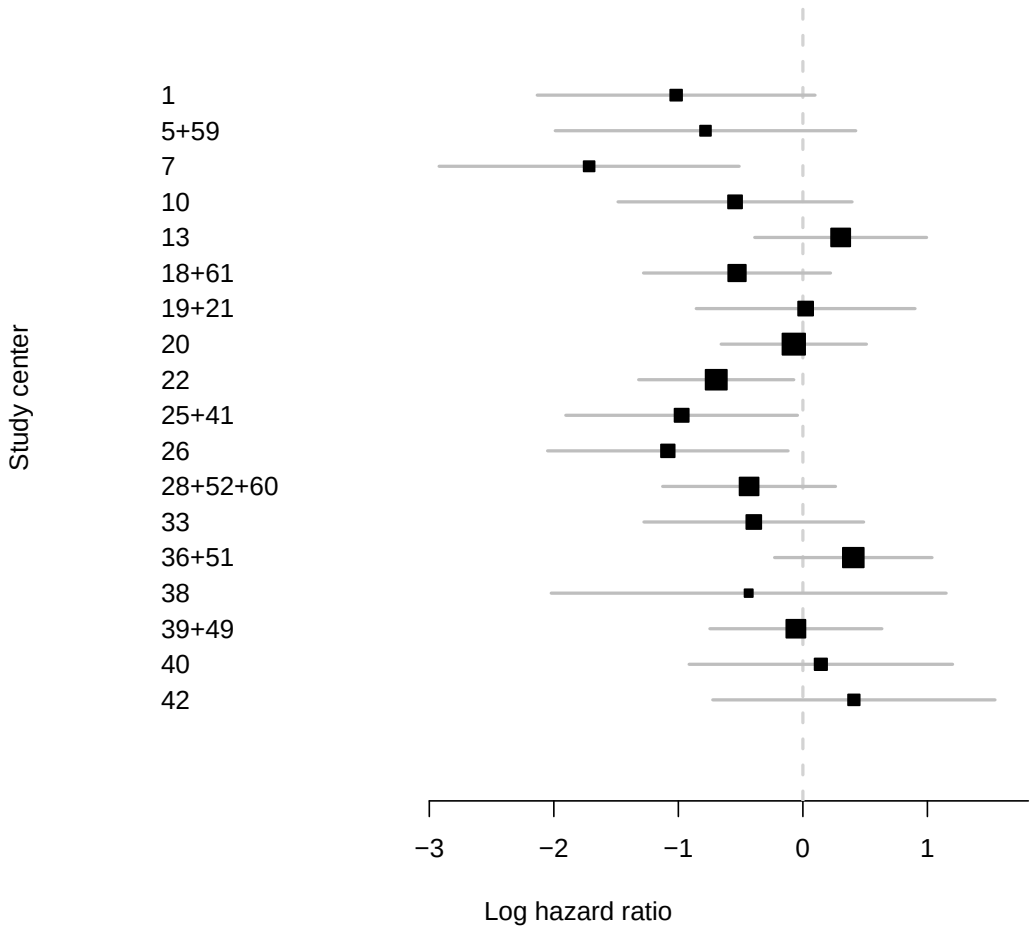
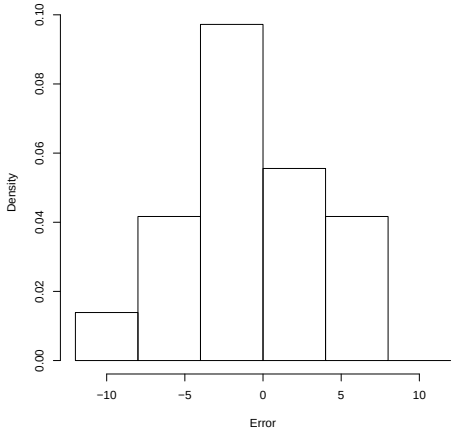
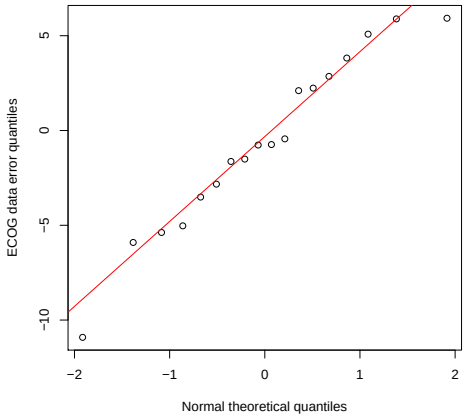


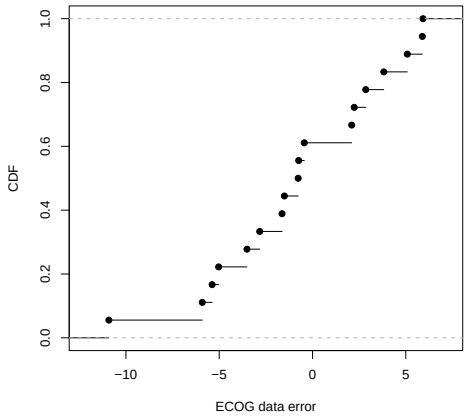
Figure 4.1: The log-hazard ratio for treatment with CAV-HEM relative to CAV. The square shows a study estimate and the size is proportional to the inverse of the variance of the log-hazard ratio.



(a) Histogram.



(b) Q-Q plot.



(c) Empirical CDF plot.

Figure 4.2: Error plots of the ECOG EST 1582 data from meta-analysis based on 18 residual error values.

4.3 SAS statements for unified analysis

The SAS code for the proportional hazards model (1.1) in Chapter 1 with Weibull baseline hazard rate (Model 2 in Table 4.2) is given below. The code for the other models in Table 4.2 can be programmed with slight modification.

```
proc nlmixed data=ECOG;
parms alpha=1 lambda=1 Treatcoff=-1 Bonecoff=1 Livercoff=1
          Performcoff=1 Weightcoff=1  s2u=1 s2v=1;
basehaz=alpha*lambda*Time**(alpha-1);
cumbasehaz=lambda*Time**alpha;
mureg=Treatcoff*treat+Bonecoff*bone+Livercoff*liver
          +Performcoff*perform+Weightcoff*weight+v*treat+u;
loglik0=-exp(mureg)*cumbasehaz;
if status=0 then loglik=loglik0;
if status=1 then loglik=log(basehaz)+mureg+loglik0;
model Time~general(loglik);
random u v ~ normal([0,0],[s2u,0,s2v]) subject=center;
run;
```

Chapter 5

Simulation

5.1 Data simulation

A simulation study was carried out to compare the performance of the unified approach (Chapter 2) and the meta-analysis approach (Chapter 3) in estimating the treatment effect of a multi-center clinical trial study. True parameters for data simulation were considered in line with the ECOG EST 1582 data used in the previous chapter. Data simulation was carried out in two steps. First, center random effects (u_i) and center-by-treatment interaction random effects (v_i) were generated independently from a normal distribution; i.e. $u_i \sim N(0, \sigma^2)$ and $v_i \sim N(0, \tau^2)$. We varied σ^2 to be 0.4 or 0.001 and τ^2 to be 0.3 or 0.16. Then independent survival times $T_{i1}, \ldots, T_{in_g}$ were generated from a Weibull distribu-

tion assuming the scale parameter $\lambda = 2$ and shape parameter $\rho = 1.4$. That is, $T_{ij} = [-\lambda^{-1} \log(a_{ij}) / \exp(\sum_{k=1}^{p} X_{ijk}\beta_k + u_i + v_i X_{ij1})]^{1/\rho}$ where a_{ij} is a random number generated from a uniform distribution U(0,1).

The number of centers (18) and the number of patients in each center was taken to be equal to those in the clustered ECOG EST 1582 dataset. Therefore the total number of patients is 570. We conducted the simulation study with and without covariates. When covariates are present they are either taken from the ECOG EST 1582 data or are simulated as independent and identically distributed binary variables with probability based on the design matrix explained in the next paragraph and the data is generated with coefficients, β_k, for bone, liver, performance and weight fixed at $\beta_2 = 0.2$, $\beta_3 = 0.3$, $\beta_4 = -0.5$ and $\beta_5 = 0.2$ respectively. These coefficients in the simulation are fixed based on the coefficients found in the ECOG EST 1582 data.

We used four different design matrices and X_{ijk} varies depending on the type of design matrix specified. For each scenario a single design matrix was fixed at the beginning of the simulation and used throughout the whole iteration. The first design matrix is the 'Original Design Matrix (ODM)', where the ECOG EST 1582 data design matrix was used for both the covariates and treatment. This means that each center in the simulation has exactly the same design matrix as the corresponding center in ECOG EST 1582 data. According to ODM, for a case j in center i in the simulation, the survival time is generated based on the X_{ij} in

the ECOG EST 1582 data and the a_{ij} are generated from a uniform distribution U(0,1). The second design matrix is 'Multinomial Design Matrix (MDM)', where for center i, the X_{ijk} are randomly generated from Bernouilli distributions with probability p_{ik} equal to the proportion of 1 for variable k in center i. In MDM treatment assignment is also randomly generated from a Bernouilli distribution with p_{i1} equal to the proportion of treated patients in center i. The third design matrix is the 'Uniform Design Matrix (UDM)', which is similar to the the second one, but the X_{ijk} for each center are randomly generated with probability p_k equal to the overall proportions of 1 for variable k in the ECOG EST 1582 data. Treatment assignment is also generated using probability p_1, which is the overall proportion of treated patients in the ECOG EST 1582 data. The fourth design matrix is the 'Absence of covariates Design Matrix (ADM),' where covariates are not included in the data generation and model fitting. The survival times were generated from $T_{ij} = [-\lambda^{-1}\log(a_{ij})/\exp(X_{ij1}\beta_1 + u_i + v_i X_{ij1})]^{1/\rho}$. Treatment assignment is taken from the ECOG EST 1582 data. ADM can be seen as the original design matrix without covariates.

The simulation programs were written using SAS 9.1 and executed on the SunOS 5.9 platform at the University of Maryland, College Park. The programs are set up in a batch mode so that they can run to completion without human interaction. We have experienced two main problems: First, the execution times were very long. A single simulation scenario with 1000 iterations and all the

four covariates included takes 14 to 18 hours. Second, the program receives a terminating signal from an unknown source and stops running before it reaches completion. In addition to these two main problems, we are unable to run more than one program at a time. When two or more programs submitted at a time, all programs stop before completing the full number of iterations. When the programs stop running before they reach the specified number of iteration, we re-run until we get the full iteration. Thus, all the simulation results in Section 5.2.1 are based on 1000 iteration and all the results in Section 5.2.2 are based on 500 iteration.

5.2 Results

The simulation results are organized in two sections, section I and section II. The first section has results from the basic simulation and each simulation is replicated 1000 times. Based on the findings in the first section, additional simulations were conducted, and results are presented in section II. In the second Section, each simulation is replicated 500 times. In both sections, the unified proportional hazards model (1.1) and random effects meta-analysis model (3.4) were fitted. For meta-analysis, first a parametric proportional hazards model without random effect (2.5) was fitted for each center. Then, the log hazard ratio and standard error from each center were collected and used to fit the random effect

meta-analysis model. For each simulation scenario, the mean, the median, the root mean squared error (RMSE) of $\hat{\beta}_1$, and the percentage of the models with significant treatment effect were collected and tabulated for both models. The percentage of the models with significant treatment effect was calculated as the proportion of simulation runs in which the upper bound of the 95% confidence interval of the log-hazard ratio was less than zero. Significant treatment effects always occurred with negative log hazard ratios.

5.2.1 Section I: Basic simulation

The results of the simulation studies using a unified model and meta-analysis model are summarized in Table 5.1 – Table 5.4. In general, the log-hazard ratio of treatment (β_1) is estimated well in both the unified and meta-analysis models for all design matrices. The results from the original design matrix are tabulated in Table 5.1. The unified model estimates the log-hazard ratio of treatment with negligible error; however, the meta-analysis model overestimates the treatment effect, and the variability of β_1, as measured by root mean square error (RMSE), is higher when compared to the unified model. The percentage of a significant treatment effect in meta-analysis is less than the percentage of a significant treatment effect in the unified analysis. For both methods, when the variance of treatment-by-center random effect is fixed higher in the data simula-

tion, the percentage of significance decreases as compared when the variance is fixed lower. For example, consider the case where the true value of β_1 is -0.4 and $var(u_i)$ is 0.3; when $var(v_i)$ is 0.3, the unified and meta-analysis models have significant treatment effects 68.4% and 66.2% of the time respectively; whereas, when $var(v_i)$ is 0.16, the unified and meta-analysis models have significant treatment effects 86.8% and 80.4% of the time respectively. The results from the multinomial design matrix are similar to the results with the original design matrix. The results are tabulated in Table 5.2.

The results of the UDM are displayed in Table 5.3. With the UDM the bias and RMSE are similar as in the cases with ODM and MDM. However, the percentage of a significant treatment effect in the meta-analysis is comparable to the unified analysis.

The results in ADM are shown in Table 5.4. The results both from the unified and meta-analysis models are similar by all criteria we compared. In both models, the treatment effect is estimated with negligible bias and with similar RMSE. The percentage of significant treatment effects is also similar in both models.

The significance of bias of $\hat{\beta}_1$ can be tested by constructing a 95% CI for β_1 as $\hat{\beta}_1 \pm (RMSE/\sqrt{R} \times 1.96)$, where R is the number of simulation iterations (1000 or 500). We constructed this confidence interval in Table 5.1 under the unified analysis, and only the first row has shown a significant bias. Under the meta-analysis all the estimated $\hat{\beta}_1$ values are clearly significantly biased. For the

rest of the simulation tables, the bias of $\hat{\beta}_1$ can be tested similarly.

The magnitude and direction of the biases of the estimate are shown graphically in Figures 5.1 and 5.2 when the true values are $\beta_1 = -0.4$, $var(u_i) = 0.4$ and $var(v_i) = 0.3$ under each design matrix. The estimated treatment effect, $\hat{\beta}_1$, values are shown by the vertical broken lines.

5.2.2 Section II: Additional simulation

In each of the additional simulations below, we fit the same unified analysis model (1.1) and random effect meta-analysis model (3.4) while varying the way the data is generated. From Section I, we have seen that when the covariates are not included, the estimates both from the unified and meta-analysis models are unbiased. However, when the covariates are included, bias is introduced in the meta-analysis estimate. Based on this initial result, we further investigated the relationship between the number of covariates and the magnitude of the bias. We conducted simulation studies in the presence of only one covariate (liver) and two covariates (liver and weight) under MDM and UDM. The results are shown in Table 5.5. The results show that as the number of covariates increase the bias in the meta-analysis estimate increases.

Treatment effect estimates, when there are one, two, and four covariates, were collected from the previous tables and displayed together in Table 5.6. We dis-

played this table for the purpose of convenience to look at the association between the number of covariates and the bias in one place.

In addition to the simulation scenarios discussed above, we conducted a comparison of the two methods by generating data with random terms in each covariate. The random terms generated independent of each other from a normal distribution with mean zero and variance 0.4, 0.3, 0.1 and 0.24 for bone, liver, performance, and weight respectively. This is a scenario in which we were fitting a misspecified model. The results are displayed in Table 5.7. The presence of random effects on covariates create bias in the unified analysis, where treatment effects are underestimated. The meta-analysis estimates appear unaffected and remain with the same magnitude and direction of bias as in section I.

Finally we simulated data with the four fixed effects, treatment fixed effect, center and treatment-by-center random effects, and bone-by-treatment random effect. The random term for the bone-by-treatment interaction is generated from a normal distribution with mean zero and variance 0.2. Multinomial and uniform design matrices used. The results are displayed in Table 5.8. Both unified and meta-analysis results are similar to the results in Section I.

Table 5.1: Simulation results with original design matrix from unified regression analysis and meta-analysis. (In the table, Unf=Unified regression, and Meta=Meta-analysis regression)

True Parameters			Estimated values					
			$\hat{\beta}_1$ (RMSE)		Median		% of sig. Effect	
β_1	σ^2	τ^2	Unf	Meta	Unf	Meta	Unf	Meta
-0.4	0.4	0.3	-0.412[a]	-0.456	-0.405	-0.451	68.4	66.2
			(0.161)	(0.197)				
-0.4	0.4	0.16	-0.398	-0.447	-0.399	-0.445	86.8	80.4
			(0.126)	(0.164)				
-0.4	0.001	0.3	-0.407	-0.452	-0.405	-0.452	71.5	64.3
			(0.161)	(0.202)				
-0.4	0.001	0.16	-0.404	-0.450	-0.405	-0.450	88.7	81.2
			(0.130)	(0.167)				
-0.2	0.4	0.3	-0.206	-0.229	-0.215	-0.233	21.9	23.3
			(0.253)	(0.261)				
-0.2	0.4	0.16	-0.197	-0.223	-0.200	-0.228	34.6	27.5
			(0.126)	(0.155)				
-0.2	0.001	0.3	-0.202	-0.231	-0.207	-0.226	28.9	24.0
			(0.158)	(0.190)				
-0.2	0.001	0.16	-0.206	-0.227	-0.203	-0.225	41.3	35.2
			(0.122)	(0.155)				

[a]This single $\hat{\beta}_1$ is significantly biased. The 95% CI is $(-0.422, -0.402)$, which does not include the true value of $\beta_1 = -0.4$.

Table 5.2: Simulation results with multinomial design matrix from unified regression analysis and meta-analysis

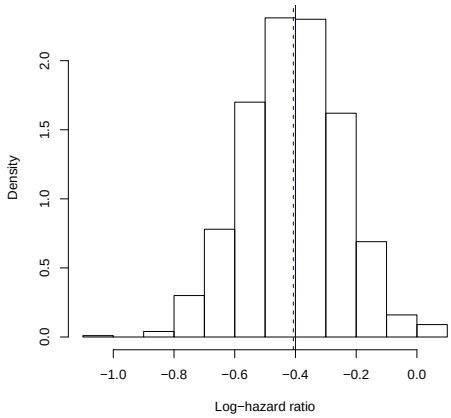
True Parameters			Estimated values					
			$\hat{\beta}_1$ (RMSE)		Median		% of sig. Effect	
β_1	σ^2	τ^2	Unf	Meta	Unf	Meta	Unf	Meta
-0.4	0.4	0.3	-0.408	-0.447	-0.409	-0.446	65.4	60.2
			(0.161)	(0.205)				
-0.4	0.4	0.16	-0.411	-0.450	-0.413	-0.453	84.1	76.6
			(0.134)	(0.173)				
-0.4	0.001	0.3	-0.409	-0.446	-0.412	-0.442	73.3	62.0
			(0.158)	(0.205)				
-0.4	0.001	0.16	-0.406	-0.449	-0.403	-0.450	87.4	77.8
			(0.130)	(0.170)				
-0.2	0.4	0.3	-0.200	-0.221	-0.199	-0.223	27.7	21.9
			(0.161)	(0.202)				
-0.2	0.4	0.16	-0.197	-0.223	-0.200	-0.228	34.6	27.5
			(0.126)	(0.155)				
-0.2	0.001	0.3	-0.206	-0.224	-0.206	-0.218	29.8	23.0
			(0.158)	(0.195)				
-0.2	0.001	0.16	-0.201	-0.212	-0.198	-0.214	32.5	25.5
			(0.134)	(0.170)				

Table 5.3: Simulation results with uniform design matrix from unified regression analysis and meta-analysis

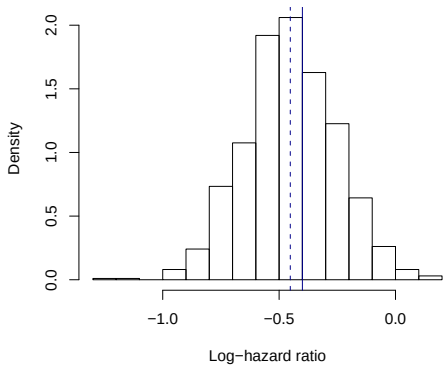
True Parameters			Estimated values					
			$\hat{\beta}_1$ (RMSE)		Median		% of sig. Effect	
β_1	σ^2	τ^2	Unf	Meta	Unf	Meta	Unf	Meta
-0.4	0.4	0.3	-0.409	-0.461	-0.406	-0.452	64.6	64.4
			(0.161)	(0.197)				
-0.4	0.4	0.16	-0.411	-0.450	-0.413	-0.453	84.1	76.6
			(0.134)	(0.173)				
-0.4	0.001	0.3	-0.405	-0.465	-0.408	-0.461	71.9	76.6
			(0.154)	(0.205)				
-0.4	0.001	0.16	-0.406	-0.461	-0.406	-0.460	85.5	78.8
			(0.138)	(0.179)				
-0.2	0.4	0.3	-0.215	-0.245	-0.216	-0.246	23.6	25.0
			(0.245)	(0.249)				
-0.2	0.4	0.16	-0.199	-0.225	-0.197	-0.222	36.2	30.2
			(0.130)	(0.164)				
-0.2	0.001	0.3	-0.208	-0.241	-0.214	-0.242	25.5	24.3
			(0.249)	(0.253)				
-0.2	0.001	0.16	-0.215	-0.250	-0.219	-0.246	38.6	36.7
			(0.134)	(0.176)				

Table 5.4: Simulation results with Absence of Covariate Design Matrix from unified regression analysis and meta-analysis

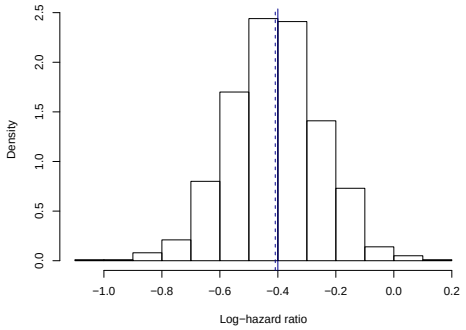
True Parameters			Estimated values					
			$\hat{\beta}_1$ (RMSE)		Median		% of sig. Effect	
β_1	σ^2	τ^2	Unf	Meta	Unf	Meta	Unf	Meta
-0.4	0.4	0.3	-0.406 (0.161)	-0.406 (0.167)	-0.407	-0.411	66.9	69.9
-0.4	0.4	0.16	-0.408 (0.134)	-0.404 (0.138)	-0.409	-0.407	82.2	83.5
-0.4	0.001	0.3	-0.406 (0.155)	-0.411 (0.164)	-0.407	-0.413	75.1	69.7
-0.4	0.001	0.16	-0.399 (0.126)	-0.403 (0.130)	-0.396	-0.399	88.5	86.1
-0.2	0.4	0.3	-0.203 (0.251)	-0.200 (0.259)	-0.201	-0.194	22.1	23.8
-0.2	0.4	0.16	-0.205 (0.235)	-0.199 (0.241)	-0.210	-0.205	31.3	32.5
-0.2	0.001	0.30	-0.201 (0.253)	-0.198 (0.261)	-0.203	-0.204	24.9	24.0
-0.2	0.001	0.16	-0.198 (0.239)	-0.195 (0.243)	-0.199	-0.197	32.40	30.90



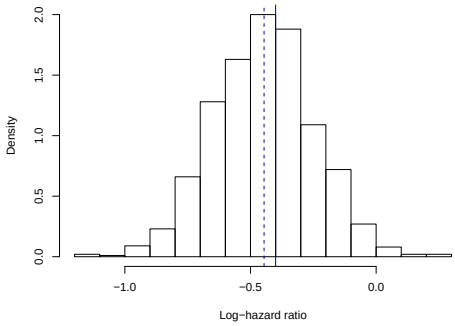
(a) ODM Unified analysis.



(b) ODM Meta-analysis.

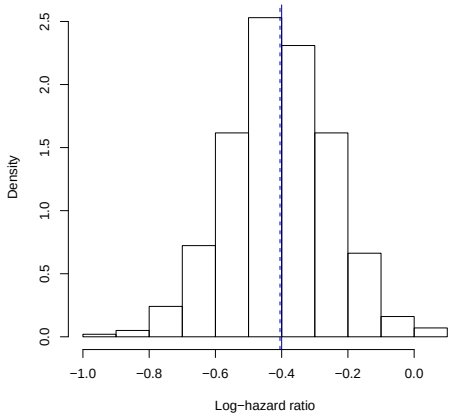


(c) MDM Unified analysis.

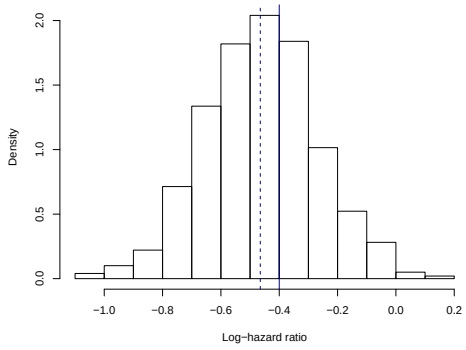


(d) MDM Meta-analysis.

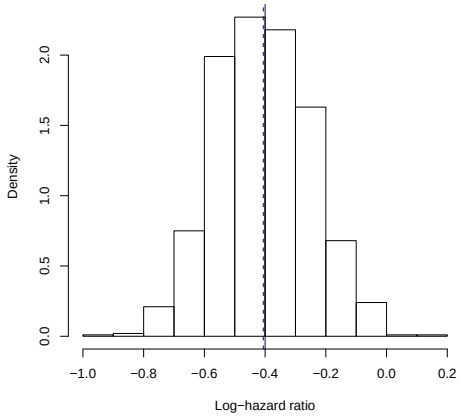
Figure 5.1: Histogram for simulation results of the log-hazard ratio for ODM and MDM. In both cases the data is generated with true values $\beta_1 = -0.4$, $\sigma^2 = 0.4$ and $\tau^2 = 0.3$. The vertical broken line shows the estimated log hazard ratio $\hat{\beta}_1$.



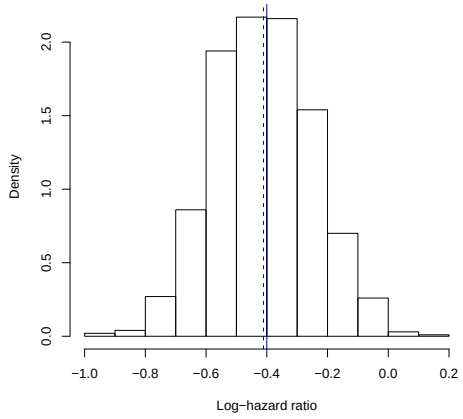
(a) UDM Unified analysis.



(b) UDM Meta-analysis.



(c) No Cov Unified analysis.



(d) No Cov Meta-analysis.

Figure 5.2: Histogram for simulation results of the log-hazard ratio for UDM and No Cov. In both cases the data is generated with true values $\beta_1 = -0.4$, $\sigma^2 = 0.4$ and $\tau^2 = 0.3$. The vertical broken line shows the estimated log hazard ratio $\hat{\beta}_1$.

Table 5.5: Parameter estimates from unified regression and meta-analysis with one and two covariates with varying design matrices.

	True Parameters			Estimated values					
				$\hat{\beta}_1$ (RMSE)		Median		% of sig. Effect	
Design Mtrx	β_1	σ^2	τ^2	Unf	Meta	Unf	Meta	Unf	Meta
One Covariate									
UDM	-0.4	0.001	0.3	-0.398	-0.409	-0.400	-0.406	72.6	66.6
				(0.158)	(0.176)				
	-0.2	0.4	0.16	-0.193	-0.195	-0.193	-0.190	33.2	29.0
				(0.134)	(0.141)				
MDM	-0.4	0.001	0.3	-0.402	-0.415	-0.398	-0.405	74.2	67.0
				(0.155)	(0.167)				
	-0.2	0.4	0.16	-0.201	-0.218	-0.206	-0.227	39.6	36.2
				(0.138)	(0.148)				
Two Covariates									
UDM	-0.4	0.001	0.3	-0.417	-0.453	-0.423	-0.458	77.0	73.6
				(0.155)	(0.182)				
	-0.2	0.4	0.16	-0.192	-0.202	-0.201	-0.197	37.0	30.6
				(0.130)	(0.148)				
MDM	-0.4	0.001	0.3	-0.411	-0.431	-0.424	-0.440	75.6	67.8
				(0.158)	(0.179)				
	-0.2	0.4	0.16	-0.202	-0.213	-0.200	-0.212	36.0	28.8
				(0.130)	(0.148)				

Table 5.6: Association between the number of covariates and bias in meta-analysis (Bias=Unf-Meta).

Design				Bias		
Mtrx	β_1	σ^2	τ^2	1 Cov	2 Cov	4 Cov
MDM	-0.4	0.001	0.3	0.013	0.020	0.037
	-0.2	0.4	0.16	0.017	0.011	0.026
UDM	-0.4	0.001	0.3	0.011	0.036	0.060
	-0.2	0.4	0.16	0.002	0.010	0.026

Table 5.7: Parameter estimates from unified regression and meta-analysis. Data generated including both fixed and all covariates random effect.

Design Mtrx	True Parameters β_1	σ^2	τ^2	Estimated values $\hat{\beta}_1$ (RMSE) Unf	Meta	Median Unf	Meta	% of sig. Effect Unf	Meta
UDM	-0.4	0.4	0.3	-0.369 (0.161)	-0.470 (0.212)	-0.364	-0.475	64.20	65.6
	-0.2	0.001	0.3	-0.182 (0.158)	-0.243 (0.205)	-0.191	-0.240	26.00	26.5
MDM	-0.4	0.4	0.16	-0.361 (0.134)	-0.446 (0.173)	-0.364	-0.453	82.4	76.40
	-0.2	0.001	0.3	-0.176 (0.158)	-0.213 (0.200)	-0.176	-0.217	26.00	20.8

Table 5.8: Parameter estimates from unified regression and meta-analysis. Data generated with all covariates fixed effect, treatment fixed effect, center and treatment-by-center random effects, and bone-by-treatment random effects.

Design Mtrx	True Parameters β_1	σ^2	τ^2	Estimated values $\hat{\beta}_1$ (RMSE) Unf	Meta	Median Unf	Meta	Treatment Effect(%) Unf	Meta
UDM	-0.4	0.001	0.3	-0.401 (0.158)	-0.445 (0.200)	-0.398	-0.437	73.4	61.8
	-0.2	0.4	0.16	-0.213 (0.141)	-0.233 (0.176)	-0.218	-0.244	40.6	31.8
MDM	-0.4	0.001	0.3	-0.425 (0.161)	-0.443 (0.197)	-0.421	-0.442	76.4	63.3
	-0.2	0.4	0.16	-0.215 (0.138)	-0.230 (0.170)	-0.222	-0.237	39.6	29.2

Chapter 6

Discussion

In this thesis we conducted data analysis and a simulation study in a multi-center clinical trial setting. The data that was used (ECOG EST 1582) comes from a multi-center clinical trial study where previous studies have shown the treatment effect to vary by center [9, 10]. The simulation study is conducted by varying parameters in models and designs chosen for similarity with the ECOG EST 1582 data.

The main goal of our research has been to evaluate meta-analysis treatment effect estimates by using a patient level data from a multi-center clinical trial study. Both on the real and simulated dataset we performed a unified analysis using the patient level data, and a meta-analysis using a summary information obtained from each center. In the meta-analysis, to imitate the usual practice,

we took the summary information from each center that might be published in a journal article and estimated overall treatment effect using meta-analysis. The estimates from the unified analysis and meta-analysis are compared in terms of bias, RMSE, and percentage of significant treatment effect.

For the unified analysis, we used a parametric proportional hazards model with Weibull baseline hazard. The model is fitted by likelihood methods using a numerical integration of the random effects via Gaussian quadrature, which is implemented in SAS NLMIXED procedure [14, 15]. For meta-analysis, the random effect meta-analysis model is used and REML estimates obtained using the PROC MIXED procedure in SAS.

Our simulation study has shown that under certain conditions (when covariates are included), meta-analysis yields slightly biased estimates for the overall treatment effect. The simulation study also showed, surprisingly, that the magnitude of the bias is directly related to the number of covariates. The reasons for this relationship between the covariates and magnitude of the bias is not known and needs further investigation. In all the scenarios we considered, when bias was present, the meta-analysis overestimated the treatment effect. If the bias had resulted in an underestimate of the treatment effect, it would have been more dangerous because of a tendency to incorrect hypothesis outcomes.

In the data analysis, we conducted the unified analysis and the meta-analysis on the ECOG EST 1582 data. The center level analysis shows that there is a

great variation in treatment effect between the centers; only a few centers have statistically significant treatment effect and some centers even have a positive point estimate although it is not statistically significant. But the overall treatment effect estimates from the unified analysis and the meta-analysis are very similar.

Based on our findings the following areas need further study: first, in our simulation study we only considered a few scenarios. It is difficult to make a general conclusion about the relationship of covariates and bias at this point. More thorough simulation studies could explore different conditions. For example, what will happen to the treatment estimates if the center and treatment-by-center random effects are correlated? How would the direction and the magnitude of the bias change if we choose different values for the center and treatment-by-center variance than we considered here when we generate the data? In this paper, we choose the center random effect variance as 0.001 or 0.4, and the treatment-by-center random effect variance as 0.16 or 0.3. Second, the simulation analysis performed in this thesis is not a theoretical proof. Thus, a theoretical explanation for the relation between the covariates and the bias needs to be conducted.

References

1. Yamaguchi T, Ohashi Y. 'Investigating center effects in a multi-center clinical trial of superficial bladder cancer', *Statistics in Medicine*, *18*, 1961-1971 (1999).

2. V. Rondeau, S. Michiels, B. Liquet and J. P. Pignon 'Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach', *Statistics in Medicine*, *27*, 1894-1910 (2007).

3. DerSimonian R, Laird N. 'Meta-Analysis in Clinical Trials', *Controlled Clinical Trials*, **7**, 177-188 (1986).

4. Sharon-Lise T. Normand 'Tutorial in biostatistics meta-analysis: Formulating, evaluating, combining, and reporting', *Statistics in Medicine*, **18**, 321-359 (1999).

5. Hard R.J., Thompson S.G. 'Detecting and describing heterogeneity in meta-

analysis.', *Statistics in Medicine*, **17**, 841-856 (1998).

6. Cox, D.R. 'Regression models and life tables(with discussions)', *Journal of the Royal Statistical Society*, Series B, **34**, 187-220 (1972).

7. Vaupel, J.W., Manton, K.G. and Stallared, E. 'The impact of heterogeneity in individual frailty on the dynamics of mortality', *Demography*, **16**, 439-454 (1979).

8. Hougaard P, 'Frailty models for survival data', *Lifetime Data Analysis*, **1**: 255-273 (1995).

9. Gray, R. 'A Bayesian analysis of institutional effects in a multicenter cancer clinical trial.', *Biometrics*, **50**, 244-253 (1994).

10. Gray, R. 'Tests for variation over groups in survival data.', *Journal of the American Statistical Association*, **90**, 198-203 (1995).

11. Ralph Bender, Thomas, Augustin, and Maria Blettner 'Generating survival times to simulate Cox Proportional hazards models.' *Statistics in Medicine*, **24**, 1713-1723 (2005).

12. Glass, G.V. 'Primary, secondary and meta-analysis of research.' *Educational Researcher*, **5**, 3-8 (1979).

13. Anne Whitehead. 'Meta-Analysis of Controlled Clinical Trials.' *John Wiley & Sons Ltd,* 2002

14. Lei Liu and Xuelin Huang 'The use of Gaussian quadrature for estimation in frailty proportional hazards models.' *Statistics in Medicine*, **27**, 2665-2683 (2008).

15. SAS Institute Inc 2004. SAS/STAT 9.1 User's Guide. Cary, NC: SAS Instite Inc.

16. John P.Klein, Melvin L.Moeschberger. 'Survival Analysis Techniques for Censored and Truncated Data, Second Edition.' *New York : Springer,* 2003

17. Hougaard P. 'Analysis of Multivariate Survival Data.' *New York: Springer* 2000

18. Steven G. Self and Kung-Yee Liang 'Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions' *Journal of the American Statistical Association*, **82**, 605-610 (1987).

19. Geert Verbeke, Geert Molenberghs 'Linear Mixed Models for Longitudinal Data' *Springer,* 2000

20. Leonard Kaufman, Peter J. Rousseeuw 'Finding groups in data : an introduction to cluster analysis' *New York : Wiley,* 1990

21. David Collett 'Modelling Survival Data in Medical Research, Second Edition.' *CHAPMAN & Hall/CRC* 2000

22. Luc Duchateau, Paul Janssen 'The Frailty Model' *New York : Springer* 2008

23. Joachim Hartung, Guido Knapp, Bimal K. Sinha 'Statistical Meta-Analysis with Applications.' *New York : Wiley* 2008