Two-Step and Likelihood Methods for Joint Models

by

Qian Ye

B.Sc., Zhejiang University, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2012

Abstract

Survival data often arise in longitudinal studies, and the survival process and the longitudinal process may be related to each other. Thus, it is desirable to jointly model the survival process and the longitudinal process to avoid possible biased and inefficient inferences from separate inferences. We consider mixed effects models (LME, GLMM, and NLME models) for the longitudinal process, and Cox models and accelerated failure time (AFT) models for the survival process. The survival model and the longitudinal model are linked through shared parameters or unobserved variables. We consider joint likelihood method and two-step methods to make joint inference for the survival model and the longitudinal model. We have proposed linear approximation methods to joint models with GLMM and NLME submodels to reduce computation burden and use existing software. Simulation studies are conducted to evaluate the performances of the joint likelihood method and two-step methods. It is concluded that the joint likelihood method outperforms the two-step methods.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank my supervisor, Dr. Lang Wu, for his guidance during my study in my master program at Department of Statistics of the University of British Columbia. Also, I would like to thank my second reader, Dr. Paul Gustafson, for his valuable comments and suggestions on this thesis.

Further more, I would like to thank Dr. John Petkau for his valuable guidance on my consulting projects. I would like to thank all the faculty in Department of Statistics for providing such a nice academic environment. I would like to thank all my friends, Betty, Huili, Can, Wen, Ciuciu, Liuna and Yingying, for making my life so enjoyable.

Most importantly, I would like to thank my parents, for their love, support and encouragement.

To my parents.

Chapter 1

Introduction

1.1 Longitudinal Studies

1.1.1 Longitudinal Data

In a longitudinal study, subjects are observed over a period of time and the same variables of each individual are repeatedly measured over time. In other words, the longitudinal data are gathered at multiple measurement times for every subject in a longitudinal study. For example, in a HIV study, viral load of each patient is repeatedly measured at several measurement times during the entire study. Usually, the number of measurement times may vary across individuals. In longitudinal studies, we can study changes of variables over time. A key feature of longitudinal data is that the repeated measurements within each individual may be correlated since they share the same characteristics of each person. For instance, if a patient's viral load is above average at the beginning, his/her viral loads may stay above average for the following time points. We assume that observations from different individuals are independent.

In longitudinal data, there may be substantial variation among repeated measurements within the same individuals and across different individuals. The variation among within-individual repeated measurements reflects changes of variables over time, while the variation across individuals reflects the differences between individuals.

1.1.2 Analysis of Longitudinal Data

When analyzing longitudinal data, the correlation between repeated measurements within each individual should be incorporated in the analysis in order to avoid bias and loss of efficiency. Hence, statistical methods for cross-sectional studies, such as classical linear regression models, cannot be directly used to analyze longitudinal data, since these methods assume all the measurements are independent.

Regression models are commonly used in longitudinal data analysis to model the relationship between the longitudinal response and covariates. The covariates in the models can be either time-dependent (i.e., the variables vary over time), or time-independent (i.e., the variables do not change across time, such as gender). Three types of regression models are commonly used for analyzing longitudinal data. They are mixed effects models, marginal models or generalized estimating equations (GEE) models, and transitional models.

Mixed effects models incorporate the correlation between within-individual measurements by introducing random effects. The random effects are individual-specific. That is, the repeated measurements within each individual share the same random effects, which suggests that the repeated measurements for the same individual are correlated. The random effects in mixed effects models not only incorporate the within-individual correlation, but also incorporate the variation between different individuals. The variation of the random effects reflects individual deviations from the population averages. Commonly used mixed effects models include linear mixed effects (LME) models, generalized linear mixed models (GLMMs), and nonlinear mixed effects (NLME) models. A mixed effects model can estimate both the population parameters which are the same for all individuals and the individual-specific parameters. Hence, one can not only make inference at population level, but also conduct individual-specific inference.

In a marginal model, we first model the mean structure of the response and then we separately model the covariance structure of the response. The mean structure and the covariance structure are assumed based on the observed data without distributional assumptions. So marginal models are robust to distributional assumptions. To estimate the parameters in a marginal model, we can solve a set of estimating equations, called generalized estimating equations (GEEs). This is the reason why marginal models are also called GEE models. It is not necessary for the assumed working covariance structure to match the unknown true covariance structure to conduct a valid analysis. GEE estimators are consistent and asymptotically normal as long as the mean structure is correctly assumed, even if the covariance structure is mis-specified.

In a transitional model, we assume a Markov structure to incorporate the correlation between within-individual measurements: we model the response at a given time-point as a function of covariates and the previous responses. We may consider the previous response values as an additional “covariate”.

1.2 Survival Studies

1.2.1 Survival Data

In longitudinal studies, we are often also interested in the time to an event, such as time to death or time to the response to a new drug. We call these types of data “event-time data” or survival data. Analysis of event-time data or survival data is called survival analysis. In the analysis of survival data, we are often interested in examining the relationship between the time to an event and some covariates of interest. For example, in medical studies, doctors may want to find the relationship between the time to the response to a new drug and demographic characteristics such as age and gender, e.g., they may be interested in checking if younger patients are more likely to have shorter response times after they take a new drug.

Some special features distinguish survival data from other types of data. The first one is that survival data are often censored, because the event of interest may not be observed for all subjects during the study period (e.g., loss of follow-up or early termination of the study). The second feature is that the distribution of survival data is often asymmetric and skewed to the right. Therefore, it is not reasonable to assume a normal distribution for survival data, unlike what we usually do for other types of continuous data.

1.2.2 Analysis of Survival Data

When analyzing survival data, we must consider the special features of survival data, so special statistical methods are needed. In survival analysis, nonparametric and semi-parametric methods are commonly used because these methods do not make assumption for the distribution of the survival data. Meanwhile, parametric models are also useful and may have more power and more efficiency than nonparametric or semi-parametric methods as long as the parametric distributional assumptions are reasonable for the survival data.

In survival analysis, the survival function, which is the probability that an individual will survive beyond a specified time, plays a similar role as the cumulative distribution function in other types of data. Non-parametric methods allow one to estimate the survival function without distributional assumption for the data. A well-known non-parametric estimator of the survival function is called the Kaplan-Meier estimator.

To assess the relationship between the time to an event and important covariates, regression models can be used. In survival regression analysis, we often model the hazard function of the event instead of the mean of the survival times as in a classical regression model. Often, we make no distributional assumption to the survival data and leave the hazard function unspecified, which leads to semi-parametric survival regression models. The most popular semi-parametric survival model is called Cox proportional hazards model. If we assume a parametric distribution to the survival data, we obtain a parametric survival regression model. Parametric models may be preferred if the distributional assumption is reasonable. The Weibull distribution plays a important role in survival analysis. For example, by assuming a Weibull distribution to the survival data, a Cox proportional hazards model becomes a Weibull proportional hazards model. Another class of popular survival regression models is the accelerated failure time (AFT) models. An AFT model may be interpreted as the speed of disease progression and this interpretation is very attractive in practice.

1.3 Joint Modeling Longitudinal Data and Survival Data

Survival data often arise in longitudinal studies. During a study, one may be interested in both longitudinal data and survival data. For example, in HIV studies, we not only repeatly measure the viral load over time for each subject, but we may also be interested in the time to an event of interest, such as time to dropout, or time to a viral load rebound.

Two situations arise frequently in practice:

- In the first situation, the longitudinal model is of primary interest, and a survival model may be secondary (e.g. it is used to model the time to dropout to avoid biased inference for longitudinal model).
- In the second situation, we are primarily interested in the analysis of survival data, with time-dependent covariates missing at failure times or with measurement errors. In this case, the survival model is of main interest and the longitudinal model is used to address the missing covariates or covariates measurement errors.

In the above two situations, one may need to model both the longitudinal process and the survival process simultaneously and make full use of the information provided by both processes. So jointly modeling longitudinal data and survival data is needed.

When modeling both the longitudinal data and survival data, the longitudinal model and the survival model are often assumed to be linked through shared parameters or shared unobserved variables. For example, in the situation where we mainly focus on longitudinal analysis with non-random dropouts, the survival model for the time to dropout may share the same random effects with the longitudinal model, since these random effects reflect the individual-specific characteristic in the longitudinal process. As another example, in the situation where we are primarily interested in the survival model with measurement errors in time-dependent covariates, the unobserved "true values" of the time-dependent covariate in the survival model are the responses of the longitudinal model, so the longitudinal model and the survival model share the same unobserved variable.

This thesis mainly focuses on joint modeling longitudinal process and survival process in the second situation, where the survival model is of primary interest and the longitudinal model is secondary.

There are several methods for joint modeling longitudinal data and survival data. These methods are reviewed in the following sections.

1.3.1 The Naive Two-step Method

For joint analysis of two models sharing same parameters or same unobserved variables, a commonly used approach is the naive two-step method:

- The first step is to fit one model (usually the secondary one) to the observed data separately, and estimate the shared parameters or shared variables, ignoring the other model.
- The second step is to substitute the shared parameters or variables in the other model by their estimates from the first step, and then make inference in the other model as if the estimated shared variables were observed data.

Although the naive two-step method is simple and straightforward, Ye, Lin, and Taylor (2008) and Albert and Shih (2009) pointed out that the naive two-step method may lead to problems in two ways:

1. the longitudinal and survival processes are associated with each other, but the naive two-step method models each process separately, ignoring the other one, so estimation based on the naive two-step method may be biased. This bias may depend on the strength of the association between the longitudinal process and the survival process;
2. Another problem is that standard errors of the parameter estimates in the primary model may be under-estimated because the uncertainty of estimation in the first step is not incorporated into the second step.

1.3.2 The Modified Two-step Method

In order to incorporate the estimation uncertainty in the first step, we can use a bootstrap method to modify the naive two-step method. This method is called the modified two-step method. The bootstrap step can be described as follows:

Step 1 generate longitudinal values from the fitted longitudinal model and generate survival times from the fitted survival model, with the unknown parameters substituted by their estimates;

Step 2 use the naive two-step method to fit the generated data from step 1 and obtain the new estimates for parameters of interest;

Step 3 repeat steps 1 and 2 many times, and then use the sample standard deviations of all the new estimates as the modified standard errors for these estimates.

This modified two-step method provides more reliable standard errors than the naive two-step method since it adjusts the standard errors of the estimates by incorporating the estimation uncertainty in the first step of the naive two-step method. But the modified two-step method still may not completely remove biases in the naive two-step method.

1.3.3 The Joint Likelihood or Joint Model Method

In order to avoid potential bias and incorporate the estimation uncertainty when jointly modeling longitudinal and survival processes, we can make statistical inference based on the joint likelihood of all the longitudinal and survival data. Maximum likelihood estimates (MLEs) of all parameters in the longitudinal model and in the survival model can be obtained simultaneously by maximizing the joint likelihood. Inference based on joint likelihood produces less biased estimates and more reliable standard errors. The MLEs have appealing properties, namely consistency, asymptotically efficient and asymptotical normality under the usual regularity conditions.

However, the computation associated with the joint likelihood inference can be quite challenging, since the joint likelihood for a longitudinal model and a survival model is often highly complicated and typically involves a high-dimensional and intractable integral due to the unobservable random effects, censoring, and the semi-parametric survival model. Monte-Carlo methods or numerical integration methods can be used for joint likelihood inference. They are computationally intensive and may arise convergence problems. Laplace approximations or Taylor approximations can be used for approximate inference. These approximate methods are computationally more efficient.

1.4 A Motivating Example - A HIV Study

This section presents a example of longitudinal and survival data which motivates our research. This example illustrates some typical features of longitudinal and survival data and also motivates the joint modeling methods.

Our example comes from a HIV study, which investigates changes in 46 patients' immunologic markers (such as CD4 and CD8 cell counts) and viral load over time after an anti-HIV treatment, as well as some other time-dependent variables. One of the main objects of this study is to examine the relationship between time-dependent covariates (eg. viral load or CD4 cell counts) and a survival response, such as the time to dropouts or time to viral load rebound. For example, we may be interested in checking whether patients with higher initial viral loads have earlier dropouts.

We consider the data within the first 60 days after the anti-HIV treatment. The viral load (in log10 scale) and the CD4 cell counts are measured repeatedly over time after the treatment. The time to dropout is recorded for each patient involved in the study. The time (both the measurement time and the survival time) is re-scaled from 0 to 1, for convenience. A description of the variables of interest in this HIV dataset is shown in Table 1.1. Table 1.2 summarize the data for viral load and CD4 measured at four selected time points.

Table 1.1: Description of variables in the HIV dataset within the first 60 days

Variable Name	Definition	Characteristics/Summaries
lgcopy	Viral load in log10 scale	Mean: 3.65, SD: 0.97
CD4	CD4 cell counts	Mean: 251.75, SD: 92.72
day	Measurement time	re-scaled from 0 to 1. The number of measurement time per patients varies from 2 to 6
timer	Survival time	re-scaled from 0 to 1
Snsign	Censoring indicator	'1' means an event is observed '0' means the event time is censored Censoring rate: 8.7%

Table 1.2: Summary statistics for viral load and CD4 at four selected measurement times

Variable	Day 2		Day 7		Day 14		Day 28	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
lgcopy	5.00	0.59	4.06	0.81	3.23	0.64	3.02	0.61
CD4	203.33	74.08	231.31	89.05	274.15	108.72	284.98	89.48

Viral load measured over time are longitudinal data. Figures 1.1 and 1.2 show viral load trajectories of all patients and five randomly selected patients respectively. We see that the viral load trajectories vary substantially across patients. That is, there exists a large variation in the data between different patients (i.e. a large between-individual variation). The repeated measurements for each patient, or the within-individual measurements, may be correlated since they share the same characteristics of each individual. These figures also illustrate some other features of the longitudinal data, e.g., the number of repeated measurements and measurement times may vary across individuals, missing data and dropouts are common.

Figure 1.1: Viral load trajectories of all patients

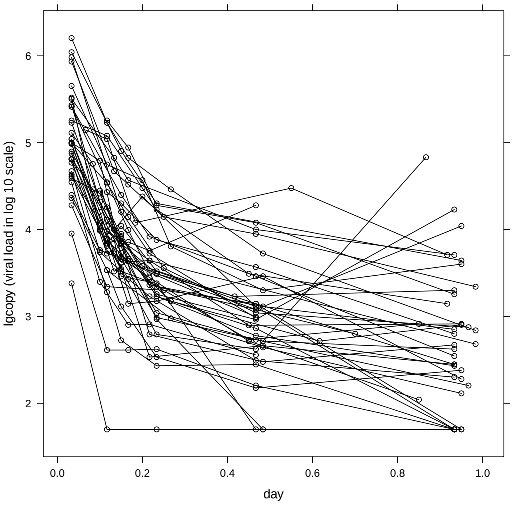
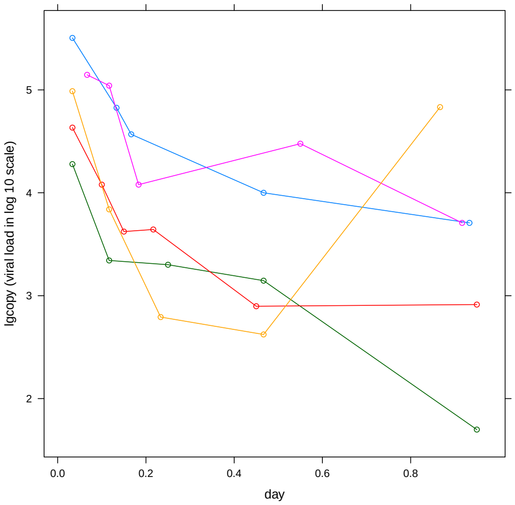


Figure 1.2: Viral load trajectories of five patients



We are also interested in the time to dropout. This type of data is event-time data or survival data. Survival data are often censored. Figure 1.3 shows the survival data for five patients in the study, with their censoring status. From Figure 1.3 we can see that three event times are censored and the other two are observed. These three censored patients in the figure may dropout during the study or they may be lost of follow-up. Figure 1.4 shows the estimated density curve of the event time. We see that the distribution of the survival data is asymmetric and skewed to right. So a normal distribution cannot be assumed to the event time in this study.

The HIV study aims in assessing the relationship between the time-dependent covariates and a survival response. The survival model is of primary interest. However, measurement errors in time-dependent covariates are common in

Figure 1.3: Survival data with censoring status

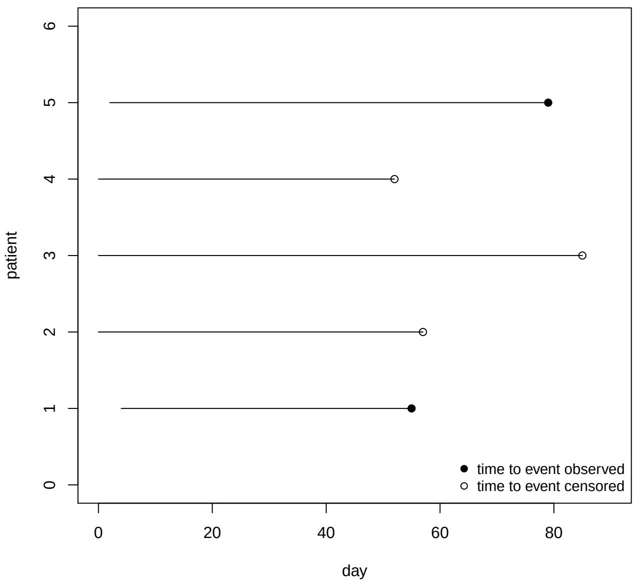
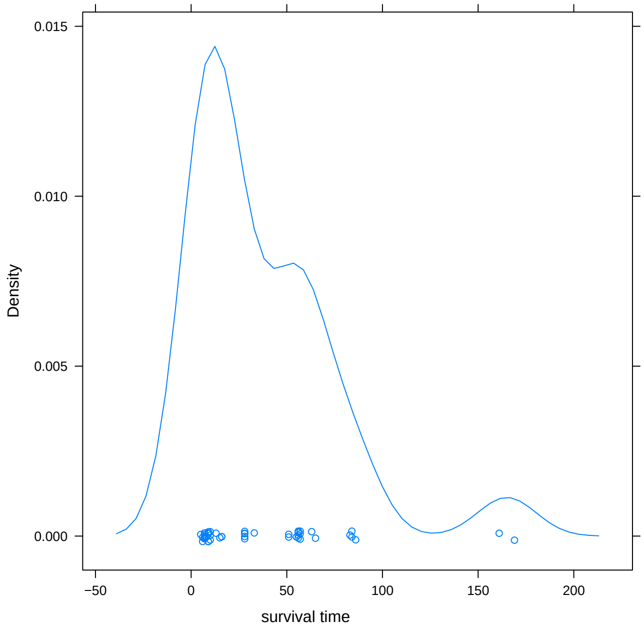


Figure 1.4: Estimated density curve of survival times



HIV studies. Thus, we also need a longitudinal model to address covariates measurement errors to avoid potential bias in the survival model, and it is secondary. In such a case, joint inference for a longitudinal model and a survival model is needed.

1.5 Literature Review

Methods jointly modeling longitudinal data and survival data have been studied in the literature. Tsiatis and Davidian (2004) reviewed the development of joint models, and described and contrasted some of earlier proposals for implementation and inference. Tseng, Hsieh, and Wang (2005) explored the joint modeling method under the accelerated failure time assumption when covariates are assumed to follow a linear mixed effects model with measurement errors. Their joint modeling method is based on maximizing the joint likelihood with random effects treated as missing data, with a Monte Carlo EM algorithm used to estimate all the unknown parameters. Wu, Hu, and Wu (2008) considered a nonlinear mixed effects model for the longitudinal process and a Cox proportional hazards model for the time-to-event process. The inference on the two models is also based on joint likelihood and allows for nonignorable data missing. Ye, Lin, and Taylor (2008) developed a two-stage semi-parametric regression calibration method to jointly model longitudinal and survival data using a semi-parametric longitudinal model and a proportional hazards model. Song and Wang (2008) proposed a local corrected score estimator and a local conditional score estimator to deal with covariate measurement error for joint modeling of a random effects model and a time-varying coefficient proportional hazards model.

The longitudinal models and the survival models are linked in many different ways. In Tseng, Hsieh, and Wang (2005), the two models shared unobserved error-free variables. Wu, Hu, and Wu (2008) considered the situation where the individual characteristics associated with the repeated measures are possible covariates of the time to an event. The Cox model assumed for survival data in Ye, Lin, and Taylor (2008) included both the current measure and the rate of change of the underlying longitudinal trajectories as covariates. In Song and Wang (2008), some covariates in the survival model were functions of the unobserved random effects in the longitudinal model.

1.6 Outline

This thesis discusses different joint inference methods for a longitudinal model and a survival model which are linked through shared unobserved variables or shared parameters. We mainly focus on the naive two-step method, the modified two-step method, and the joint likelihood method described in Section 1.3. For the longitudinal models, we mainly consider mixed effects models, including linear mixed effects (LME) models, generalized linear mixed models (GLMMs) and nonlinear mixed effects (NLME) models. For survival regression models, we consider Cox proportional hazards model and accelerated failure time (AFT) models.

Chapter 2 reviews mixed effects models for longitudinal data, including LME, GLMM and NLME models, and models for survival data, including Cox proportional hazards models and accelerated failure time models. In Chapter 3, we describe different joint inference methods for a linear or nonlinear mixed effects model and a survival model. Then, real data analysis and simulation studies are conducted to compare different joint inference methods. In Chapter 4, we consider joint inference for a generalized linear mixed effects model and a survival model. We mainly focus on the binomial and Poisson family in the class of generalized linear models. Conclusions and future work needed are discussed in Chapter 5.

Chapter 2

Review of Longitudinal Models and Survival Models

2.1 Models for Longitudinal Analysis

In this section, we briefly review some commonly used models for longitudinal data analysis. Let's first define notation that will be used in longitudinal models. Suppose there are N individuals in a longitudinal study, and n_i is the number of repeated measurements for individual i. Let y_{ij} be a response variable and $\boldsymbol{x}_{ij}$ be a $p \times 1$ vector of p covariates for individual i at time t_{ij}, $i = 1, \cdots, N$, $j = 1, \cdots, n_i$. Let $n_i \times 1$ vector $\boldsymbol{Y}_i = (y_{i1}, y_{i2}, \cdots, y_{in_i})^T$ be the n_i repeated response values for individual (or cluster) i, and let $\boldsymbol{X}_i = \{\boldsymbol{1}, (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \cdots, \boldsymbol{x}_{in_i})^T\}$ is a $n_i \times (p+1)$ design matrix containing covariates of individual i.

2.1.1 Linear Mixed Effects Models

In Section 1.1 of Chapter 1, we briefly review some key features of mixed effects models which are commonly used in the analysis of longitudinal data. Mixed effects models incorporate the correlation among within-individual measurements and the variation across different individuals by introducing random effects. A mixed effects model may be obtained by extending the corresponding standard regression model for cross-sectional data by adding random effects to appropriate parameters. Linear regression models are widely used because of their simplicity. Hence, we can extend a linear regression model to a linear mixed effects (LME) model by introducing random effects.

A general LME model can be written as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \cdots, N,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p)$ is a vector of fixed effects, $\boldsymbol{b}_i = (b_{0i}, b_{1i}, b_{2i}, \cdots, b_{qi})$ is a vector of random effects, $\boldsymbol{Z}_i$ is a $n_i \times (q+1)$ design matrix (it is often a submatrix of $\boldsymbol{X}_i$), and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{in_i})$ is a vector of random errors within individual i. We assume that

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \Sigma) \quad \boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, D_i),$$

and $\boldsymbol{b}_i$ and $\boldsymbol{\varepsilon}_i$ are independent, where Σ is a $(q+1) \times (q+1)$ covariance matrix of the random effects $\boldsymbol{b}_i$ and matrix D_i is a $n_i \times n_i$ covariance matrix of random errors $\boldsymbol{\varepsilon}_i$. The diagonal elements of Σ are the variances of $\boldsymbol{b}_i$ and they measure the variability of the longitudinal trajectories between individuals. The diagonal elements of D_i are the variances of $\boldsymbol{\varepsilon}_i$ and they measure the variability of the within-individual measurements. So, the LME model incorporates two sources of variability: the within-individual variation and the between-individual variation.

The marginal distribution of Y_i is

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\Sigma\boldsymbol{Z}_i^T + D_i).$$

The marginal mean E($\boldsymbol{Y}_i$)=$\boldsymbol{X}_i\boldsymbol{\beta}$ represents a typical longitudinal trajectory in the population. So the population-level inference is based on inference on parameters $\boldsymbol{\beta}$, while the individual-specific inference should be conducted by conditioning on the random effects $\boldsymbol{b}_i$.

Let $\boldsymbol{\theta}$ denote all parameters in the LME model. Then, the likelihood function for the observed responses $\boldsymbol{Y} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_N\}$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{Y}) &= \prod_{i=1}^{N} f(\boldsymbol{Y}_i|\boldsymbol{\theta}) \\ &= \prod_{i=1}^{N} \int f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{b}_i, \boldsymbol{\beta}, D_i) f(\boldsymbol{b}_i|\Sigma) d\boldsymbol{b}_i. \end{aligned}$$

Statistical inference for a LME model is typically based on the maximum likelihood method. The computation would be intensive when the random effects $\boldsymbol{b}_i$ have a high dimension. Methods like Monte Carlo methods and approximate methods (Lindstrom and Bates, 1990) could be considered.

2.1.2 Nonlinear Mixed Effects Models

In the previous section we describe linear mixed effects models which are widely used in longitudinal studies. However, linear models have their disadvantages: they only empirically describe the observed data but do not provide understanding of the true relationship between the covariates and the response (i.e. the underlying data-generation mechanism). Therefore, in many longitudinal studies, nonlinear models which describe the underlying mechanism of data generation are better choices when such non-linear models are available.

Compared to linear models, there are many advantages of non-linear models. First, since nonlinear models attempt to describe the data-generation mechanisms, so parameters in nonlinear models often have natural physical interpretations. Secondly, nonlinear models can provide more reliable predictions than linear models when the covariates are outside of the range of the observed data. Thirdly, non-linear models may need fewer parameters to provide reasonable fit of the observed data than linear models.

In longitudinal studies, we can obtain a nonlinear mixed effects (NLME) model by extending the corresponding nonlinear regression model through adding random effects terms to appropriate parameters.

A general NLME model can be written as

$$y_{ij} = g(t_{ij}, \boldsymbol{\beta}_i) + \varepsilon_{ij}, \quad i = 1, 2, \cdots, N \quad j = 1, 2, \cdots, n_i, \tag{2.1}$$

$$\boldsymbol{\beta_i} = h(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) \tag{2.2}$$

$$\boldsymbol{b}_i \sim N(\mathbf{0}, \Sigma), \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, D_i), \tag{2.3}$$

where $g(\cdot)$ is a known nonlinear function which specifies the mean structure for individual i, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{in_i})$ is the vector of random errors within the i^{th} individual, $\boldsymbol{\beta}_i = (\beta_{1i}, \beta_{2i}, \cdots, \beta_{pi})$ is a vector of p individual-specific parameters, $h(\cdot)$ is a p-dimensional known function, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$ is a vector of fixed effects, $\boldsymbol{b}_i = (b_{1i}, b_{21}, \cdots, b_{qi})^T$ is a vector of random effects, Σ is a covariance matrix for the random effects $\boldsymbol{b}_i$, and D_i is a covariance matrix for the random errors $\boldsymbol{\varepsilon}_i$. We assume $\boldsymbol{b}_i$ and $\boldsymbol{\varepsilon}_i$ are independent.

Let $\boldsymbol{\theta}$ denote all the parameters in the NLME model. The likelihood function of the observed responses $\boldsymbol{Y} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_N\}$ is

$$L(\boldsymbol{\theta}|\boldsymbol{Y}) = \prod_{i=1}^{N} \int f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{b}_i, \boldsymbol{\beta}, D_i) f(\boldsymbol{b}_i|\Sigma) d\boldsymbol{b}_i. \tag{2.4}$$

Since the nonlinear form and possibly high dimensions of random effects in NLME models, the likelihood function is complex and typically does not have a closed from expression, which leads to many computational problems. Three most commonly used estimation methods for NLME models are numerical or Monte Carlo methods, EM algorithms, and approximate methods. In this thesis, we mainly focus on the approximate method to the NLME model, which leads to a "working" LME model based on a Taylor series expansion about the current estimates of parameters and random effects and then update parameter estimates from this LME model, and iterate the algorithm until converge (Lindstrom and Bates, 1990).

2.1.3 Generalized Linear Mixed Effects Models

In both linear and nonlinear models, the response is assumed to be normally distributed. However, in practice, many types of responses do not necessarily follow normal distributions such as binary responses which only have two possible values or levels. In such cases, linear models and nonlinear models are not appropriate to model the data. On the other hand, generalized linear models greatly extend classical linear models and thus they can be used for the responses which follow distributions in the exponential family. The exponential family is an important class of probability distributions sharing a certain form and it includes many of the most commonly used distributions, such as normal, exponential, binomial, and Poisson distributions.

Like LME and NLME models, generalized linear mixed models (GLMMs) can be obtained by extending generalized linear models for cross-sectional data by adding random effects to appropriate parameters. Similarly, the random effects in GLME models are used to incorporate the correlation between measurements within each individual and the variation between different individuals. In a GLMM, we assume that the repeated measurements within individual i, y_{ij} $(j = 1, 2, \cdots, n_i)$ independently follow a distribution in the exponential family with a mean of μ_{ij}, conditioning on the random effects.

A general GLME model can be presented as

$$\begin{aligned} g(\mu_{ij}) &= x_{ij}^T\boldsymbol{\beta} + z_{ij}^T\boldsymbol{b}_i, \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, n_i, \\ \boldsymbol{b}_i &\sim N(\boldsymbol{0}, \Sigma) \end{aligned}$$

where $g(\cdot)$ is a monotone and differentiable function called the link function (describing how the mean response is related to a linear combination of predictors), $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T$ is a vector of fixed effects, $\boldsymbol{b}_i = (b_{0i}, b_{1i}, \cdots, b_{qi})^T (q \leq p)$ is a vector of random effects, z_{ij} is a vector containing covariates(it is often a subvector of x_{ij}), and Σ is the covariance matrix for the random effects $\boldsymbol{b}_i$.

The two most widely used generalized linear models for non-normal data are Poisson regression models and logistic regression models. If the response is a count, it may be reasonable to assume a Poisson distribution for the response. The commonly used link function for count response is log function

$$g(\mu) = \log(\mu).$$

The corresponding models are called Poisson regression models. If the response in longitudinal studies is a binary variable, it is reasonable to assume a binomial or Bernoulli distribution for the response. The most popular link function for binary response is the logit function

$$g(\mu) = \text{logit}(\mu) \equiv \log\left(\frac{\mu}{1-\mu}\right).$$

So the corresponding regression models are called logistic regression model.

Let θ denote all the parameters in a GLME model. The likelihood function of the observed data is

$$L(\boldsymbol{\theta}|\boldsymbol{Y}) = \prod_{i=1}^{N} \int f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{b}_i, \boldsymbol{\beta}) f(\boldsymbol{b}_i|\Sigma) d\boldsymbol{b}_i. \tag{2.5}$$

Similar to NLME models, the likelihood function of GLMMs typically involves an intractable integral and hence does not have a closed form expression because of the non-linearity in random effects. The most commonly used inference methods for GLMMs are the same as NLME models, including numerical or Monte Carlo methods, EM algorithms, and approximate methods. This thesis mainly focuses on the linear approximation to GLMMs which uses Taylor expansions to linearize the model and then solve the "working" LME models iteratively until converge. However, the responses in GLMMs are often discrete such as binary and count while the responses in NLME model are continuous. Therefore, the linear approximation for GLMMs may performance worse than that for NLME models. A way to improve the performance is to use a higher order Taylor expansion.

2.2 Models for Survival Analysis

In this section, we briefly review some survival models. Before describing survival models, let's first define notation that will be used. Let T be the time to an event of interest. Suppose there are N individuals in the study and t_i $(i = 1, 2, \cdots, N)$ is the observed survival time for individual i. In practice, some of the survival times t_i's may be censored. In this thesis, we only focus on right censoring. We define a censoring indicator as follows

$$\delta_i = \begin{cases} 1, & \text{if the event time is observed for individual } i; \\ 0, & \text{if the event time is right censored for individual } i. \end{cases}$$

So the observed data for individual i can be presented as (t_i, δ_i). We denote $\boldsymbol{x}_i$ as a vector of p covariates for individual i.

2.2.1 Cox Proportional Hazards Models

To assess the relationship between the time to an event of interest and important covariates, regression analysis can be used. In survival regression analysis, we often model the hazard function of the event instead of the mean of the survival times as in a classical regression model. The hazard function can be defined as

$$h(t) = lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}, t > 0$$

Since the distribution of the survival time may be complicated and hence the hazard function may be complicated, sometimes we make no distributional assumption to the survival data and leave the hazard function unspecified in the model. Then, we may use the covariates to predict the hazard function through the linear parametric predictor, which leads to a semi-parametric survival regression model. The most popular semi-parametric model is the following Cox proportional hazards model (Cox 1972)

$$h_i(t) = h_0(t)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}), \quad i = 1, 2, \cdot, N,$$

where $h_0(t)$ is an unspecified baseline hazard function, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$ is a vector of regression parameters. Although we make no parametric assumption to the survival data, we need to assume that the hazards ratio $h_i(t)/h_0(t)$ does not change over time in the Cox proportional hazards model.

Since the baseline hazard function $h_0(t)$ is unspecified, inference for $\boldsymbol{\beta}$ can be based on the partial likelihood function (Cox, 1972)

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N}\left\{\frac{\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}{\sum_{j=1}^{N} I(t_j \geq t_i)\exp(\boldsymbol{x}_j\boldsymbol{\beta})}\right\}^{\delta_i},$$

where the survival times are assumed to be independent without ties.

The Cox proportional hazards model is widely used in survival analysis since it is robust against the distributional assumptions made to the survival data. However, parametric models sometimes may be preferred if the parametric distribution assumed to the survival data is reasonable, because statistical inference based on a parametric model is more efficient than a semi-parametric model if the parametric assumption holds. If we assume a parametric distribution to the survival data and specify the baseline hazard function in a Cox proportional hazards model, it turns to a parametric survival regression model. The Weibull distribution is widely used in survival analysis. People often assume a Weibull distribution to survival data, similar to the normal distribution in linear regression models. In the following, we describe a Weibull proportional hazrds model.

Weibull Proportional Hazards Models

Let's denote the Weibull distribution with scale parameter λ and shape parameter γ as $W(\lambda, \gamma)$. The hazard function of $W(\lambda, \gamma)$ is given by

$$h(t) = \lambda\gamma t^{\gamma-1}.$$

By specifying the hazard function of a Weibull distribution as the baseline hazard function, a Cox proportional hazards model becomes a Weibull proportional hazards model. The Weibull proportional hazards model can be written as

$$h_i(t) = h_0(t)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}), \quad i = 1, 2, \cdots, N,$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$ is a vector of regression parameters, and $h_0(t)$ is the hazard function of a Weibull distribution $W(\lambda, \gamma)$, which is $h_0(t) = \lambda\gamma t^{\gamma-1}$. So

$$h_i(t) = \lambda\gamma t^{\gamma-1} \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) = \left[\lambda \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})\right] \gamma t^{\gamma-1}.$$

We can see that $h_i(t)$ is the hazard function of the Weibull distribution $W(\lambda \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}), \gamma)$, which implies the effect of covariates $\boldsymbol{x}_i$ is to change the scale parameter λ into $\lambda \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$ with the shape parameter γ unchanged.

The likelihood function of the Weibull proportional hazards model is

$$L(\boldsymbol{\beta}, \lambda, \gamma) = \prod_{i=1}^{N} (h_i(t_i))^{\delta_i} S_i(t_i),$$

where $h_i(t)$ and $S_i(t)$ are the hazard function and survival function of the Weibull distribution $W(\lambda \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}), \gamma)$ respectively. Statistical inference on regression parameters $\boldsymbol{\beta}$ can be based on standard likelihood method.

2.2.2 Accelerated Failure Time(AFT) Models

In practice, the assumption that the hazards ratio $h_i(t)/h_0(t)$ is constant over time does not necessarily hold, so the Cox proportional hazards models may not be applicable. In such cases, a popular class of survival regression models called accelerated failure time (AFT) models, which does not need the proportional hazards assumption, can be a good choice.

A widely used log-linear representation of AFT model can be written as

$$\log(T_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} + \sigma\epsilon_i, \quad i = 1, 2, \cdots, N,$$

where σ is a scale parameter and ϵ_i's are random errors.

If we make different parametric distributional assumptions to random errors ϵ_i's, we have different parametric AFT models. For example, if ϵ_i' follows $N(0, 1)$, then the survival time T_i follows a log-normal distribution; if ϵ_i' follows a logistic distribution, the survival time T_i follows a log-logistic distribution; if ϵ_i follows the Gumbel distribution, the survival time T_i follows a Weibull distribution. These three types of AFT models are the most commonly used in practice. Statistical inference for a parametric AFT model can be based on the standard likelihood method.

Chapter 3

Joint Inference for a Linear or Nonlinear Mixed Effects Model and a Survival Model

3.1 Introduction

In this chapter, we discuss joint inference methods for a linear or nonlinear mixed effects model and a survival model. We focus on the situation where the survival model with an error-prone time-dependent covariates is of primary interest and the longitudinal model is secondary. The linear or nonlinear mixed effects model model is assumed for the time-dependent covariates in the survival model to address covariates measurement errors or missing covariates. We assume the data is missing at random in this thesis. We first consider a linear mixed effects (LME) model to describe the covariate process and discuss the joint inference methods for a LME model and a survival model. Then, we consider a nonlinear mixed effects (NLME) model to describe the covariate process and discuss methods to make joint inference on a NLME model and a survival model.

The observed covariate value for individual i at time t_{ij} is denoted as $z_{ij} = z_i(t_{ij})$ $(i = 1, 2, \cdots, N, j = 1, 2, \cdots, n_i)$ and the corresponding unobserved true value of covariate is denoted as z^*_{ij}. Let $\boldsymbol{x}_i$ be a vector of other covariates without measurement errors.

3.2 Joint Inference Methods for a LME Model and a Cox Model

In the following, we describe the survival model and longitudinal model which we want to make joint inference on.

The Survival Model

Let T be the time to an event of interest. Suppose that there are N individuals in the study and t_i $(i = 1, 2, \cdots, N)$ is the observed or censored survival time for individual i. We allow some of the survival times t_i's to be right censored. For modeling survival data, we consider the following Cox model with time-dependent and time-independent covariates:

$$h_i(t) = h_0(t) \exp(z_i^*(t)\beta_1 + \boldsymbol{x}_i^T \boldsymbol{\beta}_2), \quad i = 1, 2, \cdots, N, \tag{3.1}$$

where $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2)$ are regression parameters. This model links the hazard function to the unobserved true value of covariate $z_i^*(t)$ rather than the observed covariate $z_i(t)$ which is measured with measurement error.

As discussed in Section 2.2.1 of Chapter 2, statistical inference for Cox model can be based on the partial likelihood method. To make the inference, we must know the value of the time-dependent covariate $z_i(t)$ at every event time t_i for all individuals. However, in practice, the covariate may not be measured at all event times, which leads to missing data in the time-dependent covariate in the survival model. Moreover, the true values of covariate z_i^*'s may be unobserved since there may exist measurement errors in the observed values of covariate $z_i(t)$'s. Therefore, we may consider a longitudinal model to model the covariate process in order to address both the measurement error and the missing data problems in the time-dependent covariate.

The Longitudinal Covariate Model

To address the measurement errors and missing data in the time-dependent covariate, we may assume that the covariate value changes smoothly over time, and we empirically model the covariate process. We consider the following linear mixed effects model, which is a classical measurement error model

$$\boldsymbol{z}_i = U_i \boldsymbol{\alpha} + V_i \boldsymbol{a}_i + \boldsymbol{\varepsilon}_i \equiv \boldsymbol{z}_i^* + \boldsymbol{\varepsilon}_i. \tag{3.2}$$

Thus, here we assume that the unobserved true values of covariate $\boldsymbol{z}_i^*$ are $\boldsymbol{z}_i^* = U_i\boldsymbol{\alpha} + V_i\boldsymbol{a_i}$, where U_i and V_i are design matrices (V_i is usually a submatrix of U_i), $\boldsymbol{\alpha}$ is a vector of fixed effect, $\boldsymbol{a}_i$ is a vector of random effects, and $\boldsymbol{\varepsilon}_i$ is a vector of measurement errors for individual i. We often assume that

$$\boldsymbol{a}_i \sim N(\mathbf{0}, \Sigma) \quad \boldsymbol{\varepsilon}_i \sim (\mathbf{0}, D_i),$$

and $\boldsymbol{a}_i$ and $\boldsymbol{\varepsilon}_i$ are independent.

For joint inference for a longitudinal model and a survival model, we have several approaches. A simple and widely used method is the naive two-step method, which uses one model to estimate the shared parameters or shared variables and then makes inference in the other model with the estimated shared parameters or variables as if they were observed data. Alternatively, we can use the modified two-step method, which uses a parametric bootstrap method to modify the naive two-step method in order to adjust the standard errors of the estimated parameters from the naive two-step method. Thirdly, another appealing approach is the joint likelihood or joint model method, which simultaneously obtains maximum likelihood estimates (MLEs) of all parameters by maximizing the joint likelihood. In the following, we discuss the three joint modeling methods in details.

3.2.1 The Naive Two-step Method

To joint analyzing two models sharing the same parameters or same unobserved variables, a simple and commonly used approach is the naive two-step method, which we reviewed in Section 1.3 of Chapter 1. The first step is to fit one model to the observed data separately and estimate the shared parameters or shared variables. Then, in the second step, we substitute the shared parameters or variables by their estimates from the first step and make inference in the other model with the estimated shared parameters or variables as if they were observed data.

Consider the survival model in (3.1) and the longitudinal model in (3.2). The population parameters $\boldsymbol{\beta}$ in the survival model are our main interest, and the linear mixed effects model is only used to address the measurement errors and missing data in the time-dependent covariate Z. In the first step, we estimate the true values of the covariate by fitting the linear mixed effects model (3.2) to the observed data $\{(\boldsymbol{z}_i, \boldsymbol{t_i}), \quad i = 1, 2, \cdots, N\}$, ignoring the survival model. We denote the predicted true value of the covariate at event time t_i by $\hat{z}_i^*$. Then, in the second step, we substitute the unobserved variable $z_i^*(t_i)$ by its estimate $\hat{z}_i^*$ and proceed the usual inference on the Cox mode with time-independent covariates as if the estimates were observed data. In other words, the survival model we fit in the second step is

$$h_i(t) = h_0(t)\exp(\hat{z}_i^*\beta_1 + \boldsymbol{x}_i^T\boldsymbol{\beta}_2),\ i = 1, 2, \cdots, N.$$

3.2.2 The Modified Two-step Method

The naive two-step method is simple, straightforward, and easy to understand. However, this approach may lead to two main problems. First, when conducting the two-step method, we model either process separately ignoring the other one. So bias may arise when the two processes influence each other. For example, the longitudinal covariate data may be truncated by the event. Another problem is that standard errors of the parameter estimates in the primary model may be under-estimated, because the uncertainty of the estimation in the first step is not considered into the second step. So we need a modified two-step method to adjust the under-estimated standard errors. An appealing approach is to use a bootstrap method to get more reliable standard errors.

Albert and Shih (2009) use a LME model to approximate the conditional distribution $f(\boldsymbol{z}_i|T_i)$ of the covariate Z given the event time and then generate covariate based on $f(\boldsymbol{z}_i|T_i)$. This method can remove much bias but may not completely eliminate all bias. They use simulations to simulate truncated covariate values and then treat the simulated data as observed data.

3.2.3 The Joint Likelihood or Joint Model Method

In order to avoid much bias in two-step methods when joint modeling longitudinal and survival processes, we consider an approach based on the joint likelihood of all the longitudinal data and survival data. Maximum likelihood estimates (MLEs) of all parameters in the longitudinal model and the survival model can be obtained simultaneously by maximizing the joint likelihood. Inference on the joint likelihood function produces less biased estimates and more reliable standard errors. The joint likelihood approach is quite general and can be extended to joint inference for more than two models which are linked.

Let $\boldsymbol{\theta}$ denote all parameters in the two models. The joint likelihood of the survival model (3.1) and the longitudinal model (3.2) based on all observed data $\{(\boldsymbol{z}_i, \boldsymbol{x}_i, t_i, \delta_i), \quad i = 1, 2, \cdots, N\}$ is given by

$$
\begin{aligned}
L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int & [h_0(t_i) exp(z_i^*(t_i)\beta_1 + \boldsymbol{x}_i^T \boldsymbol{\beta}_2)]^{\delta_i} \\
& \times \exp\left[- \int_0^{t_i} h_0(u) exp(z_i^*(u)\beta_1 + \boldsymbol{x}_i^T \boldsymbol{\beta}_2) du \right] \\
& \times f(\boldsymbol{z}_i | \boldsymbol{\alpha}, \boldsymbol{a}_i, D_i) f(\boldsymbol{a}_i | \Sigma) d\boldsymbol{a}_i. \qquad (3.3)
\end{aligned}
$$

We can see that the joint likelihood for a survival model and a longitudinal model is highly complicated, and involves high-dimensional and intractable integrals, due to the unobservable random effects. So the computation of joint likelihood inference can be quite challenging. To evaluate intractable integrals in the likelihood, we may consider numerical integration methods which approximate an integral by a weighted sum, with suitable points and weights. We mainly focus on the popular Gauss-Hermite quadrature method in this thesis. A R package "JM", which uses Gauss-Hermite method, is available for the computation of the joint likelihood (3.3). Evans and Swartz (2000) provided a detailed discussion of various approaches in numerical integration. In the following, we briefly review the procedure of the Gauss-Hermite quadrature method with a simple example.

The Gauss-Hermite Quadrature Method

Consider the following integral

$$I = \int g(x) f(x) dx,$$

where $g(x)$ is a continuous function and $f(x)$ is a normal density function. We illustrate the method using the $N(0, \frac{1}{2})$ distribution and let $f(x) = \exp(-x^2)$. The Gauss-Hermite quadrature method approximates the integral by

$$I = \int g(x) \exp(-x^2) dx \approx \sum_{i=1}^{k} w_i(x_i) g(x_i), \qquad (3.4)$$

where the node x_i is the i-th root of the Hermite polynomial $H_k(x)$ with degree of k. The Hermite polynomials $H_k(x)$'s are orthogonal polynomials.

The approximation in (3.4) can be arbitrarily accurate when the number of nodes k increases. When $g(x)$ is a polynomial of degree up to $2k-1$, the approximation is exact.

If $f(x)$ is the density function of a general normal distribution $N(\mu, \sigma^2)$, we may apply transformation $x = \mu + \sqrt{2}\sigma z$, and we then have

$$\begin{aligned} I = \int g(x)\exp(-x^2)dx &\approx \sum_{i=1}^{k} w_i^*(x_i) g(\mu + \sqrt{2}\sigma z_i) \\ &= \sum_{i=1}^{k} \frac{1}{\sqrt{\pi}} w_i(x_i) g(\mu + \sqrt{2}\sigma z_i). \end{aligned}$$

If $\boldsymbol{x} = (x_1, \cdots, x_h)^T$ is a h-dimensional vector, we have

$$I = \int_{R^h} g(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} \approx \sum_{i_1=1}^{k_1} w_{i_1}^{(1)} \cdots \sum_{i_h=1}^{k_h} w_{i_h}^{(h)} g\left(x_{i_1}^{(1)}, \cdots, x_{i_h}^{(h)}\right)$$

where $x_{i_j}^{(j)}$ is the i_j-th root of the Hermite polynomial with degree k_j and $w_{i_j}^{(j)}$ is the corresponding weight.

3.3 Data Analysis - A HIV Study

In the previous section, we have described three different methods for jointly analyzing a LME model and a Cox model. In this section, we use the joint likelihood method and the naive two-step method, together with a naive method, which ignores the measurement errors in covariate and fits the Cox model to the observed data directly, to analyze a real datasets from a HIV study.

Data Description and Objective

The dataset comes from the HIV study discussed in Section 1.4 of Chapter 1. It contains the data about changes in 46 patients' viral loads over time after an anti-HIV treatment, as well as some other time-dependent variables. In the mean time, this dataset also includes a survival response, i.e. the time to dropout. The objective of this data analysis is to examine the relationship between viral load trajectories and the time to dropout. More specifically, we are interested in checking whether patients with high viral loads are more likely to dropout. One thing we need to take into consideration is that there often exist substantial measurement errors in many time-dependent variables in HIV studies, such as viral load and CD4 count. Thus, measurement errors in viral load data should not be ignored when one conducts the analysis.

We consider the data in the first 60 days after the anti-HIV treatment. A summary table of the variables of interest (Table 1.1) and figures of viral load trajectories (Figure 1.1 and 1.2) can be found in Section 1.4 of Chapter 1, with a detailed discussion about data characteristics.

The Models

Since the measurement errors in viral load data cannot be ignored, we need to assume a model for the time-dependent viral load in order to address measurement errors. Since the repeated measurements of viral load for each patient are likely to be correlated, and there is a large variation between different patients, we consider a mixed effects model for the viral load (z_{ij}). Based on the trajectories shown in Figure 1.1, we use a quadratic linear mixed effects model to fit the viral load. Meanwhile, it may also be viewed as a classical measurement error model.

Random effects are used to incorporate the within-individual correlation and between-individual variarion. We use the standard model selection procedures based on the AIC and BIC values to choose an appropriate random effects specification. We find the following LME model fits the viral load data best:

$$\begin{aligned} &z_{ij} = \alpha_{0i} + \alpha_1 t_{ij} + \alpha_{2i} t_{ij}^2 + \varepsilon_{ij}, \quad i = 1, 2, \cdots, N, \; j = 1, 2, \cdots, n_i, \\ &\alpha_{0i} = \alpha_0 + a_{0i}, \quad \alpha_{2i} = \alpha_2 + a_{2i}, \end{aligned}$$

where z_{ij} is the ($\log_{10}$-transformed) viral load at time t_{ij}, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)^T$ is a vector of fixed effects, $\boldsymbol{a}_i = (a_{0i}, a_{2i})^T$ is a vector of random effects, and ε_{ij} is the measurement error. We assume that

$$\boldsymbol{a}_i \sim N(\mathbf{0}, \Sigma), \; \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 I),$$

and $\boldsymbol{a}_i$ and $\boldsymbol{\varepsilon}_i$ are independent. Let

$$z_{ij}^* = \alpha_{0i} + \alpha_1 t_{ij} + \alpha_{2i} t_{ij}^2$$

be the unobserved true value of z_{ij}.

The primary interest is to examine the relationship between viral load and the time to dropout. Let's consider a Cox model which links the hazard of the time to dropout (T_i) to the unobserved true value of viral load (z_i^*) rather than the observed z_i with measurement error.

We consider the following Cox model with a single time-dependent covariate

$$h_i(t) = h_0(t)\exp(\beta z_i^*(t)), \quad i = 1, 2, \cdots, N, \tag{3.5}$$

where $h_0(t)$ is the unspecified baseline hazard function and β is the parameter of primary interest.

Data Analysis Results

Our main interest is the effect of viral load on the time to dropout, which can be interpreted based on the inference on the parameter β. We apply the three methods, the naive method (NAIVE), the naive two-step method (NTS), and the joint likelihood or joint model (JM) method, to analyze the data.

The results from different methods are displayed in Table 3.1. We show the estimates and their standard errors of the three fixed effects α_0, α_1, α_2 in the longitudinal covariate model, and the primary parameter β which links the survival response to the covariate. We also show the estimates for the standard deviations of the random effects and measurement errors.

We can see that the NTS method and JM method give similar results for parameter estimates in the longitudinal model, i.e. estimates of α_0, α_1, α_2, σ_{11}, σ_{22} and σ, where σ_{11} and σ_{22} denote the standard deviations of the random effects. These parameter estimates are similar from the two methods and their significances at 5% level are consistent across different methods: the parameters α_0 and α_2 are significantly positive, while α_1 is significantly negative. Note that, although the NTS method and JM method give similar estimates for the mean parameters in longitudinal covariate model, the actual values of the estimates from these methods are different. This is expected, for the naive two-step method, we model the longitudinal and survival processes separately, so biases of estimated parameters in the longitudinal covariate model may arise if the longitudinal process and survival process influence each other. The standard errors obtained from the JM method are smaller than those from the NTS method. The estimates of the variance parameters σ_j's are similar across different methods.

However, the estimates of the main parameter β from the three methods are quite different. For the naive method, β is not significant, which does not suggest a significant effect of the viral load on the hazard of the time to dropout. For the NTS method, β is significantly negative, which means that the viral load negatively affects the hazard of the time to dropout. For the JM method, β is significantly positive, which implies that the viral load positively influences the hazard of the dropout time (i.e. higher viral load values are associated with high hazard of dropouts). The conclusions from the JM method should be the most reliable, as will be demonstrated from the simulation study in the next section.

Table 3.1: Summary of results from data analysis

Parameter		NAIVE[a]	NTS	JM
α_0	Estimate	N/A	4.93	5.19
	(S.E.)		(0.10)	(0.05)
α_1	Estimate	N/A	-6.88	-6.91
	(S.E.)		(0.35)	(0.19)
α_2	Estimate	N/A	4.91	5.27
	(S.E.)		(0.36)	(0.15)
$\boldsymbol{\beta}$	**Estimate**	**-0.075**	**-0.22**	**6.07**
	(S.E.)	**(0.07)**	**(0.10)**	**(0.07)**
σ_{11}[b]	Estimate	N/A	0.54	0.59
σ_{22}[c]	Estimate	N/A	0.69	0.90
σ	Estimate	N/A	0.38	0.37

Note:

[a] *The naive method does not model the longitudinal data so offers no estimates of the α_j's.*

[b] *σ_{11} is the standard deviation of random effect a_{0i}.*

[c] *σ_{22} is the standard deviation of random effect a_{2i}.*

Note that the naive method does not consider the measurement errors in the observed covariate data, and it directly uses the observed data as if they were the true values, so it may lead to a biased estimate of the main parameter β. For the NTS method, we model the covariate process and the survival process separately, so bias of the estimated β may arise if the longitudinal process and survival process influence each other. The JM method models the longitudinal process and the survival process simultaneously, so it may produce less biased estimates and may also be more efficient.

For the NTS method, the uncertainty of the estimation in the first step is not incorporated into the second step, so standard errors of $\hat{\beta}$ may be unreliable. However, the JM method makes inference based on the joint likelihood of all data, so it may produce more reliable standard error than the NTS method. To further compare and evaluate the performances of difference methods, a simulation study is conducted in the next section.

3.4 A Simulation Study

3.4.1 Introduction

In this section, we conduct a simulation study to evaluate the performance of the joint likelihood method, compared to the naive method, the naive two-step method and the modified two-step method. We compare the performances of different methods based on the biases of the estimates, and the coverage rates of the confidence intervals under several scenarios. First, we describe the models used to simulate the data. Then, we describe how we design the simulation study, including the settings of the parameter values. Finally, we compare results from different methods under different settings and draw conclusions.

3.4.2 Simulation Design

The Models

We generate the values of the time-dependent covariate Z from the following linear mixed effects model:

$$\begin{aligned} &z_{ij} = \alpha_{0i} + \alpha_{1i} t_{ij} + \varepsilon_{ij} \equiv z^*_{ij} + \varepsilon_{ij}, \quad i = 1, 2, \cdots, N, \ j = 1, 2, \cdots, n_i, \\ &\alpha_{0i} = \alpha_0 + a_{0i}, \quad \alpha_{1i} = \alpha_1 + a_{1i}, \end{aligned}$$

where z_{ij} and z^*_{ij} respectively are the observed value and unobserved true value of covariate Z for patient i at time t_{ij}, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{in_i})^T$ represents the measurement errors, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$ is a vector of fixed effects, and $\boldsymbol{a}_i = (a_{0i}, a_{1i})^T$ is a vector of random effects. We assume that $\varepsilon_{ij} \sim_{i.i.d} N(0, \sigma^2)$ and $\boldsymbol{a}_i \sim_{i.i.d} N(\mathbf{0}, \Sigma)$.

For survival data, we assume the following Cox model:

$$h_i(t) = h_0(t)\exp(\beta z_i^*(t)), \quad i = 1, 2, \cdots, N, \tag{3.6}$$

where β is the parameter of primary interest, and $h_0(t)$ is the baseline hazard function. In this simulation study, we assume that the survival time follows a Weibull distribution, so the baseline hazard function is that of a Weibull distribution $W(\lambda, \gamma)$, where λ is the scale parameter and γ is the shape parameter. We choose parameters $\lambda = 1$ and $\gamma = 5$ for convenience.

Generating Survival Times

Bender, Augustin and Blettner (2003) introduced a method to simulate survival times from a Cox proportional hazards model with time-independent covariate. Here, we extend their method to a Cox model with time-dependent covariate.

By assuming the above Weibull proportional hazard model, the survival function of the time T to an event is given by

$$S(t) = \exp[-\lambda t^\gamma \times \exp(\beta z_i^*(t))],$$

where $z_i^*(t)$ stands for the true value of the covariate z for patient i at time t. Since $z_i^*(t) = \alpha_{0i} + \alpha_{1i}t$, $S(t)$ can be written as

$$S(t) = \exp\left\{-\lambda t^\gamma \times \exp[\beta(\alpha_{0i} + \alpha_{1i}t)]\right\}.$$

The corresponding cumulative distribution function is

$$F(t) = 1 - \exp[-\lambda t^\gamma \times \exp(\beta z_i^*(t))].$$

Note that $U = F(T)$ follows a uniform distribution on the interval from 0 to 1, i.e. $U \sim \text{Unif}[0, 1]$. So the survival function $S(T) = 1 - F(T)$ also follows $\text{Unif}[0, 1]$. Therefore, we can generate the time to an event for each patient by solve the following equation

$$\exp\left\{-\lambda t^\gamma \times \exp[\beta(\alpha_{0i} + \alpha_{1i}t)]\right\} = u,$$

where u is a random number from $\text{Unif}[0, 1]$. The detailed procedure to generate the time to an event for patient i is:

Step 1: generating a random number u_i from the uniform distribution $\text{Unif}[0, 1]$;

Step 2: solving t from the equation

$$\exp\{-\lambda t^{\gamma} \times \exp[\beta(\alpha_{0i} + \alpha_{1i} t)]\} - u_i = 0,$$

which is a survival time from the Weibull model with survival function $S(t)$.

True Parameter Values

In the simulation study, the entire study period is set to be from 0 to 1 for convenience. The true values of the fixed effects $\boldsymbol{\alpha}$ is set to be $(0.05, -0.6)$, and the covariance matrix for the random effects $\boldsymbol{a_i}$ is set to be

$$\Sigma = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

The parameter of primary interest β, which measures the association between the time to an event and the time-dependent covariate, is chosen to be $\beta = -1$. The censoring rate for the survival data is controlled to be 20%. The number of patients N, the number of repeated measurements for each individual n_i, and the variance of the measurement error σ^2, are set to have several different values for comparison (see their values in the simulation results).

We will apply four different methods to the simulated datasets to evaluate the performances of these methods in terms of their biases and coverage rates of 95% confidence intervals of the parameters. The four methods are the joint likelihood method or joint model (JM) method, the naive method, the naive two-step (NTS)method, and the modified two-step (MTS) method. In the modified two-step method, we run bootstrap 500 times. The repetition times of the simulations are 500.

In the results, we show the "Estimate", "SE", "Sample SE", "Bias", and "Coverage" of the parameters. For each parameter, "Estimate" is the average of the 500 estimates from the 500 simulation runs. Similarly, "SE" is the averages of the 500 standard errors of estimates. "Sample SE" is the sample standard deviation of the 500 estimates from the 500 simulation runs. "Bias" is the difference between the "Estimate" and the corresponding true value. "Coverage" is the percentage of the confidence intervals containing the true value among the 500 confidence intervals from the 500 simulation runs.

First, we simulate samples of size $N = 50$, with 11 repeated measurements for each individual in the samples, i.e. $n_i = 11$ for $i = 1, 2, \cdots, 50$, and the gap between two consecutive observing times is 0.1. The standard deviation of the measurement error is set to be $\sigma = 0.4$. Under this setting, we only consider the situation where longitudinal data are not truncated at the event times.

Then, we change the true values of the variance of measurement errors (σ^2), the sample size (N), and the number of repeated measurements (n_i) to investigate the effects of σ^2, N and n_i on the estimation of β. Finally, we consider the situation where longitudinal data are truncated at the event times.

3.4.3 Simulation Results

Comparison of Different Methods

Table 3.2 displays simulation results from different methods for the three parameters of interest, α_0, α_1 and β. For the longitudinal model parameters α_0 and α_1, we can see that the four methods give similar results. The biases of the estimates from the four methods are almost the same, and they are all very small. The coverage rates are all around 95%, which is the nominal level. We may notice that the MTS method produces relatively lager "SE"s than the other two methods, which is not surprising because the MTS method adjusts the standard errors by bootstrapping. After all, the four methods provide pretty similar estimation performance to the longitudinal parameters α_0 and α_1.

However, the results for the main parameter β in the survival model are quite different from the four methods. For parameter β, we see that the JM method gives much smaller bias than the other three methods. It is not surprising since the JM method makes simultaneous inference based on joint likelihood for all data. The standard errors ("SE"s) from the JM method and the MTS method are larger than those from the naive method and the NTS method. This is expected because the JM method makes inference based on the joint likelihood which incorporates all the uncertainty in the longitudinal and survival data, and the MTS method adjusts the standard error of $\hat{\beta}$ by incorporating the estimation uncertainty in the first step through bootstrapping. On the other hand, the naive method ignores

Table 3.2: Simulation results ($N = 50$, $\sigma = 0.4$, $n_i = 11$)

True Parameter		JM	NAIVE[a]	NTS	MTS
$\alpha_0 = 0.05$	Estimate	0.056	N/A	0.056	0.056
	SE	0.144		0.146	0.147
	Sample SE [b]	0.143		0.143	0.143
	Bias	0.006		0.006	0.006
	Coverage[c]	0.962		0.966	0.954
$\alpha_1 = -0.6$	Estimate	-0.598	N/A	-0.598	-0.598
	SE	0.148		0.149	0.183
	Sample SE	0.148		0.148	0.148
	Bias	0.002		0.002	0.002
	Coverage	0.956		0.958	0.988
$\beta = -1$	Estimate	-1.039	-0.930	-0.907	-0.907
	SE	0.187	0.114	0.119	0.193
	Sample SE	0.201	0.186	0.187	0.187
	Bias	-0.039	0.070	0.093	0.093
	Coverage	0.934	0.710	0.714	0.908

Note:

[a] *The naive method does not model the longitudinal covariate data.*

[b] *"Sample SE" is the empirical standard deviation of the parameter estimates.*

[c] *95% coverage rate.*

the measurement error, while the NTS method does not incorporate the estimation uncertainty in the first step. So the "SE" s of the naive method and the NTS method are likely to be underestimated. This can also explain the fact that the values of "SE" and "Sample SE" are similar for the JM method and the MTS method, while for the naive method and the NTS method, the values of "Sample SE" are much larger than the values of "SE".

Because of less bias and more reliable standard errors, the coverage rate of the joint likelihood estimate of β is higher than other methods (it is 93.4%, which is close to the nominal confidence level 95%). Although the MTS method produces estimates with larger bias than that of the JM method, the coverage rate, which is 90.8%, is just next to the JM method since it adjusts the standard errors. The naive method and the NTS method have lower coverage rates because they produce larger biases and underestimated standard errors.

Table 3.3 also shows the estimates for the standard deviations of random effects (i.e. a_{0i} and a_{1i}) and measurement errors. The estimates from the three methods are all quite close to their true values. So these methods do not seem to have much effects on the estimation of variance components.

Table 3.3: Simulation results for the estimates of standard deviations

True Parameter	JM	NAIVE	NTS	MTS
$\sigma_{11} = 1$	0.995	N/A	1.005	1.005
$\sigma_{22} = 1$	0.972	N/A	0.984	0.984
$\sigma = 0.4$	0.400	N/A	0.400	0.400

Note: σ_{11} *and* σ_{22} *are the standard deviations of* a_{0i} *and* a_{1i} *respectively, and* σ *is the standard deviation of the measurement errors.*

In the following, we compare the performances of the four methods in different scenarios. We mainly focus on the estimation of the main parameter β, since these methods give similar results for the parameters in the longitudinal covariate model.

Different Magnitudes of Measurement Error

We apply different methods to simulated datasets with different variabilities of measurement error to examine the effect of measurement error on the results. We consider three standard deviations for measurement errors: $\sigma = 0.2$, $\sigma = 0.4$ and $\sigma = 0.6$. The setting of the other parameters is the same as the case in Table 3.2. The results for the case $\sigma = 0.4$ are already displayed in Table 3.2. Table 3.4 shows simulation results for cases $\sigma = 0.2$ and $\sigma = 0.6$.

From Tables 3.2 and 3.4, we see that the coverage rate of the NTS method decreases substantially as the variability of measurement errors increases, which is not surprising. A key disadvantage of the NTS method is that it does not incorporate the uncertainty in the estimation in the first step. Thus, the larger the variance of the measurement errors is, the worse the NTS method performs. We can also notice that the bias of the two-step methods increases as the variability of measurement errors increase. The JM method, on the other hand, performs well, regardless of the magnitude of measurement errors. The joint model method performs the best, in terms of both bias and coverage rate.

Table 3.4: Simulation results for estimating $\beta = -1$ ($N = 50$, $n_i = 11$)

Magnitudes of Measurement Error		JM	NAIVE	NTS	MTS
$\sigma = 0.2$	Estimate	-1.031	-0.919	-0.914	-0.914
	SE	0.182	0.117	0.118	0.190
	Sample SE	0.203	0.195	0.190	0.190
	Bias	-0.031	0.081	0.086	0.086
	Coverage	0.930	0.698	0.720	0.904
$\sigma = 0.6$	Estimate	-1.034	-0.932	-0.885	-0.885
	SE	0.194	0.111	0.119	0.196
	Sample SE	0.206	0.189	0.185	0.185
	Bias	-0.034	0.068	0.115	0.115
	Coverage	0.944	0.694	0.680	0.898

Different Sample Size

In order to check how sample size affects parameter estimation, we simulate datasets with a larger number of subjects $N = 100$. The setting of the other parameters stays the same as the case in Table 3.2. The simulation results with $N = 100$ are shown in Table 3.5. Compared with results in Table 3.2, we find that the methods for larger sample size generally produce lower coverage rates for the methods except JM method. Initially, it may seem a bit surprising since a larger sample size may generally result in better estimations. However, larger sample sizes may lead to more accurate estimation of biases and standard errors, making the differences between the methods more obvious.

Table 3.5: Simulation results for estimating $\beta = -1$ with a larger sample size $N = 100$ ($\sigma = 0.4$, $n_i = 11$)

	JM	NAIVE	NTS	MTS
Estimate	-1.014	-0.908	-0.887	-0.887
SE	0.128	0.080	0.082	0.130
Sample SE	0.134	0.123	0.123	0.123
Bias	-0.014	0.092	0.113	0.113
Coverage	0.940	0.662	0.624	0.834

Different Number of Repeated Measurements

To investigate the influence of the number of repeated measurements within individuals on parameter estimation, we apply methods to simulated datasets with a smaller number of repeated measurements 6 (i.e. $n_i = 6$ for all i, so the duration between two consecutive observing times is 0.2). The setting of the other parameters stays the same as the case in Table 3.2. The simulation results with $n_i = 6$ are shown in Table 3.6.

Compared with Table 3.2, results for less repeated measurements are associated with lower coverage rates for the JM methods, the NTS method, and the MTS method. But for the naive method, the coverage rate does not decrease. This is expected because either two-step methods or the JM method require large within-individual repeated measurements to perform well, since more repeated measurements imply more information about the longitudinal covariate process. Therefore, coverage rates of the JM method and the two-step methods decrease when the number of repeated measurements decreases. However, for the naive method, no longitudinal model is assumed and fitted, so the number of repeated measurements has no effect.

Table 3.6: Simulation results for estimating $\beta = -1$ with sparse repeated measurements($N = 50$, $\sigma = 0.4$, $n_i = 6$)

	JM	NAIVE	NTS	MTS
Estimate	-1.061	-0.948	-0.918	-0.918
SE	0.193	0.114	0.118	0.198
Sample SE	0.211	0.193	0.187	0.187
Bias	-0.061	0.052	0.082	0.082
Coverage	0.942	0.724	0.702	0.904

Longitudinal Data Truncated at Event Times

In this part, we consider the situation where longitudinal data are truncated at the event times. The true values of all the parameters are the same as the case in Table 3.2. The simulation results are shown in Table 3.7.

Compared with Table 3.2, when longitudinal data are truncated at event times, two-step methods produce much more biases and lower coverage rates than those for data not truncated. First, truncation strengthens the association between the longitudinal and survival processes. Note that the current two-step methods model the longitudinal process separately, without in-

Table 3.7: Simulation results for estimating $\beta = -1$ where the longitudinal data are truncated at event times)

	JM	NAIVE	NTS	MTS
Estimate	-1.051	-0.926	-0.828	-0.828
SE	0.200	0.114	0.122	0.182
Sample SE	0.226	0.192	0.200	0.200
Bias	-0.051	0.074	0.172	0.172
Coverage	0.930	0.706	0.596	0.776

corporating survival or event information, so biases of parameter estimates may increase when the association between the longitudinal and survival processes becomes stronger. Secondly, since the longitudinal covariate trajectory is related to the length of follow-up, some information in the longitudinal process may be lost when truncation happens. On the other hand, the JM method incorporates the association between the two processes and the naive method does not model the longitudinal process, so the truncation has no much effect on the JM method and the naive method.

In order to incorporate the association between the longitudinal and survival processes and recapture the missing measurements due to event, Albert and Shih (2009) proposed to generate the missing covariate values by incorporating the event time information. That is, generating data from the conditional distribution of the covariate given the event time $f(\boldsymbol{z}_i|T_i)$. However, their method is difficult to implement and it does not provide most efficient inference (while the JM method is asymptotically most efficient).

3.4.4 Conclusions

From the simulation results, we find that the joint model or joint likelihood method usually has the smallest biases and the largest coverage rates close to the nominal level (95%), compared to the other three methods. These results confirm that the joint likelihood method produces less biased estimates and more reliable standard errors. By adjusting the standard errors through bootstrapping, the modified two-step method produces the second largest coverage rates, but it does not correct bias. The naive method and the naive two-step method have relatively large biases and the low coverage rates. This is because the naive method ignores measurement errors, and the naive two-step method does not incorporate the uncertainty in the estimation of the longitudinal model.

The performances of the methods depend on the magnitude of measurement errors, sample size, number of within-individual measurements, and truncations of longitudinal data. Larger measurement errors lead to worse performances of the two naive methods. Larger sample size seems to lower the coverage rates of the methods except JM method. More within-individual measurements lead to better performances of the methods except the naive method. Also, all these four methods performance worse when the longitudinal data are truncated at event times, but the effect on the joint model method and the naive method is quite minor.

3.5 Joint Inference for an Non-Linear Mixed Effects Model and a Survival Model

3.5.1 Introduction

In Section 3.2, we assumed a linear mixed effects model to the longitudinal data and assume a Cox model to the survival data. Then we discussed several joint inference methods for a linear mixed effects model and a Cox model. However, linear models only empirically describe the observed data but do not provide understanding of the true relationship between the covariates and the response. Therefore, in many longitudinal studies, nonlinear models which describe the underlying mechanism of data generation are better choices when such nonlinear models are available.

In this section, we assume an nonlinear mixed effects (NLME) model for longitudinal data, and we mainly focus on joint inference methods for an NLME model and a survival model. We consider another widely used class of survival models - accelerated failure time (AFT) models for survival data. Similar to Section 3.2, we focus on the situation where the survival AFT model with an error-prone time-dependent covariate is of primary interest. Then an NLME model is assumed for the time-dependent covariate in the AFT model to address covariate measurement errors or missing covariates. The notation is the same as that in Section 3.1: we let $z_{ij} = z_i(t_{ij})$ and z^*_{ij} denote the observed covariate value and the corresponding unobserved true value for individual i at time t_{ij}, and let $\boldsymbol{x}_i$ be a vector of other covariates without measurement errors.

We again consider the naive two-step method, the modified two-step method and the joint likelihood method to make joint inference for an NLME model and an AFT model. In the naive two-step method, we first use an NLME model to fit the longitudinal data and estimate the true values of the mismeasured covariate in the first step, and then we proceed inference on the AFT model in the second step as if the estimated true covariate values were observed data. The modified two-step method uses a parametric bootstrap method to obtain adjusted standard errors of the estimates, in order to incorporate the estimation uncertainty in the first step.

For the joint likelihood method, we make simultaneous inference on the two models based on the joint likelihood of all the longitudinal and survival data. The computation of the joint likelihood is more intensive than that in Section 3.2.3 because of the nonlinearity of the mixed effects model. Thus we consider a linear approximation to the NLME model based on a Taylor series expansion to approximate the joint likelihood (Lindstrom and Bates (1990) and Pinheiro and Bates (1995)). By doing this, the computation is much more efficient and available R package "JM" for LME models can be readily incorporated.

Finally, we conduct real data analysis and a simulation study to evaluate and compare the performances of these three methods.

3.5.2 Models and Methods

In the following, we describe the survival model and longitudinal model under consideration in this section.

The Survival Model

For modeling survival data, we consider the following AFT model with time-dependent and time-independent covariates:

$$\log(T_i) = z_i^*(t)\beta_1 + \boldsymbol{x}_i^T \boldsymbol{\beta}_2 + \sigma\epsilon_i, \quad i = 1, 2, \cdots, N, \tag{3.7}$$

where $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2)$ are regression parameters, σ is a scale parameter, and ϵ_i's are random errors. This model links the survival response to the unobserved true covariate value $z_i^*(t)$ rather than the observed value $z_i(t)$ which is measured with error.

Statistical inference for an AFT model can be based on the standard likelihood method. To make inference, we encounter the measurement error and the missing covariates problems, as discussed in Section 3.2. Thus, a longitudinal model needs to be assumed for the time-dependent covariate in order to address measurement errors and missing data.

The Longitudinal Covariate Model

To address measurement errors and missing data in the time-dependent covariates, we consider the following NLME model:

$$z_{ij} = g(\boldsymbol{\alpha}, \boldsymbol{a}_i, t_{ij}) + \varepsilon_{ij} \equiv z_{ij}^* + \varepsilon_{ij}, \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, n_i, \tag{3.8}$$

where the unobserved true covariate value is assumed to be $z_{ij}^* = g(\boldsymbol{\alpha}, \boldsymbol{a}_i, t_{ij})$, $g(\cdot)$ is a specified nonlinear function, $\boldsymbol{\alpha}$ is a vector of fixed effects, $\boldsymbol{a}_i$ is a vector of random effects, and ε_{ij} is measurement error for individual i at time t_{ij}. We assume that $\boldsymbol{a}_i \sim N(\mathbf{0}, \Sigma)$, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, D_i)$, and $\boldsymbol{a}_i$ and $\boldsymbol{\varepsilon}_i$ are independent.

The Naive Two-step Method

The NLME model (3.8) and the AFT model (3.7) are linked via the unobserved variable z_{ij}^*, so the naive two-step method discussed in Section 3.2.1 can be used. In the first step, we estimate the true values of the covariate Z by fitting the NLME model (3.8) to the observed data $\{(\boldsymbol{z}_i, \boldsymbol{t}_i),\ i = 1, 2, \cdots, N\}$, ignoring the survival model. We denote the estimated true value of the covariate at event time t_i by $\hat{z}_i^*$. In the second step, we use $\hat{z}_i^*$ to substitute the unobserved covariate $z_i^*(t_i)$ and proceed the usual inference on the AFT model with time-independent covariates. In other words, the survival model used in the second step is

$$\log(T_i) = \hat{z}_i^* \beta_1 + \boldsymbol{x}_i^T \boldsymbol{\beta}_2 + \sigma \epsilon_i$$

Modified Two-step Method

The naive two-step method ignores the estimation uncertainty in the first step so it may lead to unreliable standard errors. The modified two-step method discussed in Section 3.2.2 can be used to adjust the under-estimated standard errors in the naive two-step method through bootstrapping. We first generate longitudinal covariate values from the fitted NLME model and generate survival data from the fitted AFT model, and then use the naive two-step method to fit the generated data to obtain new estimates of the parameters. After repeating this procedure many times, we can use the sample standard deviations of the new estimates to adjust the standard errors. The modified two-step method provides more reliable standard errors, but it still may not completely remove biases in the naive two-step method.

Joint Likelihood Method

For the joint likelihood method, the joint likelihood of the AFT model (3.7) and the NLME model (3.8) based on all observed data $\{(\boldsymbol{z}_i, \boldsymbol{x}_i, t_i, \delta_i), i = 1, 2, \cdots, N\}$ is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int [h_i(t_i|\boldsymbol{\beta}, \boldsymbol{x}_i, z_i^*(t))]^{\delta_i} S_i(t_i|\boldsymbol{\beta}, \boldsymbol{x}_i, z_i^*(t)) \times f(\boldsymbol{z}_i|\boldsymbol{\alpha}, \boldsymbol{a}_i, D_i) f(\boldsymbol{a}_i|\Sigma) d\boldsymbol{a}_i, \tag{3.9}$$

where $h_i(t)$ and $S_i(t)$ are the hazard function and survival function of the survival time T_i's respectively, $\boldsymbol{\theta}$ contains all the unknown parameters in the two models. Note that $f(\boldsymbol{z}_i|\boldsymbol{\alpha}, \boldsymbol{a}_i, D_i)$ is the conditional density function of $\boldsymbol{z}_i$, which is more complicated than that in (3.3) because the longitudinal model is nonlinear. Thus, the computation of the joint likelihood (3.9) is much more intensive than that for the joint likelihood (3.3).

To evaluate the joint likelihood in a computationally more efficient way, we consider a linear approximation to the NLME model, which leads to a "working" LME model based on a Taylor series expansion. By doing this, we can approximate the function $f(\boldsymbol{z}_i|\boldsymbol{\alpha}, \boldsymbol{a}_i, D_i)$ by a simpler form similar to that of a LME model. Lindstrom and Bates (1990) proposed a linear approximation based on Taylor series expansion, and it is now standard for estimation of NLME models and is used in standard software. Pinheiro and Bates (1995) evaluated the performance of Taylor approximations to NLME models through extensive simulation and concluded that the approximate methods perform well.

In the following, we briefly describe how the linear approximation to NLME model (3.8) based on Taylor series expansion works.

A Linear Approximation to a NLME model

Consider the NLME model (3.8). Following Lindstrom and Bates (1990), we take a first-order Taylor expansion about the current estimates of parameters $\hat{\boldsymbol{\alpha}}$ and the current estimates of random effects $\hat{\boldsymbol{a}}_i$ to the nonlinear function

$g(\boldsymbol{\alpha}, \boldsymbol{a}_i, t_{ij})$

$$\begin{aligned} g(\boldsymbol{\alpha}, \boldsymbol{a}_i, t_{ij}) &\approx g(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{a}}_i, t_{ij}) + \frac{\partial g}{\partial \boldsymbol{\alpha}}\big|_{\hat{\boldsymbol{\alpha}}}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) + \frac{\partial u}{\partial \boldsymbol{a}_i}\big|_{\hat{\boldsymbol{a}}_i}(\boldsymbol{a}_i - \hat{\boldsymbol{a}}_i) \\ &= \left(g(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{a}}_i, t_{ij}) - \frac{\partial g}{\partial \boldsymbol{\alpha}}\big|_{\hat{\boldsymbol{\alpha}}}\hat{\boldsymbol{\alpha}} - \frac{\partial g}{\partial \boldsymbol{a}_i}\big|_{\hat{\boldsymbol{a}}_i}\hat{\boldsymbol{a}}_i \right) + \frac{\partial g}{\partial \boldsymbol{\alpha}}\big|_{\hat{\boldsymbol{\alpha}}}\boldsymbol{\alpha} + \frac{\partial g}{\partial \boldsymbol{a}_i}\big|_{\hat{\boldsymbol{a}}_i}\boldsymbol{a}_i. \end{aligned}$$

We obtain a ‘working’ LME model:

$$\tilde{z_{ij}} = \boldsymbol{W}_{ij}\boldsymbol{\alpha} + \boldsymbol{R}_{ij}\boldsymbol{a}_i + e_{ij}, \tag{3.10}$$

where

$$\begin{aligned} \tilde{z_{ij}} &= z_{ij} - g(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{a}}_i, t_{ij}) + \boldsymbol{W}_{ij}\hat{\boldsymbol{\alpha}} + \boldsymbol{R}_{ij}\hat{\boldsymbol{a}}_i, \\ \boldsymbol{W}_{ij} = &\frac{\partial g(\boldsymbol{\alpha}, \hat{\boldsymbol{a}}_i, t_{ij})}{\partial \boldsymbol{\alpha}}\big|_{\hat{\boldsymbol{\alpha}}}, \quad \boldsymbol{R}_{ij} = \frac{\partial g(\hat{\boldsymbol{\alpha}}, \boldsymbol{a}_i, t_{ij})}{\partial \boldsymbol{a}_i}\big|_{\hat{\boldsymbol{a}}_i}. \end{aligned}$$

The linearization procedure is to iteratively solve the “working” LME model (3.10) and obtain the updated estimates $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{a}}_i)$ from the LME model at each iteration until converge. At the last iteration, we obtain the “best” working LME model. Then, in the joint likelihood method, the computation can be based on the joint likelihood of the “best” working LME model and the AFT model, which is less intensive and available R package can be readily used.

3.5.3 Data Analysis - A HIV Study

In the previous section, we have described three joint inference methods for an NLME model and an AFT model. They are the naive two-step method, the modified two-step method, and the joint likelihood method. In this section, we use the joint likelihood method and the naive two-step method, together with a naive method which ignores the measurement errors in covariate, to analyze a real datasets from a HIV study described in Section 1.4 of Chapter 1.

As described in Section 3.3, the dataset contains data for 46 patients’ viral loads measured over time and their times to dropout. The objective of this data analysis is to examine the relationship between viral load trajectories and the times to dropout. More specifically, we are interested in checking whether patients with high initial viral loads have earlier times to dropout. Here the viral load is measured with errors, and the measurement errors should not be ignored. Table 1.1 summarize the variables of interest, and Figures 1.1 and 1.2 show the viral load trajectories. We consider the data in the first 90 days after the anti-HIV treatment.

The Models

To address measurement errors in viral load data, we need to assume a model for the time-dependent viral load. Wu and Ding (1999) proposed a two-compartment exponential decay model for viral load trajectory in the early period after an anti-HIV treatment.

We use the standard model selection procedures based on the AIC and BIC values to choose an appropriate random effects specification. We obtain the following two-compartment exponential NLME model to fit the viral load data:

$$\begin{aligned} z_{ij} &= \log_{10}\left(\exp(\alpha_{1i} - \alpha_2 t_{ij}) + \exp(\alpha_{3i} - \alpha_{4i} t_{ij})\right) + \varepsilon_{ij}, \\ \alpha_{1i} &= \alpha_1 + a_{1i}, \quad \alpha_{3i} = \alpha_3 + a_{3i}, \quad \alpha_{4i} = \alpha_4 + a_{4i}, \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ is a vector of fixed effects, $\boldsymbol{a}_i = (a_{1i}, a_{3i}, a_{4i})^T$ is a vector of random effects. We assume that $\boldsymbol{a}_i \sim N(\mathbf{0}, \Sigma), \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \tau^2 I)$, and $\boldsymbol{a}_i$ and $\boldsymbol{\varepsilon}_i$ are independent. Let

$$z^*_{ij} = \log_{10}\left(\exp(\alpha_{1i} - \alpha_2 t_{ij}) + \exp(\alpha_{3i} - \alpha_{4i} t_{ij})\right)$$

be the unobserved true value of z_{ij}.

For the time to event model, we consider a parametric AFT model which links the dropout time (T_i) to the unobserved true value of viral load (z^*_i) rather than the observed z_i with measurement error. We consider the following AFT model with a single time-dependent covariate:

$$\log(T_i) = \beta_0 + \beta_1 z^*_i(t) + \sigma\epsilon_i, \quad i = 1, 2, \cdots, N. \tag{3.11}$$

where β_1 is the parameter of primary interest, σ is a scale parameter and ϵ_i's are random errors. We assume a Gumbel distribution for ϵ_i, so the survival time T_i follows a Weibull distribution.

Data Analysis Results

Our main interest is to examine the relationship between the viral load and the time to dropout, which can be interpreted based on the inference on the parameter β_1. We apply the three methods: the naive method (NAIVE), the naive two-step (NTS) method, and the joint likelihood or joint model (JM) method, to analyze the data.

The results from different methods are displayed in Table 3.8. We show the estimates and their standard errors of the four fixed effect α_1, α_2, α_3 and α_4 in the longitudinal covariate model, the intercept β_0 in the AFT model, and the primary parameter β_1 which links the survival response to the covariate. We also show the estimates for the standard deviation of the measurement errors and the scale parameter in the AFT model.

We can see that the results for estimation of the parameters in the NLME model (including α_2, α_3, α_4 and τ) are similar for the NTS method and the JM method. These parameter estimates are similar from the two methods, and their significances are consistent across different methods: the parameters α_1, α_2, α_3 and α_4 are significantly positive at 5% level. Note that, although the estimates for the fixed effects (i.e. α_1, α_2, α_3 and α_4) are similar, the actual values of the estimates from the two methods are different. For the NTS method, we model the longitudinal and survival processes separately, so biases of estimated parameters in the longitudinal covariate model may arise if the longitudinal process and survival process influence each other. Another reason may be that we use the linear approximation to the NLME model in the JM method while the NTS method does not. The standard errors obtained from the JM method and the NTS method are similar, and the estimates for the standard deviation of the measurement errors are the same across different methods.

However, the estimates of the main parameter β_1 from the three methods are quite different. For the naive method and the JM method, β_1 is not significant at 5% level, which means that the viral load has no significant effect on the time to dropout. For the NTS method, β_1 is significantly negative, which suggests that higher viral loads are associated with earlier times to dropout. The conclusion from the JM method should be the most reliable, as will be demonstrated from the simulation study in the next section.

We can see that the values of estimated β_0 and β_1 are different and the standard errors obtained from the JM method are larger than the naive method and the NTS method. These results are expected. Note that the naive method ignores the measurement errors in the observed covariate data, and it directly uses the observed data as if they were the true values, so it may lead to a biased estimate of the main parameter β_1. For the NTS method,

Table 3.8: Summary of results from data analysis

Parameter			NAIVE[a]	NTS	JM
Longitudinal Model	α_1	Estimate	N/A	12.32	12.34
		(S.E.)		(0.243)	(0.240)
	α_2	Estimate	N/A	37.46	38.03
		(S.E.)		(2.170)	(2.250)
	α_3	Estimate	N/A	7.61	7.64
		(S.E.)		(0.284)	(0.279)
	α_4	Estimate	N/A	1.88	1.93
		(S.E.)		(0.501)	(0.498)
	τ[b]	Estimate	N/A	0.29	0.29
Survival Model	β_0	Estimate	-0.88	1.27	0.31
		(S.E.)	(0.23)	(0.47)	(0.67)
	$\boldsymbol{\beta_1}$	**Estimate**	**0.025**	**-0.68**	**-0.32**
		(S.E.)	**(0.063)**	**(0.14)**	**(0.18)**
	σ[c]	Estimate	1.05	0.77	1.25

Note:

[a] *The naive method does not model the longitudinal data so there are no estimates for the α_j's.*

[b] *τ is the standard deviation of the measurement errors in the NLME model.*

[c] *σ is the scale parameter in the AFT model.*

we model the covariate process and the survival process separately, so bias of the estimated β_1 may arise if the longitudinal process and survival process influence each other. The JM method models the longitudinal process and the survival process simultaneously, so it may produce less biased estimates and may also be more efficient. For the NTS method, the uncertainty of the estimation in the first step is not incorporated into the second step, so standard errors of $\hat{\beta}_1$ may be under-estimated. However, the JM method makes inference based on the joint likelihood of all data, so it may produce more reliable standard errors than the NTS method. To further compare and evaluate the performances of difference methods, a simulation study is conducted in the next section.

3.5.4 A Simulation Study

Introduction

In this section, we conduct a simulation study to evaluate the performances of different joint inference methods for an NLME model and an AFT model. We compare the performances of different methods based on the biases of the estimates and the coverage rates of the corresponding 95% confidence intervals. First, we describe the models used to simulate the data. Then, we describe how we design the simulation study, including the settings of the parameter values. Finally, we compare results from different methods and draw conclusions.

A Simulation Design

The Models

We generate the values of the time-dependent covariate z from the following NLME model:

$$\begin{aligned} z_{ij} &= g(\boldsymbol{\alpha}, \boldsymbol{a_i}, t_{ij}) + \varepsilon_{ij} \equiv z^*_{ij} + \varepsilon_{ij}, \\ \alpha_{1i} &= \alpha_1 + a_{1i}, \quad \alpha_{2i} = \alpha_2 + a_{2i}, \\ \alpha_{3i} &= \alpha_3 + a_{3i}, \quad \alpha_{4i} = \alpha_4 + a_{4i}, \end{aligned}$$

where

$$g(\boldsymbol{\alpha}, \boldsymbol{a_i}, t_{ij}) = \log_{10}\left(\exp(\alpha_{1i} - \alpha_{2i}t_{ij}) + \exp(\alpha_{3i} - \alpha_{4i}t_{ij})\right),$$

z_{ij} and z^*_{ij} respectively are the observed value and unobserved true value of covariate Z for patient i at time t_{ij}, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{in_i})^T$ represents the measurement errors, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ is a vector of fixed effects, and $\boldsymbol{a}_i = (a_{1i}, a_{2i}, a_{3i}, a_{4i})^T$ is a vector of random effects. We assume $\varepsilon_{ij} \sim_{i.i.d} N(0, \tau^2)$ and $\boldsymbol{a}_i \sim_{i.i.d} N(\mathbf{0}, \Sigma)$.

For survival data, we assume the following AFT model

$$\log(T_i) = \beta_0 + \beta_1 z^*_i(t) + \sigma\epsilon_i, \quad i = 1, 2, \cdots, N, \tag{3.12}$$

where β_1 is the parameter of primary interest, β_0 is the intercept, σ is a scale parameter and ϵ_i's are random errors. Different choices of the distributions for ϵ_i lead to different AFT models. In this simulation study, we assume that the distribution of ϵ_i is the Gumbel distribution with survival function and hazard function given respectively by

$$S(t) = \exp(-e^t), \quad h(t) = e^t. \tag{3.13}$$

Generating Survival Times

By assuming the AFT model (3.12) and the Gumbel distribution for ϵ_i, the survival time T_i follows a Weibull distribution with parameters $\lambda = e^{-(\beta_0+\beta_1 z_i^*(t))/\sigma}$ and $\gamma = \frac{1}{\sigma}$. So the survival function of the survival time T_i is

$$\begin{aligned} S(t) &= \exp(-\lambda t^{\gamma}) \qquad (3.14) \\ &= \exp(-e^{-(\beta_0+\beta_1 z_i^*(t))/\sigma} t^{\frac{1}{\sigma}}), \end{aligned}$$

where $z_i^*(t)$ stands for the true value of the covariate for patient i at time t. Since $z_i^*(t) = g(\boldsymbol{\alpha}, \boldsymbol{a_i}, t)$, $S(t)$ can also be written as

$$S(t) = \exp(-e^{\{-\beta_0+\beta_1 g(\boldsymbol{\alpha},\boldsymbol{a_i},t)\}/\sigma} t^{\frac{1}{\sigma}}),$$

where $g(\boldsymbol{\alpha}, \boldsymbol{a_i}, t) = \log_{10}(\exp(\alpha_{1i} - \alpha_{2i}t) + \exp(\alpha_{3i} - \alpha_{4i}t))$.

As in Section 3.4, the random variable $U = S(T)$ follows a uniform distribution on the interval from 0 to 1, i.e. $U \sim \text{Unif}[0, 1]$. Therefore, we can generate the time to an event by solving the equation

$$\exp(-e^{\{-\beta_0+\beta_1 g(\boldsymbol{\alpha},\boldsymbol{a_i},t)\}/\sigma} t^{\frac{1}{\sigma}}) = U$$

The detailed procedure to generate the time to an event for individual i is:

Step 1: generating a random number u_i from the uniform distribution Unif$[0, 1]$;

Step 2: solving t from the equation

$$\exp(-e^{-\{\beta_0+\beta_1 g(\boldsymbol{\alpha},\boldsymbol{a_i},t)\}/\sigma} t^{\frac{1}{\sigma}}) - u_i = 0.$$

True Parameter Values

In the simulation study, the true values of most parameters are chosen based on the real data analysis in Section 3.5.3. The entire study period is set to be from 0 to 1. The true values of the fixed effects $\boldsymbol{\alpha}$ is set to be $(11.7, 30.5, 7.4, 1.7)$, the standard deviation of the measurement errors is set to be $\tau = 0.3$, and the covariance matrix for the random effects $\boldsymbol{a_i}$ is set to be

$$\Sigma = \begin{pmatrix} 1.00 & -1.80 & 1.12 & 0.60 \\ -1.80 & 12.96 & -3.46 & 0.11 \\ 1.12 & -3.46 & 2.56 & 2.88 \\ 0.60 & 0.11 & 2.88 & 9.00 \end{pmatrix}.$$

The parameter of primary interest β_1, which measures the association between the time to an event and the time-dependent covariate, is chosen to be $\beta_1 = -0.4$. The intercept and scale parameter in the AFT model are set to be $\beta_0 = 0.5$ and $\sigma = 1.3$ respectively. The censoring rate for the survival data is controlled to be 20%.

We simulate samples of size $N = 50$, with 11 repeated measurements for each individual, i.e. $n_i = 11$ for $i = 1, 2, \cdots, 50$, and the gap between two consecutive observing times is 0.1. We only consider the situation where the longitudinal data are not truncated at the event times. We will apply four different methods to the simulated datasets to evaluate the performances of these methods in terms of their biases and coverage rates of 95% confidence intervals of the parameters. The four methods are the joint likelihood or joint model (JM) method, the naive (NAIVE) method, the naive two-step (NTS) method, and the modified two-step (MTS) method. Due to high computing time and potential convergence problems (slow convergence or non-convergence), for the MTS method, we run bootstrap 50 times. The repetition times of the simulations are 50. Similar to Section 3.4, in the results, we show the “Estimate”, “SE”, “Sample SE”, “Bias”, and “Coverage” of the parameters.

Simulation Results and Conclusions

Table 3.9 displays simulation results from different methods for all the mean parameters of interest, including α_1, α_2, α_3, α_4, β_0 and β_1. For the fixed effects α_1, α_2, α_3 and α_4 in the longitudinal model, we can see that the biases of the estimates from the three methods are very similar, and they are all relatively small. The coverage rates in the JM method are generally lower than the two-step methods. Therefore, we may say that the JM method performances worse than the NTS and MTS methods when estimating the parameters in the NLME model. It appears that the linearization method performs worse on the NLME submodel in joint models than on NLME models alone.

However, for the parameters β_0 and β_1 in the AFT model, the JM method performances much better than the two-step methods. We see that the JM method gives much smaller bias than the naive method and the two-step methods. It is not surprising since the JM method makes simultaneous inference based on joint likelihood for all data, while the naive method used the mis-measured data directly with ignoring the measurement errors, and the NTS and MTS methods model the longitudinal data and the survival data separately, so biases may arise when the longitudinal process and the survival process influence each other. The standard errors ("SE"s) from the JM method and the MTS method are larger than that from the naive method and the NTS method. This is expected because the JM method makes inference based on the joint likelihood which incorporates all the uncertainty and the MTS method adjusts the standard errors by incorporating the estimation uncertainty in the first step through bootstrapping. On the other hand, the naive method ignores the measurement errors, and the NTS does not incorporate the estimation uncertainty in the first step, so the "SE" s of the naive method and the NTS method are likely to be underestimated.

The above results show that the linearization for NLME model in the JM method only affects estimates in the NLME model but it does not affect estimates in the AFT model. Because of less biases and more reliable standard errors, the coverage rates of β_0 and β_1 in the JM method, which both are 96%, are much higher than the naive method and the two-step methods. It also indicates that the linear approximation to the MLME model works well in the computation of the joint likelihood.

Table 3.9: Simulation results ($N = 50$, $\tau = 0.3$, $n_i = 11$)

True Parameter		JM	NAIVE[a]	NTS	MTS
$\alpha_1 = 11.7$	Estimate	11.702	N/A	11.706	11.706
	SE	0.159		0.167	0.172
	Sample SE[b]	0.161		0.161	0.161
	Bias	0.002		0.006	0.006
	Coverage[c]	0.940		0.980	0.980
$\alpha_2 = 30.5$	Estimate	23.877	N/A	23.813	23.813
	SE	1.340		1.416	1.517
	Sample SE	1.487		1.428	1.428
	Bias	-6.623		-6.687	-6.687
	Coverage	0.000		0.000	0.000
$\alpha_3 = 7.4$	Estimate	7.543	N/A	7.536	7.536
	SE	0.223		0.245	0.252
	Sample SE	0.233		0.223	0.223
	Bias	0.143		0.136	0.136
	Coverage	0.840		0.940	0.940
$\alpha_4 = 1.7$	Estimate	1.880	N/A	1.873	1.873
	SE	0.413		0.428	0.454
	Sample SE	0.486		0.496	0.496
	Bias	0.180		0.173	0.173
	Coverage	0.800		0.900	0.920
$\beta_0 = 0.5$	Estimate	-0.339	3.515	3.448	3.448
	SE	1.044	0.752	0.719	0.904
	Sample SE	0.195	0.261	0.246	0.246
	Bias	-0.839	3.015	2.948	2.948
	Coverage	0.960	0.080	0.040	0.040
$\beta_1 = -0.4$	Estimate	-0.184	-1.328	-1.293	-1.293
	SE	0.278	0.196	0.183	0.201
	Sample SE	0.195	0.261	0.246	0.246
	Bias	0.216	-0.928	-0.893	-0.893
	Coverage	0.960	0.060	0.020	0.020

Note:

[a] *The naive method does not model the longitudinal covariate data.*

[b] *"Sample SE" is the empirical standard deviation of the parameter estimates.*

[c] *95% coverage rate.*

Chapter 4

Joint Inference for a Generalized Linear Mixed Model and a Survival Model

4.1 Introduction

In Chapter 3, we have assumed a linear and nonlinear mixed effects model for the longitudinal data. In both linear and nonlinear mixed effects models, the longitudinal response is assumed to be normally distributed. However, in practice, many types of responses do not necessarily follow normal distributions, such as binary responses or count responses. In such cases, generalized linear models, which greatly extend classical linear models, can be used for the responses which follow distributions in the exponential family, such as normal, exponential, binomial, and Poisson distributions.

In this section, we consider a generalized linear mixed model (GLMM) for a non-normal longitudinal response and discuss joint inference methods for a generalized linear mixed model and a survival model. We consider the widely used accelerated failure time (AFT) model for the survival data. We focus on the situation where the GLMM and the AFT model are linked through the same unknown parameters. The observed covariate value for individual i at time t_{ij} is denoted as $z_{ij} = z_i(t_{ij})$ $(i = 1, 2, \cdots, N, j = 1, 2, \cdots, n_i)$. We assume that z_{ij}'s independently follow a distribution in the exponential family. Let $\boldsymbol{x}_i$ be a vector of other covariates.

We consider the naive two-step method, the modified two-step method and the joint likelihood method for joint inference on a GLMM and an AFT model. In the naive two-step method, we use a GLMM to fit the longitudinal data and estimate the shared parameters in the first step, and then we proceed inference on the AFT model in the second step, substituting the unknown parameters by their estimates from the first step. The modified two-step method uses a bootstrap method to adjust the standard errors in the naive two-step method, which incorporates the estimation uncertainty in the first step. For the joint likelihood method, we make simultaneous inference on the two models based on the joint likelihood of all the longitudinal and survival data. Similar to the situation in Section 3.5.2, the computation of the joint likelihood for a GLMM and a survival model is very intensive due to the nonlinearity of models. Thus, we also consider a linear approximation to the GLMM, similar to that for a NLME model. Also, available R package for joint LME and survival models can be readily incorporated. Finally, we conduct real data analysis and simulation studies to evaluate and compare the performances of these three joint inference methods.

4.2 Joint Inference Methods

In the following, we describe the longitudinal and survival models to be considered for joint inference.

The Longitudinal Model

We consider a GLMM to the time-dependent covariate z_{ij}. In the GLMM, z_{ij}'s are assumed to independently follow a distribution in the exponential family with a mean of μ_{ij}, conditioning on the random effects. Specifically, the GLMM is given by

$$\boldsymbol{\eta}_i = g(\boldsymbol{\mu_i}) = U_i\boldsymbol{\alpha} + V_i\boldsymbol{a}_i, \quad i = 1, 2, \cdots, N, \tag{4.1}$$

where $g(\cdot)$ is a monotone and differentiable function called the link function, $\boldsymbol{\mu_i} = (\mu_{i1}, \mu_{i2}, \cdots, \mu_{in_i})$ and $\mu_{ij} = E(z_{ij})$, U_i and V_i are design matrices, $\boldsymbol{\alpha}$ is a vector of fixed effect, and $\boldsymbol{a}_i$ is a vector of random effects. We assume that $\boldsymbol{a}_i \sim N(\mathbf{0}, \Sigma)$. Note that, by assuming this GLMM, the mean $\mu_{ij} = \mu_i(t_{ij})$ changes smoothly over time.

In this thesis, we focus on two most widely used generalized linear models: the Poisson regression model and the logistic regression model. If the response z_{ij} is a count, we assume a Poisson distribution for z_{ij} and choose

$$g(\mu_{ij}) = \log(\mu_{ij})$$

as the link function. If the response z_{ij} in the longitudinal data is a binary variable, we assume a binomial or Bernoulli distribution for z_{ij} and choose the logit function

$$g(\mu_{ij}) = \text{logit}(\mu_{ij}) \equiv \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right)$$

as the link function.

The Survival Model

For modeling the survival data, we consider the following AFT model with time-dependent and time-independent covariates:

$$\log(T_i) = \eta_i(t)\beta_1 + \boldsymbol{x}_i^T\boldsymbol{\beta}_2 + \sigma\epsilon_i, \quad i = 1, 2, \cdots, N, \tag{4.2}$$

where $\eta_i(t) \equiv g(\mu_i(t))$ is the linear predictor in model (4.1), $\mu_i(t) \equiv E(z_i(t))$ is the mean parameter for subject i at time t, $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2)$ are regression parameters, σ is a scale parameter, and ϵ_i's are random errors. This model links the survival time to the unknown time-dependent mean $\mu_i(t)$ rather than the observed covariate $z_i(t)$. That is, the time to event may be associated with the mean longitudinal profile, which may be practically meaningful in many situations.

4.2.1 The Naive Two-step Method

The GLMM (4.1) and the AFT model (4.2) share the same parameters, so we can use the naive two-step method discussed in Section 3.2.1 of Chapter 3 to make joint inference on these two models. In the first step, we fit the GLMM (4.1) to the observed data $\{(\boldsymbol{z}_i, \boldsymbol{t}_i), \ , i = 1, 2, \cdots, N\}$ and estimate the values of the linear predictor η_{ij}'s, ignoring the survival model. We denote the predicted value of the linear predictor at event time t_i by $\hat{\eta}_i$. Then, in the second step, we use $\hat{\eta}_i$ to substitute the unknown value of $\eta_i(t_i)$ and proceed the usual inference on the AFT model with time-independent covariates. In other words, the survival model used in the second step is

$$\log(T_i) = \hat{\eta}_i\beta_1 + \boldsymbol{x}_i^T\boldsymbol{\beta}_2 + \sigma\epsilon_i$$

4.2.2 The Modified Two-step Method

Since the standard errors from the naive two-step method may be under-estimated, we can use the modified two-step method discussed in Section 3.2.2 of Chapter 3 to adjust the under-estimated standard errors through bootstrapping. We generate longitudinal covariate values from the fitted GLMM and generate survival data from the fitted AFT model, and then use the naive two-step method to fit the generated data to obtain new estimates of the parameters. After repeating this procedure many times, we can use the sample standard deviations of the estimates as the modified standard errors for these estimates. The modified two-step method provides more reliable standard errors, but it still may not completely remove biases in the naive two-step method.

4.2.3 The Joint likelihood or Joint Model Method

For the joint likelihood method, the joint likelihood of the GLMM (4.1) and the AFT model (4.2) based on all observed data $\{(\boldsymbol{z}_i, \boldsymbol{x}_i, t_i, \delta_i), i = 1, 2, \cdots, N\}$ is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int [h_i(t_i|\boldsymbol{\beta}, \boldsymbol{x}_i, \eta_i(t))]^{\delta_i} S_i(t_i|\boldsymbol{\beta}, \boldsymbol{x}_i, \eta_i(t)) \times f(\boldsymbol{z}_i|\boldsymbol{\alpha}, \boldsymbol{a}_i) f(\boldsymbol{a}_i|\Sigma) d\boldsymbol{a}_i, \tag{4.3}$$

where $h_i(t)$ and $S_i(t)$ are the hazard function and survival function of the survival time T_i's respectively, $f(\boldsymbol{z}_i|\boldsymbol{\alpha}, \boldsymbol{a}_i)$ is the conditional density function of $\boldsymbol{z}_i$. Similar to that Section 3.5.2 of Chapter 3, $f(\boldsymbol{z}_i|\boldsymbol{\alpha}, \boldsymbol{a}_i)$ is complicated due to the nonlinear form of the mean structure. Thus, to more efficiently compute the joint likelihood, we consider a linear approximation to the GLMM, which leads to a "working" LME model, based on a Taylor series expansion or the Laplace approximation. By doing this, we can approximate $f(\boldsymbol{z}_i|\boldsymbol{\alpha}, \boldsymbol{a}_i)$ by a simpler from. Breslow and Clayton (1993), Wolfinger (1993), and McCulloch and Searle (2001) discussed such a linear approximation to a GLMM. In the following, we briefly describe this linear approximation to the GLMM (4.1).

A Linear Approximation to a GLMM

Following Breslow and Clayton (1993), approximate estimates for GLMM (4.1) can be obtained by iteratively solving the following LME model

$$\tilde{\boldsymbol{z}}_{\boldsymbol{i}} = U_i\boldsymbol{\alpha} + V_i\boldsymbol{a}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \cdots, N \tag{4.4}$$

where

$$\tilde{\boldsymbol{z}}_{\boldsymbol{i}} = U_i\hat{\boldsymbol{\alpha}} + V_i\hat{\boldsymbol{a}}_i + B_i\left(\boldsymbol{z}_i - g^{-1}(U_i\hat{\boldsymbol{\alpha}} + V_i\hat{\boldsymbol{a}}_i)\right),$$

B_i is a $n_i \times n_i$ diagonal matrix with diagonal elements $\frac{\partial g(\mu_{ij})}{\mu_{ij}}$, $\boldsymbol{\varepsilon}_i$'s independently follow a normal distribution $N(\mathbf{0}, B_iCov(\boldsymbol{z}_i|\boldsymbol{a}_i)B_i)$, $\boldsymbol{a}_i$'s are independent and follow a normal distribution $N(\mathbf{0}, \Sigma)$, and $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{a}_i$ are independent.

The linearization procedure is to iteratively solve the "working" LME model (4.4) and obtain the updated estimates $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{a}}_i)$ from the LME model at each iteration until converge. At the last iteration, we obtain the "best" working LME model. Then, the computation in the joint likelihood method can be based on the joint likelihood of the "best" working LME model and the AFT model, which is less intensive and available R package can be readily used.

If the observed response z_{ij} is a count and the link function is $g(\mu) = \log(\mu)$, the "working" response in the "working" LME model can be written as

$$\tilde{z_{ij}} = U_{ij}\hat{\boldsymbol{\alpha}} + V_{ij}\hat{\boldsymbol{a}}_i + \frac{1}{\hat{\mu}_{ij}}\left(z_{ij} - \exp(U_{ij}\hat{\boldsymbol{\alpha}} + V_{ij}\hat{\boldsymbol{a}}_i)\right).$$

If the observed response z_{ij} is a binary variable and the link function is $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$, the "working" response in the "working" LME model can be written as

$$\tilde{z_{ij}} = U_{ij}\hat{\boldsymbol{\alpha}} + V_{ij}\hat{\boldsymbol{a}}_i + \frac{1}{\hat{\mu}_{ij}(1-\hat{\mu}_{ij})}\left(z_{ij} - \frac{\exp(U_{ij}\hat{\boldsymbol{\alpha}} + V_{ij}\hat{\boldsymbol{a}}_i)}{1 + \exp(U_{ij}\hat{\boldsymbol{\alpha}} + V_{ij}\hat{\boldsymbol{a}}_i)}\right).$$

In the next few sections, we will conduct real data analysis and simulation studies to evaluate different joint inference methods for a GLMM and an AFT model. We separately consider two types of generalized linear models: Poisson regression models and logistic regression models.

4.3 Joint Inference for an AFT Model and a Poisson GLMM

In this section, we evaluate different joint inference methods for an AFT model and a GLMM with a Poisson distribution through real data analysis and a simulation study.

4.3.1 Data Analysis - A HIV Study

In Section 4.2, we have described different inference methods for a GLMM and an AFT model. In this section, we use these joint inference methods to analyze a real datasets from a HIV study described in Section 1.4 of Chapter 1.

As described in Section 3.3 of Chapter 3, the dataset contains the data for 46 patients' immunologic markers (such as CD4 and CD8 cell counts) and their times to dropout. The CD4 cell count is an important index in HIV studies. One may be interested in examining the relationship between CD4 cell count and the time to dropout. For example, we are interested in checking whether patients with high CD4 cell counts have earlier times to dropout. The CD4 cell counts are scaled in multiples of 100 counts. Table 1.1 summarize the variables of interest. We consider the data in the first 90 days after the anti-HIV treatment.

The Models

The CD4 cell counts change over time, so we assume a longitudinal model for the time-dependent CD4 cell counts. Since the values of the CD4 cell counts are discrete count numbers, we consider a GLMM, assuming the counts of CD4 cell follow a Poisson distribution.

We assume that the CD4 cell counts for patient i at time t_{ij}, denoted by y_{ij}, follow a Poisson distribution with mean μ_{ij} (i.e. $y_{ij} \sim Poi(\mu_{ij})$). Random effects are used to incorporate the within-individual correlation and the between-individual variations. We use the standard model selection procedures based on the AIC and BIC values to choose an appropriate random effects specification. We find that following GLMM fits the CD4 data best

$$\begin{aligned}
&\eta_{ij} = \log(\mu_{ij}) = \alpha_{0i} + \alpha_1 t_{ij}, \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, n_i, \\
&\alpha_{0i} = \alpha_0 + a_i,
\end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$ is a vector of fixed effects, and a_i is the random effect. We assume $a_i \sim_{i.i.d} N(0, \tau^2)$.

Our objective is to access the association between the CD4 cell counts and the time to dropout. We use a parametric accelerated failure time model which links the dropout time (T_i) to the mean of the Poisson distribution rather than the observed CD4 cell counts:

$$\log(T_i) = \beta_0 + \beta_1 \eta_i(t) + \sigma \epsilon_i, \quad i = 1, 2, \cdots, N. \tag{4.5}$$

where $\eta_i(t) = \log(\mu_i(t))$ is the linear predictor in the GLMM, $\mu_i(t)$ is the mean of the Poisson distribution, β_1 is the parameter of primary interest, σ is a scale parameter and ϵ_i's are random errors. We assume a Gumbel distribution for ϵ_i, so the survival time T_i follows a Weibull distribution.

Data Analysis Results

Our main interest is to examine the relationship between the CD4 cell mean count and the time to dropout. The effect of the CD4 cell mean count on the time to dropout can be interpreted based on the inference on the parameter β_1. We apply two joint inference methods, the naive two-step (NTS) method and the joint likelihood or joint model (JM) method, to analyze the data. Unlike the previous chapter, there is no naive method here since the models may not be interpreted as measurement error models.

The results from different methods are displayed in Table 4.1. We show the estimates and their standard errors of the fixed effects α_0, α_1 in the GLMM, the intercept β_0 in the AFT model and the primary parameter β_1. We also show the estimates for the standard deviation of the random effect and the scale parameter in the AFT model.

We can see that the NTS method and the JM method give similar results for parameters in the longitudinal model, especially for the fixed effects α_0 and α_1. These parameter estimates are similar from the two methods and their significances at 5% level are consistent across different methods: the fixed effects α_0 and α_1 are significantly positive. Although the estimates for the longitudinal parameters (i.e. α_1 and α_2) are similar, the actual values of the estimates from the two methods are different, since biases of estimated parameters in the longitudinal model based on the NTS method may arise if the longitudinal process and survival process influence each other. The linear approximation to the GLMM may also lead to biased results (Breslow and Clayton, 1993). The standard errors obtained from the JM method are smaller than that from the NTS method. The estimates of the variance parameter τ are a bit different across the two methods.

The estimates of the main parameter β_1 from the two methods are quite different. For the NTS method, β_1 is significantly positive, which suggests higher mean count of CD4 cells is associated with later time to dropout. For the JM method, β_1 is not significant, which means that the mean count of CD4 cells has no significant effect on the time to dropout. The conclusion from the JM method should be the most reliable, as will be demonstrated from the simulation study in the next section.

Table 4.1: Summary of results from data analysis

Parameter			NTS	JM
	α_0	Estimate	0.82	0.81
		(S.E.)	(0.059)	(0.047)
Longitudinal Model	α_1	Estimate	0.33	0.37
		(S.E.)	(0.12)	(0.049)
	τ[a]	Estimate	0.15	0.29
	β_0	Estimate	-6.38	-1.00
		(S.E.)	(0.78)	(0.52)
Survival Model	$\boldsymbol{\beta_1}$	**Estimate**	**5.63**	**0.22**
		(S.E.)	**(0.82)**	**(0.54)**
	σ[b]	Estimate	0.63	0.99

Note:
[a] *τ is the standard deviation of the random effect in the GLMM.*
[b] *σ is the scale parameter in the AFT model.*

We can also see that the estimates of the intercept β_0 are different, the standard errors obtained from the JM method are smaller than that of the NTS method, and the estimates of the scale parameter σ are different across the two methods.

Note that the NTS method models the covariate process and the survival process separately, so bias of the estimated β_1 may arise if the longitudinal process and survival process influence each other. For the NTS method, the uncertainty of the estimation in the first step is not incorporated into the second step, so standard errors of $\hat{\beta}_1$ may be unreliable. However, the JM method models the longitudinal process and the survival process simultaneously, and makes inference based on the joint likelihood of all data, so it may produce less biased estimates and may produce more reliable standard errors than the NTS method. To further compare and evaluate the performances of difference methods, a simulation study is conducted in the next section.

4.3.2 A Simulation Study

Introduction

In this section, we conduct a simulation study to evaluate the performances of different joint inference methods for a Poisson GLMM and an AFT model. We compare the performances of the naive two-step (NTS) method, the modified two-step (MTS) method, and the joint inference or joint model (JM) method, based on the biases of the estimates and the coverage rates of the 95% confidence intervals under several scenarios. First, we describe the models used to simulate the data. Then, we describe how we design the simulation study, including the settings of the true parameter values. Finally, we compare results from different methods under different settings and then draw conclusions.

Simulation Design

The Models

We assume that the time-dependent covariate for patient i at time t_{ij}, denoted by z_{ij}, follows a Poisson distribution with a mean of μ_{ij} (i.e. $z_{ij} \sim Poi(\mu_{ij})$). Then, we generate the mean of the Poisson distribution μ_{ij} from the following GLMM:

$$\begin{aligned} &\eta_{ij} = \log(\mu_{ij}) = \alpha_{0i} + \alpha_1 t_{ij}, \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, n_i, \\ &\alpha_{0i} = \alpha_0 + a_i \end{aligned} \tag{4.6}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$ is a vector of fixed effects, and a_i is the random effect. We assume $a_i \sim_{i.i.d} N(0, \tau^2)$.

For survival data, we assume the following AFT model:

$$\log(T_i) = \beta_0 + \beta_1 \eta_i(t) + \sigma\epsilon_i, \quad i = 1, 2, \cdots, N. \tag{4.7}$$

where $\eta_i(t) = \log(\mu_i(t))$ is the linear predictor in the GLMM (4.6), β_1 is the parameter of primary interest, β_0 is the intercept, σ is a scale parameter and ϵ_i's are random errors. Different choices of the distributions for ϵ_i lead to different AFT models. Similar to Section 3.5.4 in Chapter 3, we assume the Gumbel distribution to ϵ_i, with survival function and hazard function given by (3.13).

Generating Survival Times

By assuming the AFT model (4.7) and the Gumbel distribution for ϵ_i, the survival time T_i follows a Weibull distribution with parameters $\lambda = e^{-(\beta_0+\beta_1\eta_i(t))/\sigma}$ and $\gamma = \frac{1}{\sigma}$. So the survival function of the survival time T_i is

$$\begin{aligned} S(t) &= \exp(-\lambda t^{\gamma}) \\ &= \exp(-e^{-(\beta_0+\beta_1\eta_i(t))/\sigma} t^{\frac{1}{\sigma}}). \end{aligned} \tag{4.8}$$

Since $\eta_i(t) = \log(\mu_i(t)) = \beta_{0i} + \beta_1 t$ (from Model (4.6)), $S(t)$ can be written as

$$S(t) = \exp(-e^{-\{\beta_0+\beta_1[\alpha_{0i}+\alpha_1 t]\}/\sigma} t^{\frac{1}{\sigma}}).$$

Similar to Section 3.5.4 of Chapter 3, we can generate the survival time by solving the equation $S(t) = U$, where U is a random number from a uniform distribution on the interval from 0 to 1 i.e., Unif$[0, 1]$. The detailed procedure to generate the survival time for individual i is:

Step 1 generating a random number u_i from the uniform distribution Unif$[0, 1]$;

Step 2 solving t from the equation

$$\exp(-e^{-\{\beta_0+\beta_1[\alpha_{0i}+\alpha_1 t]\}/\sigma} t^{\frac{1}{\sigma}}) - u_i = 0,$$

which is a survival time from the AFT model with survival function $S(t)$.

True Parameter Values

In the simulation study, the true values of most parameters are chosen based on the real data analysis in Section 4.3.1. The entire study period is set to be from 0 to 1. The vector of fixed effects $\boldsymbol{\alpha}$ is set to be $(0.7, 0.5)$.

The parameter of primary interest β_1, which measures the association between the survival time and the time-dependent covariate, is chosen to be $\beta_1 = 0.5$. The intercept and scale parameter in the AFT model are set to be $\beta_0 = -1$ and $\sigma = 1$ respectively. The censoring rate for the survival data is controlled to be around 20%. We only consider the situation where longitudinal data are not truncated at the event times. The number of patients N, the number of repeated measurement times for each individual n_i, and the variance of the random effect τ^2, are all set to have several different values for comparison (see their values in the simulation results).

We will apply three different joint inference methods to simulated datasets to evaluate their performances in terms of their biases and coverage rates of 95% confidence intervals for the parameters. The three joint inference methods are the JM method, the NTS method and the MTS method. In the MTS method, we run the bootstrap 100 times. The repetition times of the simulations are 100. Similar to Section 3.4 in Chapter 3, in the results, we show the "Estimate", "SE", "Sample SE", "Bias", and "Coverage" of the parameters for each method.

First, we simulate samples of size $N = 50$, with 11 repeated measurements for each individual in the samples, i.e. $n_i = 11$ for $i = 1, 1, \cdots, 50$, and the gap between two consecutive observing times is 0.1. The standard deviation of the random effect is set to be $\tau = 0.6$. Under this setting of parameters, we compare the performances of the three methods. Then, we change the true values of the sample size (N), the number of repeated measurement times (n_i), and the variance of the random effect (τ^2), to investigate the effects of N, n_i and τ^2 on the estimation performance.

Simulation Results

Comparison of Different Methods

Table 4.2 displays simulation results for all the population parameters in the models, i.e. α_0, α_1, β_0 and β_1, from different joint inference methods. For the fixed effects α_0 and α_1 in the GLMM, we can see that the three methods give similar results. The biases of the estimates from the three methods are similar, and they are all small. The coverage rates from the three methods are all close to the nominal level 95%. In summary, the three methods provide similar estimation performances for the longitudinal parameters α_0 and α_1.

Table 4.2: Simulation results ($N = 50$, $\tau = 0.6$, $n_i = 11$)

True Parameter		JM	NTS	MTS
	Estimate	0.663	0.683	0.683
	SE	0.100	0.099	0.100
$\alpha_0 = 0.7$	Sample SE[a]	0.102	0.092	0.092
	Bias	-0.037	-0.017	-0.017
	Coverage[b]	0.930	0.970	0.960
	Estimate	0.515	0.509	0.509
	SE	0.087	0.077	0.077
$\alpha_1 = 0.5$	Sample SE	0.078	0.067	0.067
	Bias	0.015	0.009	0.009
	Coverage	0.970	0.960	0.950
	Estimate	-0.827	-1.546	-1.546
	SE	0.308	0.273	0.267
$\beta_0 = -1$	Sample SE	0.333	0.283	0.283
	Bias	0.173	-0.546	-0.546
	Coverage	0.870	0.450	0.420
	Estimate	0.491	1.063	1.063
	SE	0.286	0.247	0.247
$\beta_1 = 0.5$	Sample SE	0.333	0.283	0.283
	Bias	-0.009	0.563	0.563
	Coverage	0.900	0.340	0.360

Note:
[a] *"Sample SE" is the empirical standard deviation of the parameter estimates.*
[b] *95% coverage rate.*

However, the results for the parameters β_0 and β_1 in the survival model are quite different from the three methods. We see that the JM method gives much smaller bias than the other methods. It is not surprising since the joint likelihood method makes simultaneous inference based on joint likelihood for all data, while the NTS method and the MTS method model the longitudinal data and the survival data separately, so biases may arise in these two methods when the longitudinal process and the survival process influence each other. The standard errors ("SE"s) from the JM method are larger than those from the NTS and MTS methods. This is expected because the JM method makes inference based on the joint likelihood which incorporates all the uncertainty in the longitudinal and survival data. On the other hand, the NTS does not incorporate the estimation uncertainty in the first step, so the "SE" s of the NTS method are likely to be underestimated.

Because of less biases and more reliable standard errors, the coverage rates of the joint likelihood estimates of β_0 and β_1, which are 87% and 90% respectively, are higher than those from the other methods. The reason why the coverage rates of the JM method are lower to the nominal level 95% may be due to the linear approximation to the GLMM. The NTS method and the MTS method have lower coverage rates because they produce larger biases and unreliable (smaller) standard errors.

Table 4.3 also shows the estimates for the standard deviation of the random effect (τ) and the scale parameter (σ). The estimates from the three methods are all quite close to their true values. So these methods do not seem to have much effect on the estimation of variance componenets τ and σ.

Table 4.3: Simulation results for the estimates of the variance parameters

True Parameter	JM	NTS	MTS
$\tau = 0.6$	0.607	0.598	0.598
$\sigma = 1$	1.136	1.015	1.015

Note: *τ is the standard deviation of the random effect in the GLMM and σ is the scale parameter of the AFT model.*

In the following, we compare the performances of the three methods in different scenarios. We mainly focus on the estimation of the main parameter β_1, since these methods give similar results for the parameters in the longitudinal covariate model and the results for β_0 are similar to the results for β_1.

Different Sample Size

In order to check how sample size affects parameter estimation, we simulate datasets with a larger number of subjects $N = 100$. The setting of the other parameters stays the same as the case in Table 4.2. The simulation results with $N = 100$ are shown in Table 4.4.

We can see that the JM method still produce less biased estimate and larger standard error for β_1, so the coverage rate of the JM method is much higher than the NTS method and the MTS method. Compared with results in Table 4.2, we find that larger sample size generally produce smaller "SE" for all the methods and lead to lower coverage rates for the methods except the JM method. The reason may be that larger sample sizes may lead to more accurate estimation of biases and standard errors, making the differences between the methods more obvious.

Table 4.4: Simulation results for estimating $\beta_1 = 0.5$ with a larger sample size $N = 100$ ($\tau = 0.6$, $n_i = 11$)

	JM	NTS	MTS
Estimate	0.468	1.065	1.065
SE	0.206	0.176	0.170
Sample SE	0.218	0.188	0.188
Bias	-0.032	0.565	0.565
Coverage	0.940	0.100	0.110

Different Number of Repeated Measurements

To investigate the influence of the number of repeated measurements within individuals on parameter estimation, we apply the three methods to simulated datasets with a smaller number of repeated measurements: $n_i = 6$ for all i, and the duration between two consecutive observing times is 0.2. The setting of the other parameters stays the same as the case in Table 4.2. The simulation results with $n_i = 6$ are shown in Table 4.5.

Similar to the previous results, the JM method still produces less biased estimate and larger (more reliable) standard error for β_1, and hence the coverage rate of the JM method is much higher than the NTS method and the MTS method. Compared with Table 4.2, results for less repeated measurements are associated with larger biases for all three methods. This is expected because these three joint inference methods require larger within-individual repeated measurements to get more accurate estimates since more repeated measurements imply more information about the longitudinal process. Moreover, the linearization method requires large repeated measurements to perform well.

We also notice that the coverage rate of the NTS method decrease substantially as the number of repeated measurements decrease which implies that the NTS method is more sensitive to the number of the within-individual repeated measurements than the the MTS method. The reason may lie in that the MTS at least adjusts the standard errors through bootstrapping, although it does not correct bias.

Table 4.5: Simulation results for estimating $\beta_1 = 0.5$ with sparse repeated measurements $n_i = 6$ ($N = 50$, $\tau = 0.6$)

	JM	NTS	MTS
Estimate	0.549	1.179	1.179
SE	0.301	0.262	0.272
Sample SE	0.314	0.315	0.315
Bias	0.049	0.679	0.679
Coverage	0.910	0.270	0.380

Different Magnitudes of Random Effect

We apply the three methods to simulated datasets with different magnitudes of variability of random effect to examine the influence of random effect on the results. We consider three magnitudes of standard deviations: $\tau = 0.4$, $\tau = 0.6$ and $\tau = 0.8$. The setting of the other parameters is the same as the case in Table 4.2. The results for the case $\tau = 0.6$ are already displayed in Table 4.2. Table 4.6 shows simulation results for cases $\tau = 0.4$ and $\tau = 0.8$.

From Tables 4.2 and 4.6, we see that the biases of the NTS method and the MTS method decrease substantially as the the variance of the random effect increases, so the coverage rates of these two methods increase. Note that the variation between individuals reflects the differences between individuals, while the variation within individual reflects changes of variables over time. When the variability of the random effect becomes large, the variation between individuals dominates the variation within individual repeated measurements. Thus, most of the variation in the longitudinal covariate can be explained by the differences between individuals rather than changes of variables over time. Therefore, the NTS method and the MTS method, which treat the longitudinal covariate to be time-independent in the second step, may perform better when the magnitudes of the random effect increase. We can also see that, when the variance of the random effect is small, the coverage rate is much larger for the MTS method than the NTS method, which may be due to the fact that the MTS at least adjusts the under-estimated standard errors.

Table 4.6: Simulation results for estimating $\beta_1 = 0.5$ ($N = 50$, $n_i = 11$)

Magnitudes of Random Effect		JM	NTS	MTS
	Estimate	0.443	1.638	1.638
	SE	0.455	0.353	0.387
$\tau = 0.4$	Sample SE	0.460	0.401	0.401
	Bias	-0.057	1.138	1.138
	Coverage	0.910	0.070	0.170
	Estimate	0.431	0.790	0.790
	SE	0.220	0.200	0.195
$\tau = 0.8$	Sample SE	0.240	0.221	0.221
	Bias	-0.069	0.290	0.290
	Coverage	0.930	0.690	0.680

Conclusions

From the simulation results, we find that the joint likelihood or the joint model method usually has the smallest bias and the largest coverage rate close to the nominal level (95%), compared to the other two methods. These results confirm that the method based on the joint likelihood produces less biased estimates and more reliable standard errors. It also indicates that the linear approximation to the Poisson GLMM works well in the computation of the joint likelihood. The naive two-step method and the modified two-step method have relatively larger biases and lower coverage rates since separately modeling the longitudinal process and the survival process may lead to biases.

The performances of the three methods may depend on sample size, within-individual measurements, and the magnitude of random effect. Larger sample size seems to lower the coverage rates of the two-step methods, but not the joint likelihood method. More within-individual repeated measurements lead to less biases in all three methods. Larger magnitude of the random effect lead to better performances of the naive two-step method and the modified two-step method.

4.4 Joint Inference for an AFT Model and a Binomial GLMM

In this section, we evaluate different joint inference methods for an AFT model and a GLMM with binomial (Bernoulli) distribution through real data analysis and a simulation study.

4.4.1 Data Analysis - A HIV Study

In Section 4.2, we have described joint inference methods for a GLMM model and an AFT model. In this section, we use the joint likelihood method and the naive two-step method to analyze a real datasets described in Section 1.4 of Chapter 1.

The HIV dataset contains 46 patients' CD4 cell counts and their times to dropout. In HIV studies, CD4 cell count is an important index and it is standard to set 200 as a threshold for CD4 cell count. If the CD4 cell count for a patient is larger than 200, we say this patient has high CD4 cell count. Otherwise, if the CD4 cell count is lower than 200, we say the patient has low CD4 cell count. So, we can define a new binary variable "level of CD4" k_{ij},

$$k_{ij} = \begin{cases} 1, & \text{CD4 cell count } y_{ij} \geq 200; \\ 0, & \text{CD4 cell count } y_{ij} < 200. \end{cases}$$

One may be interested in examining the relationship between the level of CD4 cell count and the time to dropout. More specifically, we are interested in checking whether patients with high level of CD4 cell counts have earlier times to dropout. We consider the data within the first 90 days after the anti-HIV treatment. A summary of the variables of interest in this analysis is shown in Table 1.1.

The Models

Since the new binary variable k_{ij} changes over time, we consider a GLMM by assuming the levels of CD4 cell count follow a Bernoulli distribution. That is, we assume the level of CD4 cell count for patient i at time t_{ij}, k_{ij}, follows a Bernoulli distribution with probability of p_{ij} (i.e. $k_{ij} \sim Bernoulli(p_{ij})$). We use the standard model selection procedure based on the AIC and BIC values to find an appropriate random effects specification and choose the following GLMM to explain the change of p_{ij}

$$\begin{aligned} &\eta_{ij} = \mathrm{logit}(p_{ij}) = \alpha_{0i} + \alpha_{1i} t_{ij}, \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, n_i, \quad (4.9) \\ &\alpha_{0i} = \alpha_0 + a_{0i} \quad \alpha_{1i} = \alpha_1 + a_{1i} \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$ is a vector of fixed effects, and $\boldsymbol{a_i} = (a_{0i}, a_{1i})$ is a vector of random effects. We assume $\boldsymbol{a_i} \sim_{i.i.d} N(\mathbf{0}, \Sigma)$.

Our objective is to access the association between the level of CD4 cell count and the time to dropout. We consider a parametric AFT model which links the dropout time (T_i) to the probability of the Bernoulli distribution rather than the level of CD4 cell counts. The AFT model is:

$$\log(T_i) = \beta_0 + \beta_1 \eta_i(t) + \sigma\epsilon_i, \quad i = 1, 2, \cdots, N, \qquad (4.10)$$

where $\eta_i(t) = \mathrm{logit}(p_i(t))$ is the linear predictor in the GLMM (4.9), β_1 is the parameter of primary interest, β_0 is the intercept, σ is a scale parameter and ϵ_i's are random errors. We assume a Gumbel distribution for ϵ_i, so the survival time T_i follows a Weibull distribution.

Data Analysis Results

The effect of the CD4 level on the time to dropout can be interpreted based on the inference on the parameter β_1. We apply the NTS method and the JM method to analyze the data.

The results from different methods are displayed in Table 4.7. We show the estimates and standard errors of the fixed effects α_0, α_1 in the GLMM (4.9), the intercept in AFT model β_0 and the primary parameter β_1. We also show the estimates for the standard deviations of the random effects and the scale parameter in the AFT model.

Although the two methods give similar estimates for the fixed effects in the longitudinal model, their standard errors and significances at 5% level from the two methods are different: for the NTS method, α_0 is not significant and α_1 is significantly positive; for the JM method, α_0 and α_1 both are significantly positive. The reason may be that biases arise in the NTS method since the longitudinal process and survival process influence each other. Another possible reason is that the linear approximation to the GLMM with a binary response may not work well in the JM method.

The estimates of the main parameter β_1 are quite different across these two methods. For the NTS method, β_1 is significantly positive, which means that patients with high CD4 level have longer times to dropout. For the JM method, β_1 is not significant, which means that the CD4 level has no significant effect on the time to dropout. These conclusions are consistent with the data analysis results in Section 4.3.1.

We can also see that the estimates of the intercept β_0 are different, the standard errors obtained from the JM method are larger than that from the NTS method, and the estimates of the scale parameter σ are different across the two methods. To further compare and evaluate the performances of difference joint inference methods, a simulation study is conducted in the next section.

Table 4.7: Summary of results from data analysis

Parameter			NTS	JM
Longitudinal Model	α_0	Estimate	0.898	0.782
		(S.E.)	(0.518)	(0.275)
	α_1	Estimate	7.256	7.456
		(S.E.)	(2.158)	(0.662)
	σ_{11}[a]	Estimate	2.713	2.566
	σ_{22}[b]	Estimate	8.024	4.550
Survival Model	β_0	Estimate	-1.321	-0.948
		(S.E.)	(0.156)	(0.172)
	$\boldsymbol{\beta_1}$	**Estimate**	**0.117**	**0.0484**
		(S.E.)	**(0.026)**	**(0.0314)**
	σ	Estimate	0.824	1.068

Note:

[a] σ_{11} *is the standard deviation of random effect* a_{0i}.

[b] σ_{22} *is the standard deviation of random effect* a_{1i}.

4.4.2 A Simulation Study

Introduction

In this section, we conduct a simulation study to evaluate the performances of different joint inference methods for jointly analyzing a binomial (Bernoulli) GLMM and an AFT model. We compare the performances of different methods based on the biases of the estimates, and the coverage rates of the 95% confidence intervals under several scenarios. First, we describe the models used to simulate the data. Then, we introduce the design of this simulation study, including the settings of true parameters. Finally, we compare results from different methods under different settings and then draw conclusions.

Simulation Design

The Models

We assume a time-dependent binary covariate for patient i at time t_{ij}, denoted by k_{ij}, follows a Bernoulli distribution with the probability of p_{ij} (i.e. $k_{ij} \sim Bernoulli(p_{ij})$). Then, we generate the probability p_{ij} of the Bernoulli distribution from the following generalized linear mixed model:

$$\begin{aligned} &\eta_{ij} = \text{logit}(p_{ij}) = \alpha_{0i} + \alpha_1 t_{ij}, \quad i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, n_i, \quad (4.11) \\ &\alpha_{0i} = \alpha_0 + a_i \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$ is a vector of fixed effects, and a_i is the random effect. We assume $a_i \sim_{i.i.d} N(0, \tau^2)$.

For the survival data, we assume the following AFT model:

$$\log(T_i) = \beta_0 + \beta_1 \eta_i(t) + \sigma \epsilon_i, \quad i = 1, 2, \cdots, N, \tag{4.12}$$

where $\eta_i(t) = \text{logit}(p_i(t))$, β_0 is the intercept, β_1 is the parameter of primary interest, σ is a scale parameter and ϵ_i's are random errors. We assume that the distribution of ϵ_i is the Gumbel distribution.

Generating Survival Times

The procedure to generate survival times is the same as the procedure in Section (4.3.2). By assuming the AFT model 4.12 and the Gumbel distribution for the random errors, the survival function of the survival time T_i is given by

$$\begin{aligned} S(t) &= \exp(-e^{-(\beta_0 + \beta_1 \eta_i(t))/\sigma} t^{\frac{1}{\sigma}}) \\ &= \exp(-e^{-(\beta_0 + \beta_1 [\alpha_{0i} + \alpha_1 t])/\sigma} t^{\frac{1}{\sigma}}). \end{aligned} \tag{4.13}$$

Then, we can generate the survival time t by solving the equation $S(t) = U$, where U is a random number from a uniform distribution Unif$[0, 1]$. The detailed procedure to generate the time to an event for patient i is:

Step 1 generating a random number u_i from the uniform distribution $Unif[0,1]$;

Step 2 solving t from the equation

$$\exp(-e^{-(\beta_0+\beta_1[\alpha_{0i}+\alpha_1 t])/\sigma} t^{\frac{1}{\sigma}}) - u_i = 0.$$

True Parameter Values

In the simulation study, the true values of most parameters are chosen based on the real data analysis in Section 4.4.1. The entire study period is set to be from 0 to 1. The vector of fixed effects $\boldsymbol{\alpha}$ is set to be $(0.4, 5)$.

The parameter of primary interest β_1, which measures the association between the survival time and the time-dependent covariate, is chosen to be $\beta_1 = 0.07$. The intercept and scale parameter in the AFT model are set to be $\beta_0 = -0.6$ and $\sigma = 1$ respectively. The censoring rate for the survival data is controlled to be around 20%. We only consider the situation where longitudinal data are not truncated at the event times. The number of patients N, the number of repeated measurements for each individual n_i, and the variance of the random effect τ^2, are all set to have several different values for comparison (see their values in the simulation results).

We will apply three different joint inference methods to the simulated datasets to evaluate their performances in terms of biases and coverage rates of 95% confidence intervals for the parameters. The three methods are the joint likelihood or joint model (JM) method, the naive two-step (NTS) method and the modified two-step (MTS) method. In the MTS method, we run the bootstrap 100 times. The repetition times of the simulations are 100. In the results, we show “Estimate”, “SE”, “Sample SE”, “Bias”, and “Coverage” of the parameters for each method.

First, we simulate samples of size 50 ($N = 50$), with 11 repeated measurements for each individual (i.e. $n_i = 11$ for $i = 1, 2, \cdots, 50$), and the gap between two consecutive measurement times is 0.1. The standard deviation of the random effect is set to be $\tau = 6$. Under this setting, we compare the performances of the three joint inference methods.

Then, we change the true values of sample size (N), the number of repeated measurements within individuals (n_i), and the variance of the random effect (τ^2) to investigate the effects of N, n_i and τ^2 on the resulting estimates.

Simulation Results

Comparison of Different Methods

Table 4.8 displays simulation results for the population parameters in the models, i.e. α_0, α_1, β_0 and β_1, and Table 4.10 shows the estimates for the standard deviation τ of the random effect in the GLMM and the scale parameter in the AFT model. For the fixed effect α_0 in the GLMM (4.11), these three methods give similar coverage rates. For the fixed effect α_1, the bias in the JM method is larger than that in the NTS and the MTS methods, and hence the coverage rate in the JM method is much lower than the other two methods. Also, the JM method provides more biased estimate for τ than the two-step methods. Therefore, we may say that the JM method performances worse than the NTS and the MTS methods when estimating the parameters in the longitudinal GLMM. The reason may be that the linear approximation may not work well to the GLMM with a binary response, while in the two-step methods, the GLMM is fitted using R function *glmm*(). This function uses Gaussian-Hermite numerical integration method to approximate the likelihood (2.5), which can be made arbitrary accurate by increasing the number of quadrature points.

However, for the parameters β_0 and β_1 in the AFT model, the JM method performances better than the two-step methods. We see that the JM method gives much smaller bias than the two-step methods. It is not surprising since the joint likelihood method makes simultaneous inference based on joint likelihood for all data, while the NTS and MTS methods model the longitudinal data and the survival data separately, so biases may arise in these two methods when the longitudinal process and the survival process influence each other. The standard errors ("SE"s) from the JM method and the MTS method are larger than that from the NTS method. This is expected because the JM method makes inference based on the joint likelihood which incorporates all the uncertainty and the MTS method adjusts the standard errors by incorporating the estimation uncertainty in the first step through bootstrapping. On the other hand, the NTS does not incorporate the estimation uncertainty in the first step, so the "SE" s of the NTS method are likely to be underestimated.

The above estimation results show that the linearization for GLMM in the JM method only affects estimates of the GLMM but it does not affect estimates of the AFT model. Because of less biases and more reliable standard

Table 4.8: Simulation results ($N = 50$, $\tau = 6$, $n_i = 11$)

True parameter		JM	NTS	MTS
$\alpha_0 = 0.4$	Estimate	0.312	1.949	1.949
	SE	0.721	1.127	2.196
	Sample SE	1.216	2.430	2.430
	Bias	-0.088	1.549	1.549
	Coverage	0.780	0.750	0.830
$\alpha_1 = 5$	Estimate	5.171	4.929	4.929
	SE	0.316	0.767	0.873
	Sample SE	1.067	0.974	0.974
	Bias	0.171	-0.071	-0.071
	Coverage	0.510	0.900	0.920
$\beta_0 = -0.6$	Estimate	-0.598	-0.932	-0.932
	SE	0.184	0.158	0.186
	Sample SE	0.034	0.038	0.038
	Bias	0.002	-0.332	-0.332
	Coverage	0.940	0.420	0.540
$\beta_1 = 0.07$	Estimate	0.054	0.136	0.136
	SE	0.033	0.029	0.042
	Sample SE	0.034	0.038	0.038
	Bias	-0.016	0.066	0.066
	Coverage	0.890	0.430	0.680

errors, the coverage rates of β_0 and β_1 in the JM method, which are 94% and 89% respectively, are higher than the two-step methods. The MTS method performances better than the NTS method in terms of coverage rate, since it adjusts the standard errors. The NTS method produces the lowest coverage rates among these three methods because they produce larger biases and underestimate standard errors.

In the following, we compare the performances of the three methods under different settings. We focus on the estimation of the main parameter β_1.

Different Sample Size

In order to check how sample size affects parameter estimation, we simulate datasets with a larger size of samples: $N = 100$. The setting of the other parameters stays the same as the case in Table 4.8. The simulation results with $N = 100$ are shown in Table 4.10.

Table 4.9: Simulation results for the estimates of variance parameters

True Parameter	JM	NTS	MTS
$\tau = 6$	4.88	6.55	6.55
$\sigma = 1$	1.06	1.11	1.11

Note: τ *is the standard deviation of the random effect in the GLMM and* σ *is the scale parameter of the AFT model.*

We can see that the JM method produces less biased estimate, so the coverage rate in the JM method is much higher than that in the NTS and MTS method. The MTS method performances better than the NTS method in terms of the coverage rate, since it adjusts the under-estimated standard error in the NTS method. Compared with results in Table 4.8, we find that larger sample size is associated with smaller "SE" for all the methods and lead to substantially lower coverage rates for the NTS method and the MTS method. The reason may be that larger sample sizes may lead to more accurate estimation of standard errors, making the differences between the methods more obvious.

Table 4.10: Simulation results for estimating $\beta_1 = 0.07$ with a larger sample size $N = 100$ ($\tau = 6$, $n_i = 11$)

	JM	NTS	MTS
Estimate	0.053	0.136	0.136
SE	0.024	0.020	0.032
Sample SE	0.026	0.025	0.025
Bias	-0.017	0.066	0.066
Coverage	0.86	0.14	0.46

Different Number of Repeated Measurements

To investigate the influence of the number of repeated measurements within individuals on parameter estimation, we apply all three methods to simulated datasets with a smaller number of repeated measurements 6 ($n_i = 6$ for all i, and the gap between two consecutive measurement times is 0.2). The setting of the other parameters stays the same as the case in Table 4.8. The simulation results with $n_i = 6$ are shown in Table 4.11.

We can see that the JM method still produces less biased estimate than the NTS method and the MTS method. The MTS method has the highest coverage rate since it gives the largest "SE" among these three methods through bootstrapping. Compared with Table 4.8, results for less repeated measurements are associated with larger bias and lower coverage rate for the JM method. Note that the linearization procedure in the JM method requires larger within-individual repeated measurements to perform well (Breslow and Lin, 1995). Interestingly, for the NTS method and the MTS method, their coverage rates seem to increase as the number of repeated measurements decreases. Note that, in the two-step methods, the GLMM is fitted using R function *glmm*() which does not require large within-individual repeated measurements, and the estimated SE's may increase with less data, and thus lead to higher coverage rates.

Table 4.11: Simulation results for estimating $\beta_1 = 0.07$ with sparse repeated measurements $n_i = 6$ ($N = 50$, $\tau = 6$)

	JM	NTS	MTS
Estimate	0.042	0.132	0.132
SE	0.035	0.031	0.050
Sample SE	0.045	0.044	0.044
Bias	-0.028	0.062	0.062
Coverage	0.74	0.50	0.77

Different Magnitude of Random Effect

We apply the methods to simulated datasets with a smaller variability of random effect, which is $\tau = 4$, to examine the influence of random effect on the results. The setting of the other parameters is the same as the case in Table 4.8. The simulation results with $\tau = 4$ are shown in Table 4.12.

Compared with Table 4.8, we see that a smaller variation of random effect is associated with larger biases, especially for the NTS method and the MTS method. The coverage rates for these two methods decrease substantially as the variation of random effect decreases. This is expected and is consistent with the results in Table 4.6. When the variability of the random effect becomes small, most of the variation in the longitudinal covariate can be explained by changes of variables over time rather than the differences between individuals. So, the NTS method and the MTS method, which treat the longitudinal covariate to be time-independent in the second step, would performance worse when the magnitudes of the random effect decrease.

Table 4.12: Simulation results for estimating $\beta_1 = 0.07$ with a smaller variability of random effect $\tau = 4$ ($N = 50$, $n_i = 11$)

	JM	NTS	MTS
Estimate	0.047	0.177	0.177
SE	0.046	0.035	0.050
Sample SE	0.049	0.040	0.040
Bias	-0.023	0.107	0.107
Coverage	0.92	0.13	0.37

Conclusions

From the simulation results, we find that the joint likelihood or the joint model method usually has the smallest bias and the largest coverage rate, compared to the two-step methods. These results confirm that joint likelihood method produces less biased estimates and more reliable standard errors. The fact that the coverage rates in the joint likelihood method are usually much lower than the nominal level 95% may imply that the linear approximation may not work so well to the GLMM with a binary response in the computation of the joint likelihood. However, even with the linearization, the joint likelihood method still outperforms the two-step methods which do not use linearization. The naive two-step method and the modified two-step method have relatively larger biases. By adjusting the standard errors through bootstrapping, the modified two-step method performances better than the naive two-step method in terms of the coverage rate.

The performances of the methods may depend on the sample size, the number of repeated measurements within individuals, and the magnitude of random effects. Larger sample size seems to lower the coverage rates of the NTS and the MTS methods. Less repeated measurements are associated with larger bias and lower coverage rate for the JM method. Smaller variation of the random effect lead to worse performances of the naive two-step method and the modified two-step method.

Chapter 5

Conclusions and Future Research

In this thesis, we have considered joint models for a longitudinal process and a survival process which are related to each other. To model the longitudinal process, we consider linear mixed effects (LME) models, nonlinear mixed effects (NLME) models, and generalized linear mixed models (GLMMs). To model the survival process, we consider Cox models and accelerated failure time (AFT) models. We have discussed joint inference methods, including the naive two-step method, the modified two-step method, and the joint likelihood method. We have proposed linear approximation methods to joint models with GLMM and NLME submodels to reduce computation burden and use of existing software.

A real dataset from a HIV study is analyzed by different methods. We found that results from different methods may be very different. The joint likelihood method indicates that the viral load positively influences the hazard of the dropout time and the CD4 cell count has no significant effect on the time to dropout. On the other hand, the naive two-step method concludes that the viral load negatively affects the hazard of the time to dropout and the CD4 cell count positively affects the dropout time.

We have used simulation studies to evaluate which method is more reliable and we found that the joint likelihood method consistently performs the best. Thus, results based on the joint likelihood method should be most reliable. By adjusting the standard errors through bootstrapping, the modified two-step method may perform better than the naive two-step method. From the simulation study, we also find that the sample size, the number of within-individual repeated measurements, the magnitude of measurement errors and random effects, and the truncation of longitudinal data by events may influence the performance of different methods. The simulation results also indicate that the linear approximation to an nonlinear or generalized mixed effects model performs well for the joint likelihood method.

In the following, we briefly describe some possible topics for future research. First, in the survival model, we only include a time-dependent variable as a single covariate for simplicity in this thesis. It may be possible that both the current covariate value and the rate of change of the underlying covariate trajectories are associated with the survival time, or some covariates in the survival model may be functions of the unobserved random effects in the longitudinal model. Therefore, in future research, we may consider some other ways in which the longitudinal model and the survival model are linked. Secondly, we assume the longitudinal data is missing at random in this thesis. In future research, we may consider the situation where missing data in the longitudinal model are non-ignorable. For example, covariates may be missing due to drug side-effects. Thirdly, the survival models may be extended to more general cases such as multiple events or recurrent events and the longitudinal models may be extended to multivariate models such as multivariate linear mixed effects models.

References

[1] Albert, P.S. and Shih, J.H. (2009). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics*, 66(3), 983-987.

[2] Bender, R., Augustin, T., and Blettner, M. (2003). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, 24(11), 1713-1723.

[3] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9-25.

[4] Breslow, N. E. and Lin, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika*, 82, 81-91.

[5] Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of Royal Statistical Society, Series B*, 34, 187-200.

[6] Evans, M. and Swartz, T.B. (2000). *Approximating Integrals via Monte Carlo and Deterministic Methods.* Oxford University Press.

[7] Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics*, 46, 673-687.

[8] McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* New York: Wiley.

[9] Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4, 12-35.

[10] Song, X. and Wang, C.Y. (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics*, 64, 557-566.

[11] Tseng, Y.K., Hsieh, F., and Wang, J.L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92, 587-603.

[12] Tsiatis, A.A. and Davidian, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica*, 14, 793-818.

[13] Wu, H. and Ding, A. (1999). Population HIV-1 Dynamics in Vivo: Applicable Models and Inferential Tools for Virological Data from AIDS Clinical Trials. *Biometrics*, 55, 410-418.

[14] Wu, L., Hu, J., and Wu, H. (2008). Joint Inference for Nonlinear Mixed-Effects Models and Time-to-Event at the Presence of Missing Data. *Biostatistics*, 9, 308-320.

[15] Ye, W., Lin, X., and Taylor, J.M.G. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data - A two stage regression calibration approach. *Biometrics*, 64, 1238-1246.