Joint Modeling of Binary Response and Survival Data in Clinical Trials

By

Jia Wang

A practicum report submitted to
Department of Public Health Sciences
In conformity with the requirements for
the degree of Master of Science

Queen's University
Kingston, Ontario, Canada
August, 2013

Supervisor: Dr. Bingshu Chen

Abstract

In clinical trials, it is often desirable to evaluate the effect of predictive factor such as marker response on the overall survival. However, the marker response and survival could usually be associated by some unobservable factors, in which case the conventional statistic models may not be appropriate. In contrast, joint modelling of marker response and survival data provides a less biased but more efficient inference by analyzing these two processes simultaneously. In this study, we focus on a special type of marker response: binary outcome, which is investigated together with overall survival data using a joint model linked by cluster-specific multivariate random effects. A modified penalized joint likelihood approach is proposed to make statistical inference for the joint model. A series of simulation studies are conducted to assess the finite sample performance of the proposed joint model and inference method in different scenarios, which is further compared with the separate model implemented by standard statistic functions. In the end, the proposed method is applied to NCIC Clinical Trials Group's HD.6 clinical trial data to explore the predictive effect of remission on survival in patients with Hodgkin's lymphoma. From the above study, we conclude that the proposed joint model outperforms the separate model when there exists a strong underlying association between the marker response and survival data. The modified penalized joint likelihood method yields reasonably accurate parameter estimates and provides a computationally efficient alternative to the existing inference approaches. The concern about underestimation of standard errors with the proposed method is also addressed using Jackknife resampling method.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Bingshu Chen, for offering me such a great opportunity to study in the Master of Science. I feel grateful for his continuous support and invaluable suggestion throughout my course study and practicum project. His guidance not only prompted my understanding of statistics but also aroused my enormous interest in statistics learning.

I also highly appreciate the help from faculty members in both Department of Public Health Sciences and Department of Mathematics and Statistics: Dr. Paul Peng, Dr. Devon Lin, Dr. Wenyu Jiang, Dr. Miu Lam, Dr. Dongsheng Tu, Dr. Will Pickett, Dr. Kristan Aronson and Dr. Will King. They made the journey of my graduate study enlightening and enjoyable.

I am thankful for my biostatistics fellows: Sahir, Evi and Ruiqi. I will never forget the wonderful experience being together with them during the past whole year. I also thank my parents and my girlfriend for their persistent support and encouragement throughout the years.

In the end, I would like to thank NCIC Clinical Trials Group for practicum opportunity, Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support and SharcNet for computation facility.

Abbreviations

ABVD	Doxorubicin (Adriamycin), bleomycin, vinblastine and dacarbazine
AFT	Accelerated failure time
ASE	Asymptotic standard error
CI	Confidence interval
CP	Coverage probability
EM	Expectation-maximization
ESE	Empirical standard error
EST	Estimate
GHQ	Gauss-Hermite Quadrature
GLM	Generalized linear model
GLMM	Generalized linear mixed model
JK	Jackknife
MLE	Maximum likelihood estimate
MPL	Multivariate penalized likelihood
MSE	Mean square error
PH	Proportional hazards
PPL	Penalized partial likelihood
SE	Standard error

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Literature Review

1.1 Marker Response and Generalized Linear Mixed Model

In clinical trials, treatment efficacy is one of the primary interests and is usually determined by measurement of marker response and/or survival. Thanks to the rapid development of technics in the areas of immunology, proteomics and microarrays in biomedical research (Pepe et al. 2001), it is possible to predict the survival of patients based on the result of marker response. With help of the early diagnosis, doctors may revise the treatment plan accordingly and improve the clinical outcome.

Many marker responses are measured repeatedly, which generates longitudinal data, for example, CD4 counts or viral load over time (Neuhaus et al. 2009). In fact, the study in relationship between longitudinal measurement and survival data is one of the hottest topics in clinical trial study. Over the past few decades many related papers have been published. Wu et al. 2012 provided a nice overview of recent studies regarding this research topic. In the research work presented here, however, we focus on another type of marker response: binary outcome at a single time point. In some clinical trials, it may not be feasible to make the repeated measurement of the early response and therefore the measurement does not involve time factor but simply yields single binary outcome (Lai et al. 2012). For instance, In NCIC Clinical Trials Group's HD.6 clinical trial (more details in Section 1.5), after applying either chemotherapy or radiotherapy treatment, an evaluation of Hodgkin's lymphoma progression was performed and patients were classified into either "remission" or "no remission" group. This binary remission status could play a predictive role in assessing the long-term survival of patients.

Mixed effect model, which includes both fixed effects and random effects, is commonly used to analyze clinical trial data involving repeated measurement or clustered data. The random effects, either individual-specific or cluster-specific, account for unobservable characteristics that are different for each individual or group, respectively. In linear mixed effect model, the response variable is assumed to follow a normal distribution. However, in many cases, the response is not necessarily normally distributed. For instance, if the outcome is in a binary scale, it cannot be analyzed using linear mixed effect model. In this study, we apply generalized linear mixed model (GLMM) with a logit link function for clustered data to model the binary marker response (more details will be discussed in Section 1.3.3).

Model (1.1) is an example of GLMM with a logit link, where for observation $i = 1,2, \dots N$, p_i is the probability of "response"; $\boldsymbol{X_i}$ is a $p \times 1$ vector of covariates for fixed effects and $\boldsymbol{\beta}$ is a $p \times 1$ vector of fix effects; $\boldsymbol{W_i}$ represents a $q \times 1$ vector of covariates for random effects and $\boldsymbol{a_j}$ denotes a $q \times 1$ vector of random effects for cluster j, which models the dependence of outcome attributed to cluster effects and explains between-cluster variability. If W_i is set to 1, then a_j is the random intercept for cluster j. According to this model, we assume that the marker response follows a Bernoulli distribution, conditional on the cluster-specific random effects. It is noted that there is no error term in (1.1) because different from linear regression, logistic regression models probability instead of actual value and the probability itself deals with imprecision of measurement.

$$logit(p_i) = log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{X_i^T \beta} + \boldsymbol{W_i^T a_j} \qquad (1.1)$$

1.2 Survival Analysis

The primary endpoint of interest in phase III clinical trial study includes events such as death, recurrence of disease or development of a new disease (Singh and Mukhopadhyay, 2011). Researchers usually deal with the follow-up time to development of the target event, which is also known as "survival data". One distinguishing feature of survival data is that it is often

subject to “censoring” because during the period of study the event of interest may not be observed for all the patients due to loss to follow-up or early study termination.

There are mainly three types of modeling strategies for survival data: nonparametric, parametric and semi-parametric analysis. As indicated by their names, this classification is based on the extent to which they rely on the parametric assumption.

Nonparametric analysis makes no assumption about the survival distribution. Therefore, it is useful when the true distribution is unknown or hard to approximate. However, since there is no parameter with finite dimension involved, it would be difficult to model data with multiple covariates using nonparametric analysis. One typical example for nonparametric analysis is Kaplan-Meier estimation, which is used to estimate the survival function from lifetime data (Kaplan and Meier, 1958).

If the association between survival and covariates is of interest, we can apply survival regression model using parametric or semi-parametric method. For parametric analysis, all the covariates in the model are specified and the hazard function is fully characterized in the model. Its main strength is the easy interpretation and estimation. In addition, parametric method makes possible the sophisticated analysis. However, if a parametric model is misspecified, the results could be misleading. Accelerated Failure Time (AFT) model, for instance, is one widely-used parametric model (Wei, 1992).

Besides nonparametric and parametric analysis, there is another type of analytical tool for survival data called semi-parametric method, which contains both parametric and nonparametric components. For example, in Cox Proportional Hazards model (Cox PH model), the regression term is specified while the baseline hazard function is unspecified (Cox, 1972). Compared with the parametric method, semi-parametric method is more flexible and makes less assumption about the distribution of survival data. In addition, even if the baseline distribution is not specified, the semi-parametric model still allows relative hazard between covariates to be estimated so that it is possible to evaluate the effects of the

explanatory variables. For these reasons, semi-parametric model gains its popularity in survival analysis. In this project, we decide to investigate the survival data using the semi-parametric Cox PH model with random effects (frailty). More details will be given in Chapter 2.

1.3 Joint Modeling in Survival Analysis

1.3.1 Overview of Joint Modeling

In the previous two sections, we discussed about two main study interests in the process of clinical trial. In practice, it is common to collect both short-term marker response data (tumor size, remission/non remission, cancer cell count, etc.) and long-term survival data (time to event). A lot of previous methods analyzed these two types of data separately. However, the separate analysis may not be adequate and can lead to bias and inefficient estimation because in many cases these two outcomes are not independent and could be linked by certain unobserved process (Neuhaus et al. 2009).

To overcome this problem, joint modeling of marker response and survival data was introduced. The joint model analyzes these two types of outcomes simultaneously and is able to reduce the bias of parameter estimates and in the meanwhile improve the efficiency of statistical inference, which, as a result, has attracted great attentions from clinical trial researchers (Wu et al. 2012). One type of joint models that has been extensively studied in recent years is the joint analysis of longitudinal measurement and survival data (Henderson et al. 2000; Ding et al. 2008 and Li et al. 2010). Many different joint modeling approaches have been proposed so far, among which a typical strategy for model setting is to link with shared random effects or joint distribution of random effects the mixed effect model for longitudinal data and AFT or Cox frailty model for survival data (Wu et al. 2012). In this case, longitudinal data and survival data are considered to be independent conditional on the random effects and observed covariates.

As mentioned above, majorities of existing joint survival models are dealing with longitudinal measurement and survival data. Little research has addressed another typical and important marker response: binary outcome. Inoue et al. 2002 developed parametric mixture model for binary tumor response and time to event under Bayesian framework. Thereafter, Lai et al. 2012 proposed a semi-parametric model that connected the binary response and survival endpoint based on frequentist approach in sequential designing study of phase II – phase III cancer clinical trial. However, as far as we know none of the existing research has adopted the joint modeling strategy for binary response using shared or joint random effects. In this project, we propose a joint model for binary response and survival data linked through the joint distribution of cluster-specific random effects. With the proposed model, we aim to investigate the potential association between treatment and binary marker response, and association between treatment and survival. More importantly, we would like to know if the marker response could be used as potential surrogate for survival, in other words, if the early response to the treatment is predictive for the time to events. It should be noted that even if the patients who respond to the treatment have longer survival time, we cannot conclude that there is a causal relationship between the response and survival. As pointed out by Anderson et al. 2008, response to the treatment could be simply a marker of patients with prognostically favorable characteristics, which could be the true reason for longer survival.

1.3.2 Cluster-Level Random Effect

Most clinical trials nowadays are conducted in multi-centre manner. The multi-centre designing ensures the adequate sample size and generalizability of study (Glidden et al. 2004). The patients within each centre potentially share similar characteristics. For example, they could have similar socioeconomic status and expose to the same environmental factors. In addition, the treatment quality tends to be consistent within each centre while vary among different centres. All these underlying dependences within each centre are not able to be captured by fixed-effect covariates and failure to consider these dependences may cause misleading results. Accordingly, it is crucial to introduce cluster-level (or called centre-level)

random effects to account for the dependence of patients within each centre (Glidden and Vittinghoff, 2004).

Another reason for choosing centre-level random effect in this study is that individual-level random effect is unidentifiable in our proposed joint model. In the case of longitudinal data modeled by linear mixed regression, individual-level random effect is usually implemented to incorporate the correlation of repeated measurements within individual. In the context of logistic regression, however, we cannot estimate the individual-level random effect in the linear predictor like the ordinary linear regression because the variance of individual random effect is unidentified (Lancaster et al. 2004). Therefore, in our study the two regression models are linked by joint distribution of centre-level random effects instead of individual-level random effects. The centre-level random effects are assumed to follow bivariate normal distribution as a reasonable approximation (Breslow et al. 1993). In addition, normally distributed random effects allow us to handle negative association, which is not possible for the Gamma distributed frailty (Ripatti et al. 2000).

1.4 Joint Inference and Penalized Partial Likelihood (PPL) Method

One simple statistical inference approach for joint modeling with shared unobserved variables is the two-step (or called two-stage) method (Dafni and Tsiatis 1998). The first step of this method is to estimate the shared latent variables via the first sub-model based on the observed data; in the second step, the inference is made via the second sub-model with the shared latent variables substituted by their estimated values from the first step as if they are observed data. The two-stage method is computationally simple and easy to be implemented with standard software. Nevertheless, since it fails to make inference on the two processes simultaneously, the results could be biased (Ye et al. 2008).

A more robust and widely-used inference approach is the joint likelihood method, which, as

indicated by its name, utilizes the joint likelihood of all observed data to conduct statistical inference simultaneously for the joint model. It is able to reduce the bias and yield efficient inference if the joint model assumption is appropriate. However, the maximum likelihood analysis often involves intractable high-dimensional integrals due to unobservable random effects and it is more so if the frailty is normally distributed. For instance, Expectation-Maximization (EM) algorithm is a typical method to compute maximum likelihood estimates (MLE) of unknown parameters for the joint model with latent variables by iterating between expectation-step and maximization-step (Wulfsohn et al. 1997). In this method, expectation-step may involve intractable integral. Several numerical integration approaches have been developed to deal with the intractable integral, such as Gauss-Hermite Quadrature (GHQ) and Monte Carlo method, which, however, could be very computationally intensive. It is worth noting that the computation can be further complicated if the mixed effect model is non-linear (e.g. the generalized linear mixed model used in our study).

Alternatively, a more computationally efficient approximation approach: Laplace approximation can be applied to tackle the challenge of intensive computation brought by numerical integration (Breslow and Clayton, 1993; Rizopoulos et al. 2009). Compared with GHQ and Monte Carlo method, Laplace approximation provides computational advantage and if the number of repeated measurements within individual or the number of observations within each cluster is not too small, it can approximate the integral reasonably well.

Following the Laplace approximation of likelihood function, Ripatti et al. 2000 introduced penalized partial likelihood (PPL) method for multivariate frailty model to conduct statistical inference. Thereafter, Ye et al. 2008 compared PPL method and EM algorithm in the context of joint survival model and found out that the performance of the method based on penalized joint likelihood was comparable to the EM algorithm in different simulated scenarios. In the meanwhile, PPL method offers several advantages: (1) As mentioned earlier, EM algorithm may not be feasible for high dimensional random effects because the E-step has to deal with complex integration. PPL method based on Laplace approximation, however, can avoid high dimensional integration. (2) The average computing time for parameter estimation with EM

algorithm is much longer than penalized joint method. PPL method is more time-efficient. (3) Computing the asymptotic variance of parameter estimates could be extremely time-consuming with EM algorithm. However, for penalized joint method, the information matrix is simply the by-product of Newton-Raphson method, which can be easily used to compute variance estimates (Ye et al. 2008). In this study, we develop a modified penalized joint likelihood method for the proposed joint model.

1.5 A Motivating Example: Hodgkin's Lymphoma Clinical Trial

1.5.1 Data Outline

The joint model for binary response and survival data proposed in the report was motivated by NCIC Clinical Trials Group's HD.6 clinical trial (Meyer et al. 2012), which compared the treatment effect of doxorubicin (Adriamycin), bleomycin, vinblastine, and dacarbazine (ABVD) alone versus subtotal nodal radiation therapy with or without ABVD treatment on stage IA or IIA nonbulky Hodgkin's lymphoma.

This clinical trial is a "multi-centre randomized controlled trial", in which 399 eligible patients were recruited in 29 research centres located in Canada and the United States. Based on the patient risk status (defined in Meyer et al. 2005), patients were stratified into either "favorable risk cohort" or "unfavorable risk cohort". Thereafter, all the patients in each cohort were randomized into two treatment arms: ABVD alone group and radiotherapy group with median follow-up period of around 12 years. In radiotherapy group, patients with favorable risk profile received subtotal nodal radiation therapy alone while the ones with unfavorable risk profile received two cycles of ABVD treatment followed by radiation therapy; in ABVD-alone group, patients with favorable or unfavorable risk profile received four cycles of ABVD. Restaging of disease was performed at the end of treatments (six months after randomization) to evaluate the "remission status" of patients. Those who showed no clinical or radiological evidence of Hodgkin's disease were classified into

“remission” group while others were classified into “non-remission” group (Lister et al. 1989). Correspondingly, patients who had remission were considered to be “response-positive” while the ones who did not were sd“response-negative”. Information regarding the characteristics of patients such as age and gender were also recorded prior to the commencement of clinical trial. Eventually, they found that ABVD alone therapy group had higher rate of overall survival (94%) compared with radiotherapy group (87%). The higher death rate in radiation therapy group could be associated with other diseases caused by radiation treatment. Another interesting finding was that the ones who had remission displayed a higher rate of freedom from disease progression (94%) and a trend towards higher rate of overall survival (98%) as compared with the ones without complete remission (81% and 92%, respectively), which rises an intriguing question: is it possible to use the restaging of disease (remission / non-remission) as a potential surrogate for survival rate and consequently help clinicians to make subsequent treatment decision at the early stage of the trial? This is one of the main questions we will address in this project.

1.5.2 Guarantee-Time Bias

Many cancer clinical trials involve measuring response to treatment as a routine procedure during the follow-up period. In HD.6 clinical trial study, the restaging of disease was used to evaluate the response to treatment. Giobbie-Hurder et al. 2013 pointed out that if the comparison of survivals or other types of events of interest was directly made across the groups defined by treatment response, “guarantee-time bias” could be introduced. This is because patients have to live long enough to be evaluated as either responder or non-responder. If the patients who died before the classifying event are considered to be non-responders, it will lead to the bias in favor of the positive effect of response to treatment in elongating survival time. The magnitude of influence caused by guarantee-time bias is positively associated with the time when the response to treatment occurs relative to the primary outcome event (e.g. survival) and the number of the early outcome events that occur prior to the classifying event (Giobbie-Hurder et al. 2013).

There are two commonly used analytical techniques to reduce the bias: (1) extended Cox PH model with time-dependent covariate; (2) conditional landmark analysis (Anderson et al. 2008). The first method has higher statistical power than the second one. However, it requires the track of the time when classifying event occurred during the follow-up period, which, in the case of HD.6 clinical trial, is not available. Thus, in this report, we apply the second method. Landmark time here is defined as the time when the evaluation of response to treatment (or restaging of disease) occurred, which is six months after randomization. All the patients who died or were lost to follow-up before the landmark time are excluded from the study. With this method, the data analysis is only based on a subset of original data, which causes some loss of statistical power. However, the statistical inference based on conditional landmark analysis is unbiased by the guarantee time (Anderson et al. 2008, Giobbie-Hurder et al. 2013). In addition, the simplicity and easy interpretation of landmark analysis make it a useful tool to explore the role of treatment response in the presence of guarantee-time bias (Giobbie-Hurder et al. 2013).

1.6 Objective and Outline

In this project, we construct a joint random effect model to analyze the binary response and survival data simultaneously, which, as far as we know, is the first study of its kind. In addition, a modified multivariate penalized likelihood method for parameter and variance estimation is developed to make the computation easier than the existing methods and in the meanwhile yield reasonably accurate estimates.

The rest of report is organized as follows: in Chapter 2, the proposed joint model for binary response and survival data and its likelihood function are first constructed, followed by a detailed description of proposed multivariate penalized likelihood (MPL) method for parameter and variance estimation. In Chapter 3, simulated data with different parameter specifications are generated and performance of the proposed joint random effect model is compared with the separate model with random effects under different settings of parameter

values. In Chapter 4, the proposed model and inference method are applied to the actual clinical dataset from Hodgkin's lymphoma clinical trial. The roles of fixed effects and centre-level random effects are investigated. Summary and future directions are presented in Chapter 5.

Chapter 2

Joint Model and Inference Method

2.1 Joint Model for Binary Response and Survival Data

The proposed joint model consists of two sub-models: (1) Generalized linear mixed Model (GLMM) with logit link for binary treatment response; (2) Cox proportional hazards model (Cox PH model) with frailty term for survival data. These two sub-models are linked by centre-specific random effects, which are normally distributed with mean **0** and variance-covariance matrix $\sum$. Here we assume that the censoring time is non-informative and the covariate data collected from different patients are independent. The notations used in the model and corresponding interpretations are summarized as follows:

m	Total number of centres.
j	j=1,2,..,m. j-th centre.
n_j	Total number of patients in j-th centre.
i	i=1,2,..,n_j. i-th patient in j-th centre.
Y_{ij}	Treatment response for i-th patient in j-th centre (remission=1; no remission=0). It follows Bernoulli(π_{ij})
$\boldsymbol{G}_{ij}$	A $p \times 1$ vector of fixed-effect covariates in GLMM with first element being 1.
X_{ij}	Survival time for i-th patient in j-th centre. $X_{ij} = min(T_{ij}, C_{ij})$; T_{ij} is event time; C_{ij} is censoring time.
$\boldsymbol{W}_{ij}$	A $q \times 1$ vector of fixed-effect covariates in Cox frailty model with the first element being Y_{ij}.
δ_{ij}	$I(T_{ij} \leq C_{ij})$; censoring indicator for i-th patient in j-th centre.
$\lambda_0(t)$	Baseline hazard function.
$\Lambda_0(t)$	Cumulative baseline hazard function.
u_j	Random intercept for the j-th centre in GLMM. It follows a normal distribution $N(0, \sigma_u^2)$.
v_j	Random intercept for the j-th centre in Cox frailty model. It follows a normal distribution $N(0, \sigma_v^2)$.
σ_{uv}	Covariance of random effects u and v; σ_{uv}=$cov(u, v)$.
$\boldsymbol{\theta}$	Variance components; $\boldsymbol{\theta}$=$(\sigma_u^2, \sigma_v^2, \sigma_{uv})$.
Σ	Variance-covariance matrix for random effects u and v.
$\boldsymbol{\beta}$	A $p \times 1$ vector of fixed effects in GLMM.
$\boldsymbol{\gamma}$	A $q \times 1$ vector of fixed effects in Cox frailty model.
$\boldsymbol{\Psi}$	A vector of all the parameters in joint model; $\boldsymbol{\Psi}$=($\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \lambda_0(t), \Lambda_0(t)$).

Generalized linear mixed model for treatment response Y_{ij} has the structure as showed in model (2.1). It is noted that there is no error term in this model as in linear model because the variability of the binary outcome has been accounted for by modeling the probability of the event.

$$P\left(Y_{ij}=1\,\middle|\,\boldsymbol{G}_{ij},\boldsymbol{\beta},u_j\right)=\frac{e^{\boldsymbol{G}_{ij}^{T}\boldsymbol{\beta}+u_j}}{1+e^{\boldsymbol{G}_{ij}^{T}\boldsymbol{\beta}+u_j}} \qquad (2.1)$$

We assume a Cox proportional hazards model with random effect (frailty) for the survival data (model (2.2)), which is an extension of widely-used Cox PH model by introducing centre-specific random effect v_j. As mentioned above, $\boldsymbol{W}_{ij}$ is a vector of predictors with the first element being observed treatment response variable Y_{ij}.

$$\lambda\left(t\middle|\boldsymbol{W}_{ij},\boldsymbol{\gamma},v_j\right)=\lambda_0(t)e^{\boldsymbol{W}_{ij}^{T}\boldsymbol{\gamma}+v_j} \qquad (2.2)$$

The two sub-models are linked by joint centre-level random effects u_j and v_j (j=1,2,..,m). We assume that they follow a zero-mean bivariate normal distribution (2.3), where σ_u^2 and σ_v^2 determine the magnitude of variation of centre-level random effects, σ_{uv} represents the strength of association between two random effects and Σ denotes the variance-covariance matrix of random effects. When $\sigma_{uv}=0$, there is no association between two random effects and the joint model is reduced to separate model. As we can see, the association between binary treatment response and survival is determined by regression coefficient of covariate Y_{ij} in Cox frailty model as well as the covariance σ_{uv}. We assume that these two sub-models are independent given the covariates and random effects.

$$\boldsymbol{a}_j=\begin{pmatrix}u_j\\v_j\end{pmatrix}\sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma_u^2 & \sigma_{uv}\\ \sigma_{uv} & \sigma_v^2\end{pmatrix}\right)=N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\Sigma\right) \qquad (2.3)$$

2.2 Likelihood Function

The joint likelihood function for parameters based on observed data can be written as:

$$L(\boldsymbol{\Psi};\boldsymbol{Y},\boldsymbol{X})=\prod_{j=1}^{m}\prod_{i=1}^{n_j}f(Y_{ij},X_{ij}|\boldsymbol{\Psi})=\prod_{j=1}^{m}\prod_{i=1}^{n_j}\int_{\boldsymbol{a}}f\left(X_{ij}\middle|Y_{ij},\boldsymbol{a},\boldsymbol{\Psi}\right)f\left(Y_{ij}\middle|\boldsymbol{a},\boldsymbol{\Psi}\right)f(\boldsymbol{a}|\boldsymbol{\Psi})d\boldsymbol{a} \qquad (2.4)$$

where $f\left(X_{ij}\middle|Y_{ij},\boldsymbol{a},\boldsymbol{\Psi}\right)$ is the full likelihood function of Cox frailty model, $f\left(Y_{ij}\middle|\boldsymbol{a},\boldsymbol{\Psi}\right)$ is

the likelihood function of GLMM and $f(\boldsymbol{a}|\boldsymbol{\Psi})$ is the density function of bivariate normal distribution. The joint likelihood function can be rewritten as:

$$L(\boldsymbol{\Psi};\boldsymbol{Y},\boldsymbol{X}) = \int \prod_{j=1}^{m} \prod_{i=1}^{n_j} \left\{ \left[\lambda_0(X_{ij}) \exp\left(\boldsymbol{W}_{ij}^T\boldsymbol{\gamma} + v_j\right)\right]^{\delta_{ij}} \exp\left[-\Lambda_0(X_{ij}) \exp\left(\boldsymbol{W}_{ij}^T\boldsymbol{\gamma} + v_j\right)\right] \right.$$
$$\left. \left(\frac{e^{\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+u_j}}{1+e^{\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+u_j}}\right)^{Y_{ij}} \left(\frac{1}{1+e^{\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+u_j}}\right)^{1-Y_{ij}} \right\} \prod_{j=1}^{m} \left\{ \frac{1}{\sqrt{(2\pi)^2|\sum|}} \exp\left(-\frac{1}{2}\boldsymbol{a}_j^T\sum{}^{-1}\boldsymbol{a}_j\right) \right\} d\boldsymbol{a} \qquad (2.5)$$

Maximization of likelihood function (2.5) is difficult because it has to deal with integration with respect to random effects $\boldsymbol{a}$. Therefore, we apply first-order Laplace method for integral approximation of function (2.5), as described in Breslow et al. 1993. After ignoring the constant term and taking the logarithm, we obtain the approximate marginal log likelihood function (2.6). $\tilde{\boldsymbol{a}}(\tilde{u},\tilde{v})$ denotes the solution to $K'(\boldsymbol{a}) = 0$, where $-K(\boldsymbol{a}) = l_3 + l_4$ (see function 2.6). Ripatti et al. 2000 pointed out that if variance-covariance matrix $\sum$ is known and random effects $\boldsymbol{u}$ and $\boldsymbol{v}$ are treated as fixed-effect parameters, then the combination of terms l_3 and l_4 $(-K(\boldsymbol{a}))$ is actually a penalized log likelihood with the term l_4 penalizing for the extreme value of $\boldsymbol{a}$.

$$l(\boldsymbol{\Psi};\boldsymbol{Y},\boldsymbol{X}) \approx -\frac{m}{2}\log|\sum| - \frac{1}{2}log|K''(\tilde{\boldsymbol{a}})| - K(\tilde{\boldsymbol{a}})$$
$$= l_1(\sum) + l_2\left(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\Lambda_0(t)\right) + l_3\left(\boldsymbol{\beta},\boldsymbol{\gamma},\lambda_0(t),\Lambda_0(t)\right) + l_4(\boldsymbol{\theta}) \quad (2.6)$$

where

$$l_1 = -\frac{m}{2}\log|\sum|$$

$$l_2 = -\frac{1}{2}\sum_{j=1}^{m} log\left\{ \left[\sum_{i=1}^{n_j}\left[-\frac{e^{\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+\tilde{u}_j}}{\left(1+e^{\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+\tilde{u}_j}\right)^2}\right] - \frac{\sigma_v^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}\right]\left[\sum_{i=1}^{n_j}\left\{-\Lambda_0(X_{ij})\exp\left[\boldsymbol{W}_{ij}^T\boldsymbol{\gamma}+\tilde{v}_j\right]\right\} - \frac{\sigma_u^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}\right] \right.$$
$$\left. - \left[\frac{\sigma_{uv}}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}\right]^2 \right\}$$

$$l_3 = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left\{\delta_{ij}\left[\log\left(\lambda_0(X_{ij})\right) + \boldsymbol{W}_{ij}^T\boldsymbol{\gamma} + \tilde{v}_j\right] - \Lambda_0(X_{ij})\exp\left[\boldsymbol{W}_{ij}^T\boldsymbol{\gamma}+\tilde{v}_j\right] + \left[Y_{ij}\left(\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+\tilde{u}_j\right) - \log\left(1+e^{\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+\tilde{u}_j}\right)\right]\right\}$$

$$l_4 = -\frac{1}{2}\sum_{j=1}^{m}\frac{\tilde{u}_j^2\sigma_v^2 - 2\tilde{u}_j\tilde{v}_j\sigma_{uv} + \tilde{v}_j^2\sigma_u^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}$$

2.3 Parameter Estimation

In this section, we propose a multivariate penalized likelihood (MPL) method for parameter estimation in context of the proposed joint model, which is built on the penalized partial likelihood depicted in Ripatti et al. 2000 for multivariate frailty model.

According to Ripatti et al. 2000 and Ye et al. 2008, the complicated term l_2 in (2.6) has negligible effect on the estimation of parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and random effects. So it is reasonable to ignore this term to simplify the computation. In addition, with the help of standard R package, we are able to perform the nonparametric estimation of cumulative baseline hazards very easily. Therefore, we ignore term l_2 as well as l_1 (constant) in the approximate log likelihood function (2.6) and use the penalized full log likelihood (term l_3 and l_4) instead of penalized partial log likelihood to estimate $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ as well as random effects $\boldsymbol{u}$ and $\boldsymbol{v}$ (See equation (2.7)). However, we have to keep in mind that removal of term l_2 may lead to some loss of information.

$$-K(\boldsymbol{a})=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left\{\delta_{ij}\left[\log\left(\lambda_0(X_{ij})\right)+\boldsymbol{W}_{ij}^T\boldsymbol{\gamma}+v_j\right]-\Lambda_0(X_{ij})\exp\left[\boldsymbol{W}_{ij}^T\boldsymbol{\gamma}+v_j\right]+\left[Y_{ij}(\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+u_j)-\log\left(1+e^{\boldsymbol{G}_{ij}^T\boldsymbol{\beta}+u_j}\right)\right]\right\}$$

$$-\frac{1}{2}\sum_{j=1}^{m}\frac{u_j^2\sigma_v^2-2u_jv_j\sigma_{uv}+v_j^2\sigma_u^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2} \qquad (2.7)$$

For given variance components $\boldsymbol{\theta}=(\sigma_u^2,\sigma_v^2,\sigma_{uv})$ and random effects, the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ that maximize penalized full log likelihood (2.7) (denoted as $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$) can be calculated by standard R function "*glm*" and "*coxph*", respectively. The random effects are treated as offset term in each model (Ripatti et al. 2000). The cumulative baseline hazard function $\Lambda_0(t)$ is estimated with nonparametric method through "*coxph*" and is retrieved using R function "*basehaz*". For given variance components $\boldsymbol{\theta}$ and parameter estimates $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$, the estimates of random effect $\boldsymbol{u}$ and $\boldsymbol{v}$ that maximize penalized full log likelihood (2.7) (denoted as $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$) are computed with Newton-Raphson method. The score and information functions of likelihood (2.7) with respect to $\boldsymbol{u}$ and $\boldsymbol{v}$ are listed below:

(1) Random effect v_j (for j=1,2,..,m):

Score function: $\frac{\partial \ell}{\partial v_j}=\sum_{i=1}^{n_j}\{\delta_{ij}-\Lambda_0(X_{ij})\exp[\boldsymbol{W}_{ij}^T\boldsymbol{\gamma}+v_j]\}-\frac{v_j\sigma_u^2-u_j\sigma_{uv}}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}$ (2.8)

Information function: $-\frac{\partial^2 \ell}{\partial v_j^2} = -\sum_{i=1}^{n_j}\{-\Lambda_0(X_{ij})\, exp[\boldsymbol{W_{ij}^T\gamma} + v_j]\} + \frac{\sigma_u^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}$ (2.9)

For $k \neq j$, $\frac{\partial^2 \ell}{\partial v_k \partial v_j} = 0$

(2) Random effect u_j (for j=1,2,..,m):

Score function: $\frac{\partial \ell}{\partial u_j} = \sum_{i=1}^{n_j}\left(Y_{ij} - \frac{e^{G_{ij}^T\boldsymbol{\beta}+u_j}}{1+e^{G_{ij}^T\boldsymbol{\beta}+u_j}}\right) - \frac{u_j\sigma_v^2 - v_j\sigma_{uv}}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}$ (2.10)

Information function: $-\frac{\partial^2 \ell}{\partial u_j^2} = -\sum_{i=1}^{n_j}\left[-\frac{e^{G_{ij}^T\boldsymbol{\beta}+u_j}}{\left(1+e^{G_{ij}^T\boldsymbol{\beta}+u_j}\right)^2}\right] + \frac{\sigma_v^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}$ (2.11)

For $k \neq j$, $\frac{\partial^2 \ell}{\partial u_k \partial u_j} = 0$

(3) Covariance of u_k and v_j (for k, j=1,2,..,m):

$$\frac{\partial^2 \ell}{\partial u_k v_j} = \frac{\sigma_{uv}}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2} \quad (k = j) \qquad (2.12)$$

$$\frac{\partial^2 \ell}{\partial u_k v_j} = 0 \quad (k \neq j)$$

By iterating between above two steps, the converged estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and random effects $\boldsymbol{u}$ and $\boldsymbol{v}$ that maximize the penalized full log likelihood (2.7) (denoted as $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{v}}$, respectively) can eventually be determined for fixed variance components $\boldsymbol{\theta}$.

When estimating variance components $\boldsymbol{\theta}$, term l_3 in (2.6) is ignored since it does not depend on $\boldsymbol{\theta}$ and can be treated as constant. In contrast to the case when estimating $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and random effects, the terms l_1 and l_2 cannot be omitted here. Conditional on $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{v}}$ obtained from previous steps, the approximate profile log likelihood function for $\boldsymbol{\theta}$ can be expressed as:

$l(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{X}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{u}}, \widehat{\boldsymbol{v}})$

$$\approx -\frac{m}{2}\log\,|\sum| - \frac{1}{2}\sum_{j=1}^{m} log\left\{\left[\sum_{i=1}^{n_j}\left[-\frac{e^{G_{ij}^T\hat{\beta}+\hat{u}_j}}{\left(1+e^{G_{ij}^T\hat{\beta}+\hat{u}_j}\right)^2}\right] - \frac{\sigma_v^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}\right]\left[\sum_{i=1}^{n_j}\{-\Lambda_0(X_{ij})\exp[\boldsymbol{W}_{ij}^T\hat{\boldsymbol{\gamma}} + \hat{v}_j]\} - \frac{\sigma_u^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}\right] - \left[\frac{\sigma_{uv}}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2}\right]^2\right\} - \frac{1}{2}\sum_{j=1}^{m}\frac{\hat{u}_j^2\sigma_v^2 - 2\hat{u}_j\hat{v}_j\sigma_{uv} + \hat{v}_j^2\sigma_u^2}{\sigma_u^2\sigma_v^2-\sigma_{uv}^2} \qquad (2.13)$$

The estimates of variance components $\boldsymbol{\theta}$ that maximize profile log likelihood function (2.13) have closed-form (Ripatti et al. 2000), which can be written as:

$$\widehat{\boldsymbol{\theta}} = \frac{\hat{\boldsymbol{a}}^T\hat{\boldsymbol{a}} + tr(K''(\hat{\boldsymbol{a}})^{-1})}{m} \qquad (2.14)$$

After illustrating how to estimate each set of parameters, we now outline the strategy of the maximization procedure with the following three steps: (1) Initial values of variance components $\boldsymbol{\theta} = (\sigma_u^2, \sigma_v^2, \sigma_{uv})$ are first assigned and corresponding random effects are generated. The estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ that maximize the likelihood (2.7) are calculated with standard R function. (2) Given the variance components and $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$ updated from step 1, the estimates of random effects that maximize the likelihood (2.7) are computed with Newton-Raphson method. For fixed variance components, the converged estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and random effects can be found by iterating between step 1 and 2. (3) Given the updated estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and random effects obtained from the previous two steps, the estimates of variance components $\boldsymbol{\theta}$ that maximize likelihood (2.13) are calculated based on equation (2.14). The maximum likelihood estimates for all parameters can then be derived by iterating from step1 to step 3 till convergence.

2.4 Standard Error Estimation

Estimates of standard errors for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be easily drawn from the output of "*coxph*" and "*glm*" function in the last iteration when all the parameters converge. Estimation of standard errors for the variance components $\boldsymbol{\theta}$ is performed by inverting Fisher information matrix of likelihood function (2.13) (See attached R code for details). The three diagonal elements of the inverted information matrix are the estimated variances of variance components $\boldsymbol{\theta}$. It should be noted that the standard errors of parameter estimates tend to be underestimated using MPL algorithm. This issue will be investigated in simulation study.

Chapter 3

Simulation Study

In this chapter, a series of simulation studies are conducted using R software to evaluate the finite sample properties of the proposed method and also compare the performance of the proposed joint model with one reference model: separate model with centre-specific random effects. The influences of several factors on the performance of proposed joint model are also addressed: (1) magnitude of random effect variance; (2) strength of association; (3) total sample size; (4) number of centres; (5) high censoring rate.

3.1 Numerical Simulation

For simplicity, only two covariates: treatment (denoted as "arm") and remission status (denoted as "resp") are considered in simulation study. In GLMM (3.1), treatment variable Z_{ij} is the only covariate; in Cox frailty model (3.2), besides treatment variable Z_{ij} and response variable Y_{ij}, we also include the interaction term of these two variables to explore the interactive effect of treatment and response on patient's survival rate.

$$P\left(Y_{ij}=1\,\middle|Z_{ij},\boldsymbol{\beta},u_j\right)=\frac{e^{\beta_0+Z_{ij}\beta_1+u_j}}{1+e^{\beta_0+Z_{ij}\beta_1+u_j}} \qquad (3.1)$$

$$\lambda\left(t\middle|Y_{ij},Z_{ij},\boldsymbol{\gamma},v_j\right)=\lambda_0(t)e^{\gamma_1Y_{ij}+\gamma_2Z_{ij}+\gamma_3Y_{ij}Z_{ij}+v_j} \qquad (3.2)$$

The data are simulated as follows: (1) The centre-specific bivariate random effects u_j and v_j are generated from the bivariate normal distribution $N_2(\mathbf{0},\sum)$, where $\sum$ is a 2×2 variance-covariance matrix. (2) Conditional on u_j, the binary response variable Y_{ij} is simulated from GLMM (3.1), where the treatment covariate Z_{ij} is generated from a Bernoulli distribution with success probability of 0.5. (3) We assume that failure time T_{ij} follows an exponential distribution and is simulated from Cox frailty model (3.2), where the baseline hazard is set to a constant of 0.15, and the treatment covariate Z_{ij}, response covariate Y_{ij} and random effect v_j are obtained from previous two steps. According to the

relationship between survival function and cumulative distribution function, we know that the random variable $S(T)$ follows uniform distribution Unif (0, 1) (denoted as W_{ij}). Therefore, model (3.2) is rewritten in the form of (3.3), from which the failure time T_{ij} can be solved. Censoring time C_{ij} is assumed to be non-informative and is generated from uniform distribution Unif (0, 20) to make the censoring rate be around 20%. Let $\delta_{ij}= I(T_{ij} < C_{ij})$ be the censoring indicator.

$$T_{ij} = -\log(W_{ij})/\left[0.15 \times \exp\left(\gamma_1 Y_{ij} + \gamma_2 Z_{ij} + \gamma_3 Y_{ij} Z_{ij} + v_j\right)\right] \qquad (3.3)$$

The simulated data with different parameter specifications are generated to evaluate the robustness of the proposed joint model. Regression coefficients β_0, γ_2, γ_3 are fixed with values of -1.0, log(2.0) and log(2.0), respectively. Total number of observations n is set

Table 3.1 Summary of parameterizations for simulation studies

No.	β_1	γ_1	σ_u^2	σ_v^2	σ_{uv}
1	log(0.5)	log(0.5)	0.5	0.5	-0.45
2	log(0.5)	log(0.5)	0.5	0.5	0.0
3	log(0.5)	log(0.5)	0.5	0.5	0.45
4	log(0.5)	log(0.5)	1.0	1.0	-0.9
5	log(0.5)	log(0.5)	1.0	1.0	0.0
6	log(0.5)	log(0.5)	1.0	1.0	0.9
7	log(0.5)	log(2.0)	0.5	0.5	-0.45
8	log(0.5)	log(2.0)	0.5	0.5	0.0
9	log(0.5)	log(2.0)	0.5	0.5	0.45
10	log(0.5)	log(2.0)	1.0	1.0	-0.9
11	log(0.5)	log(2.0)	1.0	1.0	0.0
12	log(0.5)	log(2.0)	1.0	1.0	0.9
13	log(2.0)	log(0.5)	0.5	0.5	-0.45
14	log(2.0)	log(0.5)	0.5	0.5	0.0
15	log(2.0)	log(0.5)	0.5	0.5	0.45
16	log(2.0)	log(0.5)	1.0	1.0	-0.9
17	log(2.0)	log(0.5)	1.0	1.0	0.0
18	log(2.0)	log(0.5)	1.0	1.0	0.9
19	log(2.0)	log(2.0)	0.5	0.5	-0.45
20	log(2.0)	log(2.0)	0.5	0.5	0.0
21	log(2.0)	log(2.0)	0.5	0.5	0.45
22	log(2.0)	log(2.0)	1.0	1.0	-0.9
23	log(2.0)	log(2.0)	1.0	1.0	0.0
24	log(2.0)	log(2.0)	1.0	1.0	0.9

to 600 and number of centres m is set to 30. Other parameters including β_1, γ_1 and variance components $\boldsymbol{\theta} = (\sigma_u^2, \sigma_v^2, \sigma_{uv})$ take the values as summarized in Table 3.1. We try two different levels of magnitudes for random effect variance (0.5 or 1.0) and three different strengths of association between two random effects (positive, zero, negative) for each set of given random effect variance. A total of 500 replications are performed for each set of parameter combination. Selected results from representative parameter specifications are presented in the report and the rest are listed in Appendix 1.

3.2 Simulation Results

Several statistics are investigated to evaluate the performance of the models: (1) Bias: the difference between the true value and estimated value of parameter; (2) ASE: asymptotic standard error of parameter estimate based on Fisher information matrix method, presented as the average of 500 simulations; (3) ESE: empirical standard error of parameter estimate, calculated by taking standard deviation of 500 parameter estimates; (4) CP: coverage probability of 95% confidence interval. It represents the percentage of 95% confidence intervals that cover the true value; (5) MSE_J/MSE_S: the ratio of mean square errors (MSE) between joint model and separate model. MSE is found by formula $MSE=Bias^2 + Variance$, where variance is the square of empirical standard error (ESE). $MSE_J/MSE_S < 1$ means the joint model yields smaller MSE than the separate model. In other words, the joint model performs better than the separate model.

3.2.1 Comparison of Proposed Joint Model and Separate Model

Two models are applied to the simulated data for comparison: (1) the proposed joint model incorporated with MPL inference method; (2) the separated model consisting of generalized linear mixed model and Cox frailty model, implemented with R functions "*lme4::glmer*" and "*coxme::coxme*", respectively.

Twenty-four simulation studies with different parameterizations are conducted with both joint

model and separate model. As an example, results for simulation study # 19 are summarized in Table 3.2. It is noted that in this simulation study two random effects are highly correlated (correlation coefficient $\rho = \sigma_{uv}/\sigma_u\sigma_v$=−0.9). The biases of estimates from both models are all fairly small. Compared with the separate model, the joint model produces even smaller biases for all of parameters except for σ_v^2 and therefore yields more accurate parameter estimates. As for the empirical standard error (ESE), we find that the value of ESE for joint model is a bit smaller than separate model, which suggests that the parameter estimation of the proposed joint model is slightly more efficient. In comparison with the separate model, MSE for the joint model is generally smaller (MSE_J/MSE_S <1). More specifically MSE is

Table 3.2 Comparison of joint model and separate model (based on 500 replications)

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.008	0.181	0.181	0.952	0.017	0.183	0.184	0.944	0.960
Cox Frailty Model										
γ_1	log(2.0)	0.009	0.151	0.160	0.948	-0.037	0.162	0.163	0.940	0.918
γ_2	log(2.0)	0.007	0.120	0.126	0.936	0.012	0.125	0.133	0.942	0.885
γ_3	log(2.0)	-0.007	0.197	0.201	0.948	-0.010	0.206	0.213	0.940	0.891
Variance Components										
σ_u^2	0.5	-0.010	0.129	0.195	0.784	-0.026		0.215		0.811
σ_v^2	0.5	-0.007	0.128	0.155	0.838	0.002		0.158		0.971
σ_{uv}	-0.45	0.009	0.120	0.145	0.844					

reduced by up to around 20% if the joint model is applied. There is no obvious difference in coverage probability of 95% confidence interval between two models. CP for fixed effect parameters from both models are all close to the nominal confidence level 95%. In joint model, however, CP are relatively low (around 70 to 80%) for the variance components (separate model does not provide CP for variance components). This finding could be explained by the underestimation of standard error caused by MPL algorithm. As shown in Table 3.2, compared with the ESE which are assumed to be trustworthy estimates of "true SE"

(Hanley et al. 2003), ASE of fixed-effect parameter estimates are slightly smaller and it is more so in the case of variance components. In other words, the standard error is underestimated by joint model with MPL method, which leads to higher chance of type I error in hypothesis testing and narrower confidence interval and consequently yields lower CP. The underestimation is mainly because in MPL algorithm we estimate standard errors using only part of the likelihood function. As a result, a portion of the variation (or uncertainty) from unknown parameters is ignored. A similar phenomenon was also observed in previous literatures (Ripatti et al. 2000; Ye et al. 2008). Another possible reason for the poor estimation of standard error is related to Fisher information matrix method we used in MPL algorithm. Hsieh et al. 2006 pointed out that the application of Fisher information matrix in estimating standard error could be problematic due to the nonparametric nature of joint survival model.

In the simulation study presented above, the proposed joint model has smaller bias and lower MSE for most of the parameter estimates than the separated model. Furthermore, ESE is also slightly smaller in the case of joint model. In other words, the joint model provides more accurate and precise parameter estimation. The main concern with the proposed joint model is that in some cases the ASE underestimates the true standard error, especially for the variance components. Some methods such as Bootstrap and Jackknife can be implemented to tackle this problem. More details will be given in Section 3.2.6.

3.2.2 Magnitude of Random Effect Variance

The effect of magnitude of random effect variance is explored in this section. As an example, the results for simulation study # 22 are presented here (Table 3.3). All the other parameters are kept the same as the previous section except for variance components. The values for σ_u^2, σ_v^2 and σ_{uv} are all doubled to increase the variation of the underlying effects. The correlation coefficient remains unchanged (ρ_{uv}=-0.9).

From Table 3.3, we can see that larger variance of random effects does not obviously alter the performance of joint model or separate model in terms of bias, MSE or CP for all fixed-effect parameters. Similar to the case with a smaller variance, the MSE produced by the joint model are smaller than the separated model for most parameters, especially for parameter β_1 and γ_1, suggesting the superior performance of joint model over separate model. The results in this section provide evidence that the joint model performs equally well when the magnitude of random effect variance is modified.

Table 3.3 Simulation results for data with larger frailty variance (based on 500 replications)

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.027	0.185	0.172	0.960	0.004	0.190	0.202	0.952	0.744
Cox Frailty Model										
γ_1	log(2.0)	0.007	0.153	0.161	0.928	-0.068	0.168	0.170	0.926	0.775
γ_2	log(2.0)	0.000	0.121	0.122	0.956	0.007	0.126	0.128	0.954	0.906
γ_3	log(2.0)	-0.001	0.199	0.201	0.944	0.008	0.209	0.197	0.956	1.036
Variance Components										
σ_u^2	1.0	-0.090	0.237	0.329	0.754	-0.026		0.361		0.885
σ_v^2	1.0	-0.037	0.249	0.288	0.856	-0.001		0.303		0.921
σ_{uv}	-0.9	0.054	0.230	0.269	0.844					

3.2.3 Strength of Association

The effect of strength of association between two random effects on model performance is explored in this section. We test two scenarios of association: (1) positive correlation σ_{uv}=0.45 (instead of negative) (2) no correlation σ_{uv}=0 between two random effects. As an example, the results for simulation study #20 and #21 are presented and compared with the case of negative correlation in Section 3.2.1.

As shown in Table 3.4, when we change the value of covariance σ_{uv} from negative to positive, the joint model performs in a similar manner, which generates comparable parameter estimates and ESE as the previous case (Table 3.2). Similar to the simulation study with negative covariance, the proposed joint model produces less biased estimates and smaller MSE than the separate model in the case of positive covariance. These results indicate that the joint model is robust and is able to give decent estimates no matter whether the binary marker response is positively or negatively associated with survival data through joint random effects. Thus, it is possible to apply the proposed joint model to more diverse situations.

Table 3.4 Simulation results for data with positive covariance (based on 500 replications)

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.007	0.181	0.180	0.960	0.017	0.183	0.182	0.952	0.967
Cox Frailty Model										
γ_1	log(2.0)	0.021	0.150	0.160	0.936	0.065	0.157	0.157	0.916	0.895
γ_2	log(2.0)	0.008	0.125	0.129	0.942	-0.018	0.130	0.130	0.950	0.973
γ_3	log(2.0)	-0.010	0.197	0.205	0.936	0.000	0.205	0.213	0.948	0.927
Variance Components										
σ_u^2	0.5	-0.016	0.128	0.195	0.768	-0.028		0.201		0.928
σ_v^2	0.5	-0.008	0.128	0.149	0.860	-0.020		0.156		0.901
σ_{uv}	0.45	-0.015	0.119	0.152	0.840					

When the covariance σ_{uv} takes the value of 0, there is no association between binary marker response and survival data through joint distribution of random effects. Therefore, we assume that technically the joint model should have similar performance as the separate model with random effects. The results in Table 3.5 show that joint model and separate model give similar point estimates and ESE for most parameters. The joint model yields smaller MSE for the parameters in GLMM and variance components while the separate model gives slightly smaller MSE for the parameters in Cox frailty model. But in general, there is no obvious

evidence showing one model outperforms the other when the covariance σ_{uv} is set to 0, which is consistent with our hypothesis. Taken together with the results from previous few sections, we summarize that the advantage of the proposed joint model over separate model gets more pronounced as the association between two random effects gets stronger.

Table 3.5 Simulation results for data with "zero" covariance (based on 500 replications)

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.020	0.179	0.178	0.956	0.013	0.183	0.181	0.958	0.981
Cox Frailty Model										
γ_1	log(2.0)	0.013	0.150	0.160	0.944	-0.012	0.159	0.154	0.968	1.090
γ_2	log(2.0)	0.007	0.122	0.131	0.926	0.002	0.128	0.125	0.960	1.096
γ_3	log(2.0)	-0.007	0.197	0.203	0.946	0.016	0.205	0.196	0.966	1.067
Variance Components										
σ_u^2	0.5	-0.049	0.124	0.185	0.744	-0.022		0.196		0.934
σ_v^2	0.5	-0.015	0.126	0.142	0.860	0.005		0.160		0.796
σ_{uv}	0.0	0.001	0.087	0.122	0.862					

3.2.4 Total Sample Size

To test the effect of sample size on parameter estimation with the proposed joint model, we simulate data with the same parameter specification as simulation study #19 but with different sample size. In the first step, the sample size is reduced to 150. As shown in Table 3.6, bias, ASE and ESE of parameters for the joint model are all generally larger than the simulation study with larger sample size (n = 600) (Table 3.2). In the meanwhile, CP of all parameter estimates get slightly lower when sample size is smaller. In the condition of smaller sample size, however, the proposed joint model still outperforms the separate model in terms of bias and MSE.

Table 3.6 Simulation results for data with small sample size (based on 500 replications)

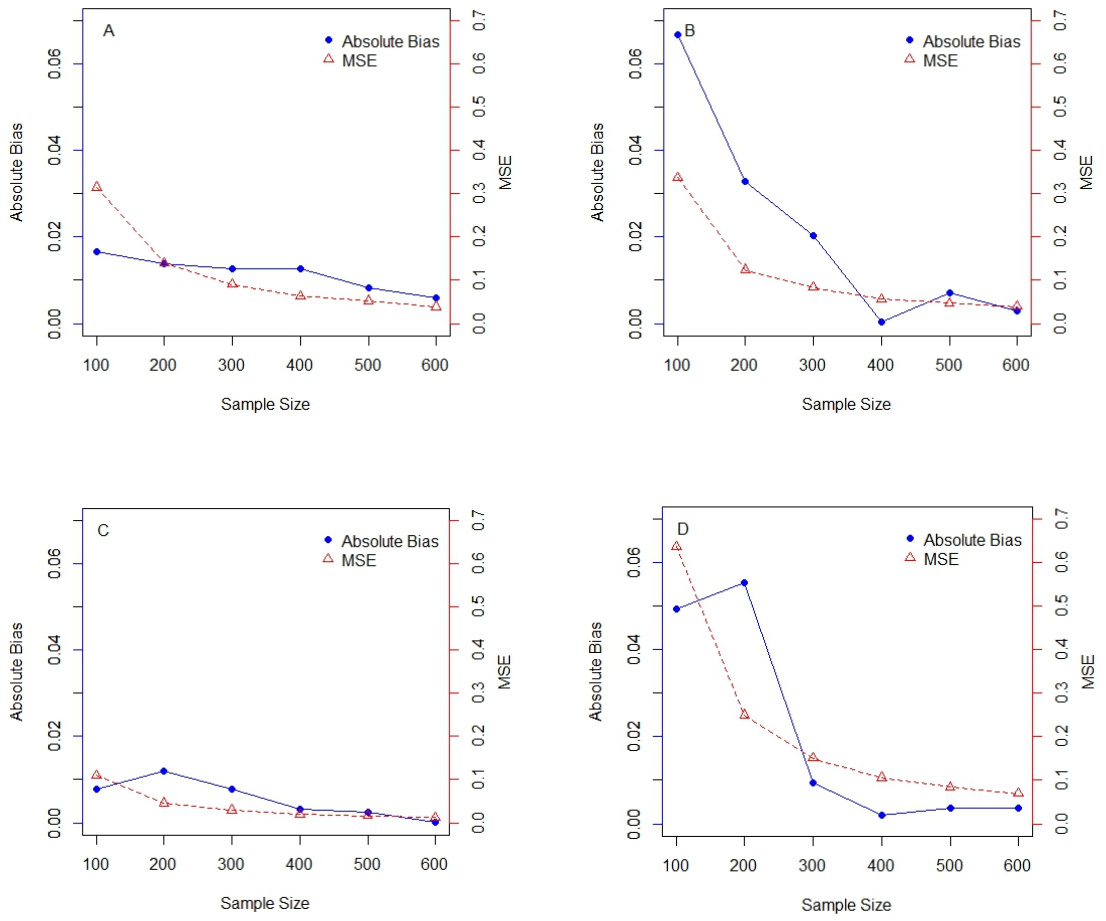
Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.025	0.362	0.385	0.938	0.011	0.370	0.394	0.944	0.957
Cox Frailty Model										
γ_1	log(2.0)	-0.021	0.308	0.351	0.916	-0.193	0.347	0.370	0.896	0.708
γ_2	log(2.0)	0.013	0.244	0.276	0.924	0.021	0.272	0.288	0.938	0.917
γ_3	log(2.0)	0.015	0.402	0.421	0.940	0.029	0.449	0.479	0.932	0.772
Variance Components										
σ_u^2	0.5	-0.002	0.144	0.314	0.620	-0.038		0.417		0.563
σ_v^2	0.5	-0.017	0.132	0.237	0.722	-0.013		0.267		0.789
σ_{uv}	-0.45	0.050	0.120	0.219	0.668					

To further explore the association between sample size and performance of joint model, we simulate data with a series of sample sizes, ranging from 100 to 600. The number of centres m is set to 20 as constant. As shown in Figure 3.1, both the biases and MSE for parameters β_1, γ_1, γ_2 and γ_3 are decreased consistently as sample size increases. This observation indicates that increasing sample size improves the accuracy of parameter estimation as well as the finite sample performance of the proposed joint model. In other words, joint model performs better as sample size increases. Two reasons could explain this observation: (1) a larger sample size is able to reduce the bias if the estimator is unbiased; (2) if we hold the number of centres to be constant, when the total sample size increases, the number of observations in each centre also increases, which will improve the accuracy of Laplace approximation (Ripatti et al. 2000; Abrahantes et al. 2005).

Note: The effect of number of centres is also investigated by holding the total number of patients while varying the number of centres in simulation study (see results in Appendix 2). No clear pattern in bias or MSE is detected as increase of centre number. Unlike the situation in total sample size, change of centre number (maintain the total number of patients) has varying effects on the performance of model. On one hand, as number of centres increases,

number of patients in each centre will decrease, which reduces the accuracy of Laplace approximation. On the other hand, since the centre-level random effects follow normal distribution, a larger number of centres will give a better representation of the distribution, which will benefit parameter estimation. Therefore, we acclaim that it is difficult to elucidate the effect of centre number on model performance under the context of our proposed joint model and inference method.

Figure 3.1 Graphic analysis of correlation between sample size and MSE / bias for four fixed-effect parameters (A) β_1 (B) γ_1 (C) γ_2 (D) γ_3.



3.2.5 High Censoring Rate

In the simulation studies discussed above, the censoring rate was set to 20%. However, in some clinical trials the disease could be chronic or curable, which will lead to large

proportion of long-term survival and probably a high censoring rate. In this section, we test the performance of the proposed joint model under the condition of high censoring rate, which is set to around 80% (Table 3.7). Compared with the case with low censoring rate (Table 3.2), parameter estimates yielded by the proposed joint model are generally more biased when censoring rate is high. It is noteworthy that high censoring rate dramatically increases the ASE and ESE for the fixed-effect parameters in Cox frailty model. Overall, the proposed joint model produces larger bias for point estimates with lower efficiency as a result of high censoring. Nevertheless, in comparison with the separated model with random effects, performance of the proposed joint model is still superior in that it gives relatively smaller ESE and MSE.

Table 3.7 Simulation results for data with high censoring rate (based on 500 replications)

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.018	0.180	0.178	0.948	0.006	0.183	0.199	0.944	0.808
Cox Frailty Model										
γ_1	log(2.0)	-0.038	0.394	0.402	0.960	-0.168	0.402	0.409	0.946	0.835
γ_2	log(2.0)	-0.003	0.292	0.284	0.966	0.045	0.296	0.302	0.946	0.868
γ_3	log(2.0)	0.026	0.460	0.456	0.958	-0.006	0.466	0.460	0.960	0.986
Variance Components										
σ_u^2	0.5	-0.036	0.124	0.184	0.766	-0.032		0.203		0.834
σ_v^2	0.5	-0.026	0.127	0.198	0.764	-0.037		0.237		0.689
σ_{uv}	-0.45	0.035	0.156	0.116	0.796					

3.2.6 A Special Topic: Improving SE Estimation by Jackknife Resampling

Traditional approaches (e.g. ordinary least squares method) of estimating bias, variance and

confidence interval depend greatly on the modeling assumption. So the validity and reliability of the statistical inference will be affected by the validity of the assumption. If we assume the sample data as a representative of the target population we want to study, then we can make statistical inference based on the "resampling" method (draw sample from the sample data) (Efron, 1982). The resampling method does not make any assumption on the distribution. In addition, it provides a much easier approach to perform statistical inference such as estimation of standard error when complicated calculation is required in traditional methods.

As discussed in early section, the true standard error is underestimated based on the proposed MPL algorithm because not all uncertainties are taken into account during variance estimation. Ripatti et al. 2000 suggested applying Bootstrap sampling as one solution. In this study, we first tried Bootstrap by sampling clusters with replacement, which, however, had convergence problem when calculating MLE of parameters (data not shown). Accordingly, instead of using Bootstrap, we decide to apply "delete-*m* Jackknife algorithm" with unequal *m* ("*m*" here represents the number of patients in each centre) to calculate standard errors of parameter estimates, which was elaborated in paper published by Busing et al. 1999. One hundred datasets are simulated based on the method described in Section 3.1 with parameterization #1 (See Table 3.1). For each dataset, observations from one out of 30 centres are sequentially removed and the remaining observations (a subset of original dataset) are applied to the proposed joint model. Calculation of Jackknife SE follows the method described in Busing et al. 1999 (See attached R code). The results are justified by comparing with the empirical SE obtained from 500 simulations.

Table 3.8 Comparison of Jackknife SE and asymptotic SE

	$\boldsymbol{\beta_1}$	$\boldsymbol{\gamma_1}$	$\boldsymbol{\gamma_2}$	$\boldsymbol{\gamma_3}$	$\boldsymbol{\sigma_u^2}$	$\boldsymbol{\sigma_v^2}$	$\boldsymbol{\sigma_{uv}}$
ESE	0.206	0.197	0.116	0.275	0.202	0.154	0.150
JK-SE	0.207	0.194	0.118	0.273	0.196	0.150	0.145
ASE	0.206	0.190	0.110	0.260	0.124	0.125	0.116

In Table 3.8, we compare three types of standard error: (1) ESE, empirical standard error. It is

considered to be a good estimate of the true SE and used as a "gold standard" here; (2) JK-SE, the average over 100 replications of SE calculated by Jackknife method; (3) ASE, asymptotic standard error estimated by MPL. As we can see from the table, the values of JK-SE for all parameters except β_1 are much closer to the "gold standard" ESE than ASE. In particular, the estimation of standard errors for variance components, which is especially problematic in the case of MPL, is greatly improved by Jackknife method. Thus, we conclude that Jackknife resampling is able to provide decent estimates of the "true SE", which can serve as a useful tool in our study to correct the underestimation of standard error caused by MPL algorithm.

Chapter 4

Hodgkin's Lymphoma Study

In Chapter 3, we evaluated the performance of the proposed joint model and MPL algorithm. In this chapter, the proposed model incorporated with MPL inference method is applied to a clinical trial dataset from NCIC Clinical Trials Group's Hodgkin's lymphoma Study (Meyer et al. 2012) to explore the potential association between the marker response and survival data and detect important predictors for patients' survival.

4.1 Data Description and Cleaning

In this example, we consider the association between marker response (denoted as "resp") and survival, and their relationship with the following categorical variables: (1) Treatment effect (denoted as "arm". ABVD alone group=0; radiotherapy group=1). (2) Risk profile of patients (denoted as "risk". favorable risk=0; unfavorable risk=1). (3) Gender of patients (denoted as "sex". Female=0; Male=1). Age is an important factor as well but we do not include it in the model because age effect has been integrated into the risk profile of patients (Meyer et al. 2005).

To remove guarantee-time bias when assessing the association between treatment response and survival data, we implement landmark analysis. The time for evaluation of response to treatment is selected as "natural landmark time", which is at the end of treatments (six months after the onset of randomization). Thus, among 399 patients, six of them who had survival time less than six months are removed from the dataset. The survival time used in data analysis is calculated by subtracting the time of response evaluation (six months) from the original survival time. Therefore, the survival data is conditional on landmark time. Unless specified otherwise, the model interpretation in this chapter is based on the patients

whose survival time, either censored or uncensored, are longer than the landmark time.

We notice that "resp" values for some observations are missing. The profile of the observations with missing response variable is summarized in Appendix 3, where the potential associations between variable "arm" and the other three variables ("event", "sex" and "risk") are checked with Fisher's exact test. The test shows that the associations between "arm" and other predictors are not statistically significant at 0.05 level, which indicates that there is no obvious unbalance between two treatment groups in terms of "event", "sex" and "risk". Therefore, a complete case analysis will be first conducted in Section 4.3 by excluding all the observations with missing "resp" variable. At the end of this chapter, a sensitivity analysis for missing data will be conducted to verify the validity of results.

4.2 Test of Proportional Hazards Assumption

In the proposed joint model, we choose Cox frailty model to analyze survival data. Before applying the proposed model to the clinical trial data, we need to verify the validity of using Cox PH model. We test the proportional hazards assumption with scaled Schoenfeld residual method, which is calculated easily with R function "*cox.zph*". The Cox frailty model for the clinical trial data is constructed in the same form as described in model (3.2) except that two extra variables "risk" and "sex" are added into the model. Table 4.1 summarizes three statistics corresponding to each covariate: (1) The correlation coefficient ρ between transformed survival time (based on Kaplan-Meier estimate of the survival function) and the scaled Schoenfeld residuals; (2) Chi-square value; (3) Two-sided p-value. It is clear that p-values for all the covariates are insignificant at 0.05 level, indicating no violation of proportional hazards assumption.

Under the PH assumption, Schoenfeld residuals are independent of time. Therefore if PH assumption is valid, the plot of Schoenfeld residuals against time should show random pattern and the corresponding fitted line should be roughly a straight line with zero slope. In

Appendix 4, scaled Schoenfled residuals for four covariates and one interaction term in Cox frailty model are plotted against the transformed survival time with a smoothing-spline fitted line. As we can see, in each plot residuals are randomly distributed and slope of fitted line is roughly around zero, which further confirms the compliance of PH assumption.

Table 4.1 Test of proportional hazards assumption

Parameter	$\boldsymbol{\rho}$	**Chisq**	**p-value**
Resp	0.208	1.330	0.249
Arm	-0.183	1.094	0.296
Risk	0.074	0.166	0.684
Sex	0.052	0.090	0.765
Resp× Arm	0.090	0.250	0.617

4.3 Data Analysis and Model Interpretation

The proposed joint model is then fitted to the data and the joint inference approach MPL is applied to calculate the point estimates for regression coefficients and frailty variance components. As discussed in simulation studies, standard errors of parameters (especially variance components) are underestimated by MPL algorithm. Therefore, in addition to MPL algorithm, we also apply Jackknife method to estimate the standard errors.

The parameter estimates based on joint model (EST), standard errors and P-values calculated through MPL or Jackknife are summarized in Table 4.2. We can see that compared with ASE calculated directly by MPL, JK-SE for most of the fixed-effect parameters are mildly larger and for variance components $\boldsymbol{\theta}=(\sigma_u^2, \sigma_v^2, \sigma_{uv})$ are dramatically larger. As a result, more parameters are classified to be statistically "insignificant" if Jackknife method is used. Based on the simulation study in Section 3.2.6, this observation suggests that Jackknife method applied here could probably be able to correct the underestimation of standard errors caused by MPL algorithm. Considering that Jackknife method generally produces more accurate estimates of standard errors, in the following sections we will only focus on the

results generated from Jackknife method. As shown in Table 4.2, covariate "arm" in the GLMM and covariate "resp" in Cox frailty model are highly significant. "risk" variable is marginally significant at level of 0.05 (p-value=0.06), whereas none of other covariates or variance components are statistically significant. It is noted that in HD.6 clinical trial the censoring rate is extremely high (around 91%), which, according to the simulation study in Section 3.2.5, could partly explain the large standard error estimates we observed here, especially for the parameters in Cox frailty model.

Table 4.2 Results of data analysis with proposed joint model

Parameter	**EST**	**MPL**		**Jackknife**	
		ASE	**P-value**	**JK-SE**	**JK P-value**
GLMM					
Arm	1.359	0.397	0.0006†	0.370	0.0002†
Risk	-0.077	0.386	0.8412	0.436	0.8592
Sex	0.126	0.349	0.7181	0.399	0.7524
Cox Frailty Model					
Resp	-2.389	0.711	0.0008†	0.674	0.0004†
Arm	0.059	0.822	0.9432	1.044	0.9552
Risk	1.908	0.733	0.0092†	1.030	0.0639
Sex	0.660	0.399	0.0985	0.467	0.1575
Resp×Arm	1.721	1.030	0.0946	1.336	0.1976
Variance components					
σ_u^2	0.122	0.052	–	0.160	–
σ_v^2	0.351	0.144	–	0.476	–
σ_{uv}	-0.201	0.084	0.0166†	0.268	0.4535

Based on the above results, the fitted joint model is presented as in Table 4.3. Due to the presence of interaction term in the second sub-model, the Cox frailty model is split into two scenarios by treatment group to facilitate the model interpretation (Arm=0: ABVD alone group; Arm=1: radiotherapy group).

In GLMM, compared with ABVD-alone group, radiotherapy group significantly increases the chance of remission after completion of treatment (odds ratio: 3.89, 95% CI: [1.88, 8.04]). Effects of "risk" and "sex" are not statistically significant in GLMM, implying that the risk profile and the gender of patients have no obvious influence on treatment response.

Table 4.3 Joint model for Hodgkin's lymphoma study

Parameter	GLMM	Cox Frailty Model	
		Arm = 0	Arm = 1
Arm	1.359	Ref.	0.059
Resp	–	-2.389	-0.667
Risk	-0.077	1.908	1.908
Sex	0.126	0.660	0.660

In the case of Cox frailty model, the interpretation of "resp" variable is conducted in two different scenarios: (1) In ABVD alone group, the treatment response (remission / no remission) is significantly correlated to hazard rate. The hazard ratio for remission is 0.09, which indicates that if the patient in ABVD alone group showed the complete remission at the end of treatment, the hazard rate would drop by as much as 91%. (2) In radiotherapy group, the hazard ratio for remission equals to 0.51, which, however, is not statistically significant (p-value=0.37). Based on the proposed model, response to treatment is an important predictive factor for hazard rate and its effect tends to be more pronounced in ABVD alone group than in radiotherapy group. However, we have to keep in mind that the different magnitudes of hazard ratio regarding treatment response between two treatment groups may not be true since the effect of interaction term is not statistically significant (p-value=0.20). On the other hand, in Cox frailty model the treatment effect is not significant (p-value=0.96), suggesting that there is no obvious difference in hazard rate (or survival rate) between two treatment groups. The "risk" effect is marginally significant, suggesting that there is a trend towards a higher hazard rate (or lower survival rate) for the patients with unfavorable risk profile compared with the ones with favorable risk profile. In the Cox frailty model, "sex" effect is not important in predicting survival of patients.

The joint random effects represent the centre-specific unobservable factors that could contribute to determine treatment response and survival rate. The correlation coefficient ρ explains the association between response to treatment and survival through the joint random effects. It could be considered as a "residual association" in addition to the association established through regression in Cox frailty model, which, taken together, explain the "true association" between these two endpoints. In this example, the covariance of two random effects is -0.201 and the correlation coefficient ρ_{uv} equals to -0.97, which implies that the two random effects tend to be correlated in a negative manner. That being said, it should be noted that in this example due to the large standard errors, the effects of variance components are not statistically significant and therefore the association between two endpoints through joint random effects is negligible.

Note: It is common practice to conduct model selection for regular statistical model in data analysis. However, this issue has not been fully understood in the context of joint modeling. Some classic model selection criteria such as AIC cannot be easily applied to the joint model due to the presence of nonparametric baseline hazard (Li et al. 2010). For this reason, we will not address the model selection issue in this study.

4.4 Sensitivity Analysis

As a simple form of sensitivity analysis, extreme case analysis of the missing data is conducted to evaluate the robustness of the results obtained from the previous section. The landmark time is kept the same as before (six months after randomization); among those observations that had survival time longer than six months, instead of eliminating the missing response values from the dataset, we set all the missing values to either "Yes" (remission) or "No" (no remission). The new sets of data are analyzed with the proposed joint model, the results of which are thereafter compared with the one described in Section 4.3 (see Table 4.2 and Appendix 5 and 6). It is noticeable that variable "arm" in GLMM and "resp" in Cox frailty model show significant effects in all three scenarios (variable "arm" is marginally

significant when the missing values are set to “No”). Compared with the first study (Table 4.2), “risk” variable and/or interaction term in Cox frailty model become statistically significant in the latter two studies. Nevertheless, most of covariates and variance components in all three scenarios show similar pattern with slightly different magnitudes. In general, the sensitivity analysis supports the conclusion made in earlier section.

Chapter 5

Summary and Future Directions

In this project, we construct a joint model of binary marker response and survival data. A modified joint inference method which is denoted as multivariate penalized likelihood (MPL) method based on the previous paper is developed. The performance of the proposed joint model and MPL method is evaluated with extensive simulation studies and then applied to an actual dataset generated from NCIC Clinical Trials Group's HD.6 clinical trial.

If the clinical trial involves multiple endpoints (e.g. early response to the treatment and time to event) and there is a reason to believe that those processes are associated, it is worthwhile to apply joint model to minimize the bias and improve the efficiency in statistical inference. Most of existing joint models discussed in literatures are focusing on the joint analysis of longitudinal measurement and survival data, which are usually linked with the shared random effects or the joint distribution of random effects. In this study, two new features are incorporated into the proposed joint model. Firstly, instead of dealing with longitudinal data, we explore the possibility for joint analysis of binary response and survival data. Secondly, in the proposed joint model the binary marker response and survival are connected through two paths: regression in Cox frailty model and joint random effects. The association between these two endpoints, therefore, is explained by both observed and unobserved (latent) effects, the latter of which, in our case, is the centre-level random effects. One thing we would like to highlight here is the use of cluster-level random effects rather than individual-level random effects in the proposed model. Under the structure of binary outcome, it is not possible to make inference for the individual random effects in the linear predictor (Lancaster et al. 2004). The creative implementation of centre-level random effects in the proposed joint model bypasses the obstacle.

Multivariate penalized likelihood (MPL) inference technic introduced in this report

circumvents the intractable integration in maximum likelihood analysis by utilizing Laplace approximation, followed by maximizing the likelihood function with respect to fixed-effect parameters, random effects (treated as parameters) and variance components in a sequential manner till all estimates converge. According to the previous study (Ye et al. 2008), the inference method based on penalized joint likelihood was computing much faster that EM algorithm and the performances of these two methods were comparable. In our case, on average it only took around 4 hours to run simulation with 500 replications under CPU Intel Xeon at 2.53 GHz. One limitation of MPL is that the two-step nature in maximization process underestimates the underlying variation, which causes the poor estimation of standard errors. Thus, we implement Jackknife method and the results indicate that it could greatly improve the estimation of standard errors, especially for the frailty variance components. Therefore, Jackknife method can be used together with MPL algorithm when reasonably accurate SE estimates are required.

Performance of the proposed joint model is compared with the separate model with random effects in simulation studies. When the covariance of random effects in the simulated data is set to zero to mimic the situation of no association between two sub-models, the overall performances of joint model and separate model are pretty comparable, and the joint model does not display any obvious benefit. However, when we increase the strength of association, no matter whether it is positive or negative, the joint model starts to show some preferable features over separate model: (1) It reduces the bias in estimates of fixed-effect parameters and variance components. (2) It decreases mean square error (MSE) and therefore increases the efficiency for parameter estimation. (3) It produces slightly smaller empirical SE, which means more precise estimation and higher power. (4) It provides a decent estimation for frailty covariance, which is not available in the case of separate model. On the other hand, there is no obvious difference between joint model and separate model in terms of coverage probability (CP) of 95% confidence interval. Performance of the proposed joint model is improved when the total sample size increases, evidenced by decreased bias and MSE; when the censoring rate is higher, the joint model yields more biased estimates with lower efficiency. Nevertheless, under these conditions, the proposed joint model consistently

outperforms the separate model.

Followed the simulation studies, we fit the proposed joint model to HD.6 clinical trial data where the potential association between treatment response (remission / no remission) and survival rate over 12 years period is of the main interest. Jackknife resampling is adopted to correct for the underestimation of standard errors caused by MPL algorithm. The joint model successfully detects the positive correlation between the marker response and survival through regression. Therefore, we can conclude that the response to treatment (remission / no remission) could be used as a potential surrogate for patient's survival. On the other hand, the "residual association" between these two endpoints through joint random effects is not statistically significant. This is related to the large standard errors of parameter estimates, which could be partly explained by the high censoring rate in this clinical trial study. In addition, some previous papers pointed out that the accuracy of Laplace approximation tends to be positively associated with the number of observations in each cluster (centre) (Ripatti et al. 2000; Abrahantes et al. 2005). As we have noticed, for some centres in the clinical trial study, there were only one or two patients, which may affect the Laplace approximation and as a result lead to less accurate parameter estimation.

In summary, the proposed joint model with MPL provides a computationally efficient and reasonably advantageous approach to jointly analyze binary marker response and survival data. It gives a better fitting than the separate model with random effects, particularly when there is a strong association between the binary response and survival data through joint random effects. In addition, the proposed joint model is capable of detecting the association between two endpoints at both fixed-effect and random-effect levels. However, the poor estimation of standard error with MPL algorithm could be problematic and needs to be corrected by other methods, such as Jackknife resampling.

Several related topics can be further explored in the future work. (1) One of the main advantages for MPL algorithm we proposed in this report is the easy computation. It is a good choice to test the new model within relatively short time period. However, as we can see from

the results, MPL method could underestimate the true SE of parameter estimates especially for the variance components and the Laplace approximation may not be appropriate when the number of observations per centre is low. An alternative inference method is EM algorithm. Although previous literature showed that the performance of the penalized joint likelihood was comparable to EM algorithm, this observation shall be explored in the context of our proposed joint model. In the future work, it may be worthwhile to apply EM algorithm to the proposed joint model and compare the performance with MPL algorithm. But it should be noted that the intensive computation of EM algorithm could be a potential problem. (2) The proposed MPL algorithm can be easily extended to other joint models with different modelling structures. One potential research direction is to study its performance in the context of joint model with response endpoint follow an ordinal distribution or a Poisson distribution. Similarly, instead of using normally-distributed multivariate random effects in the joint model, we can extend the application of MPL algorithm to the joint model with random effects follow other types of distributions such as Gamma distribution. (3) As we can see from the clinical trial data explored in this study, the censoring rate is extremely high (around 91%), which could be related to the high cure rate of Hodgkin's lymphoma. Song et al. 2012 suggested implementing cure model to handle the survival time with cure fraction. In the future, it may be worth it to substitute the Cox frailty model used in this project with the cure model to see if the fitting of joint model can be further improved for the Hodgkin's lymphoma data. (4) When analyzing the data from HD.6 clinical trial study, we only considered limited number of potential covariates for simplicity and demonstration purpose. In addition, we did not explore possible interaction effects such as risk versus treatment or risk versus response. All of these factors may influence model fitting and interpretation. Therefore, more work is required to further understand the association of survival with treatment response and other predictive factors in HD.6 clinical trial.

Reference

Abrahantes, JC, Burzykowski, T **(2005)**. A version of the EM algorithm for proportional hazard model with random effects. *Biom J*, 47, 6:847-62.

Anderson, JR, Cain, KC, Gelber, RD **(2008)**. Analysis of survival by tumor response and other comparisons of time-to-event by outcome variables. *J. Clin. Oncol.*, 26, 24:3913-5.

Breslow, NE, Clayton, DG **(1993)**. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 421:9-25.

Busing, F, Merjer, E, Leeden, RV **(1999)**. Delete-m Jackknife for unequal m. *Statistics and Computing*, 9, 3-8.

Cox, DR **(1972)**. Regression Models and Life-Tables. *Journal of the Royal Statistical Society*, Series B 34, 2: 187–220.

Dafni, UG, Tsiatis, AA **(1998)**. Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, 54, 4:1445-62.

Ding, J, Wang, JL **(2008)**. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, 64, 2:546-56.

Efron, B **(1982)**. The jackknife, the bootstrap, and other resampling plans. *Society of Industrial and Applied Mathematics CBMS-NSF Monographs*, 38.

Giobbie-Hurder, A, Gelber, RD, Regan, MM **(2013)**. Challenges of Guarantee-Time Bias. *J. Clin. Oncol*. 10.1200/JCO.2013.49.5283.

Glidden, DV, Vittinghoff, E **(2004)**. Modelling clustered survival data from multicentre clinical trials. Stat Med, 23, 3:369-88.

Hanley, JA, Negassa, A, Edwardes, MD, Forrester, JE **(2003).** Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am. J. Epidemiol.*, 157, 4:364-75.

Henderson, R, Diggle, P, Dobson, A **(2000)**. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1, 4:465-80.

Hsieh, F, Tseng, YK, Wang, JL **(2006)**. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62, 4:1037-43.

Inoue, LY, Thall, PF, Berry, DA **(2002)**. Seamlessly expanding a randomized phase II trial to phase III. *Biometrics*, 58, 4:823-31.

Kaplan, EL, Meier, P **(1958)**. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assn*. 53, 282: 457–481.

Lai, TL, Lavori, PW, Shih, MC **(2012)**. Sequential design of phase II-III cancer trials. *Stat Med*, 31, 18:1944-60.

Lancaster, T **(2004)**. An Introduction to Modern Bayesian Econometrics, *Blackwell Publishing*. ISBN 1-4051-1720-6

Li, N, Elashoff, RM, Li, G, Saver, J **(2010)**. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Stat Med*, 29, 5:546-57.

Lister, TA, Crowther, D, Sutcliffe, SB, Glatstein, E, Canellos, GP, Young, RC, Rosenberg, SA,

Coltman, CA, Tubiana, M **(1989)**. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *J. Clin. Oncol.*, 7, 11:1630-6.

Meyer, RM, Gospodarowicz, MK, Connors, JM, Pearcey, RG, Bezjak, A, Wells, WA, Burns, BF, Winter, JN, Horning, SJ, Dar, AR, Djurfeldt, MS, Ding, K, Shepherd, LE **(2005)**. Randomized comparison of ABVD chemotherapy with a strategy that includes radiation therapy in patients with limited-stage Hodgkin's lymphoma: National Cancer Institute of Canada Clinical Trials Group and the Eastern Cooperative Oncology Group. *J. Clin. Oncol.*, 23, 21:4634-42.

Meyer, RM, Gospodarowicz, MK, Connors, JM, Pearcey, RG, Wells, WA, Winter, JN, Horning, SJ, Dar, AR, Shustik, C, Stewart, DA, Crump, M, Djurfeldt, MS, Chen, BE, Shepherd, LE **(2012)**. ABVD alone versus radiation-based therapy in limited-stage Hodgkin's lymphoma. *N. Engl. J. Med.*, 366, 5:399-408.

Neuhaus, A, Augustin, T, Heumann, C, Daumer, M **(2009)**. A review on joint models in biometrical research. *Journal of Statistical Theory and Practice*. Volume 3, No.4.

Pepe, MS, Etzioni, R, Feng, Z, Potter, JD, Thompson, ML, Thornquist, M, Winget, M, Yasui, Y **(2001)**. Phases of biomarker development for early detection of cancer. *J. Natl. Cancer Inst.*, 93, 14:1054-61.

R Core Team **(2012)**. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ripatti, S, Palmgren, J **(2000)**. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56, 4:1016-22.

Rizopoulos, D, Verbeke, G, Lesaffre, E **(2009)**. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society Series*, 71, 3: 637–654.

Singh, R, Mukhopadhyay, K **(2011)**. Survival analysis in clinical trials: Basics and must know areas. *Perspect Clin Res*, 2, 4:145-8.

Song, H, Peng, YW, Tu, DS **(2012)**. A new approach for joint modelling of longitudinal measurements and survival times with a cure fraction. *The Canadian Journal of Statistics*, 40, 2: 207–224

Wei, LJ **(1992)**. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat Med*, 11, 14-15:1871-9.

Wu, L, Liu, W, Yi, GY, Huang, YX **(2012)**. Analysis of longitudinal and survival data: Joint modeling, inference Methods, and issues. *Hindawi Publishing Corporation Journal of Probability and Statistics*, Volume 2012 doi:10.1155/2012/640153.

Wulfsohn, MS, Tsiatis, AA **(1997)**. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53, 1:330-9.

Ye, W, Lin, XH, Taylor, J **(2008)**. A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and Its Interface*, 1, 33–45.

Appendix

Appendix 1.

Results of simulation studies with different parameter specifications

Simulation #1

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	-0.002	0.206	0.206	0.938	0.003	0.209	0.201	0.960	1.049
Cox Frailty Model										
γ_1	log(0.5)	0.001	0.190	0.197	0.946	-0.065	0.197	0.210	0.934	0.807
γ_2	log(2)	0.002	0.110	0.116	0.946	-0.002	0.116	0.113	0.958	1.043
γ_3	log(2)	-0.012	0.260	0.275	0.952	0.003	0.269	0.281	0.938	0.960
Variance Components										
σ_u^2	0.5	-0.032	0.124	0.202	0.724	-0.022		0.227		0.806
σ_v^2	0.5	-0.020	0.125	0.154	0.838	-0.006		0.156		0.979
σ_{uv}	-0.45	0.030	0.116	0.150	0.806					

Simulation #2

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.029	0.204	0.201	0.950	-0.004	0.209	0.211	0.950	0.927
Cox Frailty Model										
γ_1	log(0.5)	0.010	0.179	0.182	0.954	0.007	0.187	0.181	0.962	1.011
γ_2	log(2.0)	0.006	0.112	0.123	0.918	0.004	0.117	0.124	0.936	0.977
γ_3	log(2.0)	-0.010	0.248	0.258	0.944	0.009	0.258	0.259	0.952	0.995
Variance Components										
σ_u^2	0.5	-0.059	0.125	0.199	0.704	-0.049		0.220		0.847
σ_v^2	0.5	-0.017	0.126	0.142	0.858	0.000		0.150		0.912
σ_{uv}	0.0	0.005	0.086	0.126	0.818					

Simulation #3

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.017	0.206	0.203	0.950	-0.006	0.209	0.209	0.950	0.952
Cox Frailty Model										
γ_1	log(0.5)	0.015	0.170	0.180	0.930	0.047	0.180	0.178	0.942	0.955
γ_2	log(2.0)	0.011	0.114	0.123	0.926	0.011	0.119	0.118	0.954	1.082
γ_3	log(2.0)	-0.013	0.239	0.247	0.938	0.023	0.249	0.247	0.940	0.990
Variance Components										
σ_u^2	0.5	-0.013	0.129	0.207	0.730	-0.037		0.233		0.774
σ_v^2	0.5	-0.008	0.128	0.149	0.856	-0.015		0.155		0.923
σ_{uv}	0.45	-0.014	0.120	0.150	0.844					

Simulation #4

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	-0.002	0.207	0.200	0.948	0.005	0.212	0.209	0.954	0.914
Cox Frailty Model										
γ_1	log(0.5)	-0.006	0.196	0.213	0.932	-0.062	0.203	0.215	0.934	0.901
γ_2	log(2.0)	0.000	0.111	0.116	0.940	-0.005	0.117	0.119	0.946	0.953
γ_3	log(2.0)	0.002	0.265	0.293	0.924	0.001	0.274	0.291	0.940	1.016
Variance Components										
σ_u^2	1.0	-0.091	0.238	0.351	0.730	-0.053		0.367		0.954
σ_v^2	1.0	-0.045	0.247	0.286	0.860	-0.019		0.296		0.954
σ_{uv}	-0.9	0.062	0.228	0.276	0.824					

Simulation #5

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.012	0.205	0.197	0.954	-0.005	0.211	0.215	0.950	0.838
Cox Frailty Model										
γ_1	log(0.5)	0.001	0.176	0.200	0.922	-0.019	0.189	0.190	0.946	1.105
γ_2	log(2.0)	0.000	0.115	0.118	0.940	0.005	0.121	0.118	0.950	1.004
γ_3	log(2.0)	-0.003	0.244	0.267	0.926	0.030	0.255	0.247	0.952	1.154
Variance Components										
σ_u^2	1.0	-0.174	0.224	0.325	0.668	-0.059		0.411		0.790
σ_v^2	1.0	-0.038	0.249	0.272	0.896	0.021		0.330		0.686
σ_{uv}	0.0	-0.006	0.164	0.202	0.902					

Simulation #6

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.024	0.208	0.199	0.960	0.012	0.212	0.214	0.942	0.877
Cox Frailty Model										
γ_1	log(0.5)	0.010	0.165	0.175	0.946	0.048	0.178	0.188	0.922	0.818
γ_2	log(2.0)	0.009	0.118	0.125	0.930	0.006	0.123	0.120	0.972	1.084
γ_3	log(2.0)	-0.010	0.231	0.248	0.934	0.020	0.242	0.252	0.938	0.962
Variance Components										
σ_u^2	1.0	-0.066	0.244	0.339	0.790	-0.023		0.401		0.739
σ_v^2	1.0	-0.019	0.254	0.279	0.864	0.002		0.303		0.851
σ_{uv}	0.9	-0.033	0.235	0.271	0.870					

Simulation #7

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	-0.002	0.206	0.207	0.938	-0.004	0.210	0.213	0.944	0.940
Cox Frailty Model										
γ_1	log(2.0)	-0.002	0.151	0.163	0.942	-0.065	0.161	0.165	0.934	0.843
γ_2	log(2.0)	0.003	0.111	0.115	0.944	0.001	0.116	0.118	0.940	0.944
γ_3	log(2.0)	-0.005	0.223	0.232	0.934	0.020	0.233	0.233	0.946	0.988
Variance Components										
σ_u^2	0.5	-0.031	0.124	0.202	0.726	-0.017		0.227		0.801
σ_v^2	0.5	-0.017	0.126	0.152	0.864	-0.014		0.148		1.059
σ_{uv}	-0.45	0.026	0.116	0.150	0.808					

Simulation #8

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.029	0.204	0.201	0.950	0.023	0.209	0.205	0.954	0.973
Cox Frailty Model										
γ_1	log(2.0)	0.012	0.150	0.160	0.942	0.006	0.159	0.162	0.950	0.982
γ_2	log(2.0)	0.006	0.112	0.122	0.930	-0.008	0.117	0.117	0.956	1.083
γ_3	log(2.0)	-0.010	0.221	0.229	0.948	0.024	0.232	0.227	0.950	1.005
Variance Components										
σ_u^2	0.5	-0.059	0.125	0.199	0.702	-0.030		0.234		0.778
σ_v^2	0.5	-0.017	0.126	0.143	0.862	0.023		0.163		0.763
σ_{uv}	0.0	0.005	0.086	0.125	0.822					

Simulation #9

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.014	0.206	0.203	0.950	0.017	0.209	0.210	0.950	0.930
Cox Frailty Model										
γ_1	log(2.0)	0.009	0.149	0.155	0.934	0.051	0.157	0.158	0.928	0.879
γ_2	log(2.0)	0.000	0.114	0.122	0.942	0.007	0.119	0.114	0.956	1.135
γ_3	log(2.0)	0.008	0.221	0.227	0.938	0.024	0.230	0.237	0.938	0.912
Variance Components										
σ_u^2	0.5	-0.031	0.125	0.203	0.712	-0.014		0.239		0.733
σ_v^2	0.5	-0.025	0.123	0.149	0.838	-0.001		0.151		1.002
σ_{uv}	0.45	-0.032	0.115	0.150	0.778					

Simulation #10

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.024	0.208	0.212	0.936	-0.011	0.211	0.220	0.940	0.938
Cox Frailty Model										
γ_1	log(2.0)	0.003	0.153	0.170	0.912	-0.059	0.167	0.164	0.934	0.948
γ_2	log(2.0)	-0.001	0.111	0.116	0.944	0.001	0.117	0.124	0.934	0.866
γ_3	log(2.0)	0.008	0.222	0.228	0.928	0.018	0.233	0.232	0.950	0.961
Variance Components										
σ_u^2	1.0	-0.080	0.241	0.352	0.738	-0.014		0.419		0.738
σ_v^2	1.0	-0.044	0.247	0.299	0.844	0.010		0.301		1.005
σ_{uv}	-0.9	0.052	0.230	0.282	0.836					

Simulation #11

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.038	0.206	0.211	0.932	-0.029	0.212	0.213	0.940	0.994
Cox Frailty Model										
γ_1	log(2.0)	0.007	0.149	0.168	0.916	0.008	0.162	0.167	0.944	1.010
γ_2	log(2.0)	-0.001	0.114	0.123	0.942	-0.003	0.120	0.118	0.962	1.098
γ_3	log(2.0)	0.009	0.218	0.232	0.936	0.014	0.229	0.229	0.954	1.025
Variance Components										
σ_u^2	1.0	-0.174	0.224	0.321	0.670	-0.038		0.374		0.944
σ_v^2	1.0	-0.053	0.245	0.275	0.858	0.015		0.301		0.861
σ_{uv}	0	-0.017	0.162	0.190	0.900					

Simulation #12

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(0.5)	0.012	0.207	0.206	0.946	0.014	0.212	0.223	0.936	0.852
Cox Frailty Model										
γ_1	log(2.0)	0.007	0.149	0.155	0.942	0.042	0.159	0.157	0.958	0.911
γ_2	log(2.0)	-0.005	0.118	0.119	0.956	0.004	0.123	0.124	0.940	0.912
γ_3	log(2.0)	0.012	0.216	0.225	0.932	0.019	0.227	0.225	0.942	0.996
Variance Components										
σ_u^2	1.0	-0.079	0.241	0.360	0.752	-0.025		0.399		0.848
σ_v^2	1.0	-0.047	0.247	0.282	0.840	0.013		0.302		0.898
σ_{uv}	0.9	-0.052	0.230	0.277	0.834					

Simulation #13

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.006	0.181	0.186	0.946	0.007	0.183	0.193	0.942	0.925
Cox Frailty Model										
γ_1	log(0.5)	-0.008	0.190	0.197	0.928	-0.055	0.196	0.205	0.924	0.858
γ_2	log(2.0)	0.001	0.120	0.124	0.936	0.010	0.125	0.114	0.966	1.168
γ_3	log(2.0)	0.004	0.231	0.242	0.942	0.000	0.239	0.245	0.934	0.978
Variance Components										
σ_u^2	0.5	-0.031	0.124	0.190	0.740	-0.021		0.199		0.929
σ_v^2	0.5	-0.021	0.125	0.162	0.816	-0.004		0.156		1.101
σ_{uv}	-0.45	0.027	0.116	0.151	0.810					

Simulation #14

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.019	0.179	0.184	0.944	0.001	0.183	0.181	0.938	1.042
Cox Frailty Model										
γ_1	log(0.5)	-0.005	0.179	0.183	0.932	0.005	0.186	0.197	0.938	0.864
γ_2	log(2.0)	-0.001	0.122	0.127	0.946	0.003	0.128	0.124	0.964	1.045
γ_3	log(2.0)	0.005	0.222	0.230	0.938	0.005	0.229	0.239	0.936	0.932
Variance Components										
σ_u^2	0.5	-0.067	0.120	0.182	0.700	-0.014		0.218		0.784
σ_v^2	0.5	-0.027	0.123	0.149	0.842	0.016		0.163		0.848
σ_{uv}	0.0	-0.003	0.084	0.111	0.882					

Simulation #15

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.007	0.181	0.187	0.940	0.003	0.183	0.184	0.936	1.035
Cox Frailty Model										
γ_1	log(0.5)	0.000	0.170	0.175	0.932	0.068	0.179	0.187	0.912	0.772
γ_2	log(2.0)	0.003	0.124	0.130	0.938	-0.004	0.130	0.133	0.954	0.957
γ_3	log(2.0)	-0.001	0.215	0.225	0.934	-0.013	0.223	0.234	0.948	0.924
Variance Components										
σ_u^2	0.5	-0.033	0.123	0.192	0.738	-0.022		0.210		0.850
σ_v^2	0.5	-0.023	0.124	0.152	0.830	0.004		0.162		0.894
σ_{uv}	0.45	-0.029	0.115	0.149	0.786					

Simulation #16

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.010	0.186	0.187	0.954	0.008	0.190	0.197	0.946	0.904
Cox Frailty Model										
γ_1	log(0.5)	-0.008	0.196	0.209	0.930	-0.070	0.204	0.221	0.934	0.815
γ_2	log(2.0)	0.001	0.120	0.130	0.932	-0.001	0.127	0.133	0.942	0.955
γ_3	log(2.0)	0.006	0.237	0.256	0.938	0.014	0.245	0.256	0.930	0.997
Variance Components										
σ_u^2	1.0	-0.095	0.236	0.337	0.756	-0.014		0.334		1.094
σ_v^2	1.0	-0.050	0.246	0.301	0.824	-0.014		0.299		1.038
σ_{uv}	-0.9	0.061	0.227	0.281	0.830					

Simulation #17

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.024	0.184	0.184	0.956	0.000	0.190	0.194	0.962	0.909
Cox Frailty Model										
γ_1	log(0.5)	-0.004	0.177	0.191	0.926	-0.012	0.190	0.194	0.946	0.967
γ_2	log(2.0)	0.001	0.124	0.136	0.934	-0.008	0.131	0.137	0.932	0.991
γ_3	log(2.0)	0.003	0.222	0.240	0.942	0.022	0.231	0.251	0.922	0.904
Variance Components										
σ_u^2	1.0	-0.158	0.224	0.312	0.704	-0.026		0.365		0.912
σ_v^2	1.0	-0.057	0.244	0.276	0.854	-0.006		0.293		0.922
σ_{uv}	0.0	-0.011	0.163	0.190	0.924					

Simulation #18

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.005	0.186	0.196	0.946	0.010	0.190	0.195	0.938	1.007
Cox Frailty Model										
γ_1	log(0.5)	0.010	0.165	0.175	0.946	0.059	0.178	0.190	0.918	0.777
γ_2	log(2.0)	0.009	0.129	0.131	0.956	-0.008	0.136	0.146	0.934	0.804
γ_3	log(2.0)	-0.008	0.212	0.214	0.946	0.002	0.221	0.234	0.936	0.839
Variance Components										
σ_u^2	1.0	-0.057	0.246	0.334	0.782	-0.033		0.391		0.746
σ_v^2	1.0	-0.018	0.254	0.282	0.874	-0.001		0.328		0.742
σ_{uv}	0.9	-0.032	0.236	0.276	0.862					

Simulation #23

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.017	0.184	0.184	0.952	-0.003	0.189	0.194	0.948	0.903
Cox Frailty Model										
γ_1	log(2.0)	0.004	0.149	0.167	0.920	0.003	0.162	0.165	0.954	1.019
γ_2	log(2.0)	-0.004	0.125	0.123	0.964	0.002	0.131	0.135	0.944	0.833
γ_3	log(2.0)	-0.004	0.196	0.202	0.938	0.010	0.206	0.209	0.956	0.929
Variance Components										
σ_u^2	1.0	-0.145	0.228	0.313	0.712	-0.018		0.386		0.797
σ_v^2	1.0	-0.049	0.246	0.280	0.856	0.009		0.298		0.908
σ_{uv}	0.0	-0.002	0.166	0.212	0.910					

Simulation #24

Parameter	True	Joint Model				Separate Model				MSE_J/MSE_S
		Bias	ASE	ESE	CP	Bias	ASE	ESE	CP	
GLMM										
β_1	log(2.0)	-0.004	0.186	0.196	0.944	-0.008	0.190	0.193	0.944	1.037
Cox Frailty Model										
γ_1	log(2.0)	0.016	0.151	0.159	0.944	0.061	0.159	0.153	0.946	0.940
γ_2	log(2.0)	0.009	0.129	0.130	0.956	-0.007	0.135	0.132	0.960	0.975
γ_3	log(2.0)	-0.015	0.197	0.198	0.946	-0.001	0.207	0.200	0.952	0.992
Variance Components										
σ_u^2	1.0	-0.057	0.246	0.334	0.788	-0.020		0.371		0.829
σ_v^2	1.0	-0.017	0.254	0.279	0.872	-0.002		0.301		0.865
σ_{uv}	0.9	-0.032	0.236	0.275	0.866					

Appendix 2.

Absolute bias and MSE for joint model when number of centres takes different values.

Number of Centres	Absolute Bias				MSE			
	β_1	γ_1	γ_2	γ_3	β_1	γ_1	γ_2	γ_3
25	0.000	0.006	0.001	0.024	0.036	0.039	0.013	0.072
30	0.002	0.001	0.002	0.012	0.043	0.039	0.013	0.076
40	0.003	0.026	0.005	0.023	0.042	0.040	0.013	0.076
60	0.018	0.012	0.006	0.001	0.044	0.043	0.014	0.081
100	0.008	0.021	0.009	0.011	0.045	0.042	0.015	0.071

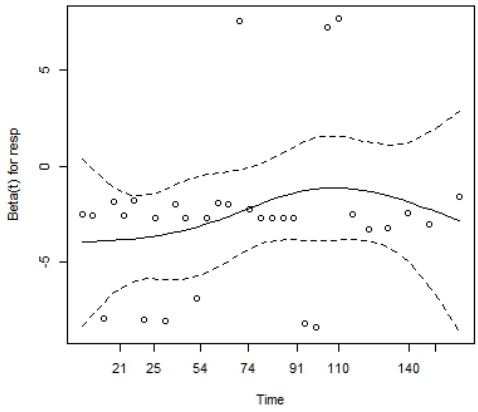
Appendix 3.

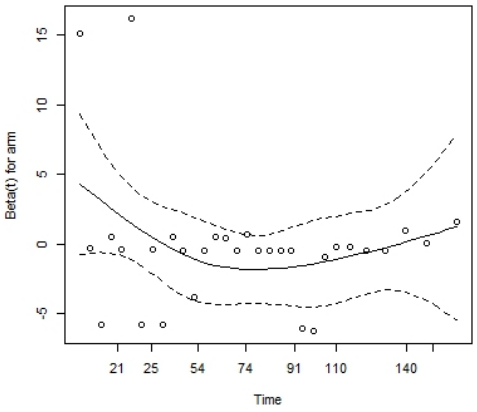
Characterization of patients with missing response data

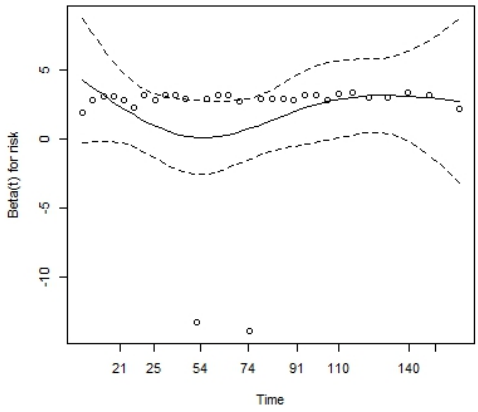
Arm	Event		Sex		Risk	
	Censored	Uncensored	Female	Male	High	Low
A	22	2	12	12	12	12
B	16	3	5	14	13	6
P-value	0.6404		0.1327		0.351	

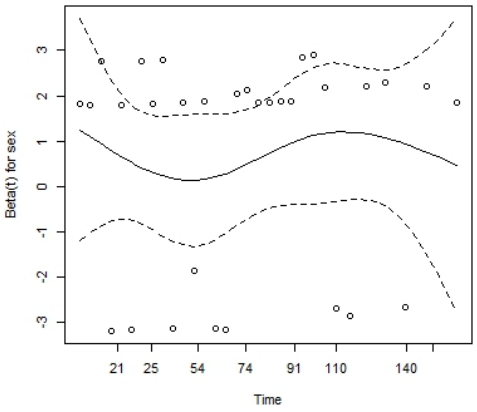
Appendix 4.

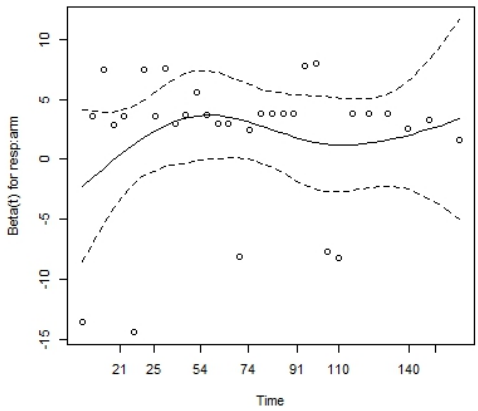
Graphic analysis of proportional hazards assumption using Schoenfeld residuals











Appendix 5.

Sensitivity analysis 1: Set all missing value to "Yes"

Parameter	EST	MPL		Jackknife	
		ASE	P-value	JK-SE	JK P-value
GLMM					
Arm	1.394	0.395	0.000 †	0.374	0.0002 †
Risk	−0.118	0.383	0.758	0.457	0.7967
Sex	0.194	0.346	0.576	0.407	0.6339
Cox Frailty Model					
Resp	−2.050	0.612	0.001 †	0.603	0.0007 †
Arm	0.005	0.823	0.995	1.015	0.9962
Risk	2.082	0.730	0.004 †	0.888	0.0190 †
Sex	0.569	0.368	0.122	0.446	0.2018
Resp×Arm	1.365	0.961	0.156	1.158	0.2384
Variance components					
σ_u^2	0.078	0.036	−	0.111	−
σ_v^2	0.328	0.137	−	0.433	−
σ_{uv}	−0.154	0.066	0.020 †	0.208	0.4584

Appendix 6.

Sensitivity analysis 2: Set all missing value to "No"

Parameter	EST	MPL		Jackknife	
		ASE	P-value	JK-SE	JK P-value
GLMM					
Arm	0.479	0.259	0.064	0.264	0.0693
Risk	0.297	0.273	0.276	0.249	0.2323
Sex	−0.084	0.261	0.746	0.252	0.7378
Cox Frailty Model					
Resp	−2.318	0.678	0.001 †	0.589	0.0001 †
Arm	−0.327	0.615	0.594	0.551	0.5522
Risk	2.102	0.729	0.004 †	0.888	0.0179 †
Sex	0.476	0.366	0.193	0.428	0.2664
Resp×Arm	2.134	0.873	0.015 †	0.623	0.0006 †
Variance components					
σ_u^2	0.031	0.017	−	0.055	−
σ_v^2	0.331	0.139	−	0.435	−
σ_{uv}	−0.093	0.042	0.026 †	0.147	0.5244 †

R Code

```
library(survival)
#### Function "rnormal" generates random effects u and v from bivariate normal distribution
library(mnormt)
rnormal<-function (ncentre,var_u,var_v,cov_uv) {
   var_cov<-matrix(c(var_u,cov_uv, cov_uv, var_v),2,2)
   ran<-rmnorm(ncentre,mean=c(0,0),varcov=var_cov)
   return (ran)
}

#### Function "MPL_step1" updates beta, gamma and random effects (treated as parameters)
## Include additional covariates if other potential predictors are of interest
MPL_step1<-function (max.iter,tol,n,u,v,ncentre,u_p,v_p,c_matrix,var_v, var_u, cov_uv,resp,
                                arm, event, time) {
   ##update beta and gamma with standard R functions "glm" and "coxph"
   logi<-glm(resp~as.factor(arm)+offset(u_p),family=binomial)
   beta<-as.numeric(logi$coefficients)
   se.beta<-as.numeric(summary(logi)$coefficients[,2])
   ph<-coxph(Surv(time, event)~as.factor(resp)+as.factor(arm)+as.factor(resp)*as.factor(arm)+
                     offset(v_p))
   gamma<-as.numeric(ph$coefficients)
   se.gamma<-as.numeric(summary(ph)$coefficients[,3])
   ##baseline hazard for each observation was obtained by basehaz
   df1<-as.data.frame(time)
   df2<-as.data.frame(basehaz(ph, center=F))
   df<-merge(df1,df2,by="time")
   lambda<-df$hazard
   lambda<-sort(lambda,decreasing=TRUE)
   ##update u and v with N-R
   s_u<-vector();i_u<-vector();s_v<-vector();i_v<-vector()
   u_new<-vector(); v_new<-vector();list_info<-list()
   flag <- 0
   for (x in 1:max.iter){
      cur=c(u,v)
      uv_centre<-cbind(u,v)
      uv_p<-c_matrix%*%uv_centre
      u_p<-uv_p[,1];v_p<-uv_p[,2]
      ##calculate score and information for u
      X1=cbind(1, arm)
      eb = exp(X1%*%beta+u_p)
      eb1 = eb/(1+eb)
      s_u = resp - eb1
```

```
    i_u = -1*eb1*(1-eb1)
    ##calculate score and information for v
    X2 = cbind(resp, arm, resp*arm)
    exp_all = lambda*exp(X2%*%gamma+v_p)
    s_v = event - exp_all
    i_v = -1*exp_all
    ##calculate centre-specific score/information function for u and v
    dinv = 1/(var_u*var_v-cov_uv^2)
    s_u_c<-t(c_matrix)%*%s_u-(u*var_v-v*cov_uv)*dinv; i_u_c<-t(c_matrix)%*%i_u-var_v*dinv
    s_v_c<-t(c_matrix)%*%s_v- (v*var_u-u*cov_uv)*dinv; i_v_c<-t(c_matrix)%*%i_v-var_u*dinv
    cov_12<-cov_uv/(var_u*var_v-cov_uv^2)
    A<-t(c_matrix)%*%i_u; B<-t(c_matrix)%*%i_v
    ##update u and v
    for (e in 1:ncentre) {
      uv<-cbind(u[e],v[e]);score<-cbind(s_u_c[e], s_v_c[e])
      info<-matrix(c(i_u_c[e],cov_12,cov_12,i_v_c[e]),2,2,byrow=TRUE)
      inverse<-solve(info)
      uv_new<-uv-score%*%inverse
      u_new[e]<-uv_new[1];v_new[e]<-uv_new[2]
      list_info[[e]]<-inverse
    }
    new=c(u_new, v_new)
    # Stop iteration if difference between current and new estimates is less than tol
    if( max(abs(cur - new)) < tol){ flag <- 1; break}
    else{u<-u_new;v<-v_new}
  }
  return (list(u=u_new, v=v_new, beta=beta, gamma=gamma,list_info=list_info,
               se.beta=se.beta,se.gamma=se.gamma,A=A,B=B))
}
```

Function "MPL_step2" finds the MLE of beta, gamma and random effects
given variance components

```
MPL_step2 <- function(max.iter,tol,n,u,v,ncentre,c_matrix, var_v, var_u, cov_uv, beta, gamma,
                      resp, arm, event, time){
  flag <- 0
  for(k in 1:max.iter){
    cur <- c(u,v,beta,gamma)
    ###assign the random effect value to each patient (uv_p means uv_patient)
    uv_centre<-cbind(u,v)
    uv_p<-c_matrix%*%uv_centre
    u_p<-uv_p[,1];v_p<-uv_p[,2]
    new_step <- MPL_step1(max.iter,tol,n,u,v,ncentre,u_p,v_p,c_matrix,var_v, var_u, cov_uv,
                          resp, arm, event, time)
    u <- new_step$u; v <- new_step$v;
```

```
    beta <- new_step$beta
    gamma <- new_step$gamma
    new <- c(u,v,beta, gamma)
    # Stop iteration if difference between current and new estimates is less than tol
    if( max(abs(cur - new)) < tol){ flag <- 1; break}
  }
  if(!flag) warning("Not converge\n")
  li_info<-new_step$list_info
  se.beta<-new_step$se.beta; se.gamma<-new_step$se.gamma
  A<-new_step$A; B<-new_step$B
  return(list(u=u,v=v,beta=beta,gamma=gamma,li_info=li_info,
                se.beta=se.beta,se.gamma=se.gamma,A=A,B=B))
}
```

Function "MPL_step3" finds MLE of variance components when fixing the other parameter
and iterates between step1~ step3 till all parameter estimates converge

```
MPL_step3<-function (max.iter2,tol2, n, ncentre,resp, arm, event, time,centre) {
  ##assign initial values to unknown parameters
  var_u<- 0.5; var_v<-0.5; cov_uv<- 0.01
  beta<-c(-1,0.5);gamma<-c(0.5,1,-1)
  max.iter<-1000
  tol<-0.0001
  ran<-normal(ncentre, var_u,var_v,cov_uv)
  u<-ran[,1]; v<-ran[,2]
  ###create the matrix which indicates the centre ID for each patient
  c_matrix<-matrix(rep(0,n*ncentre),n,ncentre)
  for (i in 1:n){
    for (j in 1:ncentre){
      if (as.numeric(centre[i])==j) {c_matrix[i,j]=1}
    }
  }
  for (l in 1:max.iter2){
    cur2 <- c(var_u,var_v,cov_uv,beta, gamma)
    data1<-MPL_step2(max.iter,tol,n,u,v,ncentre,c_matrix, var_v, var_u, cov_uv,beta, gamma,
                     resp, arm, event, time)
    u <- data1$u; v <- data1$v;
    beta <- data1$beta
    gamma <- data1$gamma
    ##update variance components
    l_info<-data1$li_info
    mat_sum<-Reduce('+',l_info)
    matrix1<-rbind(u,v)
    matrix_uv<-(matrix1%*%t(matrix1)-mat_sum)/ncentre
    var_u<-matrix_uv[1,1];var_v<-matrix_uv[2,2];cov_uv<-matrix_uv[1,2]
```

```
    new2 <- c(var_u,var_v,cov_uv,beta, gamma)
    # Stop iteration if difference between current and new estimates is less than tol
    if( max(abs(cur2 - new2)) < tol2 ){ flag <- 1; break}
  }
  if(!flag) warning("Not converge\n")
  se.beta<-data1$se.beta
  se.gamma<-data1$se.gamma
  ####manual calculation of variance of variance components using Fisher information
  A<-data1$A; B<-data1$B
  a<-var_u;b<-var_v;c<-cov_uv
  K<-A*B*(a*b-c^2)^2-B*(a*b-c^2)*b-A*(a*b-c^2)*a+(a*b-c^2)
  Iaa<-ncentre*b^2/(2*(a*b-c^2)^2)+
       1/2*sum((B*b^2+A*c^2-b)*(A*B*(2*a*b^2-2*b*c^2)-B*b^2-(2*a*b-c^2)*A+b)/(K^2))-
       1/2*sum((v^2*c^2+u^2*b^2-2*u*v*b*c)*(2*a*b^2-2*b*c^2)/((a*b-c^2)^4))
  Ibb<-ncentre*a^2/(2*(a*b-c^2)^2)+
       1/2*sum((B*c^2+A*a^2-a)*(A*B*(2*b*a^2-2*a*c^2)-A*a^2-(2*a*b-c^2)*B+a)/(K^2))-
       1/2*sum((u^2*c^2+v^2*a^2-2*u*v*a*c)*(2*b*a^2-2*a*c^2)/((a*b-c^2)^4))
  Icc<-ncentre*(a*b+c^2)/((a*b-c^2)^2)+sum(((B*b+A*a-1)*K-(B*b*c+A*a*c-c)*
       (A*B*(4*c^3-4*a*b*c)+2*B*b*c+2*A*a*c-2*c))/(K^2))-sum(((u^2*b+v^2*a-2*u*v*c)*
       (a*b-c^2)^2-(u^2*b*c+v^2*a*c-u*v*a*b-u*v*c^2)*(4*c^3-4*a*b*c))/((a*b-c^2)^4))
  Iab<-ncentre*c^2/(2*(a*b-c^2)^2)-1/2*sum(((2*B*b-1)*K-(B*b^2+A*c^2-b)*
       (A*B*(2*b*a^2-2*a*c^2)-A*a^2-(2*a*b-c^2)*B+a))/(K^2))+1/2*sum(((2*u^2*b-2*u*v*c)*
       (a*b-c^2)^2-(v^2*c^2+u^2*b^2-2*u*v*b*c)*(2*b*a^2-2*a*c^2))/((a*b-c^2)^4))
  Iac<-ncentre*(-1)*b*c/((a*b-c^2)^2)-1/2*sum((2*A*c*K-(B*b^2+A*c^2-b)*
       (A*B*(4*c^3-4*a*b*c)+2*B*b*c+2*A*a*c-2*c))/(K^2))+1/2*sum(((2*v^2*c-2*u*v*b)*
       (a*b-c^2)^2-(v^2*c^2+u^2*b^2-2*u*v*b*c)*(4*c^3-4*a*b*c))/((a*b-c^2)^4))
  Ibc<-ncentre*(-1)*a*c/((a*b-c^2)^2)-1/2*sum((2*B*c*K-(B*c^2+A*a^2-a)*
       (A*B*(4*c^3-4*a*b*c)+2*B*b*c+2*A*a*c-2*c))/(K^2))+1/2*sum(((2*u^2*c-2*u*v*a)*
       (a*b-c^2)^2-(u^2*c^2+v^2*a^2-2*u*v*a*c)*(4*c^3-4*a*b*c))/((a*b-c^2)^4))
  Iabc<-(-1)*matrix(c(Iaa,Iab, Iac, Iab, Ibb, Ibc, Iac, Ibc, Icc),3,3, byrow=TRUE)
  inv_Iabc<-solve(Iabc)
  var_sig<-c(inv_Iabc[1,1],inv_Iabc[2,2],inv_Iabc[3,3])
  return(list(var_u=var_u,var_v=var_v,cov_uv=cov_uv,beta=beta,gamma=gamma,
              se.beta=se.beta, se.gamma=se.gamma,var_sig=var_sig))
}
```

Function "simu.joint" first generates simulated data and then applies the joint model #### (with MPL) to the data

```
simu.joint <- function (n, ncentre, b0,b1, g1, g2, g3, sigma_u, sigma_v, sigma_uv, nsimu) {
  n=n
  ncentre=ncentre
  max.iter2=10000
  tol2=0.00001
  v_u<-vector();v_v<-vector(); covuv<-vector()
```

```
beta_list<-list(); gamma_list<-list()
sebeta<-list();segamma<-list()
se_v_u<-vector();se_v_v<-vector();se_covuv<-vector()
n_beta1=0; n_beta2=0; n_gamma1=0;n_gamma2=0;n_gamma3=0
n_var_u=0; n_var_v=0; n_covuv=0
##set the initial values for all the parameters that will be estimated
sim_beta<-c(b0, b1)
sim_gamma<-c(g1, g2, g3)
sim_sigma<-c(sigma_u, sigma_v, sigma_uv)
sim_var_u=sim_sigma[1];sim_var_v=sim_sigma[2];sim_cov_uv=sim_sigma[3]
##generate centre ID
sim_centre = rep(c(1:ncentre), n/ncentre)
sim_centre<-as.factor(sim_centre)
###ct_matrix records the centre information for all the patients
ct_matrix<-matrix(rep(0,n*ncentre),n,ncentre)
for (i in 1:n){
   for (j in 1:ncentre){
      if (as.numeric(sim_centre[i])==j) {ct_matrix[i,j]=1}
   }
}
for (s in 1:nsimu) {
   centre = rep(c(1:ncentre), n/ncentre)
   centre<-as.factor(centre)
   ##generate center-specific random effect u and v
   ran<-rnormal(ncentre,sim_var_u,sim_var_v,sim_cov_uv)
   u<-ran[,1]; v<-ran[,2]
   ##assign centre ID to each patient
   uv_centre<-cbind(u,v)
   uv_p<-ct_matrix%*%uv_centre
   u_p<-uv_p[,1];v_p<-uv_p[,2]
   arm<-rbinom(n,size=1,prob=0.5)        ##simulate "arm" variable
   sim_X1<-cbind(1, arm)
   ##simulate resp variable
   za<-exp(sim_X1%*%sim_beta+u_p)
   resp = rbinom(n, 1, za/(1+za))
   ##simulate the survival time
   ##assume in baseline hazard lambda=0.15, p=1 (Follow exponential distribution)
   ranuni<-runif(n,min=0,max=1)
   sim_X2<-cbind(resp, arm, resp*arm)
   stime<-log(ranuni)/exp(sim_X2%*%sim_gamma+v_p)/(-0.15)
   endstudy = runif(n,0, 20)      ###make the censoring rate around 20%
   event = ifelse(stime>endstudy, 0, 1)
   time = ifelse(stime>endstudy, endstudy, stime)
   ##put all data in a data frame and sort the data based on survival time
```

```
  dat = data.frame(time, event, resp,arm,centre)
  st = sort(dat$time, decr = T, index = T)
  idx = st$ix
  dat = dat[idx, ]
  time<-dat$time; event<-dat$event
  resp<-dat$resp; arm<-dat$arm; centre<-dat$centre
  ####calculate parameter estimates with MPL method
  simu<-MPL_step3(max.iter2,tol2, n, ncentre,resp, arm, event, time,centre)
  v_u[s]<-simu$var_u;v_v[s]<-simu$var_v; covuv[s]<-simu$cov_uv
  beta_list[[s]]<-simu$beta; gamma_list[[s]]<-simu$gamma
  sebeta[[s]]<-simu$se.beta; segamma[[s]]<-simu$se.gamma
  se_v_u[s]<-sqrt(simu$var_sig[1]);se_v_v[s]<-sqrt(simu$var_sig[2])
  se_covuv[s]<-sqrt(simu$var_sig[3])
  ###calculate the coverage probability of confidence interval at 95% level
  if (sim_beta[1]>=beta_list[[s]][1]-qnorm(0.975)*sebeta[[s]][1] &
       sim_beta[1]<=beta_list[[s]][1]+qnorm(0.975)*sebeta[[s]][1])
      {n_beta1=n_beta1+1}
  if (sim_beta[2]>=beta_list[[s]][2]-qnorm(0.975)*sebeta[[s]][2] &
       sim_beta[2]<=beta_list[[s]][2]+qnorm(0.975)*sebeta[[s]][2])
      {n_beta2=n_beta2+1}
  if (sim_gamma[1]>=gamma_list[[s]][1]-qnorm(0.975)*segamma[[s]][1] &
       sim_gamma[1]<=gamma_list[[s]][1]+qnorm(0.975)*segamma[[s]][1])
      {n_gamma1=n_gamma1+1}
  if (sim_gamma[2]>=gamma_list[[s]][2]-qnorm(0.975)*segamma[[s]][2] &
       sim_gamma[2]<=gamma_list[[s]][2]+qnorm(0.975)*segamma[[s]][2])
      {n_gamma2=n_gamma2+1}
  if (sim_gamma[3]>=gamma_list[[s]][3]-qnorm(0.975)*segamma[[s]][3] &
       sim_gamma[3]<=gamma_list[[s]][3]+qnorm(0.975)*segamma[[s]][3])
      {n_gamma3=n_gamma3+1}
  if (sim_var_u>=v_u[s]-qnorm(0.975)*se_v_u[s] & sim_var_u[1]<=v_u[s]+
       qnorm(0.975)*se_v_u[s])
      {n_var_u=n_var_u+1}
  if (sim_var_v>=v_v[s]-qnorm(0.975)*se_v_v[s] & sim_var_v<=v_v[s]+
       qnorm(0.975)*se_v_v[s])
      {n_var_v=n_var_v+1}
  if (sim_cov_uv>=covuv[s]-qnorm(0.975)*se_covuv[s] &
       sim_cov_uv<=covuv[s]+qnorm(0.975)*se_covuv[s])
      {n_covuv=n_covuv+1}
  n_beta<-c(n_beta1, n_beta2); n_gamma<-c(n_gamma1,n_gamma2,n_gamma3)
  n_sigma<-c(n_var_u, n_var_v, n_covuv)
}
return(list(v_u=v_u,v_v=v_v,covuv=covuv,beta_list=beta_list,gamma_list=gamma_list,
          sebeta=sebeta,segamma=segamma, se_v_u=se_v_u,se_v_v=se_v_v,se_covuv=se_covuv,
          n_beta=n_beta, n_gamma=n_gamma, n_sigma=n_sigma, tru_beta=sim_beta,
```

```
            tru_gamma=sim_gamma, tru_sigma=sim_sigma))
}

#### Function "datasum" summarizes the results from joint modeling
datasum<-function (result, nsimulation) {
  parameter<-c("beta1","beta2",    "gamma1","gamma2","gamma3","var_u", "var_v", "cov_uv")
  ##true parameter values
  true<-c(result$tru_beta,result$tru_gamma, result$tru_sigma)
  ##calculate the estimates of parameter
  var_u<-sum(result$v_u)/nsimulation
  var_v<-sum(result$v_v)/nsimulation
  cov_uv<-sum(result$covuv)/nsimulation
  beta<-Reduce('+',result$beta_list)/nsimulation
  gamma<-Reduce('+',result$gamma_list)/nsimulation
  est<-c(beta,gamma,var_u, var_v, cov_uv)
  ##calculate the bias
  bias<-est-true
  ##calculate empirical S.E. for beta, gamma and variance components
  ese<-vector()
  beta1<-vector(); beta2<-vector();beta3<-vector();beta4<-vector();
  gamma1<-vector();gamma2<-vector();gamma3<-vector();gamma4<-vector();gamma5<-vector()
  for (i in 1:nsimulation) {
    beta1[i]<-result$beta_list[[i]][1]
    beta2[i]<-result$beta_list[[i]][2]
    gamma1[i]<-result$gamma_list[[i]][1]
    gamma2[i]<-result$gamma_list[[i]][2]
    gamma3[i]<-result$gamma_list[[i]][3]
  }
  ese[1]<-sd(beta1);ese[2]<-sd(beta2)
  ese[3]<-sd(gamma1);ese[4]<-sd(gamma2);ese[5]<-sd(gamma3)
  ese[6]<-sd(result$v_u); ese[7]<-sd(result$v_v);ese[8]<-sd(result$covuv)
  ##calculation of asymptotical SE for beta, gamma and variance components
  se.beta<-Reduce('+',result$sebeta)/nsimulation
  se.gamma<-Reduce('+',result$segamma)/nsimulation
  se.v_u<-sum(result$se_v_u)/nsimulation; se.v_v<-sum(result$se_v_v)/nsimulation
  se.cov_uv<-sum(result$se_covuv)/nsimulation
  se<-c(se.beta, se.gamma, se.v_u, se.v_v, se.cov_uv)
  ##calculate coverage probability
  count<-c(result$n_beta, result$n_gamma, result$n_sigma)
  CP<-count/nsimulation
  output<-data.frame (parameter, true, est, bias, ese, se, CP)
  return (output)
}
```

```
#### Functions "glmer" and "coxme" are used to model binary marker response and
#### survival data with random effects separately
library(coxme)
library(lme4)
###fit generalized mixed effect model
fit1<-glmer(resp~as.factor(arm)+(1 | centre),family = binomial)
##find estimates for beta and var_u and s.e for beta
beta<-fixef(fit1)
var_u<-as.numeric(VarCorr(fit1))
se_beta<-sqrt(diag(vcov(fit1)))
### fit Cox PH model with random effect
fit2<-coxme(Surv(time, event)~as.factor(resp)+as.factor(arm)+
                    as.factor(resp)*as.factor(arm)+(1|centre))
##find estimates for gamma and var_v and SE for gamma
gamma<-fit2$coefficients
var_v<-fit2$vcoef$centre
se_gamma<-sqrt(diag(vcov(fit2)))


#### Delete-a-group Jackknife resampling
for (s in 1:sim_ncentre) {
   ##each time remove one centre from the dataset
   data<-subset(dat, dat$centre!=s)
   st = sort(data$time, decr = T, index = T)
   idx = st$ix
   dataJK = data[idx, ]
   time<-dataJK$time
   event<-dataJK$event
   resp<-dataJK$resp
   arm<-dataJK$arm
   ###convert 28 centres into 1~28 ID scale
   centre<-as.factor(as.numeric(as.factor(dataJK$centre)))
   n=nrow(dataJK)
   ncentre<-nlevels(centre)
}

#### Function "JK" calculates the Jackknife standard error
## "result" is the output of joint modeling using jackknife samples
## "jointfull" is the output of joint modeling using whole dataset
JK<-function (result, jointfull, ncentre, nsimulation,n) {
   beta_matrix<-do.call(rbind, result$beta_list)
   gamma_matrix<-do.call(rbind, result$gamma_list)
   var.u<-result$v_u; var.v<-result$v_v;cov.uv<-result$covuv
   p_matrix<-cbind(beta_matrix, gamma_matrix, var.u,var.v,cov.uv)
```

```
  ##find the number of patients in each centre
  c_matrix<-matrix(rep(0,n*ncentre),n,ncentre)
  for (i in 1:n){
    for (j in 1:ncentre){
      if (as.numeric(centre[i])==j) {c_matrix[i,j]=1}
    }
  }
  p_centre<-t(c_matrix)%*%rep(1,n)
  full<-c(jointfull[[4]],jointfull[[5]],jointfull[[1]],jointfull[[2]],jointfull[[3]])
  weight<-1-p_centre/n
  ###JK estimate of parameter
  JK_est<-as.vector(ncentre*full-t(p_matrix)%*%weight)
  ##estimate JK variance (based on the paper by Busing et al. 1999)
  h<-n/p_centre
  pseudo<-h%*%t(full)-as.vector(h-1)*p_matrix
  JK_matrix<-matrix(rep(JK_est,ncentre),ncentre, 8, byrow=TRUE)
  JK_variance<-1/ncentre*t((pseudo-JK_matrix)^2)%*%(1/(h-1))
  JK_se<-as.vector(sqrt(JK_variance))
  return (JK_se)
}
```

####Graphic analysis of correlation between sample size and MSE / bias

```
bias<-read.table("D:/Practicum project/data/bias_size1.csv",sep=",",header=T)
colnames(bias) <- c("size","beta0","beta1","gamma1","gamma2","gamma3","var_u","var_v","cov_uv")
mse<-read.table("D:/Practicum project/data/mse_size1.csv",sep=",",header=T)
colnames(mse) <- c("size","beta0","beta1","gamma1","gamma2","gamma3","var_u","var_v","cov_uv")
##beta1
par(mar=c(5.1,4.1,2.1,4.1))
plot(bias$size,abs(bias$beta1),type="o",pch=16,ylim=c(0,0.07),xlab="Sample Size",
     ylab="Absolute Bias",col="blue")
axis(2, pretty(c(0,0.07)), col="blue")
par(new=TRUE)
plot(mse$size, mse$beta1,type="o",pch=2,lty=2,xlab=" ", ylab=" ", ylim=c(0,0.7), col="red",axes=F)
axis(4, pretty(c(0,0.7)), col="red")
mtext("MSE",side=4,line=1,padj=2)
legend(420, 0.7, legend=c("Absolute Bias", "MSE"), pch=c(16,2),col=c("blue", "red"),bty="n",cex=1)
text(120,0.68, "A")
##gamma1
par(mar=c(5.1,4.1,2.1,4.1))
plot(bias$size,abs(bias$gamma1),type="o",pch=16,ylim=c(0,0.07),xlab="Sample Size",
     ylab="Absolute Bias",col="blue")
axis(2, pretty(c(0,0.07)), col="blue")
par(new=TRUE)
plot(mse$size, mse$gamma1,type="o",pch=2,lty=2,xlab=" ", ylab=" ", ylim=c(0,0.7), col="red",axes=F)
```

```
axis(4, pretty(c(0,0.7)), col="red")
mtext("MSE",side=4,line=1,padj=2)
legend(420, 0.7, legend=c("Absolute Bias", "MSE"), pch=c(16,2),col=c("blue", "red"),bty="n",cex=1)
text(120,0.68, "B")
##gamma2
par(mar=c(5.1,4.1,2.1,4.1))
plot(bias$size,abs(bias$gamma2),type="o",pch=16,ylim=c(0,0.07),xlab="Sample Size",
     ylab="Absolute Bias",col="blue")
axis(2, pretty(c(0,0.07)), col="blue")
par(new=TRUE)
plot(mse$size, mse$gamma2,type="o",pch=2,lty=2,xlab=" ", ylab=" ", ylim=c(0,0.7), col="red",axes=F)
axis(4, pretty(c(0,0.7)), col="red")
mtext("MSE",side=4,line=1,padj=2)
legend(420, 0.7, legend=c("Absolute Bias", "MSE"), pch=c(16,2),col=c("blue", "red"),bty="n",cex=1)
text(110,0.68, "C")
##gamma3
par(mar=c(5.1,4.1,2.1,4.1))
plot(bias$size,abs(bias$gamma3),type="o",pch=16,ylim=c(0,0.07),xlab="Sample Size",
     ylab="Absolute Bias",col="blue")
axis(2, pretty(c(0,0.07)), col="blue")
par(new=TRUE)
plot(mse$size, mse$gamma3,type="o",pch=2,lty=2,xlab=" ", ylab=" ", ylim=c(0,0.7), col="red",axes=F)
axis(4, pretty(c(0,0.7)), col="red")
mtext("MSE",side=4,line=1,padj=2)
legend(420, 0.7, legend=c("Absolute Bias", "MSE"), pch=c(16,2),col=c("blue", "red"),bty="n",cex=1)
text(110,0.68, "D")
```

ABVD clinical trial data and data cleaning

```
abvd<-read.table("D:/Practicum project/data/JointModel with center.csv",sep=",", skip=1)
colnames(abvd) <- c("arm", "sex", "age", "centre","resp", "risk", "time", "event")
miss<-abvd[is.na(abvd$resp),]
miss$event<-as.factor(miss$event)
summary(miss)
###delete all the observation with missing resp value
abvd<-abvd[!is.na(abvd$resp),]
##remove all the data with survival time less than 6 month
abvd<-subset(abvd,time>=6)
###convert categorical variable to numeric variable
########note: arm A=abvd + radiation; arm B=abvd alone###############
arm<-ifelse(abvd$arm=="A", 1,0)          ##abvd alone=0; abvd+radiation=1
resp<-ifelse(abvd$resp=="YES", 1,0)      ##remission=1, no remission=0
event<-as.numeric(abvd$event)
time<-abvd$time-6
centre<-as.factor(as.numeric(abvd$centre))        ##there are 29 centres in total
```

```
risk<-ifelse(abvd$risk=="High", 1,0)
sex<-ifelse(abvd$sex=="M", 1,0)
###rearrange the dataset in the descending order of survival time
dat = data.frame(time, event, resp,arm,risk, sex, centre)
st = sort(dat$time, decr = T, index = T)
idx = st$ix
dat = dat[idx, ]
time<-dat$time; event<-dat$event; resp<-dat$resp; arm<-dat$arm
risk<-dat$risk; sex<-dat$sex; centre<-dat$centre
```

###Test PH assumption using Schoenfeld residuals

```
coxmodel<-coxph(Surv(time, event)~resp+arm+risk+sex+resp*arm)
phtest<-cox.zph(coxmodel, transform="km", global=TRUE)
par(mfrow=c(3,2),mar=c(4,4,2,2))
plot(phtest)
```