Joint Inference for Longitudinal and Survival Data with Incomplete Time-dependent Covariates

by

Xu Wang

B.Sc., Zhejiang University, 2008

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 2010

Abstract

In many longitudinal studies, individual characteristics associated with their repeated measures may be covariates for the time to an event of interest. Thus, it is desirable to model both the survival process and the longitudinal process together. Statistical analysis may be complicated with missing data or measurement errors in the time-dependent covariates. This thesis considers a nonlinear mixed-effects model for the longitudinal process and the Cox proportional hazards model for the survival process. We provide a method based on the joint likelihood for nonignorable missing data, and we extend the method to the case of *time-dependent covariates*. We adapt a Monte Carlo EM algorithm to estimate the model parameters. We compare the method with the existing two-step method with some interesting findings. A real example from a recent HIV study is used as an illustration.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Lang Wu, for his excellent guidance and immense help during my study at Department of Statistics of the University of British Columbia. Without his support, expertise and patience, this thesis would not have been completed. Also, I would like to thank my second reader, Dr. Paul Gustafson, for his invaluable comments and suggestions on this thesis.

Further more, I would like to thank Dr. John Petkau for his invaluable advice on my consulting projects, which benefits me very much in the past and future. I would like to thank all the faculty and staff in Department of Statistics of the University of British Columbia for providing such a nice academic environment. I would like to thank all graduate students in the Department of Statistics for making my study and life here so enjoyable.

Most importantly, I would like to thank my parents for their great love. I also want to thank my girlfriend, Ying, for the happiness she brings to me. It is their love, constant support and encouragement that push me to be the best at everything I do.

To my parents and Ying.

Chapter 1

Introduction

1.1 Background

In many longitudinal studies, both the longitudinal process and the survival process are of interest. For example, in HIV studies, while the HIV viral load dynamics in the early period after an anti-HIV treatment are of our interest, we are also interested in the relationship between the individual-specific characteristics of the viral load process in the early period and a long term antiviral response such as the time to a viral load rebound (or a viral load suppression or death). Specifically, one such important question is to check if patients with a faster initial viral load decay rate may have an earlier viral load rebound later in the study. For longitudinal data analysis, nonlinear mixed-effects (NLME) models are often used in many cases because these models are based on underlying mechanisms which generate the observed data (Davidian and Giltinan, 1995). For survival data, the Cox proportional hazards model is often of research interest. Also in such studies, missing data are common since individuals may drop out early for various reasons such as a drug resistance. The missing data may be informative in the sense that the missing data mechanism may be related to unobserved values such as the viral load value at that time point or the initial unobservable true viral load decay rates. Also, measurement errors often appear. For example, in HIV studies, CD4 cell count, which is measured repeatedly on the same individual during the study period, may be hard to be measured accurately. Thus, the analysis of longitudinal data often involves

methods for missing data and measurement errors.

When the longitudinal process and the survival process may be related via observed variable(s) or latent variable(s), making inference based on information provided by both two process may be helpful. Methods jointly modeling longitudinal data and survival data have been studied in the literature (e.g., DeGruttola and Tu, 1994; Wulfsohn and Tsiatis 1997; Henderson et al. 2002; Guo and Carlin, 2004). Tsiatis and Davidian (2004) provides a very nice review. Wu (2008) discussed a joint model with informative missing data using baseline covariate information for statistical inference. In this thesis, we consider a joint likelihood method to jointly model longitudinal data and survival data, incorporating missing information and measurement errors using *time-dependent covariates*.

1.2 Longitudinal Data

1.2.1 Longitudinal Studies

Longitudinal studies involve repeated observations of the same individual over a long time period. Longitudinal studies are often called panel studies in economics and sociology. In longitudinal studies, individuals are followed over a period of time. For each individual, data are collected at multiple time points. These collected data are called longitudinal data, which is very common in observational studies. These repeated measurements of a variable on the same individual over time is the defining feature of longitudinal studies. These repeated measurements of a variable on the same individual may share a common characteristic and may be correlated, although measurements on different individuals could be assumed to be independent. The measurement correlation within each individual reflects the key characteristic of longitudinal data. For example, in HIV studies, the viral load of each patient is measured repeatedly over time. The viral load values of one specified patient at different time points could be correlated due to the health status of this patient.

Longitudinal studies are often compared with cross-sectional studies. Both longitudinal studies and cross-sectional studies are observational studies. The fundamental difference between cross-sectional studies and longitudinal studies is that cross-sectional studies take place at a single time point and longitudinal studies involve a series of measurements taken over a period of time. An important assumption for cross-sectional data is that all observations in the sample are independent with each other. However, in longitudinal studies, the repeated observations on the same individual are usually correlated, although observations from different individuals are regarded to be independent. Hence, applying the classical statistical methods for cross-sectional data to longitudinal data would ignore the correlation in the measurements within each individual. Longitudinal studies have an advantage over cross-sectional studies in that longitudinal studies take the correlation in measurements within the same individual into account.

Longitudinal studies are also often compared with time series analysis. Time series takes the measurement correlation within the same individual into account as well. It observes a single long series of measurements over time. When there is only one individual included in the longitudinal study, longitudinal data is reduced to a single time series. In most time series studies, only one single series is available to be used to find clues and draw conclusions. Longitudinal studies have advantage over time series analysis in that the analysis of longitudinal data can be made by borrowing information across different individuals.

1.2.2 Approaches to Longitudinal Data Analysis

For longitudinal data, there may be substantial variations in both between and within individual measurements. A main objective of statistical models for longitudinal data analysis is to address these two sources of variations. One can study variable change of a certain subject over time via modeling the within-individual variation, while one can investigate differences between individuals via modeling the between-individual variation.

Before introducing commonly used approaches to longitudinal studies, we first

define notations that will be used. Let y_{ij} be a response variable and $\mathbf{x}_{ij}$ be a $p \times 1$ vector of p explanatory variables for the jth measurement on individual i at time point t_{ij}, $i = 1,...,N$, $j = 1,...,n_i$, where N is the total number of individuals involved in the study, and n_i is the number of repeated measures for individual i. The set of repeated measurements for individual i are collected into an $n_i \times 1$ vector, $\mathbf{y}_i = (y_{i1},...,y_{in_i})^T$. The covariate matrix for individual i is denoted as $\mathbf{X}_i = (\mathbf{x}_1,...,\mathbf{x}_{n_i})^T$, an $n_i \times p$ matrix.

Three approaches are usually used in longitudinal studies. The first one is often called generalized estimating equations (GEE models) or marginal models. GEE models, which were introduced by Liang and Zeger (1986), specify the mean structure and the correlation structure separately without distributional assumptions. The primary scientific objective of GEE models is to model the mean of the response variable. The correlation structure of the response variable may be specified based on the nature of the observed data or based on simplicity, not necessarily based on any parametric distributions. Thus, GEE models could be useful when the distributional assumptions are questionable, for example, when the response variable is binary or discrete.

A transition model is another approach to longitudinal studies. A transition model specifies the measurement correlation within an individual via Markov structures. That is, one models the conditional distribution of y_{ij} given the past measurements, $y_{i,j-1},...,y_{i,1}$.

The third approach is mixed effects models (or random effects models). Mixed effects models explain the between-individual variation and the within-individual correlation by introducing random effects. In mixed effects models, the conditional expectation of y_{ij} given the individual-specific coefficient, β_i is

$$E(y_{ij}|\beta_i) = f(\mathbf{x}_{ij}, \beta_i),$$

where $f(\cdot)$ is a link function, which explains the relationship between the response variable and the explanatory variables. In practice, usually there are not enough

measurements observed for an individual, thus an efficient estimation of the regression coefficients β_i is not valid, especially when the link function has a complex form. For example, a nonlinear model. Hence, the β_i's are further assumed to be independent from some distribution with a mean of β. Then, we can write $\beta_i = \beta + \mathbf{b}_i$, where β is fixed and $\mathbf{b}_i$ is a vector of random variables with mean 0. In this way, the individual characteristics can be represented by random effects $\mathbf{b}_i$. All repeated measures of a response variable for a specified individual share a common unobserved random effect $\mathbf{b}_i$, and these responses are correlated via this common factor $\mathbf{b}_i$, although $\mathbf{b}_i$'s varies across different individuals. Mixed effects models focus on both the population parameters β and the individual characteristics $\mathbf{b}_i$'s. Hence, mixed effects models are particulary useful when inferences need to be made about both population behaviors and individual trajectories, like in HIV studies. Because of the advantage of mixed effects models in HIV studies, we will focus on mixed effects models in this thesis.

1.3 Survival Data

1.3.1 Survival Studies

In many medical studies, the time to an event is often of interest. Common time to events of interest includes the time to death, the time to drop out from a follow-up study, the time to an efficacy loss of a medical treatment, etc. These types of data are called survival data. The analysis of survival data is called survival analysis. In HIV studies, for example, the viral load of a patient would rebound some time after receiving an anti-HIV treatment. This viral load rebound may be due to the loss of the efficacy of an anti-HIV treatment. Hence, it is of interest to find possible relationship between the rebound time and covariates such as the level of CD4 cell count and the personal characteristics which would be represented by random effects if a mixed effects model is used.

Survival data, which usually refers to the time to a certain event of interest, has its own features that makes it different from other types of data. First, the

distribution of survival data is usually not symmetric and usually skewed to right. Hence, survival data may not be reasonably assumed to have a normal distribution. The second feature is that survival data are often censored. That is, the event of interest may not be observed for some individuals during the study period. The censored data may possibly be due to the dropouts of individuals, loss of follow-up, or early termination of the study. For instance, in HIV studies, due to the limited follow-up time period, the time to a viral load rebound for an individual may be censored. Therefore, at the end of the follow-up study, one could only know whether the viral load had rebounded but could know nothing for the future. With these unique features of survival data, special statistical analysis procedures are required.

1.3.2 Approaches to Survival Data Analysis

Due to the features of survival data, nonparametric and semiparametric models are popularly used, since no distributional assumptions for the survival data are made in these models. Parametric models are also valid for survival data, and they are more efficient than nonparametric or semiparametric models if the distributional assumptions hold. Fleming and Harrington (1991), Andersen er al. (1993), Collett (2003), Lawless (2003), and Wu (2009) give comprehensive discussions of survival models and methodologies. In this thesis, we focus on the Cox proportional hazards model, which are particularly popular in survival analysis.

In survival analysis, the survival function and the hazard function play an important role. Let random variable T be the time to an event of interest, called survival time. The survival function is the probability that an individual survives to some time beyond time t,

$$S(t) = P(T \geq t) = 1 - F(t), \qquad t > 0,$$

where $F(t) = P(T < t)$ is the cumulative distribution function of T. The hazard function is defined as

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

which is the risk or hazard of an event at time t. It means the probability for an individual to experience an event immediately after time t, given the individual has experienced no event or survived to time t.

In survival regression models, finding any possible relationship between the survival time and the covariates of interest is the main goal. A popular approach is to model the hazard (or risk) of an event, rather than the mean of the response as in a classical regression analysis. The hazard function could be modeled in a nonparametric way in that the hazard functions may be complicated and the distribution assumption could be avoided using nonparametric models. Then, the hazard function and the covariates $\mathbf{x}_i$ can be linked via a usual linear predictor $\mathbf{x}_i^T\beta$. This leads to a semiparametric regression model. The Cox proportional hazards model (Cox, 1972) is a widely used semiparametric survival regression model. In the Cox proportional hazards model, the hazard function and the covariates are linked in the following form:

$$h_i(t) = h_0(t)exp(\mathbf{x}_i^T\beta),$$

where $h_0(t)$ is an unspecified baseline hazard function and other notations have the same meaning as before. The baseline hazard function $h_0(t)$ could be interpreted as the hazard when all covariates equals to 0. The Cox proportional hazards model assumes the hazard ratio $\frac{h_i(t)}{h_0(t)}$ is proportional to $exp(\mathbf{x}_i^T\beta)$, which needs checking in practice. It makes no distributional assumptions for the survival data, thus it is very flexible to use.

1.4 Missing Data Problems

1.4.1 Missing Data and Measurement Errors

In both longitudinal studies and survival studies, it is usually impossible to have complete information of interest collected. In regression models, the missing data can be missing in responses, missing in covariates, or missing in both responses and covariates. For example, in HIV studies, individuals may not be able to come to medical center for measurement at every scheduled time point due to various of reasons, or they may even dropout permanently from the study due to the drug

intolerance or death. Thus, the responses, the viral load, and the covariates such as CD4 cell count, are missing in the follow-up study. Missing data is an important issue in both longitudinal studies and survival studies. Ignoring missing data or using naive methods to deal with the missing data problem may lead to invalid inferences. Standard statistical methods are usually designed for complete data, and they cannot be directly applied to the case of missing values.

Measurement errors in covariates are another form of missing data and very common in practice. For example, in HIV studies, CD4 cell count, which is measured repeatedly on the same individual during the study period, may be hard to be measured accurately, possibly due to the imprecision of medical machines. In the presence of measurement errors, the observed data are not the true values but possibly measured with errors. If we treat these mis-measured values of data as true values, statistical analysis would not be appropriate. Particularly, in regression models, if the covariates are measured with errors but treated as accurately measured, the statistical inference will be misleading, for example, a significant covariate may be found to be non-significant. Hence, the measurement errors in covariates must be taken into account for valid inference. In this thesis, the measurement errors in the covariate is taken into account.

1.4.2 Classification of Missing Mechanisms

Missing data issues make the statistical analysis for longitudinal studies and survival studies more complicated. It is important to determine the reason for the missing data mechanism (the missingness) because a valid statistical method to deal with the missing data depends on the type of missingness.

Little and Rubin (1987) and Little (1995) give a general treatment of statistical analysis with missing values. Little and Rubin (1987) classifies the missing value mechanisms into three categories: missing completely at random (MCAR), missing at random (MAR), and nonignorably or informative missing (NIM).

If the missingness is irrelevant either with the observed data or the missing

data, then the missingness is regarded as MCAR. For example, in HIV studies, if the dropout of a patient at a scheduled time point is simply because he/she forgot, the missingness at this time point could be regarded as MCAR.

Missing data are MAR if the missingness depends only on the observed data, but not on the missing data. For example, if a patient fails to come to a medical center at a scheduled because he/she is very old (suppose the age information is known at the beginning of a follow-up study), the missingness at this time point could be regarded as MAR.

Missing data are NIM if the missingness depends on missing data. NIM can further be categorized into two cases in the context of random effects models:

- The missingness depends on unobserved responses. For example, a patient fails to visit the medical center because he/she is in a very bad health state. The missingness is called outcome-based informative missingness (Little, 1995).
- The missingness depends on unknown random effects which may substantially affect the responses. For example, in HIV studies, the missingness depends on individual characteristics such as individual viral load decay rates. The missingness is called random-effect-based informative missingness (Little, 1995).

For NIM the missingness could also depend on the missing covariates if one considers a missing in covariate problem. When the missingness is NIM, the missing data mechanism must be taken into account in the likelihood inference (Little and Rubin, 1987).

1.4.3 Approaches to Missing Data Problem

Many methods have been developed to deal with the missing data problem. Simple methods, like the complete-case (CC) method, the last-value-carried-forward (LVCF) method and the mean imputation method, are widely used because they are very simple and easy to carry out; however, these simple methods may lead to

inefficient or biased results. For example, the CC method is probably the simplest method used for missing data problems. It simply discards all individuals with missing values. However, the CC method may lead biased results when the missing data is not missing completely at random. Since simple methods often lead inefficient and unbiased results, they are generally not recommended, especially when the missing rate is high. Formal methods, like likelihood inference using EM algorithms, single imputation methods with variance adjustments, multiple imputation methods, Bayesian methods are usually used for more appropriate analysis for missing data.

Little and Rubin (2002) provided an overview of missing data methods, Carroll, Ruppert, and Stefanski (2006) reviewed common methods for measurement errors, and Maronna, Martin, and Yohai (2006) discussed recent development of robust methods. Wu (2009) gives a comprehensive review of incomplete data problem in mixed effects models. In longitudinal studies, mixed effects models are widely used, and missing data problem is very common. Because the maximum likelihood method is a standard statistical inference approach for mixed effects models, in this thesis, we will use likelihood-based methods for missing data problem.

1.5 Joint Modeling

1.5.1 Motivation

In many longitudinal studies, both the longitudinal process and the survial process are of interest. For example, in HIV studies, we are interested in both the HIV viral load dynamics in the early period after an anti-HIV treatment and the long term antiviral responses such as the time to a viral load rebound. The time to a viral load rebound may possibly be related with individual characteristics of the viral load process in the longitudinal process. That is, it is of interest to check if patients with faster initial viral load decay rate may have earlier viral load rebound in the later period of the study. Also, in the analysis of survival data using time-dependent covariates with measurement errors, we may need to model the longitudinal covariate process, which is used to address the measurement errors, in addition to a survival

model. In both examples, the longitudinal process and the survival process may be related. To better understand this relationship between the two processes and to make inference based on information provided by both processes, joint modeling of the longitudinal process and the survival process is needed.

1.5.2 Approaches to Joint Modeling

The longitudinal model and the survival model are often viewed as shared parameter models since the two models are usually linked through some common unknown variables. To make statistical inference simultaneously, a simple two-step method (TS) or a two-stage method is often used. In the first step, it estimates the common unknown variables or parameters based on one model using the observed data in the second step, it estimates parameters in the other model separately, with common latent variables or unknown parameters substituted by their estimates from the first step as if the estimated values of latent variables or the unknown parameters were observed values. The two-step method is simple, and statistical softwares can be readily used, However, the simple two-step method may lead inappropriate results when the longitudinal process and the survival process are strongly associated (Tsiatis and Davidian, 2004). Also, by the simple two-step method, the uncertainty of estimation in the first step can not be incorporated in the second step.

Another approach is the joint likelihood method, where the statistical inference for the joint model is based on the joint likelihood of all observed data. Wu (2009) gives a nice review on the joint likelihood method. Joint likelihood is very appealing because it provides a valid and reliable inference and standard likelihood theory could be used. Maximum likelihood estimations of all model parameters can be obtained simultaneously by maximizing the joint likelihood. However, there are two issues related with joint likelihood methods. First, when the joint modeling contains several longitudinal process, like the HIV viral load dynamics process and the process of time-dependent covariates, there will be too many unknown parameters in the joint model, thus, the models or the parameters may possibly be non-identifiable. Another issue is that joint modeling may require high-dimensional

and intractable integrals, so the computation could be quite intensive.

1.6 A Motivating Example

Our research is motivated from HIV studies. In HIV studies, we are often interested in modeling viral load dynamics in the early period after an anti-HIV treatment. In the meantime, we are also interested in the relationship between the individual-specific characteristics of the viral load process in the early period and a long term antiviral response such as the time to a viral load rebound.

In HIV studies, a patient's viral load after an anti-HIV treatment will typically decline in the early period. Late in the follow-up period, the patient may experience a viral rebound, possibly due to an emergence of drug resistance. Some patients may even drop out before the termination of the study for various reasons such as a bad health status. NLME models have been used for modeling HIV viral load dynamics in the early period after an anti-HIV treatment, and the covariates may be used to partially explain large inter-patient variations (Wu and Ding, 1999; Wu, 2005). Ding and Wu (2001) show that some viral load dynamic parameters, such as the initial viral decay rate, may reflect the effcacy/potency of an anti-HIV treatment. It is therefore important to study if some patient-specific viral load dynamic parameters are predictive for a long term antiviral response such as the time to a viral load rebound.

1.7 Objective and Outline

In this thesis, we consider a joint likelihood method for a NLME model and a Cox proportional hazard model with informative dropouts in the response and missing or mismeasured information in covariates. By the joint likelihood method we can estimate all model parameters simultaneously. The missing responses in the NLME model are allowed to be nonignorable, which is associated with personal characteristics, while the missing covariates are assumed to be ignorable. The random effects in the NLME model, which represent individual-specific characteristics of the longitudinal process in the early period, are used as possible error-free "covari-

ates" for the proportional hazards model and for the missing response model. A Monte-Carlo EM algorithm is used for estimation. Joint modeling of longitudinal data and survival data has been studied in the literature (e.g., DeGruttola and Tu, 1994; Wulfsohn and Tsiatis 1997; Henderson et al. 2002; Guo and Carlin, 2004). Tsiatis and Davidian (2004) provides a very nice review. Wu (2008) discussed joint models with nonignorable missing data using the baseline covariate information for inference. In this thesis, we extend Wu's method to the case of *time-dependent covariate* when the covariate is measured with errors.

In Section 2, we describe the models for longitudinal data and survival data, as well as the model to describe the missing data mechanism and the model for the time-dependent covariates which are measured with errors. In Section 3, we describe the two step method for joint inference. In Section 4, we describe the joint likelihood method for simultaneous inference using a Monte-Carlo EM algorithm. In Section 5, we extend the method in Section 4 to the case of time-dependent covariates with measurement errors. A real example of HIV studies is presented in Section 6. We compare the different methods via a simulation study in Section 7. We conclude the thesis with discussions in Section 8.

Chapter 2

Statistical Models

2.1 Notation

Suppose that there are N individuals. Let y_{ij} be the response for individual i at time t_{ij}, $i = 1,...,N; j = 1,...,n_i$, and let $\mathbf{y}_i = (y_{i1},...,y_{in_i})^T$. Let $\mathbf{z}_i$ be the collection of time-independent covariates for individual i. We write $\mathbf{y}_i = (\mathbf{y}_{i,mis}, \mathbf{y}_{i,obs})^T$, where $\mathbf{y}_{i,mis}$ are a collection of missing responses and $\mathbf{y}_{i,obs}$ are a collection of observed responses, and similarly we write $\mathbf{z}_i = (\mathbf{z}_{i,mis}, \mathbf{z}_{i,obs})$. Let $\mathbf{s}_i = (s_{i1},...,s_{in_i})^T$ be a vector of missing response indicators such that $s_{ij} = 1$ if y_{ij} is missing, and $s_{ij} = 0$ if y_{ij} is observed. Let $\mathbf{r}_i = (r_{i1},...,r_{im_i})^T$ be the vector of "event" indicators for individual i. $r_{ij} = 1$ if an event has happened by time t_{ij}; $r_{ij} = 0$ if not. We assume that $r_{i1} = 0$ for all i, which means at the beginning of study, no subject experiences an event. For individual i, let T_i be the time to an event. Note that the exact event time T_i usually cannot be directly observed. However, if we observe no events at times t_{i1},...,$t_{i,k-1}$ (i.e., $r_{i1} = ... = r_{i,k-1} = 0$) but know that an event has occurred by time t_{ik} (i.e., $r_{ik} = 1$), we can conclude that the actual event time is between $t_{i,k-1}$ and t_{ik} (i.e., $t_{i,k-1} < T_i \leq t_{ik}$), $k = 1,2,...,m_i$. This type of event time data structure is referred to as interval censored event time (Lawless, 2003).

2.2 Nonlinear Mixed Effects Models

In many longitudinal studies, classical linear models are usually not appropriate, although linear models are widely used for their simplicity. In many cases, these linear models are empirical models, which means they only describe the observed data but cannot reveal the underlying mechanism of data generation. On the other hand, nonlinear models are often used in longitudinal studies when the underlying data generation mechanism can be explained by nonlinear models.

There many advantages of nonlinear models over linear models. In terms of model fitting, a nonlinear model is able to fit the observed data as well as its competing linear models but uses fewer parameters. In the aspect of interpretation, nonlinear models are often introduced based on the data generation mechanism, thus the parameters of these nonlinear models may have a natural physical meaning. Nonlinear models may also provide more reliable predictions, even outside the range of the observed data than linear models.

In HIV studies, nonlinear mixed effects models (NLMEs) are popular in that NLMEs can characterize the variation both between individuals and within an individual (Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1996). For the longitudinal process, we consider the following general NLME model which could be written as a hierarchical two-stage models (Davidian and Giltinan, 1995):

$$y_{ij} = g(t_{ij}, \beta_i) + e_{ij}, \qquad \mathbf{e}_i | \beta_i \sim N(\mathbf{0}, \sigma^2 I), \tag{2.1}$$

$$\beta_i = h(\mathbf{z}_i, \beta) + B_i \mathbf{b}_i, \qquad \mathbf{b}_i \text{ i.i.d } \sim N(\mathbf{0}, D), \qquad i = 1, ..., N; \qquad j = 1, ..., n_i, \tag{2.2}$$

where $g(\cdot)$ is a known nonlinear function, $\mathbf{e}_i = (e_{i1}, ..., e_{in_i})^T$ are random errors, $\beta_i = (\beta_{i1}, ..., \beta_{is})^T$ is a vector of individual-specific regression parameters, $\beta = (\beta_1, ..., \beta_s)^T$ is a vector of population parameters, $h(\cdot)$ is a s-dimensional vector-valued known function, $\mathbf{b}_i$ is an incidence matrix of $\mathbf{0}$'s and $\mathbf{1}$'s, $\mathbf{b}_i = (b_{i1}, ..., b_{is})^T$ is a vector of random effects and is independent of $\mathbf{e}_i$, σ^2 is the unknown within individual variance, I is the identity matrix, and D is a covariance matrix.

If there are no missing data, the probability density function for the responses $\mathbf{y}_i$ can be written as

$$f(\mathbf{y}_i|\mathbf{z}_i,\beta,\sigma,D) = \int f(\mathbf{y}_i|\mathbf{z}_i,\mathbf{b}_i,\beta,\sigma)f(\mathbf{b}_i|D)d\mathbf{b}_i. \tag{2.3}$$

Therefore, the likelihood function is

$$L(\beta,\sigma^2,D|\mathbf{y}) = \Pi_{i=1}^{N} \int f(\mathbf{y}_i|\mathbf{z}_i,\mathbf{b}_i,\beta,\sigma)f(\mathbf{b}_i|D)d\mathbf{b}_i. \tag{2.4}$$

The likelihood function is complex and generally has no closed-form expression. Thus, numerical method could be used to get exact likelihood calculations. The computation would be intensive when the dimension of random effects $\mathbf{b}_i$'s is high. Alternative methods like the Monte Carlo method and the approximate method (Lindstrom and Bates, 1990) could be considered for this intensive computation.

2.3 Covariate Models

2.3.1 Empirical Model for Time-dependent Covariate with Measurement Errors and Missing Data

Measurement errors and missing data in time-dependent covariates are very common in practice. For example, CD4 cell count is usually of interest in HIV studies. One can hardly make sure CD4 cell count could be measured at each scheduled time point for an individual because the individual may not come to the medical center every time due to various reasons. Also, CD4 cell count is often measured with errors, possibly due to the imprecision of medical machines or carelessness of physicians. Thus, it is important to model the covariate process in order to address measurement errors or missing data in the covariate.

Let z_{ikl} be the observed covariate value and z^*_{ikl} be the (unobservable) "true" value of covariate k for individual i at time u_{il}, $i = 1,...,N$; $k = 1,...,\nu$; $l = 1,...n_i$. We focus on the case where z^*_{ikl} is the current true covariate value. We allow missing data in the covariates. Let $\mathbf{z}_i = (\mathbf{z}_{i1}^T,...,\mathbf{z}_{in_i}^T)^T$, where $\mathbf{z}_{il} = (z_{i1l},...,z_{i\nu l})^T$, $l = 1,...,n_i$. Following Shah, Laird, and Schoenfeld (1997), we consider the fol-

lowing multivariate LME model to empirically describe the covariate processes

$$\mathbf{z}_{il} = U_{il}\alpha + V_{il}\mathbf{a_i} + \varepsilon_{il}, \qquad i = 1,...,N, l = 1,...,n_i, \tag{2.5}$$

where U_{il} and V_{il} are design matrices, α and $\mathbf{a}_i$ are unknown population (fixed-effects) and individual-specific (random-effects) parameter vectors, and ε_{il} are the random measurement errors for individual i at time u_{il}. Parameters in (2.5) may be regarded as nuisance parameters because they are not of our main interest.

Therefore, the true (unobservable) covariate values are assumed to be

$$\mathbf{z}^*_{il} = U_{il}\alpha + V_{il}\mathbf{a}_i.$$

We also assume that $\mathbf{a}_i$ i.i.d.$\sim N(\mathbf{0}, A)$, ε_{il} i.i.d. $\sim N(\mathbf{0}, R)$, and $\mathbf{a}_i$ and $\varepsilon_i = (\varepsilon^T_{i1}, ..., \varepsilon^T_{in_i})^T$ are independent, where A and R are unknown covariance matrices. We further assume that α and $\mathbf{a}_i$ are independent of $\mathbf{e}_i$ and $\mathbf{b}_i$ in the response model. Note that for commonly-used polynomial empirical LME models, we have $\mathbf{u}_{ik} = (1, u_{ik}, ..., u^{l-1}_{ik})^T$ and $\mathbf{v}_{ik} = (1, u_{ik}, ..., u^{r-1}_{ik})^T$.

To allow for missing data in time-dependent covariates, we recast model (2.5) in continuous time:

$$\mathbf{z}_i(t) = U_i(t)\alpha + V_i(t)\mathbf{a}_i + \varepsilon_i(t), \qquad i = 1,...,N,$$

where $\mathbf{z}_i(t)$, $U_i(t)$, and $\varepsilon_i(t)$ are the covariate values, design matrices, and measurement errors at time t respectively. At the response measurement time t_{ij}, the possibly unobserved "true" covariate values can be viewed as

$$\mathbf{z}^*_{ij} = U_{ij}\alpha + V_{ij}\mathbf{a}_i,$$

where $U_{ij} = U_i(t_{ij})$ and $V_{ij} = V_i(t_{ij})$.

Note that, without a clear understanding of the data generation mechanism for the covariates, we use an empirical model to describe the covariate process.

This empirical model only describes the observed data but cannot reveal the data generation mechanism in the covariates. We may also model the covariate process using empirical polynomial models with random effects, as Higgins et al. (1997) and Wu (2002). By standard model selection procedure, an empirical model for the covariate process can be selected in terms of AIC and BIC criteria.

2.3.2 Model for Time-independent Covariate

When the covariates are time-independent, we consider a multivariate normal distribution to model the time-independent covariates (Little and Schlucher, 1985). For example, in longitudinal studies, possible covariates of interest like gender are time-independent. Sometimes, for simplicity, time-varying covariates are only considered for their baseline values (Lee, 2009), thus they can also be regarded as time-independent covariates. To allow for both continuous and categorical covariates, the multivariate normal model for the covariates, $\mathbf{z}_i = (z_{i1}, ..., z_{ip})$, can be written as a product of one-dimensional conditional distributions (Ibrahim et. al., 1999)

$$f(\mathbf{z}_i; \alpha) = f(z_{ip}|z_{i1}, ..., z_{i,p-1}; \alpha_p)...f(z_{i,1}; \alpha_1) \tag{2.6}$$

where $\alpha = (\alpha_1^T, ..., \alpha_p^T)^T$ are nuisance parameters for the conditional models, and p are number of covariates.

2.4 Survival Model

The time to an event may possibly be related with individual characteristics, for the survival process, we assume that the distribution of T_i may depend on the random effects $\mathbf{b}_i$ which represent individual-specific longitudinal processes in the early period. For example, in HIV studies patients with a faster (or slower) viral load decay rate may be more likely to have an earlier viral load rebound, so the time to viral load rebound T_i may depend on the random effects associated with the viral load decay rates. Therefore, we consider a survival model for the distribution of T_i, which links the probability of the time to an event to the random effects $\mathbf{b}_i$ in the NLME model.

Let the survival function $S(t) = P(T > t)$ be the probability of the survival time T being larger than t. The hazard rate is

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

which means the probability of experiencing an event immediately given no event appears previously. Further we have

$$\begin{aligned} h(t) &= lim_{\Delta t \to 0} - \frac{S(t + \Delta t) - S(t)}{\Delta t S(t)} \\ &= \frac{S'(t)}{S(t)}. \end{aligned}$$

Therefore,

$$S(T) = exp(-\int_0^T h(t)dt).$$

The Cox proportional hazards model assumes the hazard proportional to covariates in an exponential form. In particular, we assume that the conditional hazard rate at time $T_i = t_i$ given the random effects $\mathbf{b}_i$ as follows

$$h(t_i | \mathbf{z}_i, \mathbf{b}_i) = h_0(t_i) exp(\gamma_1^T \mathbf{z}_i + \gamma_2^T \mathbf{b}_i), \tag{2.7}$$

where $h_0(t_i)$ is the baseline hazard function, γ_1 and γ_2 are unknown parameters linking the covariates $\mathbf{z}_i$, and random effects $\mathbf{b}_i$ to the conditional hazard rate, respectively. This assumption assumes the hazard is affected by both the covariate value and the individual characteristics.

Let

$$p_{ik} = P(\mathbf{r}_{ik} = 1 | \mathbf{r}_{il} = 0, 0 \leq l < k; \mathbf{z}_i, \mathbf{b}_i) \tag{2.8}$$

$$= 1 - P(T_i \geq t_{ik} | T_i \geq t_{i,k-1}, \mathbf{z}_i, \mathbf{b}_i) \tag{2.9}$$

$$= 1 - \frac{S(t_{ik})}{S(t_{i,k-1})}, \tag{2.10}$$

$$k = 1, 2, ..., n_i.$$

Then we have,

$$p_{ik} = 1 - exp[-exp(\gamma_{0k} + \gamma_1^T \mathbf{z}_i + \gamma_2^T \mathbf{b}_i)], \tag{2.11}$$

or,

$$log(-log(1 - p_{ik})) = \gamma_{0k} + \gamma_1^T \mathbf{z}_i + \gamma_2^T \mathbf{b}_i$$

where,

$$\gamma_{0k} = log \int_{t_{i,k-1}}^{t_{ik}} h_0(t)dt, \qquad k = 1, ..., max\{n_i\}, \qquad .$$

Let $\gamma_0 = (\gamma_{01}, ..., \gamma_{0max\{n_i\}})$ and $\gamma = (\gamma_0, \gamma_1, \gamma_2)$. The density for $\mathbf{r}_i$ can be written as

$$f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \gamma) = \Pi_{k=1}^{n_i} f(r_{ik}|r_{il} = 0, 0 \leq l < k; \mathbf{z}_i, \mathbf{b}_i, \gamma) \tag{2.12}$$

where,

$$f(r_{ik}|r_{il} = 0, 0 \leq l < k; \mathbf{z}_i, \mathbf{b}_i, \gamma) = p_{ik}^{r_{ik}}(1 - p_{ik})^{(1-r_{ik})}.$$

2.5 Model for Missing Data

When there are informative dropouts, the missing data mechanism must be taken into account for valid likelihood inference. We assume that the missing covariates are missing at random (or ignorable) in the sense that the missingness may be related to the observed data but not the missing values, so we do not need to specify a missing covariate mechanism. For the missing longitudinal responses; however, it is likely that the missingness may be nonignorable in the sense that the missingness may be related to unobserved values. For example, in HIV studies, patients with slower initial viral load decay after treatment may be more likely to dropout early or miss visits than those with faster initial viral load decay, so the missingness probability may be related to the individual-specific initial viral load decay rates. Thus, here we assume a missing longitudinal response model which allows the missing probability to possibly depend on the unobservable random ef-

fects $\mathbf{b}_i$. Such a missing data model is related to the shared-parameter models or random-effect-based dropouts (Wu and Carroll, 1988; DeGruttola and Tu, 1994; Little, 1995; Follmann and Wu, 1995; Ten Have et al., 1998). In other words, the missingness depends on both $\mathbf{y}_{mis,i}$ and $\mathbf{y}_{obs,i}$ through the random effects $\mathbf{b}_i$. For such missing responses, a model specifying the missing response mechanism must be incorporated in the likelihood inference. Note that the probability of missing responses at time t_{ij} may also depend on the missing status at the previous time point $t_{i,j-1}$.

Based on the above arguments, as an example, we may consider the following model for the missing responses:

$$logit(P(s_{ij} = 1|s_{i,j-1}, \mathbf{b}_i, \phi)) = \phi_0 + \phi_1 s_{i,j-1} + \phi_2^T \mathbf{b}_i, \tag{2.13}$$

$$f(\mathbf{s}_i|\mathbf{b}_i, \phi) = f(s_{i1}|\mathbf{b}_i, \phi)\Pi_{j=2}^{n_i} f(s_{ij}|s_{i,j-1}, \mathbf{b}_i, \phi), \tag{2.14}$$

where the parameters ϕ may be viewed as nuisance parameters and are usually not of inferential interest. More complicated missing data models may be assumed, but a too complicated missing data model may introduce too many nuisance parameters and may cause parameter identifiability problems.

Chapter 3

Two-Step Method

3.1 Simple Two-Step Method

In joint models of longitudinal data and survival data, the longitudinal model and the survival model are usually linked through some shared parameters or shared variables. For example, the following two cases often arise in practice:

- the response of a longitudinal model is a time-dependent covariate in the survival model, which often arises in survival analysis with measurement error or missing data in time-dependent covariates;
- the longitudinal model and the survival model share same parameters or random effects, which often arises in longitudinal analysis with dropouts, or when there is a latent process which governs both the longitudinal process and the survival process.

In both cases, a simple or naive two-step approach can be used. It is to first fit one model (often the secondary model) to the observed data separately, ignoring the other model, and then in the second step the shared parameters or random effects are substituted by their estimates from the first step. Then, one proceeds with the inference in a usual way as if the estimated parameters or random effects were observed data. This two-step method is closely related to the regression calibration method in measurement error literature. A major advantage of the simple or naive

two-step method is that it is simple, and standard softwares are available to use. However, such a simple or naive two-step method may lead to misleading results. In the following, we discuss the two-step method in more details.

3.2 Modified Two-Step Method

As pointed out by Ye, Lin, Taylor (2008) and Albert and Shih (2009), the simple two step method mentioned in the last section may lead to misleading results in two ways:

- (i) the covariate trajectories of subjects who experience an event (e.g., die or drop out) may be different from those who do not experience any event, so the estimation of the covariate model in the first step to all covariate data may be biased;
- (ii) inference in the second step that ignores the estimation uncertainty in the first step may lead to misleading results (e.g., under-estimating standard errors).

The bias in case (i), called bias from informative dropouts, may depend on the strength of the association between the longitudinal process and the survival process. The misleading results in case (ii) may depend on the magnitude of measurement errors in covariates. In the following, we consider a modified two-step method to address these issues.

In order to adjust the standard errors of parameter estimates in the survival model by incorporating the estimation uncertainty in the first step, we can consider a parametric bootstrap method as follows:

- Step 1: Generate covariate values based on the assumed covariate model, with unknown parameters substituted by their estimates;
- Step 2: Generate survival times from the fitted survival model;
- Step 3: For each generated bootstrap dataset from step 1 and step 2, fit the models using the two-step method and obtain new parameter estimates;

- Step 4: Repeat Step 1-3 B times (say, $B = 500$).

We can obtain the estimated standard errors for the fixed parameters from the sample covariance matrix across the B bootstrap datasets.

This modified method produces more reliable estimates of the standard errors than the naive two-step method, if the assumed models are correct. The modified two step method gets an advantages over the naive two step method in that it includes the uncertainty of latent parameters or latent variables (which are estimated in the first step). However, a limitation of this modified two-step method is that it can only deal with a dataset with no missing information. When the missing data problem appears, the previous two step method might give misleading results. In next chapters, we consider another approach based on the joint likelihood of observed data to address the issue of missing data in joint models.

Chapter 4

Joint Likelihood Inference with Time-independent Covariate

4.1 Introduction

In this chapter, we consider simultaneous likelihood inference for all parameters based on the joint likelihood of the observed data. We first focus on time-independent (or baseline) covariates. The extension to time-dependent covariates with measurement errors will be discussed in next chapter.

4.2 Joint Likelihood

We consider simultaneous likelihood inference for all parameters based on the joint likelihood of the observed data $\{(\mathbf{y}_{i,obs}, \mathbf{z}_{i,obs}, \mathbf{r}_i, \mathbf{s}_i), i = 1,2,...,N\}$. We consider time-independent (or baseline) covariates following Wu (2009). Let $f(\cdot)$ denote a generic density function and $f(y|x)$ denote the conditional distribution of y given x. Let $\theta = (\beta, \sigma, \gamma, \phi, D)$ denote the collection of all unknown parameters. We assume that $\mathbf{y}_i$ and $\mathbf{r}_i$ are conditionally independent given the random effects $\mathbf{b}_i$, i.e., $\mathbf{r}_i$ depends on $\mathbf{y}_i$ through the random effects $\mathbf{b}_i$. We also assume that $f(\mathbf{s}_i|\mathbf{y}_i, \mathbf{b}_i, \phi) = f(\mathbf{s}_i|\mathbf{b}_i, \phi)$.

Thus, we have

$$f(\mathbf{y}_i, \mathbf{r}_i, \mathbf{s}_i | \mathbf{z}_i, \mathbf{b}_i, \theta) = f(\mathbf{y}_i | \mathbf{z}_i, \mathbf{b}_i, \beta, \sigma) f(\mathbf{r}_i | \mathbf{z}_i, \mathbf{b}_i, \gamma) f(\mathbf{s}_i | \mathbf{b}_i, \phi).$$

The *joint* likelihood for the *observed* data can then be written as

$$\begin{aligned} L_0(\theta) =& \Pi_{i=1}^N \int\int\int f(\mathbf{y}_i | \mathbf{z}_i, \mathbf{b}_i, \beta, \sigma) f(\mathbf{r}_i | \mathbf{z}_i, \mathbf{b}_i, \gamma) f(\mathbf{s}_i | \mathbf{b}_i, \phi) \\ & \times f(\mathbf{z}_i | \alpha) f(\mathbf{b}_i | D) \times f(\mathbf{y}_{i,mis}, \mathbf{z}_{i,mis}, \mathbf{b}_i | \mathbf{y}_{i,obs}, \mathbf{z}_{i,obs}, \mathbf{s}_i, \mathbf{r}_i, \theta) d\mathbf{y}_{i,mis} d\mathbf{z}_{i,mis} d\mathbf{b}_i. \end{aligned}$$

4.3 A Monte Carlo EM Algorithm

Maximum likelihood estimates (MLEs) of all parameters θ can be obtained by maximizing the observed data likelihood $L_0(\theta)$. However, the observed data likelihood $L_0(\theta)$ may be difficult to evaluate because it involves intractable and high dimensional integral. In the following, we use a Monte-Carlo EM algorithm to obtain the MLEs.

If we treat the unobservable random effects $\mathbf{b}_i$ as additional "missing data", we can write the "complete data" as $\{(\mathbf{y}_i, \mathbf{z}_i, \mathbf{r}_i, \mathbf{s}_i, \mathbf{b}_i), i = 1, 2, ..., N\}$. Thus, the complete-data log-likelihood for individual i can be written as

$$\begin{aligned} l_c^{(i)} =& \log f(\mathbf{y}_i | \mathbf{z}_i, \mathbf{b}_i, \beta, \sigma) + \log f(\mathbf{z}_i | \alpha) + \log f(\mathbf{b}_i | D) \\ & + \log f(\mathbf{r}_i | \mathbf{z}_i, \mathbf{b}_i, \gamma) + \log f(\mathbf{s}_i | \mathbf{b}_i, \phi). \end{aligned}$$

The E-step at the t^{th} iteration of the EM algorithm for individual i can then be written as

$$\begin{aligned} Q_i(\theta | \theta^{(t)}) =& \int\int\int \{\log f(\mathbf{y}_i | \mathbf{z}_i, \mathbf{b}_i, \beta, \sigma) + \log f(\mathbf{z}_i | \alpha) + \log f(\mathbf{b}_i | D) \\ & + \log f(\mathbf{r}_i | \mathbf{z}_i, \mathbf{b}_i, \gamma) + \log f(\mathbf{s}_i | \mathbf{b}_i, \phi)\} \\ & \times f(\mathbf{y}_{i,mis}, \mathbf{z}_{i,mis}, \mathbf{b}_i | \mathbf{y}_{i,obs}, \mathbf{z}_{i,obs}, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)}) d\mathbf{y}_{i,mis} dz_{i,mis} d\mathbf{b}_i. \end{aligned}$$

Since it is difficult to evaluate the integral $Q_i(\theta | \theta^{(\mathbf{t})})$ analytically, we approximate the integral by the Monte-Carlo methods following Wu (2009) as follows.

Since $Q_i(\theta|\theta^{(t)})$ is a (conditional) expectation with respect to the density

$$f(\mathbf{y}_{i,mis}, \mathbf{z}_{i,mis}, \mathbf{b}_i | \mathbf{y}_{i,obs}, \mathbf{z}_{i,obs}, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)}),$$

we may approximate Q_i by its empirical mean, obtained by simulating many samples from the conditional density $f(\mathbf{y}_{i,mis}, \mathbf{z}_{i,mis}, \mathbf{b}_i | \mathbf{y}_{i,obs}, \mathbf{z}_{i,obs}, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)})$ and then replacing the expectation by an empirical mean. To generate random samples from the conditional density $f(\mathbf{y}_{i,mis}, \mathbf{z}_{i,mis}, \mathbf{b}_i | \mathbf{y}_{i,obs}, \mathbf{z}_{i,obs}, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)})$, we may use the Gibbs sampler method (Gelfand and Smith, 1990) along with the multivariate rejection method by iteratively sampling from the full conditionals $f(\mathbf{y}_{i,mis}|\mathbf{y}_{i,obs}, \mathbf{z}_i, \mathbf{b}_i, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)})$, $f(\mathbf{z}_{i,mis}|\mathbf{z}_{i,obs}, \mathbf{y}_i, \mathbf{b}_i, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)})$, and $f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)})$ in turn until the resulting Markov chain converges.

To sample these full conditionals, note that

$$f(\mathbf{y}_{i,mis}|\mathbf{y}_{i,obs}, \mathbf{z}_i, \mathbf{b}_i, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \beta^{(t)}, \sigma^{(t)}) \tag{4.1}$$

$$f(\mathbf{z}_{i,mis}|\mathbf{z}_{i,obs}, \mathbf{y}_i, \mathbf{b}_i, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \beta^{(t)}, \sigma^{(t)}) f(\mathbf{z}_i|\alpha) f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \gamma^{(t)}) \tag{4.2}$$

$$\begin{aligned} f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)}) \propto & f(\mathbf{b}_i|D^{(t)}) f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \beta^{(t)}, \sigma^{(t)}) \\ & \times f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \gamma^{(t)}) f(\mathbf{s}_i|\mathbf{b}_i, \phi^{(t)}). \end{aligned} \tag{4.3}$$

Suppose that $\{(\tilde{\mathbf{y}}_{i,mis}^{(1)}, \tilde{\mathbf{z}}_{i,mis}^{(1)}, \tilde{\mathbf{b}}_i^{(1)}), ..., (\tilde{\mathbf{y}}_{i,mis}^{(m_t)}, \tilde{\mathbf{z}}_{i,mis}^{(m_t)}, \tilde{\mathbf{b}}_i^{(m_t)})\}$ is a random sample of size m_t generated from $f(\mathbf{y}_{i,mis}, \mathbf{z}_{i,mis}, \mathbf{b}_i | \mathbf{y}_{i,obs}, \mathbf{z}_{i,obs}, \mathbf{s}_i, \mathbf{r}_i, \theta^{(t)})$. The E-step of the Monte Carlo EM algorithm at the $(t+1)^{th}$ iteration can be approximated as follows

$$Q(\theta|\theta^{(t)}) = \Sigma_{i=1}^N Q_i(\theta|\theta^{(t)}) \tag{4.4}$$

$$\begin{aligned} & \approx \Sigma_{i=1}^N \{ \frac{1}{m_t} \Sigma_{j=1}^{m_t} \log f(\mathbf{y}_{i,obs}, \tilde{\mathbf{y}}_{i,mis}^{(j)} | \mathbf{z}_{i,obs}, \tilde{\mathbf{z}}_{i,mis}^{(j)}, \tilde{\mathbf{b}}_i^{(j)}, \beta, \sigma) \\ & + \log f(\mathbf{z}_{i,obs}, \tilde{\mathbf{z}}_{i,mis}^{(j)} | \alpha) + \log f(\tilde{\mathbf{b}}_i^{(j)} | D) \\ & + \log f(\mathbf{r}_i | \mathbf{z}_{i,obs}, \tilde{\mathbf{z}}_{i,mis}^{(j)}, \tilde{\mathbf{b}}_i^{(j)}, \gamma) + \log f(\mathbf{s}_i | \tilde{\mathbf{b}}_i^{(j)}, \phi) \}. \end{aligned} \tag{4.5}$$

The above approximation can be made arbitrarily accurate by increasing m_t. The

M-step of the Monte Carlo EM algorithm is then to maximize $Q(\theta|\theta^{(t)})$, which is just like a complete data maximization, so standard optimization procedures for complete-data models such as the Newton-Raphson method can be used to obtain the updated parameters $\theta^{(t+1)}$. If we assume that the parameters in each term of $Q(\theta|\theta^{(t)})$ are distinct, we can maximize each term of $Q(\theta|\theta^{(t)})$ separately using standard methods for linear, nonlinear, and logistic regression models.

The variance covariance matrix of θ can be approximated as follows. At the convergence of the EM algorithm, let

$$S_{ij} = \partial l(\theta|\mathbf{y}_{i,obs}, \tilde{\mathbf{y}}_{i,mis}^{(j)}, \mathbf{z}_{i,obs}, \tilde{\mathbf{z}}_{i,mis}^{(j)}, \tilde{\mathbf{b}}_i^{(j)}, \mathbf{r}_i, \mathbf{s}_i)/\partial\theta$$

evaluated at $\theta = \hat{\theta}$, and

$$I(\hat{\theta}) \approx \Sigma_{i=1}^{N} \Sigma_{j=1}^{m_t} \frac{1}{m_t} S_{ij}(\hat{\theta}) S_{ij}^{T}(\hat{\theta}).$$

The approximate asymptotic variance covariance matrix of $\hat{\theta}$ is $I^{-1}(\hat{\theta})$.

4.4 Sampling Methods and Convergence

To implement the Monte-Carlo EM algorithm described in the previous section, one of the major computational steps is to sample from the full conditionals in (4.1)-(4.3). Sampling from the distribution (4.1)-(4.3) can be accomplished by rejection sampling methods. If the appropriate densities on the right hand-sides of (4.1)-(4.3) are log-concave, the adaptive rejection algorithm of Gilks and Wild (1992) may be used. If some densities are not log-concave, we may consider the multivariate rejection sampling method.

For example, suppose that we want to generate random samples from

$$f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{r}_i, \mathbf{s}_i, \theta^{(t)})$$

in (4.3). Let

$$h(\mathbf{b}_i) = f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \beta^{(t)}, \sigma^{(t)}) f(\mathbf{r}_i|\mathbf{z}_i, \mathbf{b}_i, \gamma^{(t)}) f(\mathbf{s}_i|\mathbf{b}_i, \phi^{(t)})$$

and

$$\tau = sup_b h(\mathbf{b}_i).$$

A random sample from $f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{r}_i, \mathbf{s}_i, \theta^{(t)})$ can be obtained as follows:

- Step 1: sample $\mathbf{b}_i^*$ from $f(\mathbf{b}_i|\mathbf{D}^{(t)})$, and independently, sample w from the uniform(0,1) distribution
- Step 2: if $w \leq h(\mathbf{b}_i^*)/\tau$ then accept $\mathbf{b}_i^*$; otherwise, go to step 1.

Samples from the other two full conditionals can be obtained in a similar way. Therefore, the E-step of the Monte-Carlo EM method can be accomplished by the Gibbs sampling method combined with the rejection sampling methods. To assess the convergence of the Gibbs sampler, we may use standard graphical tools such as trace plots and autocorrelations.

To implement the E-step of the Monte-Carlo EM algorithm, we should choose the numbers of Monte-Carlo samples m_t. Generally, larger values of m_t will result in more exact approximation in the E-step but the computation will be slower. To ensure convergence of the Monte-Carlo EM algorithm, we should increase m_t as the number t of EM iterations increases. Note that, for Monte-Carlo EM algorithms, the incomplete-data log-likelihood is not guaranteed to increase at each iteration due to Monte Carlo error at the E-step. However, under suitable regularity conditions, Monte-Carlo EM algorithms still converge to the maximum likelihood estimate (Fort and Moulines, 2003).

Chapter 5

Time-dependent Covariate with Measurement Error

5.1 Introduction

The method presented in Chapter 4 can be extended to the case of time-dependent covariates where the covariates may be missing (ignorable) or measured with errors. In practice, some covariates may be measured with errors, and the time-dependent covariates may also be missing due to different measurement schedules from the response measurements or other problems. For example, in HIV studies, CD4 cell count is often measured with substantial errors and may have measurement schedules different from the viral load measurement schedules. To address covariate measurement errors or missing data, we may model the time-dependent covariates empirically using linear mixed effects (LME) models as follows.

5.2 Joint Likelihood

Let z_{ikl} be the observed value and z^*_{ikl} be the (unobservable) "true" value of covariate k for the ith individual at time u_{il}, $i = 1,...,N$; $k = 1,...,\nu$; $l = 1,...m_i$. We focus on the case where z^*_{ikl} is the current true covariate value. We allow missing data in the covariates. Let $\mathbf{z}_i = (\mathbf{z}_{i1}^T,...,\mathbf{z}_{im_i}^T)^T$, where $\mathbf{z}_{il} = (z_{i1l},...,z_{i\nu l})^T$, $l = 1,...,m_i$. Following Shah, Laird, and Schoenfeld (1997), we consider the following multivariate

LME model to empirically describe the covariate processes

$$\mathbf{z}_{il} = U_{il}\alpha + V_{il}\mathbf{a_i} + \varepsilon_{il}, \qquad i = 1,...,N, l = 1,...,m_t, \tag{5.1}$$

where U_{il} and V_{il} are design matrices, α and $\mathbf{a}_i$ are unknown population (fixed-effects) and individual-specific (random-effects) parameter vectors, and ε_{il} are the random measurement errors for the ith individual at time u_{il}. The true (unobservable) covariate values are assumed to be $\mathbf{z}^*_{il} = U_{il}\alpha + V_{il}\mathbf{a}_i$. We also assume that $\mathbf{a}_i$ i.i.d.$\sim N(\mathbf{0},A)$, ε_{il} i.i.d. $\sim N(\mathbf{0},R)$, and $\mathbf{a}_i$ and $\varepsilon_i = (\varepsilon^T_{i1},...,\varepsilon^T_{im_i})^T$ are independent, where A and R are unknown covariance matrices. We further assume that α and $\mathbf{a}_i$ are independent of $\mathbf{e}_i$ and $\mathbf{b}_i$ in the response model. Note that for commonly-used polynomial empirical LME models, we have $\mathbf{u}_{ik} = (1, u_{ik}, ..., u^{l-1}_{ik})^T$ and $\mathbf{v}_{ik} = (1, u_{ik}, ..., u^{r-1}_{ik})^T$.

To allow for missing data in the time-dependent covariates, we recast model (5.1) in continuous time:

$$\mathbf{z}_i(t) = U_i(t)\alpha + V_i(t)\mathbf{a}_i + \varepsilon_i(t), \qquad i = 1,...,N, \tag{5.2}$$

where $\mathbf{z}_i(t)$, $U_i(t)$, and $\varepsilon_i(t)$ are the covariate values, design matrices, and measurement errors at time t respectively. At the response measurement time t_{ij}, the possibly unobserved "true" covariate values can be viewed as $\mathbf{z}^*_{ij} = U_{ij}\alpha + V_{ij}\mathbf{a}_i$, where $U_{ij} = U_i(t_{ij})$ and $V_{ij} = V_i(t_{ij})$.

When the covariates are measured with errors, we assume that the response and the time-to-event distributions $f(\mathbf{y}_i|\mathbf{a}_i,\mathbf{b}_i,\beta,\sigma)$ and $f(\mathbf{r}_i|\mathbf{a}_i,\mathbf{b}_i,\gamma)$ may depend on the unobserved true covariate values rather than the observed mis-measured covariate values, i.e., the distributions of $\mathbf{y}_i$ and $\mathbf{r}_i$ may depend on the random effects $\mathbf{a}_i$ and $\mathbf{b}_i$. Let θ be the collection of unknown parameters, and $\theta = (\alpha,\beta,\gamma,\phi,\sigma,\delta,D,A)$.

Therefore, the full likelihood for the observed data can thus be written as

$$\begin{aligned} L(\theta) = \Pi_{i=1}^{N} \int\int\int & f(\mathbf{y}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\beta,\sigma) f(\mathbf{r}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\gamma) \\ & \times f(\mathbf{a}_i|A) f(\mathbf{b}_i|D) f(\mathbf{s}_i|\mathbf{b}_i,\phi) f(\mathbf{z}_i|\alpha,\mathbf{a}_i,\delta) \\ & \times f(\mathbf{y}_{i,mis},\mathbf{z}_{i,mis},\mathbf{a}_i,\mathbf{b}_i|\mathbf{y}_{i,obs},\mathbf{z}_{i,obs},\mathbf{s}_i,\mathbf{r}_i,\theta) d\mathbf{y}_{i,mis} d\mathbf{z}_{i,mis} d\mathbf{a}_i d\mathbf{b}_i, \end{aligned}$$

where $z^*(\alpha,\mathbf{a}_i)$ is the true covariate value which depends on random effects $\mathbf{a}_i$ according to model (5.2), and δ^2 stands for the measurement error variability in covariate. We assume that

$$\varepsilon_{i1},\varepsilon_{i2},...,\varepsilon_{in_i} \sim_{i.i.d} N(0,\delta^2).$$

The random effects $\mathbf{a}_i$ are introduced to account for large inter-individual variations in the change of the time-dependent covariate. We assume $\mathbf{a}_i = (a_{i1},a_{i2})^T \sim N(0,A)$.

5.3 A Monte Carlo EM Algorithm

Maximum likelihood estimates (MLEs) of unknown parameters θ can be obtained by maximizing the observed data likelihood $L_0(\theta)$. However, the observed data likelihood $L_0(\theta)$ may be difficult to evaluate because it involves intractable and high dimensional integral. In the following, we use a Monte-Carlo EM algorithm to obtain the MLEs.

If we treat the unobservable random effects $\mathbf{a}_i$ and $\mathbf{b}_i$ as additional "missing data", we can write the "complete data" as $\{(\mathbf{y}_i,\mathbf{z}_i,\mathbf{r}_i,\mathbf{s}_i,\mathbf{b}_i), i = 1,2,...,N\}$. Thus, the complete-data log-likelihood for individual i can be written as

$$\begin{aligned} l_c^{(i)} = & \log f(\mathbf{y}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\beta,\sigma) + \log f(\mathbf{z}_i|\alpha,\mathbf{a}_i,\delta) \\ & + \log f(\mathbf{b}_i|D) + \log f(\mathbf{r}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\gamma) + \log f(\mathbf{s}_i|\mathbf{b}_i,\phi) + \log f(\mathbf{a}_i|A). \end{aligned}$$

The E-step at the t^{th} iteration of the EM algorithm for individual i can then be

written as

$$
\begin{aligned}
Q_i(\theta|\theta^{(t)}) = \int\int\int \{ & \log f(\mathbf{y}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\beta,\sigma) \\
& + \log f(\mathbf{z}_i|\alpha,\mathbf{a}_i,\delta) + \log f(\mathbf{b}_i|D) \\
& + \log f(\mathbf{r}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\gamma) \\
& + \log f(\mathbf{s}_i|\mathbf{b}_i,\phi) \\
& + \log f(\mathbf{a}_i|A)\} \\
& \times f(\mathbf{y}_{i,mis},\mathbf{z}_{i,mis},\mathbf{a}_i,\mathbf{b}_i|\mathbf{y}_{i,obs},\mathbf{z}_{i,obs},\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)}) d\mathbf{y}_{i,mis} d\mathbf{z}_{i,mis} d\mathbf{a}_i d\mathbf{b}_i.
\end{aligned}
$$

Since it is difficult to evaluate the integral $Q_i(\theta|\theta^{(t)})$ analytically, we approximate the integral by the Monte-Carlo methods.

Since $Q_i(\theta|\theta^{(t)})$ is a (conditional) expectation with respect to the density

$$f(\mathbf{y}_{i,mis},\mathbf{z}_{i,mis},\mathbf{a}_i,\mathbf{b}_i|\mathbf{y}_{i,obs},\mathbf{z}_{i,obs},\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)}),$$

we may approximate Q_i by its empirical mean, obtained by simulating many samples from the conditional density $f(\mathbf{y}_{i,mis},\mathbf{z}_{i,mis},\mathbf{a}_i,\mathbf{b}_i|\mathbf{y}_{i,obs},\mathbf{z}_{i,obs},\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)})$ and then replacing the expectation by an empirical mean. To generate random samples from the conditional density $f(\mathbf{y}_{i,mis},\mathbf{a}_i,\mathbf{b}_i|\mathbf{y}_{i,obs},\mathbf{z}_{i,obs},\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)})$, we may use the Gibbs sampler method (Gelfand and Smith, 1990) along with the multivariate rejection method by iteratively sampling from the full conditionals

$f(\mathbf{y}_{i,mis}|\mathbf{y}_{i,obs},\mathbf{z}_{i,obs},\mathbf{a}_i,\mathbf{b}_i,\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)})$, $f(\mathbf{z}_i|\alpha,\mathbf{a}_i,\delta)$, $f(\mathbf{a}_i|\mathbf{z}_{i,obs},\mathbf{y}_i,\mathbf{b}_i,\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)})$, and $f(\mathbf{b}_i|\mathbf{y}_i,\mathbf{z}_{i,obs},\mathbf{a}_i,\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)})$ in turn until the resulting Markov chain converges.

To sample these full conditionals, note that

$$f(\mathbf{y}_{i,mis}|\mathbf{y}_{i,obs},\mathbf{z}_{i,obs},\mathbf{a}_i,\mathbf{b}_i,\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\beta^{(t)},\sigma^{(t)}) \tag{5.3}$$

$$f(\mathbf{z}_{i,mis}|\mathbf{y}_i,\mathbf{z}_{i,obs},\mathbf{a}_i,\mathbf{b}_i,\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)}) \propto f(\mathbf{z}_i|\alpha,\mathbf{a}_i,\delta) \tag{5.4}$$

$$f(\mathbf{a}_i|\mathbf{z}_{i,obs},\mathbf{y}_i,\mathbf{b}_i,\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)}) \propto f(\mathbf{y}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\beta^{(t)},\sigma^{(t)}) \tag{5.5}$$

$$\times f(\mathbf{z}_i|\alpha,\mathbf{a}_i)f(\mathbf{r}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\gamma^{(t)})f(\mathbf{a}_i|A) \tag{5.6}$$

$$f(\mathbf{b}_i|\mathbf{y}_i,\mathbf{z}_{i,obs},\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)}) \propto f(\mathbf{b}_i|D^{(t)})f(\mathbf{y}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\beta^{(t)},\sigma^{(t)})$$

$$\times f(\mathbf{r}_i|\mathbf{z}_i^*(\mathbf{a}_i,\alpha),\mathbf{b}_i,\gamma^{(t)})f(\mathbf{s}_i|\mathbf{b}_i,\phi^{(t)}). \tag{5.7}$$

Suppose that $\{(\tilde{\mathbf{z}}_{i,mis}^{(1)},\tilde{\mathbf{z}}_{i,mis}^{(1)},\tilde{\mathbf{a}}_i^{(1)},\tilde{\mathbf{b}}_i^{(1)}),...,(\tilde{\mathbf{y}}_{i,mis}^{(m_t)},\tilde{\mathbf{z}}_{i,mis}^{(m_t)},\tilde{\mathbf{a}}_i^{(m_t)},\tilde{\mathbf{b}}_i^{(m_t)})\}$ is a random sample of size m_t generated from $f(\mathbf{y}_{i,mis},\mathbf{a}_i,\mathbf{b}_i|\mathbf{y}_{i,obs},\mathbf{z}_{i,obs},\mathbf{s}_i,\mathbf{r}_i,\theta^{(t)})$. The E-step of the Monte Carlo EM algorithm at the $(t+1)^{th}$ iteration can be approximated as follows

$$Q(\theta|\theta^{(t)}) = \Sigma_{i=1}^N Q_i(\theta|\theta^{(t)}) \tag{5.8}$$

$$\approx \Sigma_{i=1}^N\{\frac{1}{m_t}\Sigma_{j=1}^{m_t}\log f(\mathbf{y}_{i,obs},\tilde{\mathbf{y}}_{i,mis}^{(j)}|\mathbf{z}_i^*(\tilde{\mathbf{a}}_i^{(j)},\alpha),\tilde{\mathbf{b}}_i^{(j)},\beta,\sigma)$$
$$+\log f(\mathbf{z}_{i,obs},\tilde{\mathbf{z}}_{i,mis}^{(j)}|\alpha,\tilde{\mathbf{a}}_i^{(j)},\delta)+\log f(\tilde{\mathbf{a}}_i^{(j)}|A)+\log f(\tilde{\mathbf{b}}_i^{(j)}|D)$$
$$+\log f(\mathbf{r}_i|\mathbf{z}_i^*(\tilde{\mathbf{a}}_i^{(j)},\tilde{\mathbf{b}}_i^{(j)},\gamma)+\log f(\mathbf{s}_i|\tilde{\mathbf{b}}_i^{(j)},\phi)\}. \tag{5.9}$$

The above approximation can be made arbitrarily accurate by increasing m_t. The M-step of the Monte Carlo EM algorithm is then to maximize $Q(\theta|\theta^{(t)})$, which is just like a complete data maximization, so standard optimization procedures for complete-data models such as the Newton-Raphson method can be used to obtain the updated parameters $\theta^{(t+1)}$. If we assume that the parameters in each term of $Q(\theta|\theta^{(t)})$ are distinct, we can maximize each term of $Q(\theta|\theta^{(t)})$ separately using standard methods for linear, nonlinear, and logistic regression models.

The variance covariance matrix of θ can be approximated as follows. At the

convergence of the EM algorithm, let

$$S_{ij} = \partial l(\theta|\mathbf{y}_{i,obs}, \tilde{\mathbf{y}}_{i,mis}^{(j)}, \tilde{\mathbf{z}}_{i,mis}^{(j)}, \tilde{\mathbf{a}}_i^{(j)}, \tilde{\mathbf{b}}_i^{(j)}, \mathbf{r}_i, \mathbf{s}_i)/\partial\theta$$

evaluated at $\theta = \hat{\theta}$, and

$$I(\hat{\theta}) \approx \Sigma_{i=1}^{N} \Sigma_{j=1}^{m_t} \frac{1}{m_t} S_{ij}(\hat{\theta}) S_{ij}^{T}(\hat{\theta}).$$

The approximate asymptotic variance covariance matrix of $\hat{\theta}$ is $I^{-1}(\hat{\theta})$.

Chapter 6

Data Analysis

6.1 Introduction

We have discussed the two-step method and the method using the joint likelihood inference for the statistical analysis on the longitudinal process and the survival process in the previous chapters. In this chapter, we analyze a real example using the methods discussed. We describe the data set in Section 6.2. We introduce the models for longitudinal data and survival data as well as the model for missingness in Section 6.3. In Section 6.4, we analyze a real HIV dataset with some interesting findings. We discuss some computational issues in Section 6.5.

6.2 Data Description

The dataset comes from a recent HIV study. It consists of 41 HIV patients who were given an anti-HIV treatment at the beginning of the study. We consider the study within the first 400 days after the treatment since data after 400 days is likely to be influenced by the long term clinical factors. The time then is scaled from 0 to 1 for convenience. The viral load (in $\log_{10}$ scale) and the CD4 cell count are measured repeatedly over time after an anti-HIV treatment. The measurement times within a patient varies from 8 to 14 (with a mean of 10 and a standard deviation of 1.43). Note that there is a substantial variation among patients. Also some patients may not experience any viral load rebound (a viral load increase) during the study period. About 16% CD4 cell count values and 1% viral load values are missing. A summary of the HIV dataset is shown in Table 6.1.

Table 6.1: Summary of the HIV dataset

Variable	Sample mean	Sample standard deviation	Percentage of missing values
Viral load	2.16	1.189	1.17%
CD4 cell count	305.63	156.92	16.19%
No. of patients = 41			
Observations per patients from 8 to 14			
Total missing rate : 17.37%			
Rebound rate : 15.26%			

Figure 6.1 shows viral load trajectories for six randomly selected patients from the study. We see that after an anti-HIV treatment, the patients' viral loads would decline in the early period, which reflected the efficacy of the anti-HIV treatment. As time went by, the patients' viral loads may continue to decline, or become flat, or rebound. For those patients with a rebound of viral load in the latter period, the rebound might be due to the potency of the treatment in the early period and the possible drug resistance developed after the early period. The difference in the viral load among patients may be due to the individual characteristics. It is therefore interesting to study if the individual characteristics are predictive for the time to a viral load rebound.

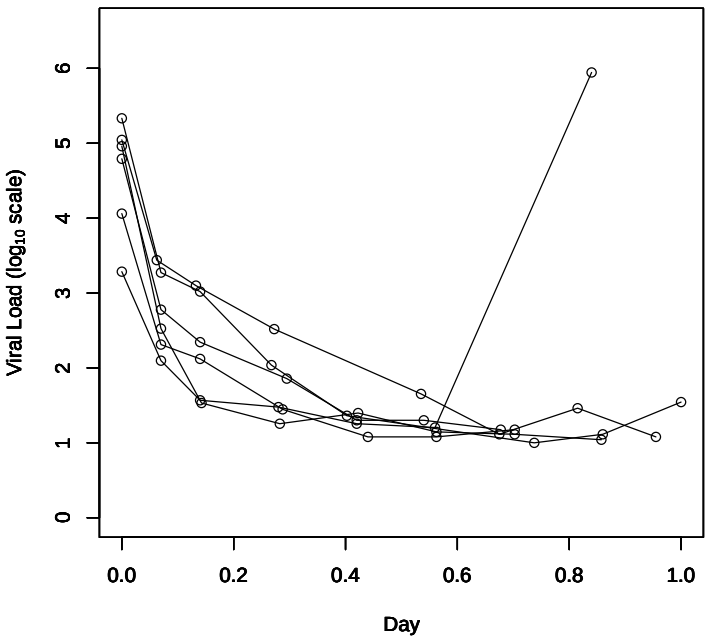

Figure 6.1: Profiles of viral load values for six randomly selected patients.

6.3 Models

6.3.1 The NLME Model for HIV Viral Dynamics

Wu and Ding (1999) proposed a two-compartment exponential decay model for viral load dynamics in the early period. They considered a NLME model for statistical inference. The random effects specifications could be various for this two-compartment model. The NLME model we used has the random effects specifications based on the standard model selection procedures. Table 6.2 shows the AIC and BIC values, and the approximate log-likelihood (logLik) values for different random effects specifications. We find that Model 3, which is without random effects specified in λ_{1i}, attains the smallest AIC value and the smallest BIC value. Further likelihood ratio test would support no significant difference between Model

1 and Model 3; however, Model 3 is simpler.

	Random Effect	df	AIC	BIC	logLik		L.Ratio	p-value
Model 1	$P_{1i}\ \lambda_{1i}\ P_{2i}\ \lambda_{2ij}$	16	305.52	364.29	-136.76			
Model 2	$\lambda_{1i}\ P_{2i}\ \lambda_{2ij}$	12	340.43	384.51	-158.21	1 vs 2	42.91	$< .0001$
Model 3	$P_{1i}\ P_{2i}\ \lambda_{2ij}$	12	294.48	338.56	-135.24	1 vs 3	3.04	0.55
Model 4	$P_{1i}\ \lambda_{1i}\ \lambda_{2ij}$	12	325.22	369.30	-150.61	1 vs 4	27.69	$< .0001$
Model 5	$P_{1i}\ \lambda_{1i}\ P_{2i}$	12	310.57	354.65	-143.28	1 vs 5	13.05	0.011

Table 6.2: Model Selection on NLME model with various random effects specifications.

Hence, we choose the NLME model with random effects specification in Model 3:

$$
\begin{aligned}
y_{ij} &= log_{10}(P_{1i}e^{-\lambda_{1i}t_{ij}} + P_{2i}e^{-\lambda_{2i}t_{ij}}) + e_{ij}, \qquad (6.1)\\
log(P_{1i}) &= \beta_1 + b_{1i}, \qquad \lambda_{1i} = \beta_2,\\
log(P_{2i}) &= \beta_3 + b_{2i}, \qquad \lambda_{2ij} = \beta_4 + \beta_5 CD4^*_{ij} + b_{3i},
\end{aligned}
$$

where y_{ij} is the $\log_{10}$ scale of the viral load measurement for the ith patient at jth measurement at t_{ij}. λ_{1i} and λ_{2ij} represent the individual-specific first and second phases of viral load decay rates, respectively, P_{1i} and P_{2i} are individual-specific baseline values, $\beta = (\beta_1, ..., \beta_5)^T$ are population parameters (fixed effects), e_{ij} represents the within individual errors. The exponential decay rates λ_{1i} and λ_{2ij} can be interpreted as the turnover rates of productively infected cells and the long-lived and/or latently infected cells, respectively. b_{ki}'s are random effects. Note that the individual characteristics of the viral load trajectories can be represented by the random effects (or individual effects) $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})^T$. We assume that $e_{ij}|\mathbf{b}_i \sim_{i.i.d} N(0, \sigma^2)$, where $\mathbf{b}_i \sim_{i.i.d} N(0, D)$. $CD4^*_{ij}$ represents the true but unobserved value of CD4 cell count for patient i at time t_{ij}. This time dependent covariate CD4 cell count is introduced to partially explain the between individual variation in the second phrase of viral load decay rate.

6.3.2 The Covariate Model

The covariate CD4 cell count changes with time and may be measured with errors. We need to model the change of CD4 cell count over the study period. In the absence of theoretical justification, we model the CD4 cell count process based on empirical polynomial linear mixed effects (LME) models. There are many specification in the random effects of LME model. Similar with the model selection process in the viral load model, we select the "best" model of covariate based on the AIC and BIC criteria. Table 6.3 shows the model selection results. It is found that the LME model with random effects in two coefficient gets the smallest AIC value and the smallest BIC value. Also, this LME model beats a quadratic model for its simplicity.

Model	Random Effect	df	AIC	BIC	logLik	Test	L.Ratio	p-value
Linear1	a_0 a_1	6	313.53	335.57	-150.76			
Linear2	a_0	4	330.60	345.29	-161.30	1 vs 2	21.07	$< .0001$
Linear3	a_1	4	620.71	635.40	-306.35	1 vs 3	311.17	$< .0001$
Quadratic	a_0 a_1 a_2	10	317.51	354.24	-148.75	1 vs 4	315.19	0.77

Table 6.3: Model Selection on covariate model in different forms (linear and quadratic).

The LME model describing the change of CD4 cell count is specified as below:

$$\begin{aligned} CD4_{ij} &= \alpha_{i0} + \alpha_{i1} t_{ij} + \varepsilon_{ij}, \\ CD4^*_{ij} &= \alpha_{i0} + \alpha_{i1} t_{ij}, \\ \alpha_{i0} &= \alpha_0 + a_{i0}, \\ \alpha_{i1} &= \alpha_1 + a_{i2}, \end{aligned} \tag{6.2}$$

δ^2 stands for the measurement error variability in CD4 cell count. $CD4^*_{ij}$ represents the true value of CD4 cell count for patient i at time t_{ij}. We assume that $\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{in_i} \sim_{i.i.d} N(0, \delta^2)$. The random effects $\mathbf{a}_i$ are introduced to account for large inter-individual variations in the change of CD4 cell count. We assume $\mathbf{a}_i = (a_{i1}, a_{i2})^T \sim N(0, A)$.

6.3.3 Survival Models

Section 2.4 has given a general discussion on the survival analysis and the Cox proportional hazards model. In particular, in this part, it is of our interest to see if the CD4 cell count or the random effects b_{i1}, b_{i2}, b_{i3} are predictive to the time of a viral load rebound. We consider the following Cox proportional hazards model of the time-to-rebound T_i with the time-dependent covariate CD4 cell count:

$$h(t_{ij}|\mathbf{z}_i, \mathbf{b}_i) = h_0(t_{ij}) exp(\gamma_1 z^*_{ij} + \gamma_2 b_{i1} + \gamma_3 b_{i2} + \gamma_4 b_{i3}), \tag{6.3}$$

where $h_0(t_{ij})$ is the baseline hazards function, and z^*_{ij} is the true value of CD4 cell count for patient i at the jth measurement.

6.3.4 The Dropout Models

Missing data appears in both CD4 cell count and viral load, which may probably be due to patients' dropouts from the follow-up study or failure to visit regularly. The missingness may be informative. The missingness probability of the responses may depend on the random effects which characterize individual differences of the viral load trajectories. Therefore, we assume the following model for the missing mechanism of the viral load in order to include the missingness in the analysis:

$$f(\mathbf{s_i}|\mathbf{b}_i, \phi) = \Pi_{j=1}^{n_i} P(s_{ij} = 1|\phi, \mathbf{b}_i)^{s_{ij}} (1 - P(s_{ij} = 1|\phi, \mathbf{b}_i))^{1-s_{ij}},$$

$$\log \frac{P(s_{ij} = 1|\phi, \mathbf{b}_i)}{1 - P(s_{ij} = 1|\phi, \mathbf{b}_i)} = \phi_0 + \phi_1 b_{1i} + \phi_2 b_{2i} + \phi_3 b_{3i}, i = 1, 2, ..., N, \tag{6.4}$$

By 6.4, we assume the missingness in the response variable depends on the unobserved value of random effects. Therefore, this missingness is informative missing.

We assume the missing in the CD4 cell count missing completely at random. Note that, although we may assume a more complicated model for the missing response mechanism, in this study we would like to avoid building too complicated a model for missing response here since too many nuisance parameters may lead to non-identifiability.

6.4 Results

We have two main interests in the analysis results. One is to see if the CD4 cell count is predictive for the decay rate of the viral load, which could be found by doing inference on parameters in the NLME model: $\beta = (\beta_1, ..., \beta_5)^T$. The other is to see if the individual characteristics are predictive for the time to a viral load rebound, which could be found by looking at parameters in the survival model: $\gamma = (\gamma_1, ..., \gamma_4)^T$. We will consider the statistical analysis with the time-dependent CD4 cell count. We apply the four methods: the naive two-step method (TS), the modified two-step method (MTS), the joint model (JM), the joint model with complete data (CC).

Table 6.4 shows the results by different methods with time-dependent CD4 cell count. All methods suggest a weak predictive power of the time-dependent CD4 cell count in the decay rate of viral load in this dataset. Also, all coefficients in the survival model are insignificant, which implies that the individual characteristics may not be predictive to the time to rebound in this dataset.

Although the four methods give similar answers to the issue of our interest, the results by different methods are different. JM considers the noninformative missingness in the inference while the other three methods assume the missingness ignorable. The estimate of β_2 by JM is higher than that by the other three methods. This difference suggests simply discarding the information containing missing data may underestimate the initial decay rate.

The standard error estimated by TS is generally smaller than the other three methods. This result is not surprising. The standard error for the estimate represents the uncertainty. In this dataset, the uncertainty comes from mainly two sources. One is the sampling variability of obtaining these observed patients. The other one comes from the uncertainty of unknown missing data and the individual characteristics. TS discarded the information containing missing values and it is well known in literature (e.g. Tsiatis and Davidian, 2004) that such a two-step method fails to include the uncertainty of the random effects. Therefore, TS leads

to an underestimation of the variability of parameters. MTS adjusts the standard errors of parameter estimates by including the uncertainty of the random effects using the parametric bootstrap method (Wu, 2009). JM, in general, gives larger standard errors probably because JM includes both the uncertainty of missing values and the unknown individual characteristics.

Table 6.4: Results summary by different methods with time-dependent covariate.

Parameter	TS		MTS			JM		CC	
	EST	SE	BSM	SEM	BSE	EST	ASE	EST	ASE
β_1	10.97	0.20	10.82	0.37	0.51	11.1	0.16	11.12	0.15
β_2	68.73	2.92	64.16	5.53	6.74	90.8	9.96	68.64	3.79
β_3	5.45	0.19	5.44	0.15	0.16	5.96	0.15	5.47	0.13
β_4	4.05	0.33	4.03	0.22	0.28	4.81	0.27	3.86	0.26
β_5	-0.05	0.19	-0.05	0.12	0.17	-0.02	0.11	0.02	0.13
γ_1	-0.03	0.20	-0.16	0.30	0.33	-0.01	0.04	-0.09	1.72
γ_2	0.43	0.41	-0.14	1.01	1.10	-0.78	0.16	-0.12	0.78
γ_3	-0.81	0.43	-0.29	0.68	0.75	-0.44	1.62	-0.76	1.45
γ_4	0.04	0.30	0.18	0.73	0.80	0.74	0.34	0.51	0.61

Note: EST is parameter estimate; SE is the estimated default standard error; BSM and BSE are the bootstrap mean and standard error; SEM is the bootstrap mean of the estimated default standard errors; ASE is the approximated asymptotic standard error.

By the analysis above, we conclude that, for this particular HIV dataset, ignoring missing data mechanisms may under-estimate the initial decay rate. Additionally, the survival process may not have been linked to the individual-specific characteristics or the CD4 cell count values. However, these conclusions are based on one single dataset; therefore simulation study is required to check the validity of these conclusions.

6.5 Computation Issues

Much of the computation issues lie in the joint model which is based on the joint likelihood inference.

6.5.1 Choice of Starting Value

The EM algorithm was used for the joint inference of the joint model in the example. The starting values for the regression coefficients in the NLME model (β) were chosen by fitting a nonlinear mixed effects model to the complete dataset, which is after removing the missing information. The regression coefficients in the survival model (γ) were chosen by fitting a Cox proportional hazards model to the complete dataset with covariates from the NLME model. The regression coefficients in the dropout model (ϕ) was chosen by fitting a logistic regression model to the original dataset, with covariates from the NLME model. The regression coefficients in the empirical model for the time-dependent covariate were chosen by fitting a linear mixed effects model to the complete dataset.

6.5.2 Convergence Criteria

The convergence criteria was based on the relative change in the parameter estimates. The EM algorithm would stop if the differences of the parameter estimates between the current step and the last step is smaller than a tolerance level which is set at beginning. In our example, the tolerance level was set to be 5%. That is, the EM algorithm would stop if the maximum difference of all differences of parameter estimate between the current step and the last step is smaller than 5%. In principle, with a smaller the tolerance level, we could get more accuracy in the parameter estimate, but we have to pay for the additional cost of computation.

6.5.3 Running Time

The running time by the joint model could be huge comparing to the native two-step method and the modified two-step method. There are mainly two reasons for the huge computation time. One is due to the use of the Gibbs sampling in generating samples from a complex probability distribution. The other one is due to the long run to reach convergence in the EM algorithm.

Chapter 7

Simulation Study

7.1 Introduction

In order to evaluate the performance of the joint model comparing to the two-step methods, and the joint method considering the missing data mechanism comparing to the methods with complete data, we conduct a simulation study in this chapter. We compare different methods in terms of their bias and the mean squared errors of the corresponding estimates. We first introduce the design of the simulation study including the setup of parameters and the data generation models. Then, we compare different methods in different scenarios.

7.2 Design of Simulation Study

7.2.1 Models

We generate the response variable y_{ij} from the NLME model as follows:

$$
\begin{aligned}
y_{ij} &= log_{10}(P_{1i}e^{-\lambda_{1i}t_{ij}} + P_{2i}e^{-\lambda_{2i}t_{ij}}) + e_{ij}, \qquad (7.1)\\
\log(P_{1i}) &= \beta_1 + b_{1i}, \qquad \lambda_{1i} = \beta_2,\\
\log(P_{2i}) &= \beta_3 + b_{2i}, \qquad \lambda_{2ij} = \beta_4 + \beta_5 CD4^*_{ij} + b_{3i},
\end{aligned}
$$

where $\beta = (\beta_1, ..., \beta_5)^T$ are the regression parameters of interest.

The true value of β is set to be $(11, 80, 5, 4, 1)$. $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})^T$ are random effects. We assume $\mathbf{b}_i \sim_{i.i.d} N(0, D)$, so $\mathbf{b}_i$ is generated from the normal distribution $N(0, D)$, where D is the variance covariance matrix of $\mathbf{b}_i$. The number of subjects, N, the measurement times for each individual n_{ij}, the variance of the error terms, σ, and the measurement error variability δ are chosen differently in later comparisons.

We generate the true value of CD4 cell count and the observed value of CD4 cell count following the linear mixed effects model as below:

$$
\begin{aligned}
CD4_{ij} &= \alpha_{i0} + \alpha_{i1}t_{ij} + \varepsilon_{ij}, \qquad (7.2)\\
CD4^*_{ij} &= \alpha_{i0} + \alpha_{i1}t_{ij},\\
\alpha_{i0} &= \alpha_0 + a_{i0},\\
\alpha_{i1} &= \alpha_1 + a_{i2},
\end{aligned}
$$

δ^2 stands for the measurement error variability in CD4 cell count. $CD4^*_{ij}$ represents the true value of CD4 cell count for patient i at time t_{ij}. We assume that $\varepsilon_{ij} \sim_{i.i.d} N(0, \delta^2)$. The random effects $\mathbf{a}_i$ are introduced to account for large inter-individual variations in the change of CD4 cell count. We assume $\mathbf{a}_i = (a_{i1}, a_{i2})^T \sim N(0, A)$.

We assign the missing values in the response y_{ij} using the dropout model as

follows:

$$\log \frac{P(s_{ij}=1|\phi,\mathbf{b}_i)}{1-P(s_{ij}=1|\phi,\mathbf{b}_i)} = \phi_0 + \phi_1 b_{1i} + \phi_2 b_{2i} + \phi_3 b_{3i}, i = 1,2,...,N, \qquad (7.3)$$

where ϕ are regression coefficients of interest. s_{ij} is the missing indicator for y_{ij}. $s_{ij}=1$ means y_{ij} is missing; $s_{ij}=0$ means y_{ij} is observed. The dropout model means the missing mechanism of response is related with the random effects. According to different missing rate, we choose different settings of ϕ in latter comparisons.

The time to a viral load rebound (event) is generated following the Cox proportional hazards model as follows:

$$P(r_{ij}=1|r_{il}=0, l<j, \gamma, \mathbf{b}_i) = 1 - exp(-exp(\gamma_{0j} + \gamma_1 z^*_{ij} + \gamma_2 b_{i1} + \gamma_3 b_{i2} + \gamma_4 b_{i3})). \qquad (7.4)$$

r_{ij} is the event indicator, which is a binary variable. $r_{ij}=1$ means there is a rebound in viral load at time t_{ij}; $r_{ij}=0$ means no rebound at time t_{ij}. The model for the time to an event suggests that the time to an event depends on the random effects and the current covariate value. We choose different settings of model coefficients and the baseline hazard according to different missing rates in the latter comparisons.

7.3 Comparison Criteria

We compare different methods in terms of bias and the mean squared errors. The criteria are made in terms of the percentage relative bias and percentage relative root of mean squared errors.

The bias for β_i is defined as

$$\text{bias}_i = |\hat{\beta}_i - \beta_i|,$$

where $\hat{\beta}_i$ is the estimate of β_i. The MSE of β_i is defined as

$$\text{MSE}_i = \text{bias}_i^2 + s_i^2,$$

where s_i is the standard deviation for $\hat{\beta}_i$.

Therefore the percentage relative bias of $\hat{\beta}_i$ is

$$100\% \times \frac{\text{bias}_i}{|\beta_i|}.$$

The percentage relative root of MSE is defined as

$$100\% \times \frac{\sqrt{MSE_i}}{|\beta_i|}.$$

In the latter paragraphs, by MSE we mean the percentage relative root of MSE, by Bias we mean the percentage relative bias.

7.4 Simulation Results

7.4.1 Comparisons of Methods in Different Missing Rates

We will apply different methods to datasets with different rate of missing values in order to check how the rate of missingness affects the estimate results by different methods. We will compare two rates of missing values, 10% and 20%. By rate of missing values, we mean the total rate of missing either in the response or the covariate or in both. The missingness is assumed to be MCAR in this part.

The variance covariance matrix for the random effects in the NLME model 7.1 is chosen as $D = diag(1,1,1)$; the standard deviation of the error term is set to be $\sigma = 0.25$; the variance covariance matrix for the random effect in the LME model 7.2 is $A = diag(0.6, 0.2)$ and $\alpha = (0.5, 0.5)$; the standard deviation of the error terms $\delta = 0.05$. We generate $N = 50$ subjects with 15 within subject measurement times. We run the simulation with 100 repetitions.

Table 7.1 shows simulation results for the missing value rate around 10%. It is

found that the bias of parameter estimation by all three methods are similar. However, JM gives quite larger MSE in β_2 than that given by the other two methods. This result makes sense considering that JM includes the uncertainty of missing values while the other two methods donot. In β_5, which is the coefficient of covariate, JM gives a smaller MSE. This finding is not unexpected since JM considers a measurement error model for the covariate and imputes the covariate value with "true" value from the covariate model.

Table 7.1: Simulation result (10% missing)

Missing	Parameter	True	MTS		JM		CC	
Rate(%)		Value	Bias	MSE	Bias	MSE	Bias	MSE
10	β_1	11	1	6	1	9	2	9
	β_2	80	7	16	12	28	11	19
	β_3	5	2	5	0	4	1	3
	β_4	4	1	10	4	8	6	11
	β_5	1	1	24	6	14	3	22
10	γ_1	-1	44	74	42	66	39	62
	γ_2	1	43	62	55	62	64	68
	γ_3	-1	41	60	37	44	40	47
	γ_4	1	43	63	49	55	48	53

Table 7.2 shows simulation results when the rate of missing values is 20%. Compared with results in Table 7.1, the results of Bias and MSE given by the other two method MTS and CC generally are not much changed. However, the MSE by JM method increases in general while the Bias of parameter estimations by JM stay similar as before. The result of increased MSE by JM is not surprising since the uncertainty of missing information gets larger as the missing rate goes up, which may bring more uncertainty to the parameter estimation for JM method.

7.4.2 Comparisons of Methods in Different Measurement Times

In order to investigate the influence of measurement times on the parameter estimates, in this part we choose 25 visits during the study period. The other setting are the same as the case of missing rate 10% in 7.4.1. The simulation results are

Table 7.2: Simulation result (20% missing)

Missing Rate(%)	Parameter	True Value	MTS		JM		CC	
			Bias	MSE	Bias	MSE	Bias	MSE
20	β_1	11	0	9	4	11	4	11
	β_2	80	7	21	11	35	10	23
	β_3	5	2	5	2	9	1	5
	β_4	4	1	11	7	14	4	10
	β_5	1	6	27	4	18	11	23
20	γ_1	-1	41	80	39	78	40	53
	γ_2	1	43	63	67	74	66	69
	γ_3	-1	37	55	49	59	41	47
	γ_4	1	40	58	60	64	55	58

shown in Table 7.3.

Comparing with Table 7.1, both Bias and MSE in parameter estimation of longitudinal process decrease, which implies that including more individual longitudinal information may give us a better understanding of the longitudinal process.

Table 7.3: Simulation results ($n_i = 25$)

Missing Rate(%)	Parameter	True Value	MTS		JM		CC	
			Bias	MSE	Bias	MSE	Bias	MSE
10	β_1	11	0	5	0	4	7	12
	β_2	80	6	12	11	23	9	13
	β_3	5	1	5	1	2	0	3
	β_4	4	1	8	3	6	3	7
	β_5	1	7	15	2	7	3	11
10	γ_1	-1	46	71	25	40	35	49
	γ_2	1	64	83	48	57	59	63
	γ_3	-1	61	73	21	35	37	44
	γ_4	1	49	59	35	44	40	45

7.4.3 Comparisons of Methods with Different Number of Patients

In practice, we can only obtain information from a limited number of patients. To judge the influence of the number of patients in parameter estimation, in this part, we run simulation with a different number of patients ($N = 500$). The setting of the other parameters are the same as the case of missing rate 10% in 7.4.1. The simulation results are shown in Table 7.4. Compared with results in Table 7.1, we find that the Bias decreases for all three methods. Particularly in parameters of survival model, the Bias decreases substantially, which shows that a larger size of subjects may help us get a better understanding of the survival process in this case.

Table 7.4: Simulation results $(N = 500)$

Missing	Parameter	True	MTS		JM		CC	
Rate(%)		Value	Bias	MSE	Bias	MSE	Bias	MSE
10	β_1	11	0	6	0	6	3	5
	β_2	80	3	12	9	17	4	11
	β_3	5	1	3	2	5	0	3
	β_4	4	2	4	4	7	4	5
	β_5	1	1	12	3	7	6	13
10	γ_1	-1	22	44	25	41	23	35
	γ_2	1	24	38	23	41	22	36
	γ_3	-1	24	35	21	38	27	41
	γ_4	1	23	36	25	40	25	34

7.4.4 Comparisons of Methods with A Larger Variance of Response

When the variance of response changes the model estimation may change as well. To judge the influence of variance of the response variable, we run simulations with an increased $\sigma = 1$. The setting about the other parameters are the same as the case of missing rate 10% in 7.4.1. The simulation results are shown in Table 7.5. Comparing to Table 7.1, we find that the MSE increases substantially both in parameters of the longitudinal model and the survival model by all three methods. The increase of MSE suggests a larger sampling variability of response, so the uncertainty of model estimation tends to be larger.

Table 7.5: Simulation results ($\sigma = 1$)

Missing Rate(%)	Parameter	True Value	MTS		JM		CC	
			Bias	MSE	Bias	MSE	Bias	MSE
10	β_1	11	7	12	4	16	3	9
	β_2	80	13	28	3	39	0	17
	β_3	5	3	6	12	25	13	24
	β_4	4	3	10	8	25	9	28
	β_5	1	3	29	7	48	12	38
10	γ_1	-1	36	55	57	64	25	42
	γ_2	1	45	62	87	91	77	78
	γ_3	-1	42	59	81	89	93	96
	γ_4	1	30	42	78	81	78	81

7.5 Conclusion

By the previous simulation study, we find that when the missingness is missing completely at random, MTS, JM, and CC tend to give similar bias in model parameter estimation but the estimated variability for the model parameter estimation are different. Besides the uncertainty of random effects, JM accounts for the uncertainty of missing values than MTS and CC, so JM gives a larger MSE in general. For the coefficient of covariate, JM gives a smaller MSE which probably because the consideration of measurement errors in the covariate.

All three methods perform relatively good when the overall missing rate is low, or when the number of subjects is large, or when the within subject measurement times is large. Also, all three methods perform better when the response has a relatively small variability.

Chapter 8

Conclusion

In this thesis, we use a joint model to describe the longitudinal process and the survival process simultaneously. The longitudinal process is characterized by a nonlinear mixed effects model, and the survival process is characterized by a Cox proportional hazards model. We introduce a method based on joint likelihood to estimate parameters in the two models. This method is able to consider time-dependent covariate which may be measured with errors and also it is able to account for informative missingness in the response. Due to the intense of likelihood computation, we use a Monte Carlo EM algorithm to get model parameter estimation.

Simulation studies are carried out to compare the performance of joint modeling and the existing modified two-step method. Our simulation results suggest the joint modeling method considering informative missingness and measurement errors in time-dependent covariates may give a more reliable results than the results given by the modified two-step method or methods with complete data. By the simulation study, we also find that the rate of missing values, the size of study subjects, the within individual visit times, the variability of response values may affect the model parameter estimation.

A real example from a recent HIV study with informative dropouts is analyzed by the joint model method and the two step methods. By the analysis result, we

find in this dataset, the CD4 cell count seems not significantly affecting the second phrase viral load decay rate. Also, the individual characteristics which are represented by random effects may not be associated with the survival time. However, the first period decay rate estimated by the joint model considering informative missing is quite larger than that by other methods. This may suggest that simply ignoring or discarding missing information may underestimate the first phrase viral load decay rate. One point needed to address is that in the joint model, we only include random effects as covariates in the missingness model for simplicity. It may be possible that CD4 cell count or other variable is associated with the missingness. Hence, in future research, we may consider a more complex missingness model; however, we should pay attention to the computational issue at the same time since a more complex model may lead a failure of model identification.

References

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman & Hall

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurements Data*. Chapman & Hall.

DeGruttola, V. and Tu, X.M. (1994). Modeling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time. *Biometrics* **50**, 1003-1014.

Ding, A. and Wu, H. (2001). Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics* **2**(1), 13 - 29.

Fort, G. and Moulines, E. (2003). Convergence of the Monte-Carlo EM for curved exponential families. *Annals of Statistics* **31**(4), 1220-1259.

Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151-168.

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

Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337-348.

Guo, X. and Carlin, B.P. Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages (2004). *The American Statistician* **58**, 1-9.

Henderson, R., Diggle, P. J. & Dobson, A. (2002). Joint modeling of longitudinal measurements and event time data. *Biostatistics* **1**, 465-480.

Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edition, Wiley Series in Probability and Statistics, Wiley-Interscience.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated measures studies. *Journal of the American Statistical Association* **90**, 1112-1121.

Shah, A., Laird, N., and Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association* **92**, 775-779.

Ten Have, T. R., Pulkstenis, E., Kunselman, A., and Landis, J. R. (1998). Mixed effects logistics regression models for longitudinal binary response data with informative dropout. *Biometrics* **54**, 367-383.

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

Wei, G. C. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association*, **85**, 699-704.

Wu, H. and Ding, A. (1999). Population HIV-1 dynamics in vivo: application models and inferential tools for virological data from AIDS clinical trials. *Biometrics*, **55**, 410-418.

Wu, H. (2005). Statistical Methods for HIV Dynamic Studies in AIDS Clinical Trials. *Statistical Methods in Medical Research*, to appear.

Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association*, **97**, 955-964.

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.

Wu, L. (2009). *Mixed effects models for the complex data*. Chapman & Hall/CRC.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **55**, 410-418. 19.

Wulfsohn, M.S. and Tsiatis, A.A. (1997). A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics* **53**, 330-339. 20.