Semiparametric Models for Joint Analysis of Longitudinal Data and Counting Processes

by
Se Hee Kim

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill
2010

Approved by:

Dr. Donglin Zeng, Advisor

Dr. David Couper, Reader

Dr. Danyu Lin, Reader

Dr. John Preisser, Reader

Dr. Ying So, Reader

Abstract

SE HEE KIM: Semiparametric Models for Joint Analysis of Longitudinal Data and Counting Processes.
(Under the direction of Dr. Donglin Zeng.)

In this dissertation, we study statistical methodology for joint modeling that correctly controls for the interplay among longitudinal and counting processes and makes the most efficient use of data. Three types of joint modeling approaches are proposed based on three different purposes of studies.

In the first topic, we develop a method for joint modeling of longitudinal data and recurrent events in the presence of an informative terminal event. We focus on data from patients who experience the same type of event at multiple times, such as multiple infection episodes or recurrent strokes, have longitudinal biomarkers, and may be subject to an event, for example death, that makes further observations impossible. To analyze such complicated data, we propose joint models based on a likelihood approach. A broad class of transformation models for the cumulative intensity of recurrent events and the cumulative hazard of the terminal event is considered. We propose to estimate all the parameters using nonparametric maximum likelihood estimators (NPMLE), and we provide computationally efficient EM algorithms to implement the proposed inference procedure. Asymptotic properties of the estimators are shown to be asymptotically normal and semiparametrically efficient. Finally, we evaluate the performance of the proposed method through extensive simulations and application to real data.

In the second topic, we develop a method for joint modeling of longitudinal and

cure-survival data. By cure-survival data, we mean time-to-event data in which a certain proportion of patients never have any event during a sufficiently long follow-up period. These patients are believed to have been cured by treatment, such as radiation therapy or an initial surgery, and are often the source of heavy tail probabilities in survival curves. To take into account the possibility of patients being cured, we propose to model time-to-event through a transformed promotion time cure model, jointly with a linear mixed effects model for longitudinal data. Due to transformations applied to the promotion time cure model, the proposed method is able to be used in cases where the proportionality assumption does not hold. All the parameters are estimated using NPMLEs, and inference procedures are implemented via a simple EM algorithm. Asymptotic properties of the proposed NPMLEs are derived based on empirical process theory. Simulation studies are conducted and the method is applied to the ARIC data in order to demonstrate the small-sample performance of the proposed method.

In the third topic, we develop a partially linear model for longitudinal data with informative censoring, where the main interest is in making inferences about the individual's trajectory of longitudinal responses, which may be informatively censored. Since a fully parameterized mean structure may be insufficient to capture the underlying patterns of longitudinal and event processes, we propose to use a partially linear model for longitudinal responses, where an unspecified underlying function is formulated along with linear covariate effects, and a transformation model is used for informative censoring times. We employ a sieve estimation for the nonparametric trajectory of longitudinal responses, where the unknown trajectory is approximated by cubic B-spline basis functions. All parameters are estimated based on a likelihood approach, and inference procedures are implemented via the EM algorithm. We also investigate a reliable way to select the number of knots and the best transformation.

Through empirical process theory, asymptotic properties of the proposed estimators are shown to provide desirable properties. The validity of the proposed method is confirmed by simulated and real data examples.

Acknowledgments

None of this would have been possible without the personal and practical support of numerous people. I would gratefully and sincerely like to thank my advisor, Dr. Donglin Zeng, for his guidance, encouragement, support, and deep consideration throughout the duration of this dissertation. He showed me different ways to approach a research problem and the need to be persistent to accomplish any goal. Furthermore, he always read and responded to the drafts of each page of my work more quickly than I could have hoped. I am grateful to my committee members, Drs. David Couper, Danyu Lin, John Preisser, and Ying So, for their valuable comments and insightful suggestions on this dissertation.

I wish to thank Drs. Ying So and Gordon Johnston, who as my supervisors and friends at SAS guided my RA work and cared deeply about me. I truly enjoyed working with them. I would also like to express my sincere appreciation to Dr. John Preisser for inspiring and supporting me to be a professor.

Special thanks go to my best friends, Joy Wu, Che-Chin Lie, and Chaeryon Kang for their support and encouragement. Last, I can never find words to express my gratitude to my fiance Seunggeun Lee and my family who are always standing right next to me.

Table of Contents

Abstract **iii**

List of Tables **xiii**

List of Figures **xv**

1 Introduction **xvi**

1.1 Joint Models of Longitudinal Data and Recurrent Events with Informative Terminal Event 2

1.2 Joint Modeling of Longitudinal Data and Cure-Survival Data 2

1.3 Partially Linear Model for Longitudinal Data with Informative Censoring 3

2 Literature Review **5**

2.1 Models for Longitudinal and Survival Data 5

2.1.1 Transformation Models for Survival Data 6

2.1.2 Partially Linear Models for Longitudinal Data 9

2.1.3 Joint Models for Longitudinal Data and Survival Event 12

2.2 Models for Cure-Survival Data 16

2.2.1 Mixture Cure Models 17

2.2.2 Promotion Time Cure Models 19

2.2.3 Transformation of Promotion Time Cure Models 20

2.3 Models for Longitudinal Data and Recurrent Events 21

2.4 Models for Recurrent and Terminal Events 23

3 Joint Models of Longitudinal Data and Recurrent Events with Informative Terminal Event 26

3.1 Introduction 27

3.2 Joint Models 29

3.3 Inference Procedure 32

3.3.1 Nonparametric Maximum Likelihood Estimation 32

3.3.2 EM Algorithm 32

3.4 Asymptotic Properties 34

3.5 Simulation Studies 36

3.6 Data Application 38

3.7 Concluding Remarks 42

3.8 E-step and M-step in EM Algorithm 43

3.9 Proof of Asymptotic Properties 47

4 Joint Modeling of Longitudinal and Cure-Survival Data 58

4.1 Introduction 59

4.2 Joint Models 61

4.3 Inference Procedure 64

4.3.1 NPMLEs for Transformation Models 64

4.3.2 EM Algorithm 65

4.4 Asymptotic Properties 69

4.5 Simulation Studies 71

4.6 Data Application 73

4.7 Concluding Remarks 75

4.8 Proof of Asymptotic Properties 77

5 Partially Linear Model for Longitudinal Data with Informative Censoring **89**

5.1 Introduction 89

5.2 Joint Models 92

5.3 Inference Procedure 94

5.3.1 Sieve Approximation 94

5.3.2 NPMLEs for Transformation Models 95

5.3.3 EM Algorithm 97

5.4 Asymptotic Properties 99

5.5 Simulation Studies 102

5.6 Data Application 105

5.7 Concluding Remarks 107

5.8 Proof of Asymptotic Properties 108

6 Summary and Future Research **123**

Bibliography **127**

List of Tables

3.1 Simulation results for $G_R(x) = G_T(x) = x$. 53
3.2 Simulation results for $G_R(x) = x$ and $G_T(x) = \log(1 + x)$. 54
3.3 Analysis results for the ARIC study. The Fisher transformation is used for testing ρ, while the 50:50 mixture of χ^2 distributions is used for testing variances. 55

4.1 Simulation results for $H(x) = x$. t_p represents the pth percentile. . . . 83
4.2 Simulation results for $H(x) = 2\log(1 + x/2)$. t_p represents the pth percentile. 84
4.3 Simulation results for $H(x) = \log(1+x)$. t_p represents the pth percentile. 85
4.4 Analysis results for the ARIC study. The 50:50 mixture of χ^2 distributions is used for testing variances. 86

5.1 Simulation results for $H(x) = x$ and $\alpha(t) = \sin(\pi t)\exp(t/2)/\{1 + \exp(t/2)\}$ based on m=6 control points of B-spline curves. τ_p represents p% of τ (study duration). 114
5.2 Simulation results for $H(x) = x$ and $\alpha(t) = \sin(\pi t)\exp(t/2)/\{1 + \exp(t/2)\}$ based on m control points of B-spline curves. τ_p represents p% of τ (study duration). 115
5.3 Simulation results for $H(x) = \log(1+x)$ and $\alpha(t) = \sin(\pi t)\exp(t/2)/\{1+ \exp(t/2)\}$ based on m=6 control points of B-spline curves. τ_p represents p% of τ (study duration). 116
5.4 Simulation results for $H(x) = \log(1+x)$ and $\alpha(t) = \sin(\pi t)\exp(t/2)/\{1+ \exp(t/2)\}$ based on m control points of B-spline curves. τ_p represents p% of τ (study duration). 117
5.5 Joint analysis results of the medical costs data. The 50:50 mixture of χ^2 distributions is used for testing variances. 118

List of Figures

3.1 Log-likelihood surface under the logarithmic transformations for the ARIC study. The x-axis and y-axis correspond to the transformation parameter γ for recurrent events and terminal event, respectively. . . 56

3.2 Predicted survival probability (a) and the expected longitudinal SBP levels (b) for a subject who had one CHD event at the 5th year of study. The solid curves are point estimates, and the dotted curves are the 95% confidence bands. 57

4.1 In the ARIC data (a) Kaplan-Meier survival curve of the entire study population; (b) Estimated survival curve of the non-immune subpopulation under the joint cure-survival model with the proportional odds structure. The solid curves are point estimates, and the dotted curves are 95% confidence bands. 87

4.2 Predicted marginal survival rates of the entire population using the results in Table 4.4. The rates beyond the cure threshold are interpreted as the immune fractions or the cure rates (CR). Reference is taken for age of 54, female, HDL-cholesterol 42 mg/dL, LDL-cholesterol 136 mg/dL, no hypertension medication use, never smoking, and no diabetes. 88

5.1 Example of basis functions (cubic B-spline) for time t in [0, 1] under 5 control points $\{0.1, 0.15, 0.2, 0.4, 0.7\}$. 119

5.2 Simulation results for the baseline coefficient function by (a) $H(x) = x$ and $\alpha(t) = \sin(\pi t)\, e^{\frac{t}{2}}/(1 + e^{\frac{t}{2}})$; (b) $H(x) = \log(1 + x)$ and $\alpha(t) = \sin(\pi t)\, e^{\frac{t}{2}}/(1 + e^{\frac{t}{2}})$; (c) $H(x) = x$ and $\alpha(t) = (t - 0.8)^2$; and (d) $H(x) = \log(1 + x)$ and $\alpha(t) = (t - 0.8)^2$. The solid curves are true values, the dashed curves are estimates under m=3, the dash-dotted curves are estimates under m=6, and the dotted curves are estimates under m=9. 120

5.3 Bayesian information criterion (BIC) for the transformation $H(x) = \log(1 + \eta x)/\eta$ and the number of control knots (m). From top to bottom, the dot-long-dashed curve is for m=8, the long-dashed curve is for m=7, the short-dashed curve is for m=4, the dot-short-dashed curve is for m=6, the dotted curve is for m=3, and the solid curve is for m=5. 121

5.4 Baseline coefficient function of hospital visit time in the medical cost data under the best fit of transformation $H(x) = 2\log(1 + 0.5x)$ and 5 control points. The solid curves are estimates from the joint model, the dashed curves are estimates from the marginal model, and the dots are residual means of $\{Y(t) - \hat{\beta}^T X_1(t)\}$. 122

Chapter 1

Introduction

Joint modeling of longitudinal data and counting processes becomes increasingly popular in a wide range of applications. In these applications, the longitudinal data serve as an outcome variable or a covariate with measurement errors, which are observed at a series of times, while the counting process often represents time to single- or multiple-endpoints, informative observation process, or informative censoring. Joint modeling starts from separate model building for each process and links the models together via correlated or common latent random effects in a variety of ways. Using the joint modeling approach, we can build a model to assess the effect of covariates on both longitudinal measures and time to events, can optimize the use of data through the information shared between the processes, and can correct the biases due to the dependence between the processes. In this dissertation, we focus on simultaneous inferences for both longitudinal measures and time to single or multiple events, while accounting for the dependence between them.

1.1 Joint Models of Longitudinal Data and Recurrent Events with Informative Terminal Event

We first consider joint modeling of longitudinal data and recurrent events along with another event that discontinues further observations, such as death. We refer to the latter event as a terminal event. Examples of recurrent events include multiple strokes, the number of bladder tumors, or informative measurement times such as emergency hospital visiting times. To model such a complicated system, we propose joint models; a linear mixed effects model is used to model longitudinal data, and a broad class of transformation models is used for the cumulative intensity and hazard functions of recurrent and terminal events, respectively. Through transformations, the proposed method is applicable more generally without the proportional hazards or odds assumption. Random effects in the longitudinal model and other dependent random effects in the recurrent event model are shared in the terminal event model, and hence they account for their respective dependencies with the terminal event.

1.2 Joint Modeling of Longitudinal Data and Cure-Survival Data

We next focus on the joint analysis of longitudinal and cure-survival data. By cure-survival data, we mean time-to-event data in which a certain proportion of patients never have any event during a sufficiently long follow-up period. These patients are believed to being cured by treatment, such as radiation therapy or an initial surgery. The potential of being cured can produce a heavy tail probability in the survival curve, and ignoring the true cure proportion may be a source of bias in the estimates of model parameters.

To take into account the possibility of patients being cured in survival data, we model time to event through the promotion time cure model, jointly with a linear mixed effects model for longitudinal data. The promotion time cure model does not separate the population into cured or uncured subpopulations intentionally, unlike other commonly used mixture cure rate models, and hence it does not involve identifiability issues. Conditional on covariates and the shared random effects between the two models, we assume longitudinal data are independent of cure-survival data. The proposed method is flexible in terms of the fact that the proportionality assumption does not need to be true for the survival event.

1.3 Partially Linear Model for Longitudinal Data with Informative Censoring

Longitudinal data analysis has been challenged by informative censoring where the censorship can provoke biases in estimating model parameters. Most existing methods for jointly modeling longitudinal data and censored event assume the full parametric specification for the mean structure of longitudinal responses. While parametric approaches are useful, questions always arise about the adequacy of the model assumptions. Apparently, many longitudinal studies, for example HIV/AIDS clinical trials, show that the parametric models are not sufficient to reveal the complicated patterns of responses with covariates in practice. This motivates us to consider a partially linear model that combines the unspecified underlying trajectory of longitudinal responses with linear covariate effects.

Specifically, we propose a partially linear model for longitudinal responses and a transformed survival model for informative censoring. This semiparametric modeling approach allows sufficient flexibility to disclose complex patterns of longitudinal re-

sponses. In the proposed method, the dependence of longitudinal data on informative censorship is modeled by shared latent effects.

Chapter 2

Literature Review

In this chapter, we review literature on statistical methods for longitudinal and survival data in Section 2.1, for longitudinal and cure-survival data in Section 2.2, for longitudinal data and recurrent events in Section 2.3, and for recurrent and terminal events in Section 2.4.

2.1 Models for Longitudinal and Survival Data

In survival analysis, the most attractive models are the Cox proportional hazards model (Cox, 1972) and the proportional odds model (Bennett, 1983), which have been fully explored in theory and extensively used in practice. For two sets of covariate values, the proportional hazards models assume that the associated ratio of the hazards to be constant over time, while the proportional odds models assume the associated odds ratio of survival to be constant over time. The two models are special cases of linear transformation models, which provide many useful alternatives. In Section 2.1.1, we review the transformation models for survival analysis. These transformation models will be one of the important features of the three topics proposed in this dissertation. In longitudinal data analysis, the main interest lies in the pattern or mean changes of responses measured at a series of observation times. To identify

the complicated trajectory of repeated measures, there has been increasing interest and activity in the general area of partially linear regression models. In Section 2.1.2, we review the methods and techniques developed for the partially linear models. The acquired knowledge and skills for the partially linear regression models will be an essential part for accomplishing the proposed work in Chapter 5. In longitudinal and survival data analysis, joint modeling approaches are one of the most popular ways to describe or control the dependence between longitudinal data and a time-to-event from the same subject. Depending on the purpose of study, various joint modeling approaches have been useful in different applications. In Section 2.1.3, we review the various joint modeling approaches for longitudinal and survival data.

2.1.1 Transformation Models for Survival Data

A class of transformation models for survival functions was proposed by Cheng et al. (1995), in which an unknown transformation of the survival time is linearly related to the covariates with completely specified error distributions. Specifically, let T be the failure time and let $\boldsymbol{Z}$ be a vector of covariates. We denote the survival function of T given $\boldsymbol{Z}$ by $S_Z(t)$. Then, the Cox proportional hazards model can be written as $\log(-\log(S_Z(t))) = H(t) + \boldsymbol{\beta}^T\boldsymbol{Z}$, and the proportional odds model can be written as $-\text{logit}(S_Z(t)) = H(t) + \boldsymbol{\beta}^T\boldsymbol{Z}$, where $H(t)$ is a completely unspecified strictly increasing function, and $\boldsymbol{\beta}$ is a vector of unknown regression coefficients. A natural generalization of these models is

$$g(S_Z(t)) = H(t) + \boldsymbol{\beta}^T\boldsymbol{Z},$$

where g is a known continuous and decreasing function. It is easy to see that the above equation is equivalent to the linear transformation model by Cheng et al. (1995),

$$H(t) = -\boldsymbol{\beta}^T \boldsymbol{Z} + \epsilon, \tag{2.1}$$

where ϵ is a random error with a known distribution function F, where $F = 1 - g^{-1}$. If F is the extreme value distribution $F(s) = 1 - \exp\{-\exp(s)\}$, (2.1) is the proportional hazards model, while if F is the standard logistic distribution, that is $P[\epsilon > s] = \exp(s)/\{1 + \exp(s)\}$, (2.1) is the proportional odds model. We note that model (2.1) is appealing in that it is a familiar linear model and includes the proportional hazards and the proportional odds models as special cases. However, model (2.1) cannot handle time-dependent covariates or cannot be generalized to counting processes such as recurrent events.

Zeng and Lin (2006) proposed a class of semiparametric transformation models for general counting processes to accommodate time-varying covariates on the intensity functions of recurrent events. In particular, let $N^*(t)$ be the number of events that occurred by time t, and let $\boldsymbol{Z}(\cdot)$ be a vector of time-varying covariates. Then, the cumulative intensity function for $N^*(t)$ conditional on $\{\boldsymbol{Z}(s);\, s \leqslant t\}$, denoted by $\Lambda_Z(t)$, takes the form

$$\Lambda_Z(t) = G\left(\int_0^t R^*(s)\, e^{\boldsymbol{\beta}^T \boldsymbol{Z}(s)}\, d\Lambda(s)\right), \tag{2.2}$$

where $R^*(\cdot)$ is the indicator process for the risk set, $\Lambda(\cdot)$ is an arbitrary increasing function, and G is a continuously differentiable and strictly increasing function with $G(0) = 0$, $G(\infty) = \infty$ and $G'(0) > 0$. As examples of the transformation function

$G(\cdot)$, the class of Box-Cox transformations,

$$G(x) = \frac{(1+x)^{\rho} - 1}{\rho}, \quad \rho \geqslant 0$$

with $\rho = 0$ corresponding to $G(x) = \log(1+x)$ and the class of logarithmic transformations

$$G(x) = \frac{\log(1+\gamma x)}{\gamma}, \quad \gamma \geqslant 0$$

with $\gamma = 0$ corresponding to $G(x) = x$ can be considered. In both cases, the choice of $G(x) = x$ yields the proportional intensity or hazards models, while $G(x) = \log(1+x)$ leads to the proportional odds models. We note that when $N^*(t)$ has a single jump at the survival time T and $\boldsymbol{Z}$ is time-invariant, (2.2) reduces to the linear transformation model (2.1) in that

$$\log \Lambda(T) = -\boldsymbol{\beta}^T \boldsymbol{Z} + \log G^{-1}(-\log(\epsilon^*)),$$

where ϵ^* has a uniform distribution.

Zeng and Lin (2007a) further extended the class of semiparametric transformation models for the intensity function of counting process with random effects, which allows non-proportional intensity and various frailty distributions. By introducing the random effects, the proposed models account for the dependence of the recurrent event times within the same subject. Let $\boldsymbol{X}(\cdot)$ and $\boldsymbol{Z}(\cdot)$ be vectors of possibly time-dependent covariates associated with the fixed and random effects, respectively. Conditional on $\{\boldsymbol{Z}(s), \boldsymbol{X}(s), b\,;\, s \leqslant t\}$, the cumulative intensity function for $N^*(t)$ has the form of

$$\Lambda(t|\boldsymbol{X}, \boldsymbol{Z}; b) = G\left(\int_0^t R^*(s)\, e^{\boldsymbol{\beta}^T \boldsymbol{X}(s) + b^T \boldsymbol{Z}(s)}\, d\Lambda(s)\right), \tag{2.3}$$

where b is a set of random effects with a parametric density function. These models are substantially flexible in the sense that one can have a wide variety of options for the transformation G as well as the distribution of the random effects.

2.1.2 Partially Linear Models for Longitudinal Data

Parametric regression models for longitudinal data have received tremendous attention, and the related methods have been well developed. However, a major limitation of these methods is that the fully parameterized mean structure may be insufficient in modeling the complicated relationship between the responses and covariates in various longitudinal studies. Examples include trajectories of CD4 cell counts in HIV/AIDS research (Zeger and Diggle, 1994; Lin and Ying, 2001; Huang et al., 2002; Brown et al., 2005); time-varying effects of gender and HIV status on the growth of infants born from HIV infected mothers (Hoover et al., 1998); age effects on childhood respiratory disease (Diggle et al., 2002); and treatment effects on the longitudinal number of bladder tumors (Sun et al., 2005; Liang et al., 2009). These practical applications encouraged significant developments of nonparametric regression methods for longitudinal data, in which unspecified functions of time or covariates provide enough flexibility to reflect the complicated relationship between longitudinal outcomes and covariates. Despite the fact that, a semiparametric partially linear regression model is more desirable than modeling every covariate effect nonparametrically in many cases, only limited work has been done on semiparametric regression for correlated data. We review three ways of estimating parameters in the semiparametric regression models using kernal smoothing, smoothing splines, and regression splines.

Kernal smoothing was considered by Zeger and Diggle (1994) and Lin and Carroll (2001) for models with linear covariate effects and a nonparametric function of time with correlated observations, among others. Let $Y_{ij} = Y_i(t_{ij})$ $(i = 1, \ldots, n;\ j =$

$1, \ldots, m_i$) be the jth outcome of the ith subject at time t_{ij}. Zeger and Diggle (1994) and Moyeed and Diggle (1994) proposed a semiparametric mixed effects model for longitudinal data

$$Y_{ij} = \mu(t_{ij}) + \boldsymbol{X}_{ij}^T \boldsymbol{\beta} + W_i(t_{ij}) + \epsilon_{ij}, \tag{2.4}$$

where $\mu(t)$ is a twice-differentiable smooth function of time t, $\boldsymbol{\beta}$ is a vector of regression coefficients associated with covariates $\boldsymbol{X}_{ij}$, $W_i(t)$ is a subject-specific stationary Gaussian process with mean zero, and ϵ_{ij} is a white measurement noise with constant variance σ^2. They suggested a backfitting procedure which initially estimates $\mu(t)$ by a kernel smoother with the bandwidth parameter chosen via cross-validation, and then iteratively estimates $\mu(t)$ and $\boldsymbol{\beta}$ using generalized least squares. For clustered data, Lin and Carroll (2001) considered a marginal partially generalized linear model and the profile-kernel method where the nonparametric function is estimated using the local linear kernel regression and the regression coefficients are estimated using the profile estimating equations. Surprisingly, the resulting regression parameter estimators by the conventional profile-kernel method failed to achieve semiparametric efficiency.

A smoothing spline can be an alternative choice of the nonparametric estimation of $\mu(t)$, which uses a piecewise polynomial function with all the observation times used as knots and smoothness constraints imposed at the knots. The most commonly used smoothing spline is the natural cubic smoothing spline, which approximates $\mu(t)$ by a piecewise cubic function with boundary constraints. The natural cubic smoothing spline was studied by Zhang et al. (1998) to estimate the nonparametric function of time in the partially linear model which was expanded from (2.4) with the addition of subject-specific random effect terms. They estimated $\boldsymbol{\beta}$ and $\mu(t)$ as a natural cubic

spline by maximizing the penalized likelihood function with the penalty term

$$\frac{\lambda}{2}\int_{T_1}^{T_2}[\mu''(t)]^2\,dt = \frac{\lambda}{2}\,\boldsymbol{\mu}^T\boldsymbol{K}\boldsymbol{\mu},$$

where $\lambda \geqslant 0$ is a smoothing parameter controlling the balance between the goodness of fit and the roughness of the estimated $\mu(t)$, T_1 and T_2 specify the range of t, $\boldsymbol{\mu} = (\mu(t_{11}), \ldots, \mu(t_{n,m_n}))^T$, and $\boldsymbol{K}$ is the nonnegative definite smoothing matrix defined in the equation (2.3) of Green and Silverman (1994). A key feature of this approach is that the proposed semiparametric model can be represented as a modified parametric linear mixed model. Therefore, the smoothing parameter and variance components can be estimated simultaneously using the restricted maximum likelihood estimator.

Another attractive method to estimate the nonparametric function is regression splines. The smoothing spline has the merit of not involving the knot selection issue since it uses all the observation points as knots. However, when the sample size is large, computational demands substantially grow and make it difficult to work properly. In contrast, a key advantage of regression splines is its computational simplicity. The regression splines is a basis function-based nonparametric regression method, which uses a small number of knots and implements a parametric regression using the bases. The most commonly used basis function for regression splines is the B-spline basis. Rice and Wu (2001) adopted the B-spline basis with equally spaced knots in estimating $\mu(t)$ and a smooth random function $W_i(t)$ in (2.4). The approximated mean function is,

$$\mu(t) = \sum_{k=1}^{p}\xi_k B_k(t),$$

where $\{B_k(\cdot)\}$ is a basis for spline function on the time range with a fixed knot

sequence. Similarly, the random function for the ith subject can be approximated with splines

$$W_i(t) = \sum_{k=1}^{q} \nu_{ik} \tilde{B}_k(t),$$

where $\{\tilde{B}_k(\cdot)\}$ is a basis for random spline function, which may be a different basis than $\{B_k(\cdot)\}$, and ν_{ik} are random coefficients with mean zero and covariance matrix $\mathbf{V}$. Then, conditional on p and q, the approximated model is a classical linear mixed effects model. Estimation of the parameters $\boldsymbol{\beta}$, ξ, σ^2, and the covariance matrix $\mathbf{V}$ can be accomplished by the EM algorithm. In terms of regression splines method, choices of the number and location of the knots for the splines are critical since estimation of $\mu(t)$ and $W_i(t)$ could be very sensitive to these choices. Rice and Wu (2001) suggested using model selection techniques such as Akaike information criterion (AIC) and Bayesian information criterion (BIC), and leave-one-subject-out cross-validation.

2.1.3 Joint Models for Longitudinal Data and Survival Event

Analysis of longitudinal and survival data can be classified into three categories, depending on how one factors the joint distribution of repeated measurements and an event time to meet the study objective. A joint model of the vector of repeated measurement Y and the event time T corresponds to the factorization

$$f(Y, T) = f(Y|T)f(T) = f(T|Y)f(Y),$$

where $f(.)$ denotes a density function. The three categories are referred to as a selection model, a pattern-mixture model, and a simultaneous model. First, in selection models, time-to-event is the endpoint of interest, and the common primary objective of the study is to assess the relationship between the event time and longitudinal covariate process with measurement error. One example is modeling the probability

of death given trajectory of CD4 cell counts, that is $f(T|Y)$. Second, in pattern-mixture models, the repeated measures are the primary endpoint, and investigators are focusing on modeling $f(Y|T)$ and mainly interested in the effect of covariates on the longitudinal outcomes, while accounting for possible correlation with an event such as non-ignorable dropout or death. In these cases, the longitudinal process is subject to right-censoring because it is unobservable after the censoring time. Third, in simultaneous models, repeated measures and survival time are both important outcomes, namely $f(Y, T)$ are focused. The primary goal of the joint analysis is to evaluate simultaneously the effect of covariates on the two types of outcomes, while accounting for the relationship between longitudinal and event time data. In all three types of joint models, it is commonly assumed that observation times of the longitudinal outcomes are usually not informative because they are measured at scheduled follow-up visits. Recent literature is briefly reviewed in the subsequent paragraphs.

Selection models The association of longitudinal covariates with a failure time as the primary endpoint can be assessed through joint modeling of the Cox proportional hazards model of the failure time and a random process model of the longitudinal covariates when the longitudinal covariates are intermittently measured with errors. In this situation, the longitudinal covariates may not be observed at the failure times. The presence of random error in a measured covariate causes the parameter estimators to be biased toward the null (Prentice, 1982). A naive approach is to substitute the closest observed covariate value prior to the failure time, often termed 'last value carried forward', in the Cox partial likelihood for each subject at each failure time. However, it is well known that substituting mismeasured values for true covariates in the Cox model leads to biased estimation (Prentice, 1982). Various approaches have been proposed to deal with measurement error.

Tsiatis et al. (1995) proposed use of a two-stage method where, in the first stage, empirical Bayes estimates of the random covariates are computed, and in the second stage, they are imputed into the partial likelihood of the Cox model as true values of time-dependent covariates at each event time point. However, the two-stage model did not use any survival information in modeling the covariate process, and hence information is not utilized as efficiently as it could be. In addition, the estimated covariate values from stage one are regarded as fixed in stage two, thus the approach does not convey uncertainty from stage one to stage two. Instead of simply utilizing the predicted covariate values to find the parameters in the Cox model, Wulfsohn and Tsiatis (1997) studied the two-stage method in a different way to maximize the joint likelihood of the covariate process and survival data. The joint maximization makes more efficient use of the data by utilizing information from both the longitudinal covariates and survival simultaneously. Wulfsohn and Tsiatis (1997) used the EM algorithm to estimate all the parameters in covariate and survival processes together, assuming random effects that characterize the longitudinal covariate process are normally distributed. An attractive feature of this likelihood-based approach is its robustness against departure from the normal random effects assumption. Hsieh et al. (2006) confirmed that the likelihood-based procedure with normal random effects can be very efficient and robust as long as the longitudinal data are not too sparse or do not carry too large measurement errors.

In contrast, considering other situations where the normality assumption on random effects is violated or regarded as being too strict, Tsiatis and Davidian (2001) proposed conditional score estimators. The underlying idea of the conditional score approach was to treat the random effects as nuisance parameters for which a sufficient statistic may be derived and construct a set of estimating equations conditioning on the sufficient statistic. Then, the resulting estimating equations can be free of the

random effects. The proposed model is semiparametric in the sense that it does not require any distributional assumption on the random effects. Song et al. (2002) also proposed another semiparametric model in which parametric assumptions on the distribution of random effects may be relaxed to those following a smooth density. They took a likelihood-based approach with the EM algorithm for the estimation procedure. An important feature of this procedure, in contrast to the conditional score approach, is that the investigation of robustness to parametric assumptions on the random effects is possible. Song and Wang (2008) proposed an even more flexible semiparametric model by adapting time-varying coefficients to the proportional hazards model of the failure time, which allows the effect of coefficients to vary over time, in addition to no distributional assumptions on the underlying longitudinal covariate process. An estimation procedure was constructed based on the conditional score estimators, and asymptotic properties of the estimators were derived based on martingale and empirical process theories.

Pattern-mixture models Vonesh et al. (2006) presented a joint model of longitudinal and survival data, focusing on the estimation and comparison of serial trends over time while adjusting for possible informative censoring due to patient dropout. They strongly addressed the need for accounting for non-ignorable dropout/death through extensive simulation studies. They used the generalized linear mixed effects model for repeated measurements and a family of accelerated failure time (AFT) models for conditioning the event time. The presented joint model was relatively flexible in that the family of AFT models includes various proportional hazards models (e.g. Weibull, extreme value, piecewise exponential) and non-proportional hazards models (e.g. log-normal). An alternative joint model was introduced by Liu et al. (2007) for medical cost repeatedly recorded at fixed time intervals in the presence of a terminating event, such as death. Both Vonesh et al. (2006) and Liu et al. (2007)

modeled the terminal event as a function of covariates and linked the terminal event to the pattern of repeated measures through shared random effects by the longitudinal and survival components. Taking the likelihood-based approach, Vonesh et al. (2006) used maximum likelihood (ML) estimation with the approximated observed log-likelihood through the second-order Laplace's method, while Liu et al. (2007) used the ML estimation through the EM algorithm.

Simultaneous models Henderson et al. (2000) considered both longitudinal data and recurrent or single event time to be equally important endpoints and jointly formulated them via correlated latent Gaussian processes. For clustered data, Ratcliffe et al. (2004) proposed a joint model for longitudinal and survival outcomes of interest, which linked the two outcomes at the cluster-level random effects. In their method, repeated measures were modeled using a mixed effects model that incorporates both subject-level and cluster-level random effects, and survival data were modeled using a Cox model with the cluster-level random effects to allow for between-cluster heterogeneity. While most of the joint models associated repeated measures with survival data via common random effects or latent processes, Zeng and Cai (2005a) allowed every unobserved random factor to differently affect the longitudinal measure and survival time. Commonly, ML estimation was used with EM algorithm in Henderson et al. (2000), Ratcliffe et al. (2004), and Zeng and Cai (2005a). However, the asymptotic properties of the proposed ML estimators were established for the first time by Zeng and Cai (2005a).

2.2 Models for Cure-Survival Data

A cure model is applicable when there exist 'immunes' or 'long-term survivors' in survival data. As a result of cure, cured subjects never experience an event endpoint

but are censored because cure can never be observed. On the other hand, susceptible subjects would eventually develop the endpoint if followed for long enough. The primary interest in such studies can be on the effect of covariates on the cure rate as well as on the time-to-event. In this section, we review the approaches of modeling cure in survival analysis, which do not involve any longitudinal data, in Sections 2.2.1 - 2.2.3.

2.2.1 Mixture Cure Models

One of the commonly used cure models is the so-called 'mixture model', named after the basic concept that the underlying population consists of two subpopulations of the cured and non-cured. The mixture cure model is the mixture of a certain proportion $\pi(\boldsymbol{Z}_i)$ belonging to the cured subpopulation and the remaining fraction $1 - \pi(\boldsymbol{Z}_i)$ being not cured, such that

$$S_{pop}(t \mid \boldsymbol{Z}_i) = \pi(\boldsymbol{Z}_i) + \{1 - \pi(\boldsymbol{Z}_i)\}S_{uc}(t),$$

where $\boldsymbol{Z}_i$ is the vector of covariates, and $S_{uc}(t)$ is the conditional survival function for the uncured population. It is assumed that all patients in the non-cured subpopulation will eventually experience the event while those in the cured subpopulation will never. Early work on such models was done by Berkson and Gage (1952), Farewell (1982, 1986), and Yamaguchi (1992) under completely specified parametric models. Berkson and Gage (1952) used a mixture of exponential distributions with a constant cure fraction $\pi(\boldsymbol{Z}_i) = \pi$. Farewell (1982) adopted the Weibull regression for survival and the logistic regression for the cure fraction give by

$$\pi(\boldsymbol{Z}_i) = \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)/(1 + \exp((\boldsymbol{\beta}^T \boldsymbol{Z}_i)). \tag{2.5}$$

Yamaguchi (1992) applied a cure model with a logistic mixture probability model (2.5) and an accelerated failure time model with generalized gamma distribution.

Laska and Meisner (1992) extensively studied the cure model, specifically nonparametric failure time models, adapting Kaplan and Meier (1958) estimation. More recent work has focused on semiparametric approaches, mixtures of the cure fraction modeled through a logistic link (2.5) and the survival distribution with a complete or partial nonparametric component. Taylor (1995) introduced a more flexible mixture cure model, an extension of Farewell (1982), by retaining the conditional survival distribution for uncured individuals as a completely unspecified function. To investigate the effects of covariates on the time to event, other semiparametric mixture models have been proposed (Kuk and Chen, 1992; Sy and Taylor, 2000; Peng and Dear, 2000; Lu and Ying, 2004). Kuk and Chen (1992) estimated the regression parameters first by eliminating the baseline survival function via a Monte Carlo approximation of a marginal likelihood, and then estimated the baseline survival function using an EM algorithm, given the regression parameter estimates. However, Sy and Taylor (2000) and Peng and Dear (2000) studied alternative estimation techniques using the classic EM algorithm, to compute estimates for both the parametric and nonparametric components. The theoretical properties of the resulting estimators for the proportional hazards cure model remain to be established. Lu and Ying (2004) considered a class of transformation models for the event time. They proposed to use generalized estimating equations for parameter estimation, and the asymptotic properties were established by the usual counting process and its associated martingale theory. However, their approach was limited to only time-independent covariates due to the form of transformations. Although the mixture cure model is intuitively appealing, it involves several unresolved issues discussed by Farewell (1986), Laska and Meisner (1992), Taylor (1995), Chen et al. (1999) and Ibrahim et al. (2001). One problem

associated with the mixture model is identifiability. This arises due to the lack of information at the end of the follow-up period, caused by a significant proportion of censored subjects before the end of study. As a result, we can have difficulties in distinguishing whether the information from the censored subjects should be a part of cured group or susceptible group.

2.2.2 Promotion Time Cure Models

An alternative way to incorporate the cure fraction in survival analysis is the promotion time cure model, or referred to as the bounded cumulative hazard model (Yakovlev et al., 1996). The literature existing on the promotion time cure models is mainly the Bayesian context since the population survival function is improper. These models have been proposed and studied by Yakovlev et al. (1996), Tsodikov (1998), and Chen et al. (1999), among others. The promotion time cure model was motivated by cancer clinical trials under the biological assumption that a patient has N metastatic tumor cells remaining after treatment. Let N_i be the number of metastatic cancerous cells of the ith patient, which is an unobservable latent variable. The N_i's are assumed to have a Poisson distribution with mean $\pi(Z_i)$. We denote the time for the kth metastatic cancer cell to produce a detectable tumor (promotion time) by $\tilde{T}_k$ $(k = 1, \ldots, N_i)$ and assume that, conditional on N_i, $\tilde{T}_k$'s are identically independently distributed with the cumulative distribution function $F(t)$. If we understand $F(t) = 1 - S_{uc}(t)$, it can be interpreted similarly to the distribution function for the uncured patients in the mixture model. Then, the time to relapse of cancer for the ith patient, defined by $T_i = min\{\tilde{T}_1, \ldots, \tilde{T}_{N_i}\}$, has a form of the population

survival function

$$\begin{aligned} S_{pop}(t|\boldsymbol{Z}_i) &= P\left[N_i = 0\right] + \sum_{k \geqslant 1} P\left[\tilde{T}_1 > t, \ldots, \tilde{T}_{N_i} > t \,|\, N_i = k\right] P\left[N_i = k\right] \\ &= \exp\{-\pi(\boldsymbol{Z}_i)\} + \sum_{k \geqslant 1} \{1 - F(t)\}^k \, \frac{\pi(\boldsymbol{Z}_i)^k \exp\{-\pi(\boldsymbol{Z}_i)\}}{k!} \\ &= \exp\{-\pi(\boldsymbol{Z}_i) F(t)\}. \end{aligned} \tag{2.6}$$

In the promotion time cure model (2.6), the survival function is integrated into one formulation regardless of cured or uncured. The hazard function is given by $\pi(\boldsymbol{Z}_i) f(t)$, where $f(t) = dF(t)/dt$. Thus, we can see that the model (2.6) retains the proportional hazards structure when the covariates $\boldsymbol{Z}_i$ are formulated through $\pi(\boldsymbol{Z}_i) = \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)$. Moreover, if the regression coefficients $\boldsymbol{\beta}$ include an intercept term, say β_0, the baseline cumulative hazard function is equal to $\exp(\beta_0) F(t)$, which implies that the model (2.6) becomes the Cox's proportional hazards model with a bounded baseline cumulative hazard. For the cured patients, the survival rate at $t = \infty$ can be interpreted as the cure rate, i.e., the cure rate is $S_{pop}(\infty) = \exp\{-\pi(\boldsymbol{Z}_i)\} \neq 0$, leading to an improper survival function.

2.2.3 Transformation of Promotion Time Cure Models

In model (2.6), the independent assumption on $\{\tilde{T}_k \,|\, N_i;\, k = 1, \ldots, N_i\}$ may not be realistic in practice since they have common features shared by the same patient, such as the patient's underlying health condition or dietary habits. As a solution to adjust the correlated cancer progression times, Zeng et al. (2006) have introduced a subject-specific frailty ζ_i and assumed that, given (N_i, ζ_i), $\tilde{T}_k$'s are mutually independent with the distribution function $F(t)$. Moreover, ζ_i makes the most of an opportunity to reflect the underlying heterogeneity for the rate of metastatic cancer cells by the

assumption that N_i follows the Poisson distribution with mean $\zeta_i \pi(\boldsymbol{Z}_i)$, conditional on (Z_i, ζ_i). Following the similar derivation to (2.6), the resulting survival function for the time to relapse T takes a form

$$S(t|\boldsymbol{Z}_i) = E_{\zeta_i}\left[\exp\{-\pi(\boldsymbol{Z}_i)F(t)\zeta_i\}\right], \tag{2.7}$$

where E_{ζ_i} denotes the expectation with respect to ζ_i. Explicitly specifying the distribution for ζ_i as a gamma distribution with unit mean and variance η, we can express (2.7) as

$$\begin{aligned} S(t|\boldsymbol{Z}_i) &= [1 + \eta\pi(\boldsymbol{Z}_i)F(t)]^{-1/\eta} \\ &= G_\eta\left(\pi(\boldsymbol{Z}_i)F(t)\right), \end{aligned} \tag{2.8}$$

where $G_\eta(.)$ has a form of transformations with a parameter η such that

$$G_\eta(x) = \begin{cases} (1+\eta x)^{-1/\eta}, & \eta > 0 \\ \exp(-x), & \eta = 0. \end{cases}$$

This class of transformations includes the proportional hazards model (when $\eta = 0$) and the proportional odds model (when $\eta = 1$) as special cases.

2.3 Models for Longitudinal Data and Recurrent Events

For the analysis of longitudinal data with informative observation times, a variety of joint models have been developed. Instead of considering a common set of observation times across all subjects, Lin and Ying (2001), Lin et al. (2004), and Sun et al.

(2005), among others, proposed to use counting processes to describe arbitrary observation times. The counting process approach allowed subject-specific observation times through directly adjusted covariate effects, thereby providing a flexible tool for modeling the observation process. For the longitudinal component, Lin and Ying (2001) and Sun et al. (2005) modeled the pattern of longitudinal outcomes using a partially linear model, whereas Lin et al. (2004) modeled that using a nonparametric function of linear coviariate effects.

In these models, different assumptions have been made for the longitudinal outcome and observation processes. In Lin and Ying (2001), the observation process is assumed to be independent of the longitudinal outcome process after adjusting for some external covariates. In Lin et al. (2004), the intensity of the observation at time t is assumed to be independent of the longitudinal outcomes at that time point given the past observed data; whereas in Sun et al. (2005), the longitudinal outcome at time t is assumed to be dependent only on some external covariates and the past observation history such as the total and recent numbers of observations. Among them, the commonly used approaches were the marginal models based on estimating equations for both longitudinal data and time processes. Under these marginal approaches, it is challenging to obtain efficient estimators and also impossible to predict future outcomes of an individual given the past information.

An alternative approach was suggested by Liang et al. (2009). Based on the idea that the observation process may be correlated with the longitudinal outcomes through some unmeasured confounders even after conditioning on external covariates in practice, they studied the joint modeling approach using random effects. The longitudinal outcomes with irregular observation times were modeled through a partially linear mixed model and the informative observation process was modeled by adopting a frailty nonhomogeneous Poisson process structure. However, their method is

limited to the case where both the distribution of frailty and the conditional linear mean structure between the random effects in longitudinal and observation processes can be specified.

2.4 Models for Recurrent and Terminal Events

In this section, we review previous research on joint modeling of recurrent and terminal events. Statistical methodology and theory for analyzing recurrent event data are typically developed based on non-informative censoring times. In many applications, however, when a failure event serves as a part of the censoring mechanism, meaning that the failure event terminates observing further recurrent events (so-called informative censoring), the independent censoring assumption can be violated. For example, if the rate of recurrent tumors is high in a patient, this patient is also subject to increased risk of death. The most popular solution to model or control the dependence of recurrent events with a terminal event or informative censoring is a joint modeling approach.

Joint (or shared) frailty (or random effects) models have been studied by several authors. In these models, the dependence between recurrent and terminal events were specified via a common frailty variable allowed to have a multiplicative effect on their respective rates. The most popular distributional assumption on the frailty was a gamma distribution with unit mean to avoid the non-identifiability issue (Lancaster and Intrator, 1998; Liu et al., 2004; Ye et al., 2007; Huang and Liu, 2007). Lancaster and Intrator (1998) considered joint parametric modeling of recurrent event and survival data, using Poisson processes for the rate functions of the recurrent and terminal events. Liu et al. (2004) considered proportional hazards frailty models where the recurrent and terminal event processes were jointly modeled by a shared gamma frailty.

The frailty effect was allowed to be different for the two processes and time-dependent covariates could be incorporated separately in both processes. A Monte Carlo EM algorithm with a Metropolis-Hastings sampler in the E-step was adapted to obtain the maximum likelihood estimators. However, the Monte Carlo EM algorithm is often computationally inefficient and less accurate than the standard EM algorithm. Instead, Rondeau et al. (2007) studied a penalized likelihood approach, to estimate parameters in the model proposed by Liu et al. (2004), adopting the sum of squared norms of the second derivatives of the intensity and hazard functions as the roughness penalty.

Without distributional assumptions on the latent variables and censoring time, Wang et al. (2001) modeled the occurrence rate function for recurrent events with informative censoring in semiparametric and nonparametric ways. They assumed a subject-specific nonstationary Poisson process via a latent variable. However, the proposed model is not applicable to situations where inferences for both the recurrent and terminal events are of interest. To overcome this limitation, Huang and Wang (2004) presented a joint model for recurrent event process and a failure time, while informative censoring is allowed for observing both the recurrent events and failure times. They assumed that the recurrent, failure, and censoring events are mutually independent conditioning on the covariates and latent variables. They proposed a "borrow-strength estimation" procedure, in which first the value of the latent variable was estimated from recurrent event data, then the estimated value was used in the failure time model. Since the proposed approach did not utilize the information of the failure times in estimating the latent effect, it might be less efficient than it could be. The semiparametric models proposed by Wang et al. (2001) and Huang and Wang (2004) are flexible in that no parametric assumption was imposed on the frailty by treating as a nuisance parameter, however, these models are not applicable

to time-dependent covariates.

Most of the existing work required the proportional intensity or hazards assumption and assumed time-independent covariates. Recently, Zeng and Lin (2009) developed transformation models for the recurrent and terminal events that can deal with non-proportional hazards as well as time-varying covariates. Their proposed models are flexible enough as one can choose different forms of transformation for the respective events and as the class of transformations includes a variety of models of interest such as the proportional hazards model and the proportional odds model. Also, there is a wide range of choices open for the distribution of the shared random effects as long as satisfying the imposed conditions.

Chapter 3

Joint Models of Longitudinal Data and Recurrent Events with Informative Terminal Event

3.1 Introduction

In many biomedical studies, data are collected from patients who experience the same type of event multiple times, such as repeated hospital admissions or medical emergency episodes, recurrent strokes, multiple infection episodes, or tumor recurrences. At the same time, some longitudinal biomarkers are observed either at the time of occurrence of the event or at regular clinic visits. In addition, some subjects may experience a terminal event such as death. As longitudinal markers, recurrent events, and death are dependent on and informative of one another, analyzing one or two of these processes but ignoring the dependence from the other processes may lead to bias or result in inefficient inference. Therefore, it is important to jointly model longitudinal markers, recurrent events, and death altogether. In this way, we will be able to make the most efficient use of all data and identify the effects of variables

after correctly controlling the interplay among these processes.

There is scant literature considering the dependence of a terminal event in modeling both repeated measures and recurrent event processes. Most recently, Liu and Huang (2009) have developed a joint model for repeatedly measured CD4 cell counts and related opportunistic infection recurrences while associating their relationship with the mortality of HIV patients in the CPCRA (AIDS) study. In this study, since the CD4 cell counts were observed at scheduled visits, the observation times were non-informative. However, when the CD4 cell counts are measured at emergency admissions or unexpected hospitalizations, the information of the number and times of observations is critical, and hence it should be taken into account in modeling. By treating the hospital visits as recurrent events, Liu et al. (2008) presented a joint model of the medical cost process for chronic heart failure patients in the presence of informative observation times and a dependent death event. The joint modeling approaches by Liu et al. (2008) and Liu and Huang (2009) required the proportionality assumption for both recurrent and terminal events. In cases where the proportionality assumption does not hold, their joint models may yield biased estimators. In their inference procedures, the piecewise constant functions were adopted for estimating the underlying baseline intensity and hazards functions; however, there was no general rule for selecting the number of knots that led to the best reflection of the underlying baseline intensity functions. Moreover, the theoretical properties of the suggested estimators have not been established.

In this paper, we will use general transformation models for both the recurrent events and the terminal event, while accounting for the dependence among these two event processes and longitudinal data. Our transformation models include the proportional hazards models and the proportional odds models as special cases. We will propose efficient estimates and establish their asymptotic properties. The rest

of the chapter is organized as follows. In Section 3.2, we introduce joint models for longitudinal measurements and recurrent events in the presence of a terminal event. In Section 3.3, we estimate all the parameters using the nonparametric maximum likelihood estimation (NPMLE) and provide computationally simple algorithms to implement the proposed inference procedure. The theoretical work that shows the weak convergence and efficiency of the proposed NPMLEs is given in Section 3.4. Section 3.5 evaluates the performance of the proposed method through extensive simulation studies, and the application to the Atherosclerosis Risk in Communities (ARIC) data is reported in Section 3.6. We conclude with some remarks in Section 3.7. In Section 3.8, the EM algoritm is described in more detail, and the proof of the established asymptotic properties are provided in Section 3.9.

3.2 Joint Models

Let $Y(t)$ denote the longitudinal outcome measured at time t, $N^*(t)$ denote the number of the recurrent events occurring by time t, and T be the time to the terminal event. We introduce latent random effects to account for the association among these processes. Particularly, let $b = (b_1^T, b_2^T)^T$ denote the subject-specific random effects following a multivariate normal distribution with mean zeros and covariance matrix Σ_b. Let $Z(t)$ be a vector of external covariates, possibly time-varying at time t, plus the unit component. We assume that $Y(\cdot)$, $N^*(\cdot)$, and T are independent conditional on $Z(\cdot)$ and b. We then propose the following joint models:

$$
\begin{aligned}
Y(t|Z;\, b) &= \beta_1^T Z_1(t) + b_1^T \tilde{Z}_1(t) + \epsilon(t), && (3.1)\\
\Lambda_R(t|Z;\, b) &= G_R\Big(\int_0^t \exp\Big\{\beta_2^T Z_2(s) + b_2^T \tilde{Z}_2(s)\Big\}\, d\Lambda_R(s)\Big), && (3.2)\\
\Lambda_T(t|Z;\, b) &= G_T\Big(\int_0^t \exp\Big\{\beta_3^T Z_3(s) + (b \circ \phi)^T \tilde{Z}_3(s)\Big\}\, d\Lambda_T(s)\Big), && (3.3)
\end{aligned}
$$

where $\beta^T = (\beta_1^T, \beta_2^T, \beta_2^T)$ is a vector of unknown regression parameters, $\Lambda_R(\cdot)$ and $\Lambda_T(\cdot)$ are unspecified increasing functions, $\phi^T = (\phi_1^T, \phi_2^T)$ is a set of unknown constants, and $b \circ \phi$ denotes the component-wise product of b and ϕ. Both $Z_i(t)$ and $\tilde{Z}_i(t)$ $(i = 1, 2, 3)$ are some subsets of $Z(t)$, but $Z_2(t)$ and $Z_3(t)$ do not have the unit component. This allows that each of three outcomes $(Y(\cdot), N^*(\cdot), T)$ can depend on different sets of predictors. Additionally, $\epsilon(t)$ is a white noise process with mean zero and variance σ_e^2. Both G_R and G_T are continuously differentiable and strictly increasing transformation functions to be specified in the analysis. For example, $G_R(x)$ and $G_T(x)$ can take a form of the logarithmic transformation,

$$\begin{cases} \log(1+\gamma x)/\gamma, & \gamma > 0 \\ x, & \gamma = 0, \end{cases}$$

or a form of the Box-Cox transformation,

$$\begin{cases} \{(1+x)^\gamma - 1\}/\gamma, & \gamma > 0 \\ \log(1+x), & \gamma = 0. \end{cases}$$

According to the choice of γ in both classes of transformations, the transformation model can be the proportional hazards model or the proportional odds model.

Note in (3.3) that the hazards of the terminal event depends on longitudinal measures through the shared b_1 and on recurrent events through the shared b_2, respectively. Thus, ϕ characterizes a degree of dependence explained by unobserved latent factors in (3.1) and (3.2); $\phi = 0$ implies that the dependence can be fully explained by the observed covariates. In addition, longitudinal measures are related to recurrent events through correlations between b_1 and b_2.

Let C be the non-informative censoring time assumed to be independent of $(Y(\cdot)$, $N^*(\cdot)$, T, $b)$ given Z, and let $X = \min(T, C)$ denote the observed terminal event

time. The observed data for the ith subject with m_i repeated measurements are $\{Y_i(t_{ik}),\ N_i(t),\ X_i,\ \Delta_i,\ Z(t)\,;\ t_{ik} \leqslant X_i,\ t \leqslant X_i,\ i = 1, \ldots, n,\ k = 1, \ldots, m_i\}$, where $N_i(t) = N_i^*(t \wedge X_i)$, $\Delta_i = I(T_i \leqslant C_i)$ with $I(\cdot)$ being the indicator function. Under models (3.1)-(3.3), the log-likelihood function of the observed data from a random sample of n subjects is given by

$$
\begin{aligned}
\sum_{i=1}^{n} \log \int_b \prod_{k=1}^{m_i} \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ \frac{-(Y_i(t_{ik}) - \beta_1^T Z_{1i}(t_{ik}) - b_1^T \tilde{Z}_{1i}(t_{ik}))^2}{2\sigma_e^2} \right\} \right] \\
\times \prod_t \left[\lambda_R(t)\, e^{\beta_2^T Z_{2i}(t) + b_2^T \tilde{Z}_{2i}(t)}\, G_R'\Big(\int_0^t e^{\beta_2^T Z_{2i}(s) + b_2^T \tilde{Z}_{2i}(s)}\, d\Lambda_R(s) \Big) \right]^{R_i(t)\Delta N_i^*(t)} \\
\times \exp\left\{ -G_R\Big(\int_0^{X_i} e^{\beta_2^T Z_{2i}(t) + b_2^T \tilde{Z}_{2i}(t)}\, d\Lambda_R(t) \Big) \right\} \\
\times \left[\lambda_T(X_i)\, e^{\beta_3^T Z_{3i}(X_i) + (b \circ \phi)^T \tilde{Z}_{3i}(X_i)}\, G_T'\Big(\int_0^{X_i} e^{\beta_3^T Z_{3i}(t) + (b \circ \phi)^T \tilde{Z}_{3i}(t)}\, d\Lambda_T(t) \Big) \right]^{\Delta_i} \\
\times \exp\left\{ -G_T\Big(\int_0^{X_i} e^{\beta_3^T Z_{3i}(t) + (b \circ \phi)^T \tilde{Z}_{3i}(t)}\, d\Lambda_T(t) \Big) \right\} \times f(b;\, \Sigma_b)\, db,
\end{aligned}
$$

where $R_i(t) = I(X_i \geqslant t)$ is the indicator for the risk set, $\Delta N_i^*(t)$ denotes the jump size of the underlying recurrent event at time t, $f(b;\, \Sigma_b)$ denotes the multivariate normal density function of b with covariance matrix Σ_b, and $\lambda_R(t) = \Lambda_R'(t)$ and $\lambda_T(t) = \Lambda_T'(t)$ are the derivatives of Λ_R and Λ_T, respectively. Note in (3.1) and (3.2) that the obervation times of longitudinal outcomes do not need to be the same as the recurrent event times. Instead, the longitudinal measures may be observed at some scheduled visits or at the times when the recurrent events occur.

3.3 Inference Procedure

3.3.1 Nonparametric Maximum Likelihood Estimation

We propose to use the nonparametric maximum likelihood estimation (NPMLE) for estimating parameters (β, ϕ, σ_e^2, Σ_b) and infinite-dimensional parameters $\Lambda_R(t)$ and $\Lambda_T(t)$. In the log-likelihood, we assume the cumulative intensity function $\Lambda_R(t)$ and the cumulative hazards function $\Lambda_T(t)$ to be step functions with the jumps at the observed event times, and we replace the intensity $\lambda_R(t)$ and the hazards $\lambda_T(t)$ with the jump size of Λ_R and Λ_T at time t, denoted by $\Lambda_R\{t\}$ and $\Lambda_T\{t\}$, respectively. The modified log-likelihood function is given by

$$
\begin{aligned}
& l_n(\beta,\, \phi,\, \sigma_e^2,\, \Sigma_b,\, \Lambda_R,\, \Lambda_T) \\
&= \sum_{i=1}^{n} \log \int_b \prod_{k=1}^{m_i} \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ \frac{-(Y_i(t_{ik}) - \beta_1^T Z_{1i}(t_{ik}) - b_1^T \tilde{Z}_{1i}(t_{ik}))^2}{2\sigma_e^2} \right\} \right] \\
&\quad \times \prod_t \left[\Lambda_R\{t\}\, e^{\beta_2^T Z_{2i}(t) + b_2^T \tilde{Z}_{2i}(t)}\, G_R'\Big(\int_0^t e^{\beta_2^T Z_{2i}(s) + b_2^T \tilde{Z}_{2i}(s)}\, d\Lambda_R(s) \Big) \right]^{R_i(t)\Delta N_i^*(t)} \\
&\quad \times \exp\left\{ -G_R\Big(\int_0^{X_i} e^{\beta_2^T Z_{2i}(t) + b_2^T \tilde{Z}_{2i}(t)}\, d\Lambda_R(t) \Big) \right\} \\
&\quad \times \left[\Lambda_T\{X_i\}\, e^{\beta_3^T Z_{3i}(X_i) + (b \circ \phi)^T \tilde{Z}_{3i}(X_i)}\, G_T'\Big(\int_0^{X_i} e^{\beta_3^T Z_{3i}(t) + (b \circ \phi)^T \tilde{Z}_{3i}(t)}\, d\Lambda_T(t) \Big) \right]^{\Delta_i} \\
&\quad \times \exp\left\{ -G_T\Big(\int_0^{X_i} e^{\beta_3^T Z_{3i}(t) + (b \circ \phi)^T \tilde{Z}_{3i}(t)}\, d\Lambda_T(t) \Big) \right\} \times f(b;\, \Sigma_b)\, db. \qquad (3.4)
\end{aligned}
$$

Hence the likelihood can be expressed as a function of a finite number of parameters, which include (β, ϕ, σ_e^2, Σ_b) and the jump sizes of Λ_R and Λ_T.

3.3.2 EM Algorithm

To obtain the NPMLEs and their variance estimators, we use the expectation-maximization (EM) algorithm (Dempster et al., 1977), treating the subject-specific random effects

b_i as missing data. In the E-step, we compute the conditional expectations of the log-likelihood for the complete data, given the observed data and the current parameter estimates. Particularly, using numerical approximation methods such as the Gaussian quadrature, we can evaluate the integration of certain functions of b_i, say $g(b_i)$. We denote such expectation by $\hat{E}[g(b_i) \,|\, Y_i(t), N_i(t), X_i, \Delta_i, Z(t)]$, hereafter abbreviated as $\hat{E}[g(b_i)]$. In the M-step, we maximize the conditional expectation of the complete-data log-likelihood function given the observed data. Specifically, the closed-forms of the maximizers exist for $(\beta_1,\, \sigma_e^2,\, \Sigma_b)$ as follows:

$$
\begin{aligned}
\hat{\beta}_1 &= (Z_1^T Z_1)^{-1} Z_1^T (Y - \hat{E}[\tilde{Z}_1 b_1]), \\
\hat{\sigma}_e^2 &= \hat{E}[(Y - Z_1\beta_1 - \tilde{Z}_1 b_1)^T (Y - Z_1\beta_1 - \tilde{Z}_1 b_1)] \;/\; \sum_{i=1}^{n} m_i, \\
\hat{\Sigma}_b &= \hat{E}[b_1\, b_1^T],
\end{aligned}
$$

where Y denotes the vector of longitudinal measurements at the observed times, and Z_1 and $\tilde{Z}_1$ denote matrices with each row equal to the observed covariates $Z_1(t)^T$ and $\tilde{Z}_1(t)^T$ at the same times, respectively. For the rest of parameters $(\beta_2,\, \beta_3,\, \phi,\, \Lambda_R\{.\},\, \Lambda_T\{.\})$, the quasi-Newton algorithm is used to update the parameter estimates at each M-step.

When covariates of the recurrent and terminal events $(Z_2, \tilde{Z}_2, Z_3, \tilde{Z}_3)$ are time-independent, we propose to use recursive formulae, provided in Section 3.8, in order to reduce the number of parameters to be maximized to a very small set of parameters. Basic ideas of the recursive formulae can be described as follows. In the forward recursive formula, since $\Lambda_R(t)$ and $\Lambda_T(t)$ can be calculated from the jumps which are observed before time t, only $(\lambda_{1R},\, \lambda_{1T})$ are involved in the quasi-Newton iteration, where λ_{1R} and λ_{1T} are the jump sizes at the first observed event times of the recurrent and terminal events, respectively. In the backward recursive formula, similarly, $\Lambda_R(t)$

and $\Lambda_T(t)$ can be expressed as a function of the jumps which are observed after time t and the sum of all jumps. Thus, the backward recursive formula requires to maximize only the last jump sizes and the sums of all observed jump sizes of the recurrent and terminal events.

To estimate the variances and covariances of the NPMLEs, we compute the observed information matrix via the Louis formula (Louis, 1982) as given in Section 3.8. Then, the inverse of the observation information is the estimator of the covariance matrix of the NPMLEs.

3.4 Asymptotic Properties

Let θ be the vector of (β, ϕ, σ_e^2, Vec(Σ_b)) and let (θ_0, $\Lambda_{0R}(t)$, $\Lambda_{0T}(t)$) be the true parameter values of (θ, $\Lambda_R(t)$, $\Lambda_T(t)$), where Vec(Σ_b) denotes the vector consisting of the upper triangular elements of Σ_b. We then establish the asymptotic properties of the NPMLEs under the following conditions:

(A1) The parameter value θ_0 belongs to the interior of a compact set Θ within the domain of θ. Additionally, $\Lambda'_{0R}(t) > 0$ and $\Lambda'_{0T}(t) > 0$, for all $t \in [0, \tau]$, where τ is the duration of the study.

(A2) With probability 1, $Z(.)$ is left-continuous with uniformly bounded left and right derivatives in $[0, \tau]$.

(A3) For some constant δ_0, $P(C \geqslant \tau | Z) > \delta_0 > 0$ with probability 1.

(A4) $E[N^*(\tau)] < \infty$ with probability 1.

(A5) For some positive constant M_0, $M_0^{-1} < \sigma_{0e}^2 < M_0$ and $M_0^{-1} < c^T \Sigma_{0b}\, c < M_0$ for any constant vector $\|c\| = 1$.

(A6) The transformation functions $G_R(.)$ and $G_T(.)$ are four-times differentiable with

$G_R(0) = G_T(0) = 0$, $G'_R(0) > 0$, and $G'_T(0) > 0$. In addition, there exist positive constants μ_0 and κ_0 such that for any integer $m \geqslant 0$ and for any sequence $0 < x_1 < ... < x_m \leqslant y$,

$$\begin{aligned} \prod_{j=1}^{m} \{(1 + x_j)\, G'_R(x_j)\} \exp\{-G_R(y)\} &\leqslant \mu_0^m (1 + y)^{-\kappa_0}, \quad \text{and} \\ (1 + x)\, G'_T(x) \exp\{-G_T(x)\} &\leqslant \mu_0 (1 + x)^{-\kappa_0}. \end{aligned}$$

Furthermore, there exists a constant $\rho_0 > 0$ such that

$$\sup_x \left\{ \frac{|G''_R(x)| + |G^{(3)}_R(x)| + |G^{(4)}_R(x)|}{G'_R(x)\,(1 + x)^{\rho_0}} \right\} + \sup_x \left\{ \frac{|G''_T(x)| + |G^{(3)}_T(x)| + |G^{(4)}_T(x)|}{G'_T(x)\,(1 + x)^{\rho_0}} \right\} < \infty,$$

where $G^{(3)}_R$, $G^{(4)}_R$, $G^{(3)}_T$, and $G^{(4)}_T$ are the third and fourth derivatives.

(A7) For some $t \in [0, \tau]$, if there exist a deterministic function $c(t)$ and v such that $c(t) + v^T Z(t) = 0$ with probability 1, then $c(t) = 0$ and $v = 0$.

(A8) For some $t \in [0, \tau]$, $\tilde{Z}_i^T(t)\tilde{Z}_i(t)$ $(i = 1, 2)$ has full rank with some positive probability.

(A9) Let K be the number of repeated measures and let d_b be the dimension of b_1. With probability one, $P(K > d_b) > 0$.

Conditions (A1) - (A3) are the standard assumptions for survival analysis. Conditions (A4) - (A5) are necessary to prove the existence of the NPMLEs. It can be easily verified that Condition (A6) holds for all transformations commonly used, including the classes of Box-Cox and logarithmic transformations described in Section 3.2. Conditions (A7) - (A8) entail the linear independence of covariates for the fixed and random effects. Condition (A9) prescribes that some subjects have at least d_b repeated measures.

Under the above conditions, the following theorem shows the consistency of the NPMLEs $(\hat{\theta}, \hat{\Lambda}_R, \hat{\Lambda}_T)$.

Theorem 3.1. *Under Conditions (A1) - (A9),*

$$|\hat{\theta} - \theta_0| \to 0, \quad \sup_{t\in[0,\tau]} |\hat{\Lambda}_R(t) - \Lambda_{0R}(t)| \to 0, \quad \sup_{t\in[0,\tau]} |\hat{\Lambda}_T(t) - \Lambda_{0T}(t)| \to 0, \quad a.s.$$

Theorem 3.1 then leads to the following results on the asymptotic normality of $(\hat{\theta}, \hat{\Lambda}_R, \hat{\Lambda}_T)$ and the asymptotic efficiency of $\hat{\theta}$.

Theorem 3.2. *Under Conditions (A1) - (A9),* $\sqrt{n}\,(\hat{\theta}-\theta_0,\ \hat{\Lambda}_R-\Lambda_{0R},\ \hat{\Lambda}_T-\Lambda_{0T})$ *weakly converges to a zero-mean Gaussian process in* $R^{d_\theta} \times BV[0,\tau] \times BV[0,\tau]$*, where* d_θ *is the dimension of* θ *and* $BV[0,\tau]$ *denotes the space of all functions with bounded variations in* $[0,\tau]$*. Furthermore, the asymptotic covariance matrix of* $\sqrt{n}\,(\hat{\theta} - \theta_0)$ *achieves the semiparametric efficiency bound for* θ_0*.*

Furthermore, in Section 3.9, we show that the inverse of the observed information matrix is a consistent estimator of the asymptotic covariance matrix of the NPMLEs. This result allows us to make inference for any functional of $(\theta, \Lambda_R, \Lambda_T)$. To prove Theorems 3.1 - 3.2, we apply the general asymptotic theory of Zeng and Lin (2007). The desired asymptotic properties of the NPMLEs are established followed by the arguments in Appendix B of Zeng and Lin (2007) if we can verify that their regularity conditions hold for our joint model setting. Checking the regularity conditions, however, is challenging in our cases. The detailed proofs are provided in Section 3.9.

3.5 Simulation Studies

In this section, we examined the performance of the proposed methods through extensive simulation studies. We considered a dichotomous covariate of Z_1 taking the

value of 0 or 1 with the equal probability of 0.5 and a continuous covariate of Z_2 randomly sampled from the uniform distribution on [-1, 1]. We generated data for the longitudinal outcomes from $Y(t \mid Z_1, Z_2; b_1) = 0.7 + Z_1 + 0.5Z_2 + b_1 + \epsilon(t)$, where $\epsilon(t) \sim N(0, \sigma_e^2)$ with $\sigma_e^2 = 1$, the recurrent event process from the cumulative intensity of $\Lambda_R(t \mid Z_1, Z_2; b_2) = G_R(e^{Z_1+0.5Z_2+b_2}\Lambda_R(t))$, where $\Lambda_R(t) = \nu_1 t$, and the terminal event time from the cumulative hazards of $\Lambda_T(t \mid Z_1, Z_2; b_1, b_2) = G_T(e^{Z_1+0.5Z_2+b_1\phi_1+b_2\phi_2}\Lambda_T(t))$, where $\Lambda_T(t) = \nu_2 t^2$.

For each subject, the correlation within repeated measures was reflected by a random effect of $b_1 \sim N(0, \sigma_1^2)$, and the correlation within recurrent event times was reflected by another random effect of $b_2 \sim N(0, \sigma_2^2)$. In addition, the dependence between the longitudinal measures and the recurrent event times was given by ρ, which was the correlation between (b_1, b_2). Particularly, we chose $\sigma_1^2 = \sigma_2^2 = \rho = 0.5$. We considered two cases of $\phi = (0.5, 0.2)$ and $(0, 0.2)$, where we simulated some positive dependence between the longitudinal measures and the terminal event (i.e., $\phi_1 = 0.5$) or no dependence explained by random effect b_1 (i.e., $\phi_1 = 0$) in the latter. Also, combinations of the proportional intensity or hazards models and the proportional odds models were considered to be the transformations for $G_R(.)$ and $G_T(.)$.

The non-informative censoring time C_i was randomly sampled from the uniform distribution on [1, 5], and (ν_1, ν_2) was chosen according to the considered transformation models in order to achieve the desired total number of recurrent event times of 2$\sim$ 3 and the desired censoring rate of 35%, on average. We set longitudinal observation times to be fixed intervals so that a subject had about six longitudinal measurements, on average.

The results presented in Table 3.1 and Table 3.2 are based on 1000 replications for n=200 and n=400. Table 3.1 - 3.2 include the difference between the estimate and true

parameter (Bias), the sample standard deviation of the parameter estimators (SE), and the average of the standard error estimators (SEE), and the coverage probability of 95% confidence intervals (CP). The confidence intervals for $\Lambda_R(.)$ and $\Lambda_T(.)$ are constructed based on the log transformation, and those for ρ are based on the Fisher transformation. In addition, we use the Satterthwaite approximation to compute the confidence intervals of σ_e^2, σ_1^2, and σ_2^2.

Table 3.1 shows that the NPMLEs under $G_R(x) = x$ and $G_T(x) = x$ are noticeably unbiased, the standard error estimators calculated via the Louis formula well reflect the true variations of the proposed estimators, and the coverage probabilities are in a reasonable range, even with a moderate sample size of 200. As the sample size increases to 400, the biases slightly increase for some estimates; however, they are still very small comparing to the sizes of true parameter values and the variations of the parameter estimators become smaller, and hence the coverage probabilities still lie in a reasonable range. The simulation results shown in Table 3.2 are similar to those for Table 3.1, indicating that the proposed method seems to work well for the transformation models of $G_R(x) = x$ and $G_T(x) = \log(1 + x)$. We also studied other combinations of transformations such as $(G_R(x), G_T(x)) = (x, \log(1 + x))$ and $(G_R(x), G_T(x)) = (\log(1 + x), \log(1 + x))$, and the results are similar and hence omitted here.

3.6 Data Application

We apply the proposed method to the data from the cohort components of the Atherosclerosis Risk in Communities (ARIC) study. The cohort component was designed to investigate the trends in rates of hospitalized myocardial infarction (MI) and fatal coronary heart diseases (CHD) in men and women aged 45-64 years from

four US communities; Minneapolis suburbs (Minnesota), Forsyth County (North Carolina), Washington County (Maryland), and Jackson County (Missisippi). It is well known that some risk factors for coronary heart diseases differ considerably by race and gender, therefore, our research focuses on a total of 1651 subjects who are white males living in Forsyth County.

The existing studies (Chambless et al., 2003; Wattanakit et al., 2005) found that systolic blood pressure (SBP) is an important risk factor for both incidence and recurrence of CHD event in the ARIC data. Also, we observe from the preliminary analysis that patients who have experience more recurrent CHD events are likely to be in a higher risk of death. Thus, the primary objective of this analysis is to characterize these relationships between SBP changes over time, recurrent CHD events, and death, to assess the effects of baseline covariates on these three outcomes, and to utilize the final models for the accurate prediction of risk of recurrent CHD events and death. To model such a complicated system, we propose a joint transformation random effect model for the main outcomes consisting of three components: (a) longitudinal systolic blood pressure (SBP) measures, (b) recurrent CHD events, and (c) death.

Beginning with the first screen examination (baseline) in 1987-89, longitudinal measures were collected at approximately three-year intervals, in 1990-92, 1993-95, and 1996-98. The recurrent event of interest is the (multiple) occurrences of CHD events including definite MI, probable MI, definite fatal CHD, definite fatal MI, or possible fatal CHD events, which are classified based on Mortality and Morbidity Classification Committee (MMCC) reviews or computer algorithm if MMCC reviews are not required. Follow-up process for the recurrent CHD events and death continued until 2005 through reviewing death certificates and hospital discharge records and investigating out-of-hospital deaths, while the follow-up for longitudinal measures ended with each patient's last examination (up to 1998). The median follow-up time

was 16.8 years with the largest follow-up time being 19 years, and 24% of patients died during the study period. 221, 41, and 15 patients have experienced one, two, and more than two CHD events, respectively.

In our joint model, we included the baseline covariates of age, Body Mass Index (BMI), SBP, use of hypertension lowering medications, and existence of diabetes, and visiting time in years, which were significant variables from preliminary studies fitting separate models for each of the three outcomes of interest. The subject-specific random intercepts b_1 and b_2 are included in the joint model to cope with correlations within and between three outcomes. We also considered the class of transformations $\log(1+\gamma x)/\gamma$ for $G_R(x)$ and $G_T(x)$. This class includes the proportional intensity/hazards and proportional odds models. We used the Akaike information criterion (AIC) to determine the best transformation model. Figure 3.1 shows the surface of the log-likelihood function corresponding to the combination of transformations for $G_R(x)$ and $G_T(x)$. The combination of G_R and G_T with the largest log-likelihood value corresponds to the proportional intensity model for the recurrent CHD events and the proportional odds model for death.

Table 3.3 summarizes the estimation results under the selected best model. Both age at entry and baseline measures of SBP were significant for all three outcomes. Elder patients with higher levels of SBP at baseline had the elevated SBP levels over time, a higher intensity rate of CHD occurrences, and a higher risk of death. Surprisingly, baseline BMI affected all three outcomes jointly with other factors such as a patient's age and baseline SBP level. Interactions between baseline BMI and a patient's age had a significant effect on the rates of CHD and death, while those between baseline BMI and SBP were statistically significant in explaining SBP changes over time. If a patient who had SBP higher than 152 mmHg and took hypertension medications at baseline, then the patient was likely to have lower longitudinal SBP

levels over time than those who did not take the medications. We also find that patients diagnosed with diabetes were at higher risks of CHD events and death.

For the model association, the observed covariates in the fitted model seemed to fully explain dependence between the longitudinal observations of SBP and CHD events as well as dependence between repeated measures of SBP and death. However, there seemed to be some positive correlation between recurrence of CHD and death due to the unobserved random factor. This result coincided with the expectation that patients who get admitted to hospital more frequently with CHD are at even higher risk of death. These findings may lead us some interesting application points of the analysis: 1) to predict the survival distribution after the incidence of CHD at a fixed time s, and 2) to estimate the expected SBP levels over time after the incidence of CHD at a fixed time s. To answer the question 1), the conditional survival distribution can be calculated as

$$
\begin{aligned}
&P[T > t \,|\, Z, \Delta N^*(s) = N^*(s) = 1, T > s;\, t > s] \\
&= \frac{\int_b e^{-G_T\left(\int_0^t e^{\beta_3^T Z_3(u) + (b \circ \phi)^T \tilde{Z}_3(u)}\, d\Lambda_T(u)\right)} f(\Delta N^*(s) = N^*(s) = 1, T > s | Z, b)\, f(b; \Sigma_b)\, db}{\int_b f(\Delta N^*(s) = N^*(s) = 1, T > s | Z, b)\, f(b; \Sigma_b)\, db}.
\end{aligned}
$$

Also, for the question 2), the conditional expectation of longitudinal SBPs is given by, for $t > s$,

$$
\begin{aligned}
&E[Y(t) \,|\, Z, \Delta N^*(s) = N^*(s) = 1, T > s;\, t > s] \\
&= \frac{\int_b \left\{\beta_1^T Z_1(t) + b_1^T \tilde{Z}_1(t)\right\} f(\Delta N^*(s) = N^*(s) = 1, T > s | Z, b)\, f(b; \Sigma_b)\, db}{\int_b f(\Delta N^*(s) = N^*(s) = 1, T > s | Z, b)\, f(b; \Sigma_b)\, db}.
\end{aligned}
$$

For illustration purposes, the predicted survival distribution and the longitudinal SBP levels for a subject who had one CHD event at study year 5 and average baseline measures of age, BMI, and SBP with neither hypertension nor diabetes are displayed

in Figure 3.2, along with the point-wise 95% confidence intervals. The confidence intervals are obtained by applying the functional delta method and evaluating at the NPMLEs.

3.7 Concluding Remarks

We have presented joint transformation models for repeated measures and recurrent event times with an informative terminal event. We have provided an efficient EM algorithm to compute the maximum likelihood estimators of the cumulative intensity and hazards functions as well as regression parameters. The nonparametric maximum likelihood estimators are shown to be consistent, asymptotically normal, and asymptotically efficient. The proposed approach has been applied to the ARIC data, and the resulting joint models can be used in predicting a patient's future survival rate and longitudinal measures given his/her past history.

To obtain the variance estimates of $(\hat{\theta}, \hat{\Lambda}_R(t), \hat{\Lambda}_T(t))$, we have used the inverse of the observed information matrix evaluated at the NPMLEs. Even if this approach yields consistent variance estimators, inverting such a large dimensional matrix may be intimidating if there are a large number of observations. This limitation can be overcome by using a profile likelihood (Murphy and van der Vaart, 2000) rather than a full likelihood if the parameter of interest is only θ. Particularly, let $p\ell_n(\theta)$ be the profile log-likelihood function for θ, expressed as $p\ell_n(\theta) = \sup_{\{\Lambda_R, \Lambda_T\}} \sum_{i=1}^{n} \ell_i(\theta, \Lambda_R, \Lambda_T)$, where $\ell_i(\theta, \Lambda_R, \Lambda_T)$ is the observed log-likelihood function for the ith subject. Then the negative second-order numerical difference of $p\ell_n(\theta)$ at $\theta = \hat{\theta}$ can approximate the inverse of the asymptotic variance of $\hat{\theta}$ (Zeng and Cai, 2005b). In this approach, we need to compute $\hat{\Lambda}_{R\theta}$ and $\hat{\Lambda}_{T\theta}$ which maximize the observed log-likelihood function for a fixed θ in the neighborhood of $\hat{\theta}$. By utilizing the EM algorithm in Section 3.3.2,

but holding θ fixed all the time in both the E-step and M-step, we can calculate $\hat{\Lambda}_{R\theta}$ and $\hat{\Lambda}_{T\theta}$.

Our model assumes that longitudinal measures are linearly related to all covariates considered. Where it is believed that the longitudinal measures are nonlinearly related to some predictors, we can increase the flexibility of our joint models by including some nonparametric functions of those predictors additively in the longitudinal components. We can also extend our model to multiple types of recurrent and/or terminal events.

3.8 E-step and M-step in EM Algorithm

In this section, we describe numerical algorithms to obtain the NPMLEs and their variance estimators along with recursive formulae for the jumps of the cumulative intensity and hazards functions. For simplicity, we consider time-invariant covariates in the models for recurrent and terminal events, and we define $q_{2i} = e^{\beta_2^T Z_{2i} + b_2^T \tilde{Z}_{2i}}$ and $q_{3i} = e^{\beta_3^T Z_{3i} + (b \circ \phi)^T \tilde{Z}_{3i}}$. In the E-step, we calculate the conditional expectation of $g(b)$, $b = (b_1^T, b_2^T)^T$ given the observed data $O_i = \{Y_i(t_{ik}), N_i(t), X_i, \Delta_i, Z(t)\,;\, t_{ik} \leqslant X_i,\, t \leqslant X_i,\, i = 1, \ldots, n,\, k = 1, \ldots, m_i\}$ and current parameter estimates as follows:

$$\hat{E}[g(b)] = \frac{\int_b g(b) h(O_i|\, b) f(b; \Sigma_b)\, db}{\int_b h(O_i|\, b) f(b; \Sigma_b)\, db},$$

where

$$\begin{aligned} h(O_i|\, b) = \exp &\left\{ -\sum_{k=1}^{m_i} \left[(b_1^T \tilde{Z}_{1i}(t_{ik}))^2 - 2(Y_i(t_{ik}) - \beta_1^T Z_{1i}(t_{ik})) b_1^T \tilde{Z}_{1i}(t_{ik}) \right] / 2\sigma_e^2 \right\} \\ &\times \exp \left\{ \sum_t \left[R_i(t) \Delta N_i^*(t) \{ b_2^T \tilde{Z}_{2i} + \log G_R'(q_{2i} \Lambda_R(t)) \right] - G_R(q_{2i} \Lambda_R(X_i)) \right\} \\ &\times \exp \left\{ \Delta_i [b \circ \phi + \log G_T'(q_{3i} \Lambda_T(X_i))] - G_T(q_{3i} \Lambda_T(X_i)) \right\}. \end{aligned}$$

An appropriate numerical approximation, such as Gaussian quadrature with Hermite orthogonal polynomial, can be considered to evaluate the integrals in conditional expectations. In the M-step, we maximize the following objective function of the expected log-likelihood for complete data:

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left\{-\log\sigma_e^2/2-\hat{E}\left[(Y_i(t_{ik})-\beta_1^T Z_{1i}(t_{ik})-b_1^T\tilde{Z}_{1i})^2/2\sigma_e^2\right]\right\}\\
&+\sum_{i=1}^{n}\int R_i(t)\left\{\log\Lambda_R\{t\}+\beta_2^T Z_{2i}+\hat{E}[b_2]^T\tilde{Z}_{2i}+\hat{E}[\log G_R'(q_{2i}\Lambda_R(t))]\right\}dN_i^*(t)\\
&+\sum_{i=1}^{n}\Delta_i\left\{\log\Lambda_T\{X_i\}+\beta_3^T Z_{3i}+\hat{E}[b\circ\phi]^T\tilde{Z}_{3i}+\hat{E}[\log G_T'(q_{3i}\Lambda_T(X_i))]\right\}\\
&-\sum_{i=1}^{n}\left\{\hat{E}[G_R(q_{2i}\Lambda_R(X_i))+G_T(q_{3i}\Lambda_T(X_i))\right\}\\
&+\sum_{i=1}^{n}\hat{E}[\log f(b;\Sigma_b)]. \qquad (3.5)
\end{aligned}$$

Maximizing (3.5) over (β_1, σ_e^2, Σ_b) is simple, while the rest of parameters do not yield the closed-form of maximizers, and hence it is required to involve a reliable numerical approach. We consider one-step Quasi-Newton algorithm to find a maximazer of the objective function over a set of parameters (β_2, β_3, ϕ, $\Lambda_R\{.\}$, $\Lambda_T\{.\}$).

Maximizing over a high dimension of $\Lambda_R\{.\}$ and $\Lambda_T\{.\}$ may cause a computation problem. Accordingly, we introduce a forward recursive formula and a backward recursive formula that lessen the burden on maximization. Let $w_{1r} < w_{2r} < ... < w_{m_r r}$ be the ordered recurrent event times observed and $\lambda_{1R}, \lambda_{2R}, ..., \lambda_{m_r R}$ be the jump sizes of Λ_R corresponding to those ordered time points, where m_r is the total number of the observed recurrent events. In a similar way, let $w_{1t} < w_{2t} < ... < w_{m_t t}$ be the observed terminal event times and $\lambda_{1T}, \lambda_{2T}, ..., \lambda_{m_t T}$ be the jump sizes of Λ_T corresponding to those ordered time points, where m_t is the total number of the observed terminal events. By differentiating (3.5) with respect to λ_{jR} ($j = 1, ..., m_r$) and setting the

derivative to be zero, we have

$$0 = \frac{1}{\lambda_{jR}} + \sum_{i=1}^{n} \int R_i(t)\, \hat{E}\left[I(t \geqslant w_{jr}) \frac{G_R^{''}(q_{2i}\Lambda_R(t))\, q_{2i}}{G_R^{'}(q_{2i}\Lambda_R(t))} \right] dN_i^*(t)$$
$$- \sum_{i=1}^{n} \hat{E}\left[I(X_i \geqslant w_{jr})\, G_R^{'}(q_{2i}\Lambda_R(X_i))\, q_{2i} \right],$$

where $G_R^{''}(x) = d^2 G_R(x)/dx^2$. Since the derivatives with respect to λ_{jR} and $\lambda_{j+1,R}$ are both equal to zero, we obtain the following equation of

$$\frac{1}{\lambda_{j+1,R}} = \frac{1}{\lambda_{jR}} + \sum_{i=1}^{n} \int R_i(t)\, \hat{E}\left[I(w_{jr} \leqslant t < w_{j+1,r}) \frac{G_R^{''}(q_{2i}\Lambda_R(t))\, q_{2i}}{G_R^{'}(q_{2i}\Lambda_R(t))} \right] dN_i^*(t)$$
$$- \sum_{i=1}^{n} \hat{E}\left[I(w_{jr} \leqslant X_i < w_{j+1,r})\, G_R^{'}(q_{2i}\Lambda_R(X_i))\, q_{2i} \right], \tag{3.6}$$

for $j = 1, ..., (m_r - 1)$. The fact that $\Lambda_R(s) = \lambda_{1R} + \lambda_{2R} + ... + \lambda_{jR}$ for time s such that $w_{jr} \leqslant s < w_{j+1,r}$ yields a forward recursive formula of (3.6) that calculates $\lambda_{j+1,R}$ from (λ_{1R}, ..., λ_{jR}). To obtain a backward recursive formula, we define $\zeta_r = \Lambda_R(w_{m_r r}) = \sum_{j=1}^{m_r} \lambda_{jR}$ and $f_{jr} = \lambda_{jR}/\zeta_r$, then f_{jr} is calculated from ($f_{j+1,r}$, ..., $f_{m_r r}$) and ζ_r. The backward recursive fomula for the cumulative intensity function is given by

$$\frac{1}{f_{jr}} = \frac{1}{f_{j+1,r}} - \zeta_r \sum_{i=1}^{n} \sum_{l=1}^{n_i} \hat{E}\left[I(w_{jr} \leqslant T_{il} < w_{j+1,r}) \frac{G_R^{''}(q_{2i}\check{\Lambda}_R(T_{il}))\, q_{2i}}{G_R^{'}(q_{2i}\check{\Lambda}_R(T_{il}))} \right]$$
$$+ \zeta_r \sum_{i=1}^{n} \hat{E}\left[I(w_{jr} \leqslant X_i < w_{j+1,r})\, G_R^{'}(q_{2i}\check{\Lambda}_R(X_i))\, q_{2i} \right], \tag{3.7}$$

where T_{il} is the lth observed recurrent event time, n_i is the total number of the recurrent events for the ith subject, and $\check{\Lambda}_R(s) = \zeta_r(1 - \sum_{l=j+1}^{m_r} f_{lr})$ for $w_{jr} \leqslant s < w_{j+1,r}$. Similarly, the forward recursive formula for $\lambda_{j+1,T}$ and the backward recursive

formula for f_{jt} take a form of

$$\frac{1}{\lambda_{j+1,T}} = \frac{1}{\lambda_{jT}} + \sum_{i=1}^{n} \Delta_i \hat{E}\left[I(w_{jt} \leqslant X_i < w_{j+1,t}) \frac{G_T''(q_{3i}\Lambda_T(X_i))\, q_{3i}}{G_T'(q_{3i}\Lambda_T(X_i))} \right] - \sum_{i=1}^{n} \hat{E}\left[I(w_{jt} \leqslant X_i < w_{j+1,t})\, G_T'(q_{3i}\Lambda_T(X_i))\, q_{3i} \right], \tag{3.8}$$

where $G_T''(x) = d^2 G_T(x)/dx^2$ and $\Lambda_T(s) = \sum_{l=1}^{j} \lambda_{lT}$ for $w_{jt} \leqslant s < w_{j+1,t}$, and

$$\frac{1}{f_{jt}} = \frac{1}{f_{j+1,t}} - \zeta_t \sum_{i=1}^{n} \Delta_i \hat{E}\left[I(w_{jt} \leqslant X_i < w_{j+1,t}) \frac{G_T''(q_{3i}\check{\Lambda}_T(X_i))\, q_{3i}}{G_T'(q_{3i}\check{\Lambda}_T(X_i))} \right] + \zeta_t \sum_{i=1}^{n} \hat{E}\left[I(w_{jt} \leqslant X_i < w_{j+1,t})\, G_T'(q_{3i}\check{\Lambda}_T(X_i))\, q_{3i} \right], \tag{3.9}$$

where $\zeta_t = \Lambda_T(w_{m_t,t}) = \sum_{j=1}^{m_t} \lambda_{jT}$, $f_{jt} = \lambda_{jT}/\zeta_t$, and $\check{\Lambda}_T(s) = \zeta_t(1 - \sum_{l=j+1}^{m_t} f_{lt})$ for $w_{jt} \leqslant s < w_{j+1,t}$. We note that the forward recursive equations (3.6) and (3.8) diminish parameter sets for $\Lambda_R\{.\}$ and $\Lambda_T\{.\}$ needed to be maximized in (3.5) from $(\lambda_{1R}, ..., \lambda_{m_r R}, \lambda_{1T}, ..., \lambda_{m_t T})$ to $(\lambda_{1R}, \lambda_{1T})$, and so do the backward recursive equations (3.7) and (3.9) from $(\lambda_{1R}, ..., \lambda_{m_r R}, \lambda_{1T}, ..., \lambda_{m_t T})$ to $(\zeta_r, f_{m_r r}, \zeta_t, f_{m_t t})$.

To obtain the NPMLEs, we iterate the E-step and M-step until the maximizers converge at a certain rate. The variances of the NPMLEs can be estimated from the inverse of the observed information matrix for $(\beta, \phi, \sigma_e^2, \Sigma_b, \lambda_{1R}, ..., \lambda_{m_r R}, \lambda_{1T}, ..., \lambda_{m_t T})$. The observation information matrix can be computed from the complete-data log-likelihood function denoted by ℓ_i^c for the ith subject using the following formula of

$$-\sum_{i=1}^{n} \hat{E}[\nabla^2 \ell_i^c(b_i) \,|\, O_i] - \sum_{i=1}^{n} \left\{ \hat{E}[\nabla \ell_i^c(b_i)^{\otimes 2} \,|\, O_i] - \hat{E}[\nabla \ell_i^c(b_i) \,|\, O_i]^{\otimes 2} \right\},$$

where $u^{\otimes 2} = uu^T$, ∇ and ∇^2 denote the first and the second derivatives with respect to parameters, and $\hat{E}$ denotes the conditional expectation of a function of b given the observed data.

3.9 Proof of Asymptotic Properties

This section proves Theorems 3.1 - 3.2 stated in Section 3.4 by applying the general asymptotic theory of Zeng and Lin (2007). Specifically, it is easy to see that our conditions (A1) - (A9) imply (C1) - (C4), (C6), (C8) of Zeng and Lin (2007b), and it remains to prove the two identifiability conditions (C5) and (C7) of Zeng and Lin (2007b). The first identifiability is the key step to prove the consistency of the NPMLEs, and the second is to entail the invertibility of the observed information matrix at the true parameters for the proof of the asymptotic normality.

Proof of the First Identifiability

Proof. First, we verify the first identifiability condition (C5) in Appendix B of Zeng and Lin (2007b). Suppose that the likelihood function for (β, ϕ, σ_e^2, Σ_b, Λ_R, Λ_T) is the same as that for the true parameter values (β_0, ϕ_0, σ_{0e}^2, Σ_{0b}, Λ_{0R}, Λ_{0T}). That is,

$$
\begin{aligned}
&\int_b \prod_{k=1}^{m} \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ -\frac{(Y(t_k) - \beta_1^T Z_1(t_k) - b_1^T \tilde{Z}_1(t_k))^2}{2\sigma_e^2} \right\} \right] \\
&\times \prod_t \left[\lambda_R(t)\, e^{\beta_2^T Z_2(t) + b_2^T \tilde{Z}_2(t)}\, G_R'(q_2(t)) \right]^{R(t)\Delta N^*(t)} \\
&\times \left[\lambda_T(X)\, e^{\beta_3^T Z_3(X) + (b\circ\phi)^T \tilde{Z}_3(X)}\, G_T'(q_3(X)) \right]^{\Delta} \\
&\times \exp\{-G_R(q_2(X)) - G_T(q_3(X))\}\ f(b;\Sigma_b)\, db \\
\\
= \quad &\int_b \prod_{k=1}^{m} \left[\frac{1}{\sqrt{2\pi\sigma_{0e}^2}} \exp\left\{ -\frac{(Y(t_k) - \beta_{01}^T Z_1(t_k) - b_1^T \tilde{Z}_1(t_k))^2}{2\sigma_{0e}^2} \right\} \right] \\
&\times \prod_t \left[\lambda_{0R}(t)\, e^{\beta_{02}^T Z_2(t) + b_2^T \tilde{Z}_2(t)} G_R'(q_{02}(t)) \right]^{R(t)\Delta N^*(t)} \\
&\times \left[\lambda_{0T}(X)\, e^{\beta_{03}^T Z_3(X) + (b\circ\phi_0)^T \tilde{Z}_3(X)} G_T'(q_{03}(X)) \right]^{\Delta} \\
&\times \exp\{-G_R(q_{02}(X)) - G_T(q_{03}(X))\}\ f(b;\Sigma_{0b})\, db, \qquad (3.10)
\end{aligned}
$$

where $q_2(t) = \int_0^t e^{\beta_2^T Z_2(s)+b_2^T \tilde{Z}_2(s)}\, d\Lambda_R(s)$, $q_3(t) = \int_0^t e^{\beta_3^T Z_3(s)+(b\circ\phi)^T \tilde{Z}_3(s)}\, d\Lambda_T(s)$, and $q_{02}(t)$, $q_{03}(t)$ are $q_2(t)$, $q_3(t)$ evaluated at the true parameter values, and $f(b;\Sigma_b)$ is the density function of the (multivariate) normal distribution with mean zeros and covariance matrix Σ_b. From now, we take the following actions on both sides of (3.10).

Step 1. For the proof of the identifiability of the longitudinal component, we consider a case with the observed longitudinal measures at time $s_{11}, \ldots, s_{1K}$ for arbitrary K, $R(t) = 1$, $\Delta N^*(t) = 0$, $\Delta = 0$, and $X \approx 0$.

Considering $E[Y(s_{1k})]$ and using the fact that

$$\int b_1\, f(b_1;\Sigma_b)\, db_1 = \int b_1\, f(b_1;\Sigma_{0b})\, db_1 = 0,$$

we have $\beta_1^T Z_1(s_{1k}) = \beta_{01}^T Z_1(s_{1k})$, for $k = 1, \ldots, K$. By Condition (A7), we prove $\beta_1 = \beta_{01}$. Now, we consider $\mathrm{Var}(Y(s_{1k}))$ and obtain

$$\begin{aligned} & \int_b \left\{\sigma_e^2 + b_1^T \tilde{Z}_1(s_{1k})\tilde{Z}_1(s_{1k})^T b_1\right\} f(b;\Sigma_b)\, db \\ = & \int_b \left\{\sigma_{0e}^2 + b_1^T \tilde{Z}_1(s_{1k})\tilde{Z}_1(s_{1k})^T b_1\right\} f(b;\Sigma_{0b})\, db, \end{aligned}$$

for $k = 1, \ldots, K$. Then, we have $\sigma_e^2 = \sigma_{0e}^2$ from (A8).

Step 2. For the recurrent events and the terminal event components, set $R(\cdot) = 1$, $\Delta = 0$, $X = t_3$, and suppose that $N^*(\cdot)$ has jumps at $s_{21}, \ldots, s_{2K}$, $s_{21}^{'}, \ldots, s_{2M}^{'}$ for any arbitrary $(K + M)$ in $[0, t_3]$.

For simplicity, we can drop the longitudinal component in (3.10) by integrating over $y(t)$ for the observation times t. We integrate (3.10) with respect to $s_{21}, \ldots, s_{2K}$ from 0 to $t_{21}, \ldots, t_{2K}$, while integrating with respect to $s_{21}^{'}, \ldots, s_{2M}^{'}$ from 0 to t_3. Since $s_{21}, \ldots, s_{2K}$ are arbitrary, multiplying $\frac{1}{M!}\prod_{k=1}^{K}(iw_k)^{a_k}/a_k!$ and taking summation over

M and $a_1, \ldots, a_k$ from 0, 1, $\ldots$, ∞ results in

$$\begin{aligned} & \int_b \exp\left\{\sum_{k=1}^{K} iw_k G_R(q_2(t_{2k})) - G_T(q_3(t_3))\right\} f(b;\Sigma_b)\, db \\ = & \int_b \exp\left\{\sum_{k=1}^{K} iw_k G_R(q_{02}(t_{2k})) - G_T(q_{03}(t_3))\right\} f(b;\Sigma_{0b})\, db. \end{aligned}$$

This equation implies that $G_R(q_2(t_{2k}))$ and $G_R(q_{02}(t_{2k}))$ for $k = 1, \ldots, K$ are the same in distribution as a function of $b \sim e^{-G_T(q_3(t_3))} f(b;\Sigma_b)$ and $b \sim e^{-G_T(q_{03}(t_3))} f(b;\Sigma_{0b})$, respectively. By the one-to-one mapping of G_R and logarithm function, the distribution of $\log(\frac{d}{dt_{2k}} q_2(t_{2k}))$ and $\log(\frac{d}{dt_{2k}} q_{02}(t_{2k}))$ are the same for $k = 1, \ldots, K$. That is, for $\forall(t_2, t_3)$ such that $t_2 < t_3$,

$$\begin{aligned} & \int_b \left\{\log \lambda_R(t_2) + \beta_2^T Z_2(t_2) + b_2^T \tilde{Z}_2(t_2)\right\} e^{-G_T(q_3(t_3))} f(b;\Sigma_b)\, db \\ = & \int_b \left\{\log \lambda_{0R}(t_2) + \beta_{02}^T Z_2(t_2) + b_2^T \tilde{Z}_2(t_2)\right\} e^{-G_T(q_{03}(t_3))} f(b;\Sigma_{0b})\, db. \end{aligned}$$

From the equality of $\int_b e^{-G_T(q_3(t_3))} f(b;\Sigma_b)\, db = \int_b e^{-G_T(q_{03}(t_3))} f(b;\Sigma_{0b})\, db$ and the condition (A7), we can show $\beta_2 = \beta_{02}$ and $\lambda_R(t) = \lambda_{0R}(t)$, and hence

$$\exp\{-G_T(q_3(t_3))\} f(b;\Sigma_b) = \exp\{-G_T(q_{03}(t_3))\} f(b;\Sigma_{0b}).$$

By setting $t_3 = 0$ in the foregoing equation, we obtain $f(b;\Sigma_b) = f(b;\Sigma_{0b})$ followed by $\Sigma_b = \Sigma_{0b}$ because b is normally distributed. Subsequently, it is true that $\exp\{-G_T(q_3(t_3))\} = \exp\{-G_T(q_{03}(t_3))\}$. Applying similar arguments above, we can obtain the following equation

$$\begin{aligned} & \log \lambda_T(t_3) + \beta_3^T Z_3(t_3) + b^T(\phi \circ \tilde{Z}_3(t_3)) \\ = & \log \lambda_{0T}(t_3) + \beta_{03}^T Z_3(t_3) + b^T(\phi_0 \circ \tilde{Z}_3(t_3)). \end{aligned}$$

Since $E[b] = 0$ on both sides, we conclude that $\beta_3 = \beta_{03}$ and $\log \lambda_T(t) = \log \lambda_{0T}(t)$ by (A7). Clearly, $\phi = \phi_0$ by (A8). □

Proof of the Second Identifiability

Proof. Next, we verify the second identifiability condition (C7) in Appendix B of Zeng and Lin (2007b). It starts from the score equation along with the path ($\beta_{01} + \eta\nu_1$, $\beta_{02}+\eta\nu_2$, $\beta_{03}+\eta\nu_3$, $\sigma^2_{0e}+\eta\nu_4$, $\phi_0+\eta\nu_\phi$, $\Sigma_{0b}+\eta\nu_b$, $\Lambda_{0R}+\eta\int h_1\, d\Lambda_{0R}$, $\Lambda_{0T}+\eta\int h_2\, d\Lambda_{0T}$).

Step 1. To make the score equation simple for the proofs of $\nu_1 = 0$ and $\nu_4 = 0$, we consider the same setting as used in *Step 1* of the first identifiability proof. We define

$$V_b^{-1} = \Sigma_{0b_1}^{-1} + \sum_k \tilde{Z}_1(s_{1k})\tilde{Z}_1^T(s_{1k})/\sigma^2_{0e}, \quad \text{and} \quad \mu_b = \sum_k c_k \tilde{Z}_1(s_{1k})/\sigma^2_{0e},$$

then, the score equation is given by

$$\begin{aligned}
&\left(\frac{1}{\sqrt{2\pi\sigma^2_{0e}}}\right)^K |\Sigma_{0b_1}|^{-\frac{1}{2}}|V_b|^{\frac{1}{2}} \exp\left\{-\sum_{k=1}^K c_k^2/(2\sigma^2_{0e}) + \mu_b^T V_b \mu_b/2\right\} \\
\times\ &[\sum_{k=1}^K \frac{\nu_4}{2\sigma^4_{0e}}\left\{c_k^2 - 2\mu_b^T V_b \tilde{Z}_1(s_{1k})c_k + \mu_b^T V_b \tilde{Z}_1(s_{1k})\tilde{Z}_1^T(s_{1k})V_b\mu_b + \tilde{Z}_1^T(s_{1k})V_b\tilde{Z}_1(s_{1k})\right\} \\
&\quad -K\frac{\nu_4}{2\sigma^2_{0e}} + \sum_{k=1}^K \frac{\nu_1^T Z_1(s_{1k})}{\sigma^2_{0e}}\left\{c_k - \mu_b^T V_b \tilde{Z}_1(s_{1k})\right\}], \qquad (3.11)
\end{aligned}$$

where $c_k = y(s_{1k}) - \beta_{01}^T Z_1(s_{1k})$, and K is the number of repeated measures.

Since $\forall \epsilon > 0$, $Pr\left[\,|Y(t) - \beta_{01}^T X_1(t)| < \epsilon\right] > 0$, we have

$$(2\sigma^2_{0e})^{-1}\nu_4\left[K - \sum_k \tilde{Z}_1^T(s_k)V_b\tilde{Z}_1(s_k)/\sigma^2_{0e}\right] = 0.$$

Under (A9), we conclude $\nu_4 = 0$ because $tr[\sum_k \tilde{Z}_1^T(s_k)V_b\tilde{Z}_1(s_k)/\sigma^2_{0e}] < d_b$, where d_b

stands for the dimension of b_1. Then, (3.11) becomes equivalent to

$$\frac{\nu_1^T Z_1^T}{\sigma_{0e}^2}\left[\, I - \frac{\tilde{Z}_1 V_b \tilde{Z}_1{}^T}{\sigma_{0e}^2}\,\right](Y - Z_1\beta_{01}) = 0,$$

where $Y^T = (y(s_{11}), \ldots, y(s_{1K}))$, $Z_1^T = (Z_1(s_{11}), \ldots, Z_1(s_{1K}))$, and $\tilde{Z}_1^T = (\tilde{Z}_1(s_{11}),$ $\ldots,$ $\tilde{Z}_1(s_{1K}))$. Followed by the fact that $\left[\, I - \tilde{Z}_1 V_b \tilde{Z}_1{}^T/\sigma_{0e}^2 \,\right]$ is positive definite, we just prove $\nu_1 = 0$.

Step 2. For the second identifiability of the recurrent and terminal events, we repeat the same process as *Step 2* of the first identifiability proof with the score equation and obtain

$$\int_b \left[\sum_{k=1}^{K} C_1(t_{2k}, b) - C_2(t_3, b) + \frac{f'(b;\Sigma_{0b})^T \nu_b}{f(b;\Sigma_{0b})}\right] \times \exp\left\{\sum_{k=1}^{K} i w_k G_R(q_{02}(t_{2k}) - G_T(q_{03}(t_3))\right\} f(b;\Sigma_{0b})\, db = 0, \quad (3.12)$$

where

$$\begin{aligned} C_1(t,b) &= \int_0^t \left(\nu_2^T Z_2(s) + h_1(s)\right) q_{02}'(s)\, ds\; G_R'(q_{02}(t))/G_R(q_{02}(t)), \text{ and} \\ C_2(t,b) &= \int_0^t \left(\nu_3^T Z_3(s) + (b \circ \nu_\phi)^T \tilde{Z}_3(s) + h_2(s)\right) q_{03}'(s)\, ds\; G_T'(q_{03}(t))/G_T(q_{03}(t)). \end{aligned}$$

Application of the Fourier transformation to (3.12) results in

$$\begin{aligned} &\textstyle\sum_k E_b\left[C_1(t_{2k}, b)\, e^{-G_T(q_{03}(t_3))} \,|\, G_R(q_{02}(t_{21})) = \xi_1, \ldots, G_R(q_{02}(t_{2K})) = \xi_K\right] f_\xi \\ &- E_b\left[C_2(t_3, b)\, e^{-G_T(q_{03}(t_3))} \,|\, G_R(q_{02}(t_{21})) = \xi_1, \ldots, G_R(q_{02}(t_{2K})) = \xi_K\right] f_\xi \\ &+ E_b\left[\frac{f'(b;\Sigma_{0b})^T \nu_b}{f(b;\Sigma_{0b})}\, e^{-G_T(q_{03}(t_3))} \,|\, G_R(q_{02}(t_{21})) = \xi_1, \ldots, G_R(q_{02}(t_{2K})) = \xi_K\right] f_\xi \\ &= \quad 0, \end{aligned} \quad (3.13)$$

where $f_\xi = f(\xi_1, \ldots, \xi_K)$ is the joint density of $\{G_R(q_{02}(t_{21})), \ldots, G_R(q_{02}(t_{2K}))\}$. By integrating out $(\xi_1, \ldots, \xi_K)$ and setting $t_3 = 0$, we make each term of (3.13) zero. From the first term $E_b\left[C_1(t_{2k}, b)\right] = 0$ $(k = 1, \ldots, K)$, it follows that $\nu_2^T Z_2(t_{2k}) + h_1(t_{2k})$ has a trivial solution. Under the assumption (A7), we show that $\nu_2 = 0$ and $h_1 = 0$. Subsequently, the similar argument for $t_3 > 0$ leads (3.13) to

$$-C_2(t_3, b) + \frac{f'(b; \Sigma_{0b})^T \nu_b}{f(b; \Sigma_{0b})} = 0.$$

Clearly, $f'(b; \Sigma_{0b})^T \nu_b = 0$ yields $\nu_b = 0$ because f' is the derivative of the normal density. Finally, we attain to $\nu_3^T Z_3(t_3) + (b \circ \nu_\phi)^T \tilde{Z}_3(t_3) + h_2(t_3) = 0$ followed by $\nu_3 = 0$, $\nu_\phi = 0$, and $h_2 = 0$. We complete the proofs of Theorems 3.1 - 3.2 by Theorems 1 - 2 in Zeng and Lin (2007b). □

Let I_n denote the negative Hessian matrix of the observed log-likelihood function with respect to $(\theta, \Lambda_R\{\cdot\}, \Lambda_T\{\cdot\})$. As a remark, by following Theorem 3 in Zeng and Lin (2007b), we can show that I_n is invertible for large n, and

$$(\nu^T, U_R^T, U_T^T)\, n I_n^{-1} (\nu^T, U_R^T, U_T^T)^T$$

is the consistent estimator of the asymptotic variance of

$$\sqrt{n}\left\{\nu^T(\hat{\theta} - \theta_0) + \int u_R(t)\, d(\hat{\Lambda}_R - \Lambda_{0R}) + \int u_T(t)\, d(\hat{\Lambda}_T - \Lambda_{0T})\right\},$$

where U_R and U_T are vectors of $u_R(\cdot)$ and $u_T(\cdot)$ at the observed recurrent and terminal event times, respectively.

Table 3.1: Simulation results for $G_R(x) = G_T(x) = x$.

		$N = 200$				$N = 400$			
Parameter	*True*	*Bias*	*SE*	*SEE*	*CP*	*Bias*	*SE*	*SEE*	*CP*
					$\phi = (0.5, 0.2)$				
β_1	0.7	-.011	.083	.081	.934	-.010	.059	.058	.944
	1.0	-.012	.123	.118	.935	-.012	.086	.083	.946
	0.5	-.008	.102	.108	.947	-.007	.072	.072	.953
σ_e^2	1.0	.002	.041	.041	.958	-.000	.031	.029	.940
β_2	1.0	-.012	.138	.138	.951	-.021	.094	.097	.955
	0.5	-.007	.118	.119	.952	-.009	.084	.083	.949
$\Lambda_R(\tau/4)$	1.0	-.014	.111	.112	.948	-.012	.077	.079	.955
$\Lambda_R(\tau/2)$	2.0	-.033	.213	.215	.942	-.026	.149	.152	.946
β_3	1.0	.019	.199	.204	.962	-.003	.137	.141	.957
	0.5	.012	.175	.171	.954	-.000	.118	.118	.954
ϕ	0.5	.028	.214	.211	.960	.005	.147	.143	.944
	0.2	-.016	.219	.219	.960	-.010	.153	.149	.945
$\Lambda_T(\tau/4)$	0.1	-.003	.022	.022	.962	-.002	.016	.016	.953
$\Lambda_T(\tau/2)$	0.4	-.009	.069	.069	.953	-.003	.048	.049	.960
σ_1^2	0.5	-.013	.068	.067	.951	-.006	.049	.048	.951
σ_2^2	0.5	.013	.095	.092	.946	.013	.063	.064	.947
ρ	0.5	-.006	.091	.095	.955	-.005	.066	.066	.961
					$\phi = (0, 0.2)$				
β_1	0.7	-.009	.082	.082	.939	-.009	.057	.058	.950
	1.0	-.014	.116	.118	.947	-.016	.084	.083	.936
	0.5	-.001	.101	.103	.947	-.005	.072	.072	.952
σ_e^2	1.0	.001	.042	.041	.952	-.000	.029	.029	.949
β_2	1.0	-.016	.139	.139	.940	-.017	.099	.097	.932
	0.5	-.004	.121	.118	.947	-.008	.082	.083	.949
$\Lambda_R(\tau/4)$	1.0	-.012	.108	.113	.955	-.015	.082	.079	.938
$\Lambda_R(\tau/2)$	2.0	-.028	.209	.216	.949	-.031	.159	.152	.942
β_3	1.0	.015	.196	.191	.950	.003	.135	.133	.947
	0.5	.009	.163	.160	.946	.003	.107	.110	.960
ϕ	0.0	.001	.204	.200	.954	.007	.139	.135	.953
	0.2	-.008	.234	.221	.944	-.018	.157	.148	.947
$\Lambda_T(\tau/4)$	0.1	-.002	.022	.022	.958	-.001	.015	.015	.941
$\Lambda_T(\tau/2)$	0.4	-.001	.068	.067	.960	-.001	.047	.047	.944
σ_1^2	0.5	-.007	.069	.068	.954	-.004	.046	.047	.960
σ_2^2	0.5	.010	.092	.093	.946	.010	.063	.064	.948
ρ	0.5	-.002	.096	.094	.947	-.003	.066	.065	.941

Table 3.2: Simulation results for $G_R(x) = x$ and $G_T(x) = \log(1 + x)$.

		$N = 200$				$N = 400$			
Parameter	*True*	*Bias*	*SE*	*SEE*	*CP*	*Bias*	*SE*	*SEE*	*CP*
					$\phi = (0.5, 0.2)$				
β_1	0.7	-.009	.084	.082	.935	-.008	.058	.058	.949
	1.0	-.003	.120	.119	.956	-.006	.085	.084	.939
	0.5	-.007	.104	.103	.943	-.002	.074	.073	.940
σ_e^2	1.0	.002	.041	.041	.955	.001	.030	.029	.946
β_2	1.0	-.003	.157	.156	.957	-.006	.107	.108	.956
	0.5	-.009	.135	.132	.950	-.001	.091	.092	.949
$\Lambda_R(\tau/4)$	0.6	-.006	.081	.080	.949	-.006	.056	.056	.943
$\Lambda_R(\tau/2)$	1.2	-.012	.156	.154	.953	-.016	.109	.107	.939
β_3	1.0	.010	.282	.284	.952	.001	.198	.198	.954
	0.5	.004	.228	.241	.966	.005	.168	.169	.951
ϕ	0.5	-.009	.349	.367	.969	.007	.245	.246	.957
	0.2	-.012	.436	.444	.969	-.011	.301	.299	.949
$\Lambda_T(\tau/4)$	0.1	.004	.036	.037	.965	-.002	.026	.026	.966
$\Lambda_T(\tau/2)$	0.4	-.000	.136	.131	.953	.000	.093	.092	.948
σ_1^2	0.5	-.009	.070	.069	.956	-.004	.049	.049	.949
σ_2^2	0.5	.015	.145	.112	.945	.009	.078	.078	.941
ρ	0.5	-.002	.108	.108	.961	-.004	.073	.075	.953
					$\phi = (0, 0.2)$				
β_1	0.7	-.008	.082	.082	.950	-.008	.057	.058	.954
	1.0	-.001	.120	.119	.953	-.005	.084	.084	.948
	0.5	.001	.106	.103	.945	-.006	.071	.072	.958
σ_e^2	1.0	.001	.041	.041	.947	.001	.028	.029	.953
β_2	1.0	-.011	.151	.152	.956	-.005	.111	.108	.938
	0.5	.005	.134	.130	.944	-.002	.092	.092	.944
$\Lambda_R(\tau/4)$	0.15	-.001	.081	.079	.944	-.007	.056	.055	.943
$\Lambda_R(\tau/2)$	0.6	-.005	.151	.151	.947	-.019	.106	.106	.949
β_3	1.0	.013	.276	.276	.951	.023	.193	.192	.947
	0.5	.009	.234	.235	.950	.003	.163	.164	.947
ϕ	0.0	.014	.367	.357	.963	.008	.246	.241	.956
	0.2	-.029	.459	.440	.956	-.045	.297	.293	.942
$\Lambda_T(\tau/4)$	0.1	-.004	.036	.036	.955	-.004	.025	.025	.949
$\Lambda_T(\tau/2)$	0.4	.004	.126	.128	.959	-.010	.085	.087	.952
σ_1^2	0.5	-.006	.070	.068	.951	-.005	.049	.048	.955
σ_2^2	0.5	.007	.112	.110	.943	.011	.077	.077	.946
ρ	0.5	-.007	.108	.107	.959	-.004	.073	.074	.959

Table 3.3: Analysis results for the ARIC study. The Fisher transformation is used for testing ρ, while the 50:50 mixture of χ^2 distributions is used for testing variances.

Effect	*Estimate*	*Std.Error*	*p-value*
	Longitudinal measures of SBP		
Intercept	-0.093	0.025	0.0002
Age	0.126	0.024	$< .0001$
BMI (kg/m^2)	0.009	0.015	0.5441
SBP (baseline, $mmHg$)	0.615	0.020	$< .0001$
Hypertension medication (yes vs. no)	0.238	0.040	$< .0001$
Visit Year	0.038	0.003	$< .0001$
BMI * SBP	-0.059	0.015	$< .0001$
SBP * Hypertension medication	-0.149	0.042	0.0004
σ_e^2	0.280	0.008	$< .0001$
	Recurrent CHD event		
Age	0.974	0.160	$< .0001$
BMI (kg/m^2)	0.115	0.066	0.0786
SBP (baseline, $mmHg$)	0.222	0.093	0.0164
Diabetes (yes vs. no)	1.168	0.225	$< .0001$
Age * BMI	-0.238	0.089	0.0074
	Terminal event		
Age	2.521	0.667	0.0002
BMI (kg/m^2)	0.206	0.118	0.0806
SBP (baseline, $mmHg$)	0.264	0.152	0.0819
Diabetes (yes vs. no)	1.273	0.460	0.0057
Age * BMI	-0.439	0.160	0.0062
ϕ_1	-0.141	0.329	0.6684
ϕ_2	1.962	0.498	$< .0001$
	Variance components for random effect		
σ_1^2	0.183	0.011	$< .0001$
σ_2^2	2.905	0.511	$< .0001$
ρ	0.059	0.055	0.2819

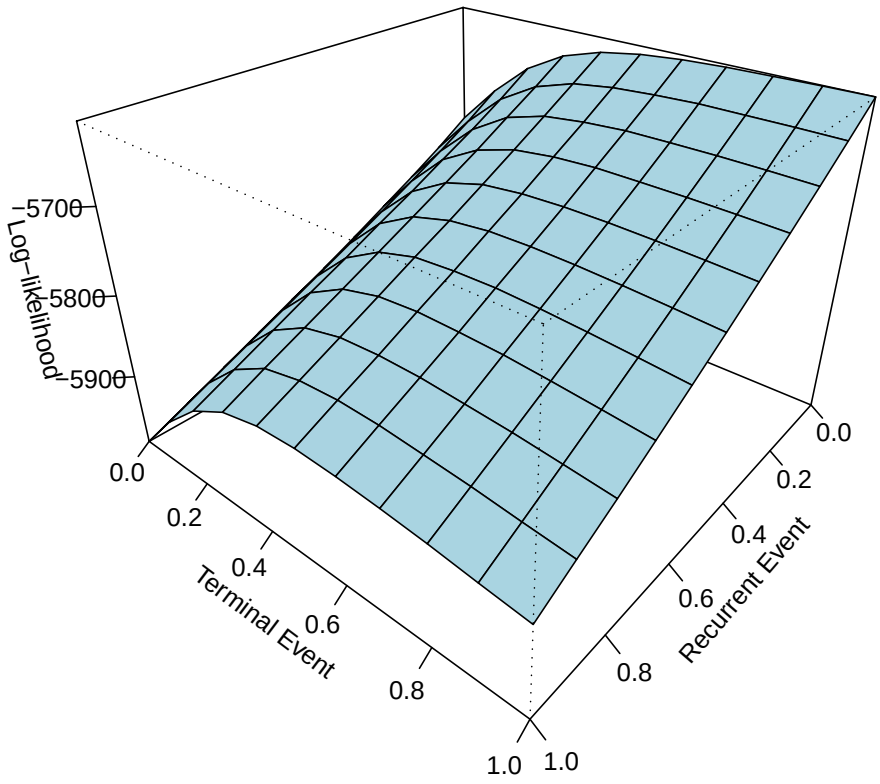

Figure 3.1: Log-likelihood surface under the logarithmic transformations for the ARIC study. The x-axis and y-axis correspond to the transformation parameter γ for recurrent events and terminal event, respectively.

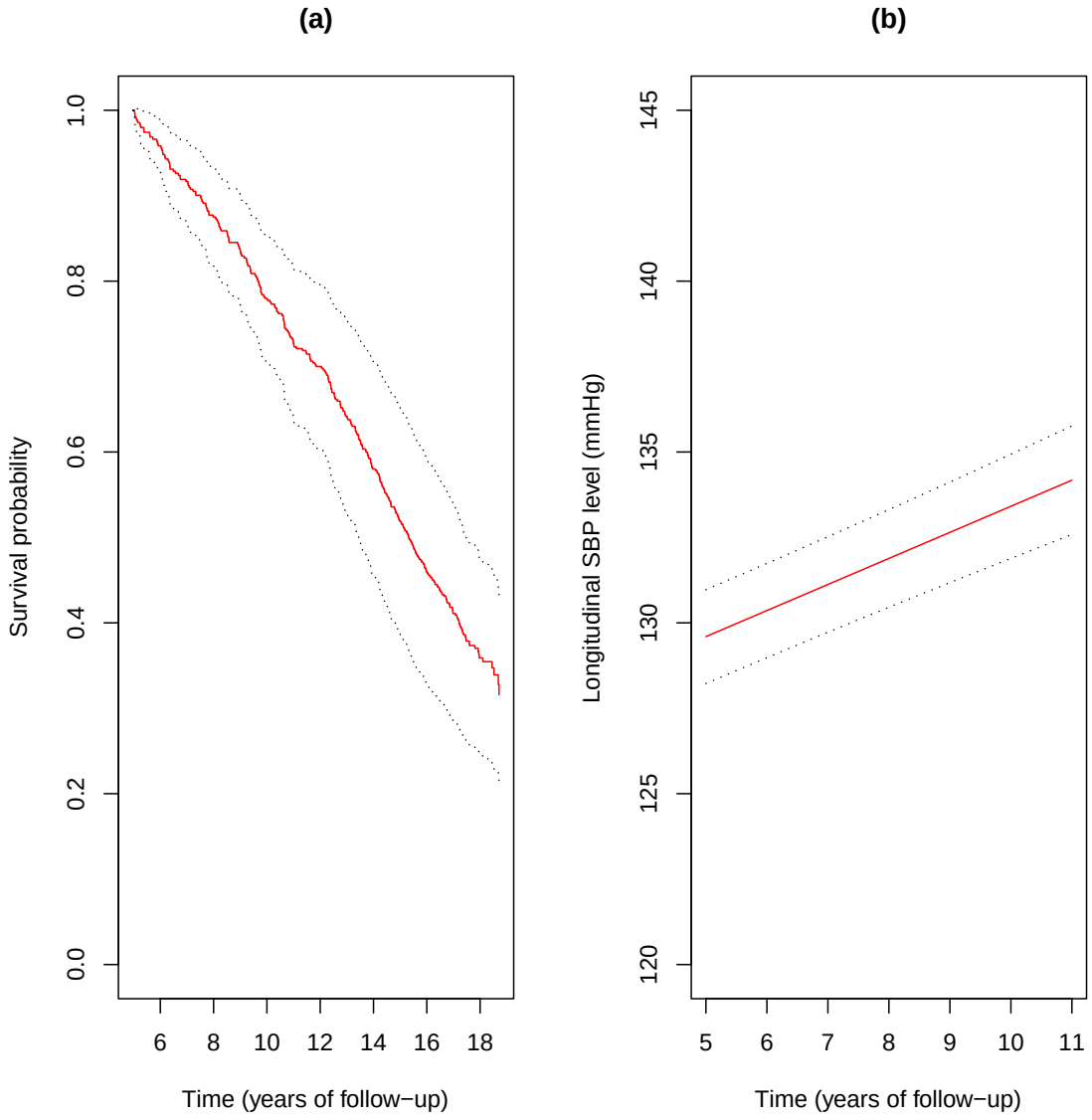

Figure 3.2: Predicted survival probability (a) and the expected longitudinal SBP levels (b) for a subject who had one CHD event at the 5th year of study. The solid curves are point estimates, and the dotted curves are the 95% confidence bands.

Chapter 4

Joint Modeling of Longitudinal and Cure-Survival Data

4.1 Introduction

In many medical studies, patients are repeatedly followed-up with a series of clinical markers measured until some survival event occurs. To investigate the association between such longitudinal biomarkers and time to the event, joint modeling approaches have become more and more popular (Wulfsohn and Tsiatis, 1997; Xu and Zeger, 2001; Tsiatis and Davidian, 2001, 2004; Hsieh et al., 2006; Vonesh et al., 2006; Liu et al., 2007; Song and Wang, 2008). When the event of interest is an event other than death, it is not uncommon that a certain proportion of subjects might never experience such an event; these patients are considered cured. For example, patients can be cured when cancer is treated by radiation therapy or an initial surgical intervention if the treatment removes all the cancer cells. The survival data with cured subjects generally present a heavy censoring rate at the end of a long follow-up, and the joint modeling approaches based on the classic survival models may not always be appropriate. It has been shown by Brown and Ibrahim (2003b) that joint cure

rate models performed better than joint models ignoring cure in terms of biasedness of regression coefficients when a true cure proportion existed in the study population.

Two classes of cure rate models are commonly used to analyze cure-survival data: mixture cure model and promotion time cure model. For a detailed examination of these models, we refer readers to Section 2.2. There has been scant literature about jointly modeling longitudinal and survival data with a cure fraction. The mixture cure model with a longitudinal disease progression marker as a covariate has been studied by Law et al. (2002), Yu et al. (2004), and Yu et al. (2008) to model a clinical recurrence, for which a fraction of patients are cured by the treatment and are immune from recurrence. Specifically, they used a mixed-effects model and a time-dependent Cox proportional hazards model conditioning on the unobserved random effects to build a model for the patients who are susceptible to recurrence in the uncured subpopulation. In these models, different assumptions have been made for the baseline hazard function. Law et al. (2002) assumed the baseline hazard to be nonparametric and obtained maximum likelihood estimators of the parameters via a Monte Carlo EM algorithm. In contrast, Yu et al. (2004) took a Bayesian approach with the baseline hazard following a Weibull distribution. The Weibull assumption was made for the computational simplicity in using Markov chain Monte Carlo methods. Their estimation results were similar to those achieved by the maximum likelihood estimation in Law et al. (2002). Yu et al. (2008) extended the use of a Weibull baseline hazard to a generalized Weibull baseline hazard in order to allow more flexibility.

On the other hand, Brown and Ibrahim (2003b) and Chen et al. (2004) have proposed joint models with a proportional hazards structure for longitudinal and survival data with a cure fraction, where the promotion time cure models were used to fit event times. Brown and Ibrahim (2003b) focused on modeling the association between the longitudinal markers and time to survival endpoint with the possibility

of cured patients. Chen et al. (2004) employed a mixed-effects model for longitudinal biomarkers and the promotion time cure model with a proportional hazards structure for survival times. In this model, random effects were shared by the longitudinal and survival components to account for the correlation between them. In both Brown and Ibrahim (2003b) and Chen et al. (2004), a piecewise exponential distribution was considered to estimate the baseline distribution function, and Bayesian approaches were used for inference.

In this paper, to correctly handle the heavy tail of survival distribution by long-term survivors, we propose a mixed-effects model for longitudinal data and a transformed promotion time cure model for survival times in data where a portion of the patients can be cured. These two models share the same frailty, but with different magnitude. Using transformation gives a flexible way to fit survival data. We propose to use nonparametric maximum likelihood estimation for efficient inference. In Section 4.2, we introduce the proposed joint models. In Section 4.3, we provide the nonparametric maximum likelihood estimators (NPMLE) and describe a simple algorithm used for implementing the proposed inference procedure. In Section 4.4, we establish asymptotic properties of the proposed estimators. We evaluate the numerical performance of the proposed method through both simulation studies in Section 4.5 and an application to the Atherosclerosis Risk in Communities (ARIC) data in Section 4.6. Finally, concluding remarks are given in Section 4.7 and proofs of asymptotic properties are provided in Section 4.8.

4.2 Joint Models

Let $Y(t)$ be the longitudinal measurement at time t, T be the time to the survival event, and $\mathcal{Z} = \{Z(t); t \geqslant 0\}$ be the covariate process, where $Z(t)$ is the vector

of external covariates at time t, possibly time-varying. We introduce latent random effects to account for the correlation between longitudinal and survival components on the same subject. Particularly, let b denote the subject-specific random effects following a multivariate normal distribution with mean zeros and covariance matrix Σ_b. We further assume that $Y(t)$ and T are independent, conditional on $\mathcal{Z}$ and b. Then, the proposed joint models for the longitudinal data $Y(\cdot)$ and the population survival function of T with a cure fraction are given by

$$\begin{aligned} Y(t \mid \mathcal{Z}, b) &= \alpha^T Z_1(t) + b^T \tilde{Z}_1(t) + \epsilon(t), \\ S(t \mid \mathcal{Z}, b) &= \exp\left\{-H\left(\int_0^t e^{\beta^T Z_2(u) + (\psi \circ b)^T \tilde{Z}_2(u)}\, dF(u)\right)\right\}, \end{aligned} \tag{4.1}$$

where α and β are vectors of unknown regression parameters in the longitudinal and survival components, respectively, $Z_k(t)$ and $\tilde{Z}_k(t)$ $(k = 1, 2)$ are subsets of $Z(t)$ plus the unit component, and $F(t)$ is an unspecified distribution function of the event times. In addition, $\epsilon(t)$ is a white noise process with mean zero and variance σ_e^2, ψ is a set of unknown constants with the same number of elements as b, and $\psi \circ b$ denotes the component-wise product of ψ and b. Note in (4.1) that the correlation among the longitudinal outcomes is formulated through the latent random effects b, and that the association between longitudinal outcomes and the event time is characterized by ψ with the shared latent variables b. Thus, for a fixed covariate $\mathcal{Z}$, $\psi > 0$ implies the larger longitudinal measures are, the higher hazard rate of the event is. On the other hand, $\psi = 0$ implies that the association can be fully explained by the common covariates in both longitudinal and survival components.

In the model (4.1), $H(\cdot)$ represents a transformation function of the conditional cumulative hazard function, which is required to be pre-specified in the analysis. The transformation functions are assumed to be continuously differentiable and strictly

increasing. For example, $H(x)$ can take a form of the logarithmic transformation,

$$H(x) = \begin{cases} \log(1+\eta x)/\eta, & \eta > 0 \\ x, & \eta = 0. \end{cases}$$

The choices of $\eta = 0$ and $\eta = 1$ lead to the proportional hazards structure and the proportional odds structure, respectively.

We notice that the survival model for the entire population in (4.1) includes the cure rate model by letting $t = \infty$. That is, the cure rate model can be expressed as

$$\lim_{t\to\infty} S(t \,|\, \mathcal{Z}, b) = \exp\left\{-H\left(\int_0^\infty e^{\beta^T Z_2(u) + (\psi \circ b)^T \tilde{Z}_2(u)}\, dF(u)\right)\right\}.$$

Thus, our joint cure-survival model (4.1) allows us to explore a link between the longitudinal measures and the probability of being cured through the shared random effects as well as covariates. Especially, when $Z_2(t)$ and $\tilde{Z}_2(t)$ are time-independent covariates, z_2 and $\tilde{z}_2$, respectively, the cure rate can be simplified to

$$\exp[-H(\exp\{\beta^T z_2 + (\psi \circ b)^T \tilde{z}_2\})].$$

In fact, it is always true that the conditional cure rate is $\lim_{t\to\infty} S(t \,|\, \mathcal{Z}, b) > 0$ (improper survival function), because H is assumed to be finite.

Let C be the non-informative censoring time which is independent of $(Y(\cdot), T, b)$ given $\mathcal{Z}$, and let $X = \min(T, C)$ denote the observed event time. The observed data for the ith subject with m_i repeated measurements are defined as $O_i = \{Y_i(t_{ik}), X_i, \Delta_i, Z(t); t_{ik} \leqslant X_i, t \leqslant X_i, i = 1, \ldots, n, k = 1, \ldots, m_i\}$, where $\Delta_i = I(T_i \leqslant C_i)$ with $I(\cdot)$ being the indicator function. Under the model (4.1), the log-likelihood function

for the observed data is given by

$$\sum_{i=1}^{n} \log \int_b \prod_{k=1}^{m_i} \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ \frac{-(Y_i(t_{ik}) - \alpha^T Z_{1i}(t_{ik}) - b^T \tilde{Z}_{1i}(t_{ik}))^2}{2\sigma_e^2} \right\} \right]$$
$$\times \left[f(X_i)\, e^{\beta^T Z_{2i}(X_i) + (\psi \circ b)^T \tilde{Z}_{2i}(X_i)}\, H'\Big(\int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)}\, dF(u) \Big) \right]^{\Delta_i}$$
$$\times \exp\left\{ -H\Big(\int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)}\, dF(u) \Big) \right\} \times f(b;\, \Sigma_b)\, db, \qquad (4.2)$$

where $f(b;\, \Sigma_b)$ is the density function of b with the parameters Σ_b, and $f(t) = dF(t)/dt$ and $H'(x) = dH(x)/dx$ are the first derivatives of $F(t)$ and $H(x)$, respectively.

4.3 Inference Procedure

4.3.1 NPMLEs for Transformation Models

We propose to use the nonparametric maximum likelihood estimation (NPMLE) for estimating parameters $\theta = (\alpha,\, \beta,\, \psi,\, \sigma_e^2,\, \mathrm{Vec}(\Sigma_b))$ and infinite-dimensional parameter $F(t)$, where $\mathrm{Vec}(\Sigma_b)$ denotes the vector consisting of the upper triangular elements of Σ_b. To obtain the NPMLEs, in the log-likelihood function (4.2), we treat F as a step function with jumps only at the observed failure times and replace $f(t)$ by the jump size of F at t, which is denoted by $F\{t\}$.

For commonly used transformation functions such as a logarithmic transformation, $\exp\{-H(x)\}$ can be expressed as the Laplace transformation of some function $\phi(t)$, $t \geqslant 0$, such that

$$\exp\{-H(x)\} = \int_0^{\infty} \exp(-xt)\, \phi(t)\, dt.$$

For example, if we choose $\phi(t) = t^{1/\eta - 1} \exp(-t/\eta)/\{\Gamma(1/\eta)\, \eta^{1/\eta}\}$, then it is true that $H(x) = \log(1 + \eta x)/\eta$. Applying the Laplace transformation with a subject-specific

frailty ζ_i and using the fact that

$$H'(x)\exp\{-H(x)\} = \int_0^\infty \zeta \exp(-x\zeta)\,\phi(\zeta)\,d\zeta,$$

the observed log-likelihood function (4.2) can be rewritten as

$$\begin{aligned}
& l_n(\theta,\, F\{\cdot\}) \\
= \quad & \sum_{i=1}^{n} \log \int_b \prod_{k=1}^{m_i} \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ \frac{-(Y_i(t_{ik}) - \alpha^T Z_{1i}(t_{ik}) - b^T \tilde{Z}_{1i}(t_{ik}))^2}{2\sigma_e^2} \right\} \right] \\
& \qquad \times \int_{\zeta_i} \left[\zeta_i\, F\{X_i\}\, e^{\beta^T Z_{2i}(X_i) + (\psi \circ b)^T \tilde{Z}_{2i}(X_i)} \right]^{\Delta_i} \\
& \qquad\qquad \times \exp\left\{ - \int_0^{X_i} \zeta_i\, e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)}\, dF(u) \right\} \phi(\zeta_i)\, d\zeta_i \\
& \qquad \times f(b;\, \Sigma_b)\, db, \qquad (4.3)
\end{aligned}$$

where we assume that ζ_i and b are independent. The most attractive feature about taking transformation in this way is that the modified log-likelihood (4.3) can be seen as the proportional hazards frailty model with the conditional hazard function

$$\lambda(t|Z(t), \zeta_i, b_i) = \zeta_i\, f(t) \exp\{\beta^T Z_{2i}(t) + (\psi \circ b)^T \tilde{Z}_{2i}(t)\}.$$

This makes the algorithm more stable and computationally efficient.

Now, the computation of the NPMLEs is identical to maximizing the modified log-likelihood function with respect to θ and all jump sizes of F at the observed failure times. This maximization can be carried out through the following EM algorithm.

4.3.2 EM Algorithm

We describe the EM algorithm, treating ζ_i and b as missing data to compute the NPMLEs of $(\theta,\, F\{\cdot\})$. In the E-step, we calculate the conditional expectation of

the log-likelihood function for the complete data, given the observed data O_i and the current parameter estimates. Particularly, we need to evaluate the integration of certain functions of $(\zeta_i,\, b)$, say $\hat{E}[\zeta_i\, g_i(b)\,|\,O_i]$. Hereafter, we drop the conditional part on the observed data and the current parameter estimates, and abbreviate such expectation $\hat{E}[\zeta_i\, g_i(b)\,|\,O_i]$ as $\hat{E}[\zeta_i\, g_i(b)]$. Computation of this expectation can become doable by first obtaining the *nested* conditional expectation of ζ_i, given b and the observed data. That is, $\hat{E}[\zeta_i\, g_i(b)]$ can be calculated as $\hat{E}_b[\,\hat{E}_{\zeta_i}[\zeta_i\,|\,b]\, g_i(b)]$. With the fact that the conditional distribution of ζ_i given b is proportional to

$$h(\zeta_i, b) = \zeta_i^{\Delta_i} \exp\left\{ - \int_0^{X_i} \zeta_i\, e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)}\, dF(u) \right\},$$

and the useful relationships by the Laplace transformation, the conditional expectation of ζ_i given b has the form of

$$\hat{E}_{\zeta_i}[\zeta_i\,|\,b] = \int \zeta_i\, \frac{h(\zeta_i, b)\, \phi(\zeta_i)}{\int h(\zeta_i, b)\, \phi(\zeta_i)\, d\zeta_i}\, d\zeta_i = H'(\tilde{x}_i(b)) - \left[\frac{H''(\tilde{x}_i(b))}{H'(\tilde{x}_i(b))} \right]^{\Delta_i},$$

where $\tilde{x}_i(b) = \int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)}\, dF(u)$. Once $\hat{E}_{\zeta_i}[\zeta_i\,|\,b]$ is calculated, which is a function of b, the conditional expectation $\hat{E}[\zeta_i\, g_i(b)]$ can be computed using numerical approximation methods such as the Gaussian quadrature with Hermite orthogonal polynomial. Since the conditional distribution of b given O_i is proportional to $\Gamma(O_i|\, b) f(b; \Sigma_b)$, the conditional expectation is calculated by

$$\hat{E}[\zeta_i\, g_i(b)] = \int_b \hat{E}_{\zeta_i}[\zeta_i\,|\,b]\, g_i(b)\, \frac{\Gamma(O_i|\, b) f(b; \Sigma_b)}{\int_b \Gamma(O_i|\, b) f(b; \Sigma_b)\, db}\, db,$$

where

$$\Gamma(O_i|\, b) = \exp\left\{ -\sum_{k=1}^{m_i} \left[(b^T \tilde{Z}_{1i}(t_{ik}))^2 - 2(Y_i(t_{ik}) - \alpha_1^T Z_{1i}(t_{ik}))\, b^T \tilde{Z}_{1i}(t_{ik}) \right] / (2\sigma_e^2) \right\}$$
$$\times \exp\left\{ \Delta_i \left[(\psi \circ b)^T \tilde{Z}_{2i}(X_i) + \log H'\left(\int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)}\, dF(u) \right) \right] \right\}$$
$$\times \exp\left\{ -H\left(\int_0^{X_i} e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)}\, dF(u) \right) \right\}.$$

In the M-step, we maximize the following objective function of the expected log-likelihood for the complete data:

$$\sum_{i=1}^{n} \sum_{k=1}^{m_i} \left\{ -\log \sigma_e^2/2 - \hat{E}\left[(Y_i(t_{ik}) - \alpha^T Z_{1i}(t_{ik}) - b^T \tilde{Z}_{1i}(t_{ik}))^2/(2\sigma_e^2) \right] \right\}$$
$$+ \sum_{i=1}^{n} \Delta_i \left\{ \log \zeta_i + \log F\{X_i\} + \beta^T Z_{2i}(X_i) + \hat{E}[\psi \circ b]^T \tilde{Z}_{2i}(X_i) \right\}$$
$$+ \sum_{i=1}^{n} \left\{ -\hat{E}\left[\int_0^{X_i} \zeta_i\, e^{\beta^T Z_{2i}(u) + (\psi \circ b)^T \tilde{Z}_{2i}(u)}\, dF(u) \right] + \hat{E}\left[\log \phi(\zeta_i) + \log f(b; \Sigma_b) \right] \right\},$$

under the restriction of $\sum_{i=1}^{n} \Delta_i F\{X_i\} = 1$. Maximizing the above objective function over (α, σ_e^2, Σ_b) is simple; whereas the rest of parameters (β, ψ, $F\{.\}$) do not yield the closed-form of maximizers, and hence it is required to involve a reliable numerical approach. By introducing the Lagrange multiplier μ, we solve the following equation for β:

$$\sum_{i=1}^{n} \Delta_i \left\{ Z_{2i}(X_i) - \frac{\sum_{j=1}^{n} R_j(X_i) Z_{2j}(X_i) \hat{E}\left[\zeta_j\, e^{q_{2j}(X_i)} \right]}{\sum_{j=1}^{n} R_j(X_i) \hat{E}\left[\zeta_j\, e^{q_{2j}(X_i)} \right] + \mu} \right\} = 0, \tag{4.4}$$

the following equation for ψ:

$$\sum_{i=1}^{n} \Delta_i \left\{ \hat{E}\left[b \circ \tilde{Z}_{2i}(X_i) \right] - \frac{\sum_{j=1}^{n} R_j(X_i) \hat{E}\left[\zeta_j\, e^{q_{2j}(X_i)} (b \circ \tilde{Z}_{2j}(X_i)) \right]}{\sum_{j=1}^{n} R_j(X_i) \hat{E}\left[\zeta_j\, e^{q_{2j}(X_i)} \right] + \mu} \right\} = 0, \tag{4.5}$$

and the following equation for μ:

$$\sum_{i=1}^{n} \Delta_i F\{X_i\} = 1, \tag{4.6}$$

where $R_j(t) = I(X_j \geqslant t)$ and $q_{2j}(t) = \beta^T Z_{2j}(t) + (\psi \circ b)^T \tilde{Z}_{2j}(t)$. In addition, F is estimated as a step function with the following jump size at X_i :

$$F\{X_i\} = \frac{\Delta_i}{\sum_{j=1}^{n} R_j(X_i)\hat{E}\left[\zeta_j \, e^{q_{2j}(X_i)}\right] + \mu}. \tag{4.7}$$

To solve these equations at each M-step, we consider a two-step optimization. In the first step, we estimate μ using the bisection method based on the equation (4.6) and the fact $F\{X_i\} > 0$ $(i = 1, \ldots, n)$. Since the left side of (4.6) is a monotone decreasing function of μ by considering $F\{X_i\}$ as a function of μ in (4.7), the solution always exists. In the second step, to update β and ψ, we plug the estimates $\hat{\mu}$ into equations (4.4) and (4.5), treat them as the functions of $\hat{\mu}$, and solve the equations using one-step Newton-Raphson algorithm. Updating the jump sizes of F can be easily done by the equation (4.7) with $\hat{\mu}$.

To obtain the NPMLEs, we iterate the E-step and M-step until the parameter estimates converge. The variances of the NPMLEs can be estimated from the inverse of the observed information matrix for all parameters of (θ, $F\{\cdot\}$), under the restriction of $\sum_{i=1}^{n} \Delta_i F\{X_i\} = 1$. The observation information matrix can be computed from the complete data log-likelihood function denoted by ℓ_i^c for the ith subject using the following Louis formula (Louis, 1982) of

$$-\sum_{i=1}^{n} \hat{E}[\nabla^2 \ell_i^c(b) \,|\, O_i] - \sum_{i=1}^{n} \left\{ \hat{E}[\nabla \ell_i^c(b)^{\otimes 2} \,|\, O_i] - \hat{E}[\nabla \ell_i^c(b) \,|\, O_i]^{\otimes 2} \right\},$$

where $u^{\otimes 2} = uu^T$, ∇ and ∇^2 denote the first and the second derivatives with respect

to parameters, and $\hat{E}$ denotes the conditional expectation of a function of b given the observed data and is evaluated at the NPMLEs.

4.4 Asymptotic Properties

Let $(\hat{\theta}, \hat{F})$ denote the NPMLEs and (θ_0, F_0) denote the true parameter values of (θ, F). Under the regularity conditions, we will establish the asymptotic properties of the NPMLEs under the following conditions:

(A1) The true parameter value θ_0 belongs to the interior of a compact set Θ within the domain of θ.

(A2) With probability 1, $Z(t)$ is left-continuous with uniformly bounded left and right derivatives in $[0, \infty]$.

(A3) For some constant δ_0, $P(C = \infty | \mathcal{Z}) > \delta_0 > 0$ with probability 1.

(A4) For some positive constant M_0, $M_0^{-1} < \sigma_{0e}^2 < M_0$ and $M_0^{-1} < c^T \Sigma_{0b}\, c < M_0$ for any constant vector $\|c\| = 1$.

(A5) The transformation functions $H(\cdot)$ are four-times differentiable with $H(0) = 0$ and $H'(0) > 0$. In addition, there exist positive constants μ_0 and κ_0 such that

$$(1 + x)\, H'(x) \exp\{-H(x)\} \leqslant \mu_0 (1 + x)^{-\kappa_0}.$$

Furthermore, there exists a constant $\rho_0 > 0$ such that

$$\sup_x \left\{ \frac{|H''(x)| + |H^{(3)}(x)| + |H^{(4)}(x)|}{H'(x)\,(1 + x)^{\rho_0}} \right\} < \infty,$$

where $H^{(3)}$ and $H^{(4)}$ are the third and fourth derivatives.

(A6) For some $t \in [0, \infty]$, if there exist a deterministic function $c(t)$ and v such that

$c(t) + v^T Z(t) = 0$ with probability 1, then $c(t) = 0$ and $v = 0$.

(A7) With some positive probability, $\tilde{\boldsymbol{Z}}_1^T \tilde{\boldsymbol{Z}}_1$ has full rank, where $\tilde{\boldsymbol{Z}}_1$ denotes a matrix with each row equal to the observed covariate $\tilde{Z}_1(t)^T$ at the time of each measurement.

(A8) Let K be the number of repeated measures and let d_b be the dimension of b. With probability one, $P(K > d_b | \mathcal{Z}, X) > 0$.

Conditions (A1) - (A3) are the standard assumptions in survival analysis. Condition (A4) is necessary to prove the existence of the NPMLEs. It can be easily verified that Condition (A5) holds for all transformations commonly used, including the logarithmic transformations described in Section 4.2. Conditions (A6) - (A7) entail the linear independence of design matrices of covariates for the fixed and random effects. Condition (A8) prescribes that some subjects have at least d_b repeated measures.

Under the above conditions, the following theorem shows the consistency of the NPMLEs $(\hat{\theta}, \hat{F})$.

Theorem 4.1. *Under Conditions (A1) - (A8),*

$$|\hat{\theta} - \theta_0| \to 0, \quad \sup_{t \in [0,\infty]} |\hat{F}(t) - F_0(t)| \to 0, \quad a.s.$$

Theorem 4.1 then leads to the following results on the asymptotic normality of $(\hat{\theta}, \hat{F})$ and the asymptotic efficiency of $\hat{\theta}$.

Theorem 4.2. *Under Conditions (A1) - (A8), $\sqrt{n}\,(\hat{\theta} - \theta_0,\ \hat{F}(t) - F_0(t))$ weakly converges to a zero-mean Gaussian process in $R^{d_\theta} \times BV[0,\infty]$, where d_θ is the dimension of θ and $BV[0,\infty]$ denotes the space of all functions with bounded variations in $[0,\infty]$. Furthermore, the asymptotic covariance matrix of $\sqrt{n}\,(\hat{\theta} - \theta_0)$ achieves the semiparametric efficiency bound for θ_0.*

Furthermore, in Section 4.8, we show that the inverse of the observed information matrix is a consistent estimator of the asymptotic covariance matrix of the NPMLEs. This result allows us to make inference for any functional of $(\theta,\ F(t))$. To prove Theorems 4.1 - 4.2, we apply the general asymptotic theory of Zeng and Lin (2007). The desired asymptotic properties of the NPMLEs are established followed by the arguments in Appendix B of Zeng and Lin (2007) if we can verify that their regularity conditions hold for our joint cure-survival model setting. Checking the regularity conditions, however, is challenging in our cases. The detailed proofs are provided in Section 4.8.

4.5 Simulation Studies

In this section, we demonstrate the small-sample performance of the proposed method through extensive simulation studies. The longitudinal data are generated from

$$Y(t \mid z_1, z_2, b) = 0.7 + z_1 - 0.5z_2 + b + \epsilon(t),$$

and the survival data with a cure proportion are generated from transformation models

$$S(t \mid z_1, z_2, b) = \exp\left\{-H\left(e^{0.5z_1 - z_2 + \psi b}\, F(t)\right)\right\},$$

where z_1 is a dichotomous covariate taking the value of 0 or 1 with the equal probability of 0.5, z_2 is a continuous covariate generated from a uniform distribution on [-1, 1], and $\epsilon(t) \sim N(0, \sigma_e^2)$ is assumed with $\sigma_e^2 = 1$. The true failure distribution function in the uncured subpopulation is set to be $F(t) = 1 - \exp(-t)$.

For each subject, the correlation within repeated measures is reflected by the subject-specific random intercept $b \sim N(0, \sigma_b^2)$ with $\sigma_b^2 = 0.5$, and the negative, no,

and positive dependences between the longitudinal measures and the cure-survival rate are simulated through different ψ values of -0.3, 0, and 0.3, respectively. For the cure-survival model, we consider three types of transformations $H(\cdot)$ representing the proportional hazards structure ($\eta = 0$), the proportional odds structure ($\eta = 1$), and a transformation in the middle of them with $\eta = 0.5$.

The non-informative censoring time C_i is generated from a uniform distribution with varying rates, depending on the chosen transformation, to design a 30$\sim$45% chance of being right-censored and a 20% chance of being cured. We set longitudinal measures to be observed every 0.2 unit of time so that each individual can have about 3 repeated measures, on average.

The results based on 1000 replications are presented in Tables 4.1 - 4.3 for n=200 and n=400. Tables 4.1 - 4.3 include the average of the differences between the true parameter and the estimates (Bias), the sample standard deviation of the parameter estimators (SE), and the average of the standard error estimators (SEE), and the coverage probability of 95% confidence intervals (CP). The confidence intervals for σ_e^2 and σ_b^2 are constructed based on the the Satterthwaite approximation.

Table 4.1 shows that the NPMLEs under the proportional hazards structure $H(x) = x$ are noticeably unbiased, the standard error estimators calculated via the Louis formula well reflect the true variations of the proposed estimators, and the coverage probabilities are in a reasonable range, even with a moderate sample size of 200. As the sample size increases to 400, the biases slightly increase for some estimates; however, they are still very small comparing to the sizes of true parameter values and the variations of the parameter estimators become smaller, and hence the coverage probabilities still lie in a reasonable range. The simulation results shown in Tables 4.2 - 4.3 are similar to those for Table 4.1, indicating that the proposed method seems to work well for $H(x) = 2\log(1 + x/2)$ and $H(x) = \log(1 + x)$.

4.6 Data Application

The proposed method is applied to the data from the Atherosclerosis Risk in Communities (ARIC) study. The main interest of this analysis is to jointly model the longitudinal pattern of systolic blood pressure (SBP) with the incidence of hospitalized myocardial infarction (MI) or fatal coronary heart diseases (CHD) in patients aged 44-66 years. Our research focuses on a total of 870 white patients living in Forsyth County, who were diagnosed with hypertension at the first examination in 1987-89. This is because the effect of race, known to be a critical factor for the incidence, is nested within the center effect in the ARIC study.

For each subject, SBP was measured four times at approximately three-year intervals, in 1987-89, 1990-92, 1993-95 and 1996-98. On the other hand, the event time of interest, defined as the first time to have either MI or fatal CHD by hospital discharge records, was followed up from 1987 to 2005 with a median follow-up of 16.6 years. During 19 years of the study period, we observe that 158 patients had experienced MI or fatal CHD.

Based on the Kaplan-Meier survival curve in Figure 4.1 (a), we find that the estimated survival rate at the end of study is very high (79%) even after a sufficient follow-up period (19 years). We presume that the high tail probability of the survival curve is due to the patients who are immune to cardiovascular disease, and to adjust this high censoring rate at the end of study, we believe that a cure rate model is a suitable approach for the ARIC data. In fitting the cure rate model, we treat patients as being "cured" or "immune" if they were censored beyond 17.5 years (the largest observed failure time), and 198 patients were considered immune ($X_i = \infty$) in the ARIC data. We note that some of patients who were right-censored before 17.5 years might indeed have been immune to cardiovascular disease, but it is inconclusive due to the right-censoring.

We fit the proposed joint model for the longitudinal SBP trend and MI or fatal CHD event with the patient's baseline information; age at entry, gender, low density lipoprotein (LDL) cholesterol, high density lipoprotein (HDL) cholesterol, and indicators for hypertension lowering medication use, diabetes (with fasting glucose $\geqslant$126 mg/dL) and ever smoker were included as covariates. Among them, LDL- and HDL-cholesterols are standardized at mean 0 and standard deviation 1, and age variable is centered at the mean age of 54 and divided by 10 to represent a decade. In addition, a subject-specific random intercept is included in both longitudinal and survival models to account for the correlation between these two outcomes. We notice that the shared random intercept can also quantify the association between the longitudinal SBPs and the event, given covariates.

We apply transformation models $H(x) = \log(1 + \eta)/\eta$ to cure-survival data by varying values of η in [0, 1]. This class of transformation allows us to explore the possibility of the proportional hazards and the proportional odds structures in cure-survival data. We use the Akaike information criterion (AIC) to determine the best form of transformation (i.e. η). Using the proposed method in Section 4.3, the NPMLEs for the regression coefficients are computed under each transformation model. The observed log-likelihood function increasing over [0, 1] indicates that the proportional odds cure model is the best fit to the data.

Under the selected best transformation model, Table 4.4 summarizes the estimation results and Figure 4.1 (b) displays the estimated survival distribution for the uncured patients ($X_i < \infty$), i.e., $\hat{S}(t) = 1 - \hat{F}(t)$, along with their pointwise 95% confidence bands. We note that the tail probability in the estimated survival curve reaches zero in Figure 4.1 (b). The results in Table 4.4 show that age and the use of hypertension lowering medication are significant factors to explain the pattern of longitudinal SBP, while age, LDL-cholesterol, and diabetes status are significant fac-

tors for MI or fatal CHD events. In more detail, older patients who do not take hypertension medications tend to have higher SBP levels. In terms of MI or fatal CHD events, the risk grows with age and LDL-cholesterol level. Male patients who have diabetes are exposed to higher risk of MI or fatal CHD events. For the random effect b, through the significant variance component $\hat{\sigma_b}^2$, we note that heterogeneity between patients exists in repeated measures. We also notice that the highly significant $\hat{\psi}$ suggests that there is a strong association between the longitudinal SBP levels and MI or fatal CHD - patients with increased SBP levels are likely to have the higher risk of MI or fatal CHD event.

As an example of quantitative interpretation of the results, the marginal survival rates ($\leqslant$17.5 years) and immune fractions (>17.5 years) for the whole population are given in Figure 4.2. By comparing each curve to reference (age of 54, female, HDL-cholesterol 42 mg/dL, LDL-cholesterol 136 mg/dL, no hypertension medication use, never smoking, no diabetes), we can see that the immunity ratios of (age of 74, male, LDL-cholesterol 170 mg/dL, diabetes) relative to reference are (0.76, 0.82, 0.88, 0.74), respectively.

Lastly, we compare the results from our proposed joint model (in Table 4.4) with the one from two separate marginal models (i.e., a linear mixed effects model and a proportional hazards cure model). We find that ignoring the cure fraction and the correlation between the longitudinal and survival components (ψ) and fitting separate marginal models shows substantial variations in the model estimates.

4.7 Concluding Remarks

We have proposed the joint transformation model for longitudinal and survival data with a cure fraction. Ignoring a heavy tail probability in the survival function, caused

by a true cure proportion in the study population, will produce biased estimates for both regression coefficients and survival prediction. The unique features of the proposed model, compared to the existing joint models, are that it takes the possibility of patients being cured or immune to disease into account and it allows us to explore a feasible proportional hazards or proportional odds structure in the cure rate model through varied forms of transformations. We have used the NPMLEs to make inferences on the model parameters, and the NPMLEs have been shown to be asymptotically efficient. The new EM algorithm has offered a simpler and more stable way to compute the NPMLEs. In addition, the proposed method has been well evaluated through simulation studies and illustrated using the ARIC data.

The proposed approach has the advantage of handling time-varying covariates in the cure-survival model. Lu and Ying (2004) generalized transformation cure models based on a type of mixture cure model. However, their approach is limited to only time-independent covariates due to the form of transformations. Furthermore, the extra constraint on the tail of the estimated transformation function is necessary to resolve the identifiability issue of cure parameters with a finite sample size, likewise in the usual mixture cure models (Taylor, 1995; Sy and Taylor, 2000; Peng and Dear, 2000).

In this paper, we assumed that the number of observations of repeated measures are independent of survival data. Our joint cure model can be extended in a way to account for the informative observation process. The AIC was used to determine the best transformation, but there exist other criteria for model selection such as the Bayes information criterion and cross-validation ('leave-one-subject-out'). Further method development for model checking would be useful for the practical application of the joint models.

4.8 Proof of Asymptotic Properties

This section proves Theorems 4.1 - 4.2 stated in Section 4.4 by applying the general asymptotic theory of Zeng and Lin (2007). Specifically, it is easy to see that our conditions (A1) - (A8) imply (C1) - (C4), (C6), (C8) of Zeng and Lin (2007b), and it remains to prove the two identifiability conditions (C5) and (C7) of Zeng and Lin (2007b). The first identifiability is the key step to prove the consistency of the NPMLEs, and the second is to entail the invertibility of the observed information matrix at the true parameters for the proof of the asymptotic normality.

Proof of the First Identifiability

Proof. First, we verify the first identifiability condition (C5) in Appendix B of Zeng and Lin (2007b). Suppose that the likelihood function for (α, β, ψ, σ_e^2, Vec(Σ_b)) is the same as that for the true parameter values (α_0, β_0, ψ_0, σ_{0e}^2, Vec(Σ_{0b})). That is, for arbitrary $K > 0$,

$$\int_b (2\pi\sigma_e^2)^{-\frac{K}{2}} \exp\left\{-\frac{(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha-\tilde{\boldsymbol{Z}}_1 b)^T(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha-\tilde{\boldsymbol{Z}}_1 b)}{2\sigma_e^2}\right\}$$
$$\times \left[f(x)\, e^{\beta^T Z_2(x)+(b\circ\psi)^T \tilde{Z}_2(x)}\, H'(q(x))\right]^{\Delta} \exp\{-H(q(x))\}\, f(b;\Sigma_b)\, db$$

$$= \int_b (2\pi\sigma_{0e}^2)^{-\frac{K}{2}} \exp\left\{-\frac{(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha_0-\tilde{\boldsymbol{Z}}_1 b)^T(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha_0-\tilde{\boldsymbol{Z}}_1 b)}{2\sigma_{0e}^2}\right\}$$
$$\times \left[f_0(x)\, e^{\beta_0^T Z_2(x)+(b\circ\psi_0)^T \tilde{Z}_2(x)}\, H'(q_0(x))\right]^{\Delta} \exp\{-H(q_0(x))\}\, f(b;\Sigma_{0b})\, db, \quad (4.8)$$

where bold $\boldsymbol{Y}$ denotes the vector of the observed longitudinal measures at time $s_1, \ldots, s_K$, and $\boldsymbol{Z}_1$ and $\tilde{\boldsymbol{Z}}_1$ in bold type denote matrices with each row equal to the observed covariate $Z_1(s_k)^T$ and $\tilde{Z}_1(s_k)^T$ at $k = 1, \ldots, K$, respectively. In addition,

$q(t) = \int_0^t e^{\beta^T Z_2(u) + (b \circ \psi)^T \tilde{Z}_2(u)} \, dF(u)$, and $q_0(t)$ is $q(t)$ evaluated at the true parameter values, and $f(b; \Sigma_b)$ is the density function of the (multivariate) normal distribution with mean zeros and covariance matrix Σ_b. From now, we take the following actions on both sides of (4.8).

Step 1. For the proof of the identifiability of the longitudinal component, we consider a case $\Delta = 0$ and $X \approx 0$.

Using the fact that $\int b \, f(b; \Sigma_b) \, db = \int b \, f(b; \Sigma_{0b}) \, db = 0$ and considering $E[Y(s_k)]$ conditional on b, we have $\alpha^T Z_1(s_k) = \alpha_0^T Z_1(s_k)$, for $k = 1, \ldots, K$. By Condition (A6), we prove $\alpha = \alpha_0$. Similarly, we consider $E[Y(s_k)Y(s_{k'})]$ and $\mathrm{Var}(Y(s_k))$, given b, and obtain for $k \neq k'$

$$\begin{aligned} & \int_b \left\{ \alpha_0^T Z_1(s_k) + b^T \tilde{Z}_1(s_k) \right\} \left\{ \alpha_0^T Z_1(s_{k'}) + b^T \tilde{Z}_1(s_{k'}) \right\} f(b; \Sigma_b) \, db \\ = & \int_b \left\{ \alpha_0^T Z_1(s_k) + b^T \tilde{Z}_1(s_k) \right\} \left\{ \alpha_0^T Z_1(s_{k'}) + b^T \tilde{Z}_1(s_{k'}) \right\} f(b; \Sigma_{0b}) \, db, \end{aligned}$$

followed by the proof of $\Sigma_b = \Sigma_{0b}$ from (A6), and

$$\begin{aligned} & \int_b \left\{ \sigma_e^2 + b^T \tilde{Z}_1(s_k) \tilde{Z}_1(s_k)^T b \right\} f(b; \Sigma_b) \, db \\ = & \int_b \left\{ \sigma_{0e}^2 + b^T \tilde{Z}_1(s_k) \tilde{Z}_1(s_k)^T b \right\} f(b; \Sigma_{0b}) \, db, \end{aligned}$$

for $k = 1, \ldots, K$. Accordingly, we have that $\sigma_e^2 = \sigma_{0e}^2$.

Step 2. For the survival component, suppose $\Delta = 0$ and $X = t$. Then, (4.8) implies

$$E_b \left[\exp \left\{ -H(q(t)) \right\} \right] = E_b \left[\exp \left\{ -H(q_0(t)) \right\} \right],$$

where b follows a normal distribution with mean $\mu_b = V_b \tilde{\boldsymbol{Z}}_1^T (\boldsymbol{Y} - \boldsymbol{Z}_1 \alpha_0)/\sigma_{0e}^2$ and covariance matrix $V_b = [\Sigma_{0b}^{-1} + \tilde{\boldsymbol{Z}}_1^T \tilde{\boldsymbol{Z}}_1/\sigma_{0e}^2]^{-1}$. For fixed $\boldsymbol{Y}$, $\boldsymbol{Z}_1$, and $\tilde{\boldsymbol{Z}}_1$, since b is the complete statistic for μ_b, we can have that

$$\begin{aligned} & \exp\left\{-H(\int_0^t e^{\beta^T Z_2(u)+(b\circ\psi)^T \tilde{Z}_2(u)}\, dF(u))\right\} \\ = \; & \exp\left\{-H(\int_0^t e^{\beta_0^T Z_2(u)+(b\circ\psi_0)^T \tilde{Z}_2(u)}\, dF_0(u))\right\}. \end{aligned}$$

Furthermore, it is followed from the one-to-one mapping of H and exponential function that

$$\log(f(t)) + \beta^T Z_2(t) + b^T(\psi \circ \tilde{Z}_2(t)) = \log(f_0(t)) + \beta_0^T Z_2(t) + b^T(\psi_0 \circ \tilde{Z}_2(t)),$$

with probability 1. By taking the expectation with respect to b for fixed $\boldsymbol{Y}$, $\boldsymbol{Z}_1$, and $\tilde{\boldsymbol{Z}}_1$, we conclude that $\beta = \beta_0$, $f(t) = f_0(t)$ and $\psi = \psi_0$ from the Condition (A6). □

Proof of the Second Identifiability

Proof. Next, we verify the second identifiability condition (C7) in Appendix B of Zeng and Lin (2007b). It starts from the score equation along with the path $(\alpha_0 + \xi\nu_1,\ \beta_0 + \xi\nu_2,\ \psi_0 + \xi\nu_3,\ \sigma_{0e}^2 + \xi\nu_4,\ \mathrm{Vec}(\Sigma_{0b}) + \xi\nu_b,\ F_0 + \xi \int h\, dF_0)$. We define D_b as the symmetric matrix such that $\mathrm{Vec}(D_b) = \nu_b$.

Step 1. To make the score equation simple for the proofs of $\nu_1 = 0$, $\nu_4 = 0$ and $D_b = 0$, we consider the same case $\Delta = 0$ and $X \approx 0$ as used in *Step 1* of the first identifiability proof. We define

$$V_b^{-1} = \Sigma_{0b}^{-1} + \tilde{\boldsymbol{Z}}_1^T \tilde{\boldsymbol{Z}}_1/\sigma_{0e}^2, \quad \text{and} \quad \mu_b = V_b\, \tilde{\boldsymbol{Z}}_1^T (\boldsymbol{Y} - \boldsymbol{Z}_1 \alpha_0)/\sigma_{0e}^2,$$

then, the score equation is given by

$$\begin{aligned} 0 &= -\frac{1}{2}\mathrm{Tr}(\Sigma_{0b}^{-1}D_b) + \frac{1}{2}\mu_b^T\Sigma_{0b}^{-1}D_b\,\Sigma_{0b}^{-1}\mu_b + \frac{1}{2}\mathrm{Tr}(\Sigma_{0b}^{-1}D_b\,\Sigma_{0b}^{-1}V_b) + \frac{\nu_4}{2\sigma_{0e}^4}\mathrm{Tr}(\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1V_b) \\ &\quad -\frac{\nu_4\,K}{2\sigma_{0e}^2} + \frac{\nu_1^T\boldsymbol{Z}_1^T(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha_0-\tilde{\boldsymbol{Z}}_1\mu_b)}{\sigma_{0e}^2} \\ &\quad +\frac{\nu_4}{2\sigma_{0e}^4}\left\{(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha_0)^T(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha_0) - 2(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha_0)\tilde{\boldsymbol{Z}}_1\mu_b + \mu_b^T\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1\mu_b\right\}. \quad (4.9) \end{aligned}$$

By comparing coefficients for the constant, linear and quadratic terms of $(\boldsymbol{Y}-\boldsymbol{Z}_1\alpha_0)$, we have that

$$0 = \frac{\nu_4}{2\sigma_{0e}^2}[K - \frac{\mathrm{Tr}(\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1V_b)}{\sigma_{0e}^2}] + \frac{1}{2}\mathrm{Tr}(\Sigma_{0b}^{-1}D_b) - \frac{1}{2}\mathrm{Tr}(\Sigma_{0b}^{-1}D_b\,\Sigma_{0b}^{-1}V_b), \quad (4.10)$$

$$0 = \frac{\nu_1^T\boldsymbol{Z}_1^T}{\sigma_{0e}^2}[I - \frac{\tilde{\boldsymbol{Z}}_1V_b\tilde{\boldsymbol{Z}}_1^T}{\sigma_{0e}^2}], \quad (4.11)$$

$$0 = \frac{\nu_4}{2\sigma_{0e}^4}[I - \frac{2\tilde{\boldsymbol{Z}}_1V_b\tilde{\boldsymbol{Z}}_1^T}{\sigma_{0e}^2} + \frac{\tilde{\boldsymbol{Z}}_1V_b\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1V_b\tilde{\boldsymbol{Z}}_1^T}{\sigma_{0e}^4}] + \frac{\tilde{\boldsymbol{Z}}_1V_b\Sigma_{0b}^{-1}D_b\,\Sigma_{0b}^{-1}V_b\tilde{\boldsymbol{Z}}_1^T}{2\sigma_{0e}^4}. \quad (4.12)$$

Since $[I - \tilde{\boldsymbol{Z}}_1V_b\tilde{\boldsymbol{Z}}_1^T/\sigma_{0e}^2]$ is positive definite, we can see that $\nu_1 = 0$ in (4.11). To simplify (4.12), we multiply $\tilde{\boldsymbol{Z}}_1^T$ from the left, $\tilde{\boldsymbol{Z}}_1$ from the right, and then $[\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1]^{-1}$ from the right on both sides of (4.12). Using the fact that $\Sigma_{0b}^{-1}D_b = I - \tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1V_b/\sigma_{0e}^2$, the equation (4.12) becomes

$$\frac{\nu_4}{2\sigma_{0e}^2}\left[I - \frac{\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1V_b}{\sigma_{0e}^2}\right] + \frac{\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1V_b\,\Sigma_{0b}^{-1}D_b}{2\sigma_{0e}^2} = 0, \quad (4.13)$$

and the equation (4.10) becomes

$$\frac{\nu_4}{2\sigma_{0e}^2}\left[K - \frac{\mathrm{Tr}(\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1V_b)}{\sigma_{0e}^2}\right] + \frac{1}{2\sigma_{0e}^2}\mathrm{Tr}(\tilde{\boldsymbol{Z}}_1^T\tilde{\boldsymbol{Z}}_1V_b\,\Sigma_{0b}^{-1}D_b) = 0. \quad (4.14)$$

After taking the trace of (4.13) and subtracting from the equation (4.14), we obtain that

$$\frac{\nu_4}{2\sigma_{0e}^2}(K - d_b) = 0,$$

where d_b stands for the dimension of b. Based on Condition (A8), we conclude that $\nu_4 = 0$, and hence $D_b = 0$ in (4.13) by Condition (A7).

Step 2. For the second identifiability of the survival component, we set $\Delta = 0$ and $X = t$. Then, the score equation can be written as

$$E_b \left[\exp\{-H(q_0(t))\}\, H(q_0(t))\, \dot{q}_0(t)\right] = 0, \tag{4.15}$$

where $\dot{q}_0(t) = \int_0^t \{h(u) + \nu_2^T Z_2(u) + (\nu_3 \circ b)^T \tilde{Z}_2(u)\}\, e^{\beta_0^T Z_2(u) + (b \circ \psi_0)^T \tilde{Z}_2(u)}\, dF_0(u)$, and b is normally distributed with mean μ_b and covariance matrix V_b. By the completeness of the exponential family of b, we can have

$$\exp\{-H(q_0(t))\}\, H(q_0(t))\, \dot{q}_0(t) = 0,$$

for any fixed $\boldsymbol{Y}$, $\boldsymbol{Z}_1$ and $\tilde{\boldsymbol{Z}}_1$ with probability 1. Since $H(q_0(t)) > 0$ for $\forall t > 0$ from (A5), we can obtain $\dot{q}_0(t) = 0$, and hence

$$h(t) + \nu_2^T Z_2(t) + (\nu_3 \circ b)^T \tilde{Z}_2(t) = 0.$$

Clearly, we attain $\nu_2 = 0$, $\nu_3 = 0$ and $h = 0$ by (A6). □

Finally, we complete the proofs of Theorems 4.1 - 4.2 by Theorems 1 - 2 in Zeng and Lin (2007b). Let I_n denote the negative Hessian matrix of the observed log-likelihood function with respect to $(\theta, F\{\cdot\})$. As a remark, by following Theorem 3 in Zeng and Lin (2007b), we can show that I_n is invertible for large n, and $(\nu^T, U^T)\, n I_n^{-1} (\nu^T, U^T)^T$

is the consistent estimator of the asymptotic variance of

$$\sqrt{n}\left\{\nu^T(\hat{\theta}-\theta_0)+\int u(t)\,d(\hat{F}-F_0)\right\},$$

where U is the vector of $u(\cdot)$ at the observed failure times.

Table 4.1: Simulation results for $H(x) = x$. t_p represents the pth percentile.

		$N = 200$				$N = 400$			
Parameter	*True*	*Bias*	*SE*	*SEE*	*CP*	*Bias*	*SE*	*SEE*	*CP*
					$\psi = -0.3$				
α	0.7	-0.009	0.101	0.099	0.944	-0.012	0.073	0.074	0.952
	1.0	0.002	0.143	0.142	0.952	-0.004	0.104	0.103	0.947
	-0.5	0.002	0.125	0.129	0.954	-0.000	0.093	0.096	0.951
σ_e^2	1.0	-0.001	0.064	0.064	0.956	-0.000	0.042	0.041	0.946
β	0.5	0.009	0.133	0.136	0.951	0.009	0.096	0.094	0.946
	-1.0	-0.020	0.165	0.168	0.946	-0.016	0.120	0.118	0.945
ψ	-0.3	-0.010	0.192	0.204	0.968	-0.026	0.148	0.144	0.952
$F(t_{25})$	0.25	-0.005	0.033	0.035	0.949	-0.003	0.025	0.025	0.946
$F(t_{50})$	0.50	-0.004	0.048	0.051	0.954	-0.005	0.035	0.035	0.951
$F(t_{75})$	0.75	-0.003	0.053	0.054	0.942	-0.002	0.036	0.036	0.952
σ_b^2	0.5	-0.008	0.088	0.090	0.963	0.006	0.066	0.064	0.942
					$\psi = 0.0$				
α	0.7	-0.008	0.102	0.098	0.936	-0.010	0.070	0.070	0.941
	1.0	-0.002	0.147	0.141	0.943	-0.001	0.099	0.099	0.959
	-0.5	0.002	0.129	0.127	0.942	0.008	0.089	0.090	0.953
σ_e^2	1.0	0.001	0.062	0.064	0.963	-0.002	0.046	0.045	0.947
β	0.5	0.007	0.130	0.129	0.955	0.004	0.089	0.090	0.956
	-1.0	-0.013	0.166	0.160	0.941	-0.009	0.116	0.112	0.935
ψ	0.0	-0.017	0.186	0.190	0.967	-0.018	0.128	0.130	0.952
$F(t_{25})$	0.25	-0.003	0.034	0.034	0.941	-0.002	0.023	0.024	0.945
$F(t_{50})$	0.50	-0.004	0.050	0.049	0.936	-0.002	0.035	0.034	0.944
$F(t_{75})$	0.75	-0.002	0.054	0.052	0.937	-0.001	0.037	0.037	0.948
σ_b^2	0.5	-0.009	0.084	0.088	0.965	-0.003	0.062	0.062	0.961
					$\psi = 0.3$				
α	0.7	-0.011	0.098	0.099	0.955	-0.011	0.071	0.070	0.943
	1.0	0.002	0.141	0.141	0.944	-0.001	0.100	0.099	0.944
	-0.5	0.001	0.122	0.129	0.959	0.006	0.094	0.090	0.944
σ_e^2	1.0	0.001	0.062	0.065	0.960	-0.001	0.046	0.046	0.946
β	0.5	-0.003	0.133	0.135	0.948	-0.003	0.093	0.093	0.953
	-1.0	-0.015	0.166	0.168	0.954	-0.002	0.112	0.115	0.958
ψ	0.3	-0.003	0.200	0.202	0.953	-0.023	0.137	0.136	0.955
$F(t_{25})$	0.25	-0.002	0.033	0.035	0.961	-0.000	0.023	0.025	0.953
$F(t_{50})$	0.50	-0.001	0.050	0.051	0.950	-0.000	0.034	0.035	0.955
$F(t_{75})$	0.75	-0.000	0.053	0.053	0.937	0.000	0.037	0.037	0.950
σ_b^2	0.5	-0.011	0.085	0.089	0.966	-0.007	0.060	0.062	0.958

Table 4.2: Simulation results for $H(x) = 2\log(1 + x/2)$. t_p represents the pth percentile.

		$N = 200$				$N = 400$			
Parameter	*True*	*Bias*	*SE*	*SEE*	*CP*	*Bias*	*SE*	*SEE*	*CP*
					$\psi = -0.3$				
α	0.7	-0.001	0.096	0.099	0.953	-0.008	0.070	0.070	0.944
	1.0	-0.004	0.137	0.138	0.954	-0.002	0.097	0.098	0.942
	-0.5	0.003	0.126	0.125	0.952	0.002	0.087	0.088	0.956
σ_e^2	1.0	0.004	0.063	0.063	0.943	-0.000	0.045	0.044	0.938
β	0.5	0.009	0.174	0.174	0.944	0.010	0.121	0.121	0.955
	-1.0	-0.021	0.209	0.209	0.950	-0.014	0.149	0.146	0.945
ψ	-0.3	-0.034	0.286	0.278	0.955	-0.034	0.199	0.192	0.944
$F(t_{25})$	0.25	-0.005	0.036	0.038	0.949	-0.003	0.027	0.026	0.944
$F(t_{50})$	0.50	-0.005	0.054	0.054	0.948	-0.004	0.038	0.038	0.950
$F(t_{75})$	0.75	-0.004	0.059	0.057	0.932	-0.002	0.040	0.040	0.948
σ_b^2	0.5	-0.008	0.091	0.087	0.949	-0.002	0.064	0.061	0.950
					$\psi = 0.0$				
α	0.7	-0.008	0.104	0.099	0.930	-0.005	0.070	0.070	0.952
	1.0	-0.005	0.144	0.138	0.950	-0.002	0.099	0.098	0.943
	-0.5	0.011	0.121	0.124	0.954	0.006	0.084	0.088	0.954
σ_e^2	1.0	-0.001	0.064	0.062	0.945	-0.001	0.044	0.044	0.946
β	0.5	0.005	0.171	0.170	0.945	0.003	0.120	0.118	0.949
	-1.0	-0.012	0.211	0.204	0.951	-0.007	0.152	0.143	0.939
ψ	0.0	-0.031	0.286	0.270	0.936	-0.027	0.191	0.188	0.943
$F(t_{25})$	0.25	-0.005	0.035	0.037	0.945	-0.003	0.027	0.026	0.942
$F(t_{50})$	0.50	-0.007	0.051	0.053	0.953	-0.003	0.038	0.038	0.952
$F(t_{75})$	0.75	-0.004	0.056	0.056	0.939	-0.001	0.040	0.040	0.948
σ_b^2	0.5	-0.009	0.082	0.085	0.966	-0.007	0.059	0.060	0.958
					$\psi = 0.3$				
α	0.7	-0.005	0.102	0.098	0.943	-0.006	0.067	0.070	0.961
	1.0	-0.005	0.141	0.138	0.936	-0.004	0.095	0.098	0.951
	-0.5	0.006	0.123	0.124	0.951	0.003	0.086	0.088	0.961
σ_e^2	1.0	-0.002	0.062	0.062	0.952	0.000	0.043	0.044	0.952
β	0.5	0.000	0.169	0.173	0.954	-0.007	0.121	0.120	0.947
	-1.0	-0.018	0.212	0.208	0.949	0.001	0.144	0.145	0.958
ψ	0.3	-0.011	0.287	0.275	0.950	-0.028	0.191	0.190	0.946
$F(t_{25})$	0.25	-0.004	0.035	0.037	0.949	0.000	0.027	0.026	0.939
$F(t_{50})$	0.50	-0.002	0.053	0.054	0.950	0.000	0.038	0.038	0.953
$F(t_{75})$	0.75	-0.001	0.057	0.057	0.949	0.001	0.041	0.040	0.951
σ_b^2	0.5	-0.010	0.083	0.086	0.964	-0.004	0.061	0.061	0.957

Table 4.3: Simulation results for $H(x) = \log(1 + x)$. t_p represents the pth percentile.

		$N = 200$				$N = 400$			
Parameter	*True*	*Bias*	*SE*	*SEE*	*CP*	*Bias*	*SE*	*SEE*	*CP*
					$\psi = -0.3$				
α	0.7	-0.002	0.099	0.098	0.945	-0.002	0.071	0.070	0.942
	1.0	-0.008	0.137	0.136	0.949	-0.005	0.095	0.096	0.958
	-0.5	0.007	0.125	0.122	0.944	0.000	0.088	0.086	0.947
σ_e^2	1.0	-0.003	0.062	0.061	0.952	-0.001	0.043	0.043	0.953
β	0.5	0.009	0.213	0.207	0.948	0.003	0.139	0.144	0.960
	-1.0	-0.030	0.255	0.245	0.938	-0.008	0.175	0.170	0.950
ψ	-0.3	-0.048	0.359	0.348	0.943	-0.028	0.254	0.243	0.944
$F(t_{25})$	0.25	-0.006	0.040	0.040	0.939	-0.003	0.028	0.028	0.951
$F(t_{50})$	0.50	-0.006	0.059	0.058	0.941	-0.000	0.041	0.041	0.950
$F(t_{75})$	0.75	-0.003	0.063	0.061	0.940	-0.000	0.044	0.043	0.942
σ_b^2	0.5	-0.008	0.083	0.085	0.965	-0.004	0.063	0.060	0.940
					$\psi = 0.0$				
α	0.7	-0.008	0.102	0.099	0.933	-0.004	0.072	0.070	0.952
	1.0	0.000	0.142	0.136	0.947	0.000	0.097	0.097	0.956
	-0.5	0.004	0.122	0.122	0.957	-0.001	0.087	0.086	0.944
σ_e^2	1.0	0.003	0.060	0.061	0.955	-0.001	0.043	0.043	0.946
β	0.5	0.005	0.205	0.203	0.954	-0.001	0.144	0.142	0.953
	-1.0	-0.016	0.239	0.241	0.948	-0.009	0.166	0.168	0.947
ψ	0.0	-0.031	0.359	0.349	0.942	-0.025	0.244	0.240	0.945
$F(t_{25})$	0.25	-0.005	0.041	0.040	0.940	-0.002	0.027	0.028	0.953
$F(t_{50})$	0.50	-0.003	0.057	0.058	0.944	-0.000	0.041	0.041	0.952
$F(t_{75})$	0.75	-0.001	0.061	0.060	0.934	-0.000	0.043	0.043	0.933
σ_b^2	0.5	-0.015	0.085	0.083	0.957	-0.003	0.061	0.060	0.949
					$\psi = 0.3$				
α	0.7	-0.001	0.097	0.099	0.947	-0.008	0.069	0.070	0.949
	1.0	-0.008	0.138	0.136	0.939	0.005	0.096	0.096	0.951
	-0.5	-0.001	0.121	0.122	0.958	0.007	0.085	0.086	0.949
σ_e^2	1.0	0.002	0.062	0.061	0.946	-0.001	0.042	0.043	0.959
β	0.5	-0.007	0.205	0.205	0.956	-0.004	0.146	0.143	0.943
	-1.0	-0.019	0.253	0.243	0.941	0.002	0.173	0.170	0.942
ψ	0.3	-0.035	0.352	0.350	0.956	-0.031	0.246	0.243	0.951
$F(t_{25})$	0.25	-0.004	0.039	0.040	0.943	-0.000	0.029	0.028	0.936
$F(t_{50})$	0.50	-0.004	0.058	0.058	0.951	0.000	0.042	0.041	0.949
$F(t_{75})$	0.75	-0.003	0.062	0.061	0.931	-0.000	0.043	0.043	0.942
σ_b^2	0.5	-0.012	0.083	0.084	0.960	-0.007	0.059	0.060	0.950

Table 4.4: Analysis results for the ARIC study. The 50:50 mixture of χ^2 distributions is used for testing variances.

Effect	*Estimate*	*Std.Error*	*p-value*
Longitudinal measures of SBP			
Intercept	1.057	0.064	$< .0001$
Age	0.334	0.045	$< .0001$
Male	-0.015	0.057	0.7910
LDL-cholesterol (mg/dL)	0.011	0.023	0.6409
HDL-cholesterol (mg/dL)	0.025	0.021	0.2260
Hypertension medication (yes vs. no)	-0.561	0.055	$< .0001$
Ever smoker (yes vs. no)	-0.052	0.053	0.3306
Diabetes	0.100	0.077	0.1975
σ_e^2	0.500	0.016	$< .0001$
σ_b^2	0.337	0.025	$< .0001$
MI/fatal CHD event			
Intercept	-2.231	0.271	$< .0001$
Age	0.526	0.174	0.0025
Male	0.700	0.224	0.0018
LDL-cholesterol (mg/dL)	0.248	0.085	0.0034
HDL-cholesterol (mg/dL)	-0.130	0.091	0.1502
Hypertension medication (yes vs. no)	-0.030	0.207	0.8853
Ever smoker (yes vs. no)	0.308	0.207	0.1368
Diabetes ($\geqslant 126$ vs. <126 mg/dL)	1.152	0.231	$< .0001$
ψ	0.487	0.206	0.0182

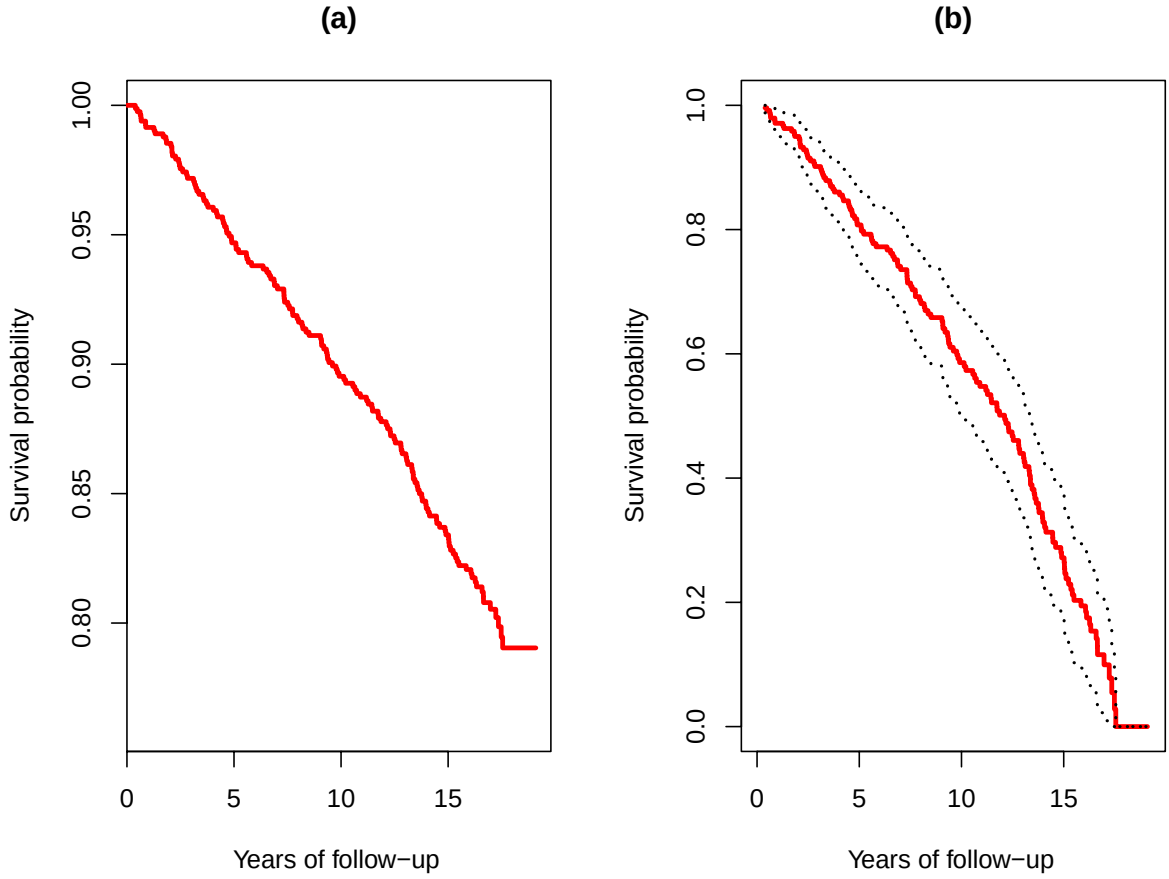

Figure 4.1: In the ARIC data (a) Kaplan-Meier survival curve of the entire study population; (b) Estimated survival curve of the non-immune subpopulation under the joint cure-survival model with the proportional odds structure. The solid curves are point estimates, and the dotted curves are 95% confidence bands.

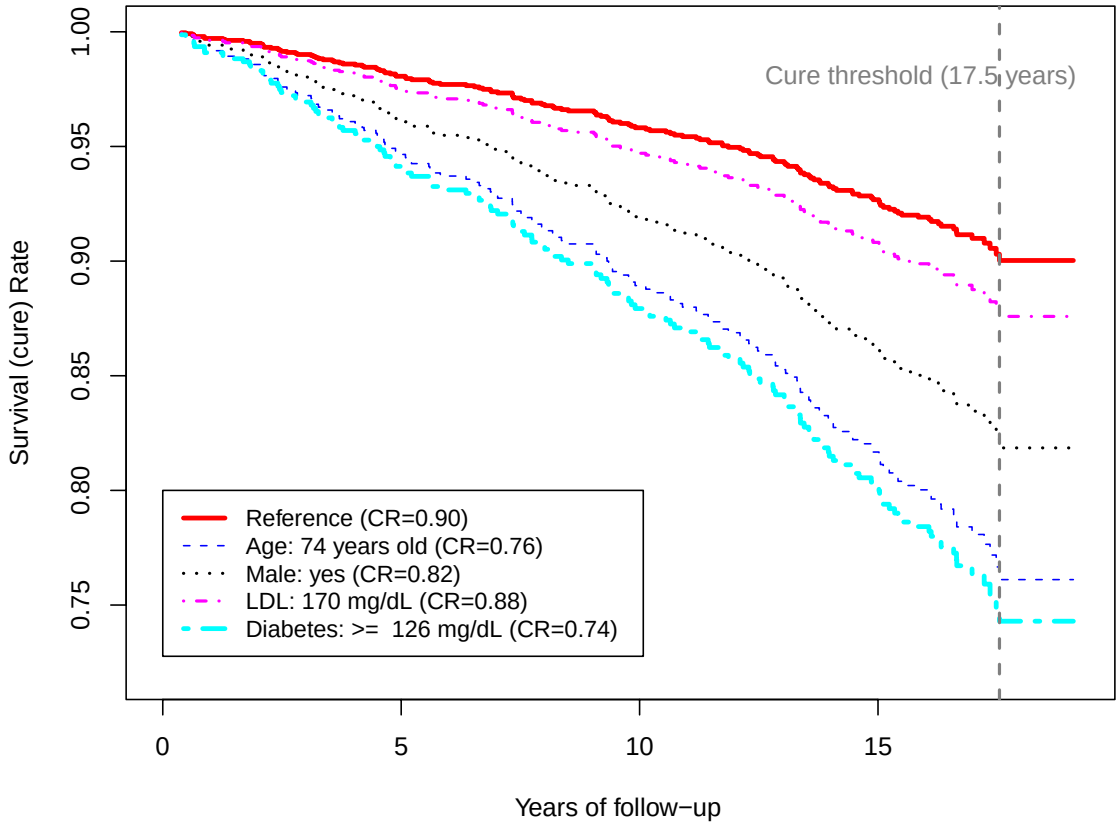

Figure 4.2: Predicted marginal survival rates of the entire population using the results in Table 4.4. The rates beyond the cure threshold are interpreted as the immune fractions or the cure rates (CR). Reference is taken for age of 54, female, HDL-cholesterol 42 mg/dL, LDL-cholesterol 136 mg/dL, no hypertension medication use, never smoking, and no diabetes.

Chapter 5

Partially Linear Model for Longitudinal Data with Informative Censoring

5.1 Introduction

Joint analysis of longitudinal and survival data, where the longitudinal data are repeatedly measured at irregular times, but are subject to informative right-censoring, has received considerable attention in recent literature. Methods and theory for the joint analysis with the fully parameterized mean structure of longitudinal responses have been well developed (Zeng and Cai, 2005a,b), while some work has considered relaxing distributional assumptions on random effects (Tsiatis and Davidian, 2001; Song et al., 2002; Brown and Ibrahim, 2003a). However, in some longitudinal data, the full parametric specification between longitudinal responses and covariates may be insufficient to reflect the complicated patterns. Examples include longitudinal trajectories of CD4 cell counts (Zeger and Diggle, 1994; Lin and Ying, 2001; Huang et al., 2002; Brown et al., 2005) and time-varying treatment effects (Hogan et al., 2004) in

HIV/AIDS research; time-varying effects of gender and HIV status on the growth of infants born from HIV infected mothers (Hoover et al., 1998); age effects on childhood respiratory disease (Diggle et al., 2002); and treatment effects on the number of bladder tumors over time (Sun et al., 2005; Liang et al., 2009). In other applications, it was shown that these longitudinal measures were affected by informative drop-out of patients, for instance, due to death or side effects of treatments (Vonesh et al., 2006; Liu et al., 2007). Motivated by these practical examples, we focus on modeling trajectories of longitudinal responses with an unspecified smooth function and linear covariate effects, while taking informative right-censoring into account. We treat the informative censorship as an event that terminates subsequent observations such as death and withdrawal. Advantages of the semiparametric approach are that it allows for an easier interpretation of the covariate effects, compared to standard nonparametric regression models, and is a natural way to assess the effect of covariates on the censoring process.

In the absence of informative censoring, partially linear models for repeated measurements were developed by Zeger and Diggle (1994), Moyeed and Diggle (1994), Zhang et al. (1998), Rice and Wu (2001), and Lin and Carroll (2001), among others. For estimation, they adapted different nonparametric regression techniques such as kernal smoothing, smoothing splines, and regression splines. For a more detailed literature review, we refer readers to Section 2.1.2.

In the presence of informative censoring, Hogan et al. (2004) developed a varying-coefficient mixture model of longitudinal data, conditional on a discrete or continuous censoring time. In the conditional model, covariate effects were allowed to depend on informative censoring time through an unspecified coefficient function of the censoring time, and step functions or cubic smoothing splines (depending on the continuity of the underlying censoring time) were used for estimation of coefficient functions.

Due to the conditional structure, use of their model is restricted to investigators who need to assess how the covariates affect differently longitudinal data and the censoring process. In contrast, a joint model was proposed by Brown et al. (2005), where patterns of longitudinal markers using a partially linear model and an event time using a proportional hazards model were simultaneously modeled. They used cubic B-splines to estimate the nonparametric function in the longitudinal component, while using piecewise constants to estimate the baseline hazard function. Estimation procedures were implemented through a Bayesian approach.

In this paper, we propose a partially linear model for longitudinal data while allowing simultaneously the underlying censoring times to be possibly dependent on covariates through general transformation models, including the proportional hazards and the proportional odds models as special cases. We use B-splines to estimate the baseline function of repeated measurements. A key advantage of B-splines is its computational simplicity in the use of a small number of knots and implementation of a parametric regression using a fixed number of base functions. Use of the B-spline method counterbalances the model complexity in the joint modeling approach. Through the flexible but readily interpretable modeling approach, we can detect significant changes in the level or direction of the trajectory, which may not be found in linear mixed effects models due to the underlying parametric model constraints, and we can also correct biases caused by informative censoring.

The rest of the chapter is organized as follows. We introduce the partially linear model of longitudinal data with informative censoring in Section 5.2, and we describe efficient inference procedures based on maximum likelihood estimation (MLE) via a simple EM algorithm in Section 5.3. In Section 5.4, we establish asymptotic properties of the proposed MLEs. We assess the validity of the proposed method through simulated datasets in Section 5.5 and an example of medical costs data in Section

5.6. Finally, we conclude with some remarks in Section 5.7 and provide proofs of the asymptotic properties in Section 5.8.

5.2 Joint Models

Let $Y(t)$ be the longitudinal response at time t, and let T be the informative censoring time. We define $\mathcal{X} = \{X(t);\, t \geqslant 0\}$ and $\mathcal{Z} = \{Z(t);\, t \geqslant 0\}$ as the covariate processes of fixed and random effects, respectively, where $X(t)$ and $Z(t)$ are the vectors of external covariates at time t, possibly time-varying. We introduce a common latent variable b to account for the correlation between $Y(t)$ and T, assuming b follows a (multivariate) normal distribution with mean zeros and covariance matrix Σ_b. We further assume that $Y(t)$ and T are independent, conditional on $\mathcal{X}$, $\mathcal{Z}$ and b. We then propose to jointly model $Y(t)$ through a partially linear model

$$Y(t \,|\, \mathcal{X}, \mathcal{Z}, b) \;=\; \alpha(t) + \beta^T X_1(t) + b^T Z_1(t) + \epsilon(t), \tag{5.1}$$

and T through the transformed Cox model with the cumulative hazard function

$$\Lambda(t \,|\, \mathcal{X}, \mathcal{Z}, b) \;=\; H\left(\int_0^t \exp\{\gamma^T X_2(u) + (\phi \circ b)^T Z_2(u)\}\, d\Lambda(u) \right), \tag{5.2}$$

where $\alpha(t)$ is the underlying nonparametric trajectory, β and γ are the vectors of unknown regression coefficients, and $\Lambda(\cdot)$ is an unspecified increasing function. In the models, $X_i(t)$ and $Z_i(t)$ $(i = 1, 2)$ are subsets of $X(t)$ and $Z(t)$, respectively, and $\epsilon(t)$ is a white noise process with variance σ_e^2. In the model (5.2), ϕ is a set of unknown constants with the same number of elements as b, and $\phi \circ b$ denotes the component-wise product of ϕ and b. We note that informative censorship in longitudinal data is adjusted by ϕ in (5.2) with the shared frailty b and the common covariates between

the longitudinal and censoring models.

In the model (5.2), $H(\cdot)$ represents a transformation function of the cumulative hazard function of the censoring time, which is required to be specified in the analysis. The transformation function is assumed to be continuously differentiable and strictly increasing. For example, $H(x)$ can take a form of the logarithmic transformation,

$$H(x) = \begin{cases} \log(1+\eta x)/\eta, & \eta > 0 \\ x, & \eta = 0. \end{cases}$$

The choices of $\eta = 0$ and $\eta = 1$ lead to the proportional hazards model and the proportional odds model, respectively.

Let C be the non-informative censoring time assumed to be independent of ($Y(\cdot)$, T, b) given $\mathcal{X}$ and $\mathcal{Z}$, and let $V = \min(T, C)$ denote the observed censoring time. The observed data for the ith subject with n_i repeated measurements are denoted by O_i= $\{Y_i(t_{ij}),\, V_i,\, \Delta_i,\, X(t),\, Z(t);\, t_{ij} \leqslant V_i,\, t \leqslant V_i,\, i = 1, \ldots, n,\, j = 1, \ldots, n_i\}$, where $\Delta_i = I(T_i \leqslant C_i)$ with $I(\cdot)$ being the indicator function. The log-likelihood function for the observed data is given by

$$\begin{aligned} \sum_{i=1}^{n} \log \int_b \prod_{j=1}^{n_i} \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ \frac{-(Y_i(t_{ij}) - \alpha(t_{ij}) - \beta^T X_{1i}(t_{ij}) - b^T Z_{1i}(t_{ij}))^2}{2\sigma_e^2} \right\} \right] \\ \times \left[\lambda(V_i)\, e^{\gamma^T X_{2i}(V_i) + (\phi \circ b)^T Z_{2i}(V_i)}\, H'\Big(\int_0^{V_i} e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u) \Big) \right]^{\Delta_i} \\ \times \exp\left\{ -H\Big(\int_0^{V_i} e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u) \Big) \right\} \times f(b;\, \Sigma_b)\, db, \end{aligned} \tag{5.3}$$

where $f(b;\, \Sigma_b)$ is the density function of b with the parameters Σ_b, and $\lambda(t) = d\Lambda(t)/dt$ and $H'(x) = dH(x)/dx$ are the first derivatives of $\Lambda(t)$ and $H(x)$, respectively.

5.3 Inference Procedure

We propose to use sieve maximum likelihood estimation (Shen, 1997) for infinite-dimensional parameter $\alpha(t)$, and nonparametric maximum likelihood estimation for parameters $\theta = (\beta, \gamma, \phi, \sigma_e^2, \text{Vec}(\Sigma_b))$ and infinite-dimensional parameter $F(t)$, where $\text{Vec}(\Sigma_b)$ denotes the vector consisting of the upper triangular elements of Σ_b.

5.3.1 Sieve Approximation

Suppose that the subjects are followed up for a fixed time τ. We approximate $\alpha(t)$ in (5.1) through a finite number of basis functions in a sieve space of t in $\mathbb{T} = [0, \tau]$, as follows:

$$\alpha(t) \simeq \sum_{k=1}^{m+p+1} \zeta_k B_k^p(t),$$

where $\{B_k^p(\cdot)\}$ is a basis function of t with the degree p, ζ_k is the regression coefficient with a fixed knot sequence, and m is the number of control (interior) points in the sieve space. Specifically, the sieve space for $\alpha(t)$ is defined as

$$S_n(p, m, M_n) = \left\{ \alpha(t) : \ \alpha(t) = \sum_{k=1}^{m+p+1} \zeta_k B_k^p(t), \ \sum_{k=1}^{m+p+1} |\zeta_k| \leqslant M_n \right\},$$

on a finite partition of $\mathbb{T}$

$$\left\{ 0 = s_1 = \cdots = s_{p+1} < s_{p+2} < \cdots < s_{m+p+1} < s_{m+p+2} = \cdots = s_{m+2(p+1)} = \tau \right\}.$$

The second condition in $S_n(p, m, M_n)$ guarantees the sieve space is a bounded set in a finite dimensional space. Unlike parametric regression, the number of knots and the basis function at each knot k need to be estimated from the data. In particular,

we use cubic B-spline functions ($p = 3$),

$$B_k^p(t) = \frac{t - s_k}{s_{k+p} - s_k} B_k^{p-1}(t) + \frac{s_{k+p+1} - t}{s_{k+p+1} - s_{k+1}} B_{k+1}^{p-1}(t), \qquad \text{for } t \in [0, \tau]$$

and $B_k^0(t) = I(s_k \leqslant t < s_{k+1})$ for $k = 1, \ldots, (m + p + 1)$. Figure 5.1 displays one example of the cubic B-spline curves $\{B_k^p(t)\}_{k=1}^9$ for $t \in [0, 1]$ with 5 control points $\{0.1, 0.15, 0.2, 0.4, 0.7\}$. We notice in Figure 5.1 that, for a given t value, only at most 4 basis functions among $\{B_k^p(t)\}$ are nonzero, therefore, $\alpha(t)$ is approximated by a linear combination of $\{B_k^p(t)\}$ on $(p + 1)$ nearest knot points at any point t. We let the data choose the control point $\{s_k\}$ by using percentiles of the observed longitudinal measurement times. It can prevent an optimization problem in $\{\zeta_k\}$ estimation, caused by the sparse data from informative right-censoring in the later study period. Conditional on m, we can use the methodology that has been developed for the parametric longitudinal data analysis in this nonparametric context.

5.3.2 NPMLEs for Transformation Models

To obtain the NPMLEs, in the log-likelihood function (5.3), we treat Λ as a step function with jumps only at the observed failure times and replace $\lambda(t)$ by the jump size of Λ at t, which is denoted by $\Lambda\{t\}$.

For commonly used transformation functions such as a logarithmic transformation, $\exp\{-H(x)\}$ can be expressed as a Laplace transformation of some function $\delta(\xi)$ for $\xi \geqslant 0$. For example, if we choose a gamma frailty ξ with the density function $\delta(\xi) = \xi^{1/\eta - 1} \exp(-\xi/\eta)/\{\Gamma(1/\eta)\, \eta^{1/\eta}\}$, we can show that the transformation H in Section 5.2 can be expressed by

$$\int_0^\infty \exp(-x\xi)\, \delta(\xi)\, d\xi = (1 + \eta x)^{-1/\eta} = \exp\{-H(x)\}.$$

Applying the Laplace transformation on $H(\cdot)$ and the sieve approximation for $\alpha(t)$ and using the fact that

$$H'(x)\exp\{-H(x)\} = \int_0^\infty \xi \exp(-x\xi)\,\delta(\xi)\,d\xi,$$

the observed log-likelihood function (5.3) can be rewritten as

$$\begin{aligned}
&l_n(\zeta,\,\theta,\,\Lambda\{\cdot\}) \\
&= \sum_{i=1}^{n} \log \int_b (2\pi\sigma_e^2)^{-\frac{n_i}{2}} \exp\left[-\sum_{j=1}^{n_i} \frac{\left\{Y_i(t_{ij}) - \zeta^T B^p(t_{ij}) - \beta^T X_{1i}(t_{ij}) - b^T Z_{1i}(t_{ij})\right\}^2}{2\sigma_e^2}\right] \\
&\qquad \times \int_\xi \left[\xi\,\Lambda\{V_i\}\,\exp\{\gamma^T X_{2i}(V_i) + (\phi\circ b)^T Z_{2i}(V_i)\}\,\right]^{\Delta_i} \\
&\qquad\quad \times \exp\left\{-\int_0^{V_i} \xi\, e^{\gamma^T X_{2i}(u) + (\phi\circ b)^T Z_{2i}(u)}\, d\Lambda(u)\right\} \delta(\xi)\,d\xi \times f(b;\,\Sigma_b)\,db, \qquad (5.4)
\end{aligned}$$

where $\zeta = (\zeta_1, \ldots, \zeta_m)^T$, $B^p(t) = (B_1^p(t), \ldots, B_m^p(t))^T$, and ξ is assumed to be independent of b. The most attractive feature about using the Laplace transformation is that the modified log-likelihood (5.4) can be seen as the proportional hazards frailty model with the conditional hazard function

$$\lambda(t\,|\,\mathcal{X}, \mathcal{Z},\,\xi,\,b) = \xi\,\lambda(t)\exp\{\gamma^T X_{2i}(t) + (\phi\circ b)^T Z_{2i}(t)\}.$$

This makes the algorithm more stable and computationally efficient. Now, the computation of the MLEs is identical to maximizing the modified log-likelihood function over $S_n(p, m, M_n)$, θ and all jump sizes of Λ at the observed failure times. This maximization can be carried out through the following EM algorithm.

5.3.3 EM Algorithm

We describe the EM algorithm, treating ξ and b as missing data to compute the MLEs of $(\zeta,\, \theta,\, \Lambda\{\cdot\})$. In the E-step, we calculate the conditional expectation of the log-likelihood function for the complete data, given the observed data O_i and the current parameter estimates. In other words, we need to evaluate the integration of certain functions of $(\xi,\, b)$, say $\hat{E}[\xi\, g_i(b)\,|\, O_i]$. Hereafter, we drop the conditional part of the observed data and the current parameter estimates, and abbreviate such expectation $\hat{E}[\xi\, g_i(b)\,|\, O_i]$ as $\hat{E}[\xi\, g_i(b)]$. Computation of this expectation can become doable by first obtaining the *nested* conditional expectation of ξ, given b and the observed data. That is, $\hat{E}[\xi\, g_i(b)]$ can be calculated as $\hat{E}_b[\,\hat{E}_\xi[\xi\,|\,b]\; g_i(b)]$. With the fact that the conditional distribution of ξ given b is proportional to

$$h(\xi, b) = \xi^{\Delta_i} \exp\left\{ - \int_0^{V_i} \xi\, e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u) \right\},$$

and by application of the Laplace transformation, the conditional expectation of ξ given b has the form of

$$\hat{E}_\xi\,[\xi\,|\,b] = \int \xi\, \frac{h(\xi, b)\, \delta(\xi)}{\int h(\xi, b)\, \delta(\xi)\, d\xi}\, d\xi = H'(\tilde{x}_i(b)) - \left[\frac{H''(\tilde{x}_i(b))}{H'(\tilde{x}_i(b))} \right]^{\Delta_i},$$

where $\tilde{x}_i(b) = \int_0^{V_i} e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u)$. Once $\hat{E}_\xi\,[\xi\,|\,b]$ is calculated, which is a function of b, the conditional expectation $\hat{E}[\xi\, g_i(b)]$ can be computed using numerical approximation methods such as the Gaussian quadrature with Hermite orthogonal polynomial. Since the conditional distribution of b given O_i is proportional to $\Gamma(O_i|\, b) f(b; \Sigma_b)$, the conditional expectation is calculated by

$$\hat{E}[\xi\, g_i(b)] = \int_b \hat{E}_\xi[\xi\,|\,b]\, g_i(b)\, \frac{\Gamma(O_i|\, b) f(b; \Sigma_b)}{\int_b \Gamma(O_i|\, b) f(b; \Sigma_b)\, db}\, db,$$

where

$$\Gamma(O_i|\, b) = \exp\left\{-\sum_{j=1}^{n_i} \frac{\{b^T Z_{1i}(t_{ij})\}^2 - 2b^T Z_{1i}(t_{ij})[Y_i(t_{ij}) - \zeta^T B^p(t_{ij}) - \beta^T X_{1i}(t_{ij})]}{2\sigma_e^2}\right\}$$
$$\times \exp\left\{\Delta_i \left[(\phi \circ b)^T Z_{2i}(V_i) + \log H'\left(\int_0^{V_i} e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u)\right)\right]\right\}$$
$$\times \exp\left\{-H\left(\int_0^{V_i} e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u)\right)\right\}.$$

In the M-step, we maximize the following objective function of the expected log-likelihood for the complete data:

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{-\log\sigma_e^2/2 - \hat{E}\left[\{Y_i(t_{ij}) - \zeta^T B^p(t_{ij}) - \beta^T X_{1i}(t_{ij}) - b^T Z_{1i}(t_{ij})\}^2/(2\sigma_e^2)\right]\right\}$$
$$+\sum_{i=1}^{n}\Delta_i\left\{\log\xi + \log\Lambda\{V_i\} + \gamma^T X_{2i}(V_i) + \hat{E}[\phi \circ b]^T Z_{2i}(V_i)\right\}$$
$$+\sum_{i=1}^{n}\left\{-\hat{E}\left[\int_0^{V_i} \xi\, e^{\gamma^T X_{2i}(u) + (\phi \circ b)^T Z_{2i}(u)}\, d\Lambda(u)\right] + \hat{E}\left[\log\delta(\xi) + \log f(b;\Sigma_b)\right]\right\}.$$

Maximizing the above objective function over (ζ, β, σ_e^2, Σ_b) is simple as a classic linear regression; whereas the rest of parameters (γ, ϕ, $\Lambda\{.\}$) do not yield the closed-form of maximizers. Involving a reliable numerical approach, we solve the following equation for γ:

$$\sum_{i=1}^{n}\Delta_i\left\{X_{2i}(V_i) - \frac{\sum_{j=1}^{n} R_j(V_i) X_{2j}(V_i)\, \hat{E}\left[\xi\, e^{q_{2j}(V_i)}\right]}{\sum_{j=1}^{n} R_j(V_i)\, \hat{E}\left[\xi\, e^{q_{2j}(V_i)}\right]}\right\} = 0, \tag{5.5}$$

and the following equation for ϕ:

$$\sum_{i=1}^{n}\Delta_i\left\{\hat{E}\left[b \circ Z_{2i}(V_i)\right] - \frac{\sum_{j=1}^{n} R_j(V_i)\, \hat{E}\left[\xi\, e^{q_{2j}(V_i)}(b \circ Z_{2j}(V_i))\right]}{\sum_{j=1}^{n} R_j(V_i)\, \hat{E}\left[\xi\, e^{q_{2j}(V_i)}\right]}\right\} = 0, \tag{5.6}$$

where $R_j(t) = I(V_j \geqslant t)$ and $q_{2j}(t) = \gamma^T X_{2j}(t) + (\phi \circ b)^T Z_{2j}(t)$. In addition, Λ is

estimated as a step function with the following jump size at V_i:

$$\Lambda\{V_i\} = \frac{\Delta_i}{\sum_{j=1}^{n} R_j(V_i)\,\hat{E}\left[\xi\, e^{q_{2j}(V_i)}\right]}. \tag{5.7}$$

At each M-step, we update γ and ϕ by solving the equations (5.5) and (5.6) through one-step Newton-Raphson algorithm. Updating the jump sizes of Λ can be easily done by the equation (5.7).

To obtain the MLEs, we iterate the E-step and M-step until the parameter estimates converge. The variances of the NPMLEs can be estimated from the inverse of the observed information matrix for all parameters of $(\zeta,\, \theta,\, \Lambda\{\cdot\})$. The observation information matrix can be computed from the complete data log-likelihood function denoted by ℓ_i^c for the ith subject using the following Louis formula (Louis, 1982) of

$$-\sum_{i=1}^{n} \hat{E}[\nabla^2 \ell_i^c(b) \,|\, O_i] - \sum_{i=1}^{n} \left\{ \hat{E}[\nabla \ell_i^c(b)^{\otimes 2} \,|\, O_i] - \hat{E}[\nabla \ell_i^c(b) \,|\, O_i]^{\otimes 2} \right\},$$

where $u^{\otimes 2} = uu^T$, ∇ and ∇^2 denote the first and the second derivatives with respect to parameters, and $\hat{E}$ denotes the conditional expectation of a function of b given the observed data.

5.4 Asymptotic Properties

Let $(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})$ denote the MLEs and $(\alpha_0,\, \theta_0,\, \Lambda_0)$ denote the true parameter values of $(\alpha,\, \theta,\, \Lambda)$. Suppose the study duration is $\mathbb{T} = [0,\, \tau]$. Under the regularity conditions, we will establish the asymptotic properties of the MLEs under the following conditions:

(A1) The true parameter value θ_0 belongs to the interior of a compact set Θ within

the domain of θ.

(A2) With probability 1, $X(t)$ and $Z(t)$ is left-continuous with uniformly bounded left and right derivatives in $[0, \tau]$.

(A3) For some constant δ_0, $P(C \geqslant \tau | \mathcal{X}, \mathcal{Z}) > \delta_0 > 0$ with probability 1.

(A4) For some positive constant M_0, $M_0^{-1} < \sigma_{0e}^2 < M_0$ and $M_0^{-1} < c^T \Sigma_{0b} c < M_0$ for any $\|c\| = 1$.

(A5) The transformation functions $H(.)$ are four-times differentiable with $H(0) = 0$ and $H'(0) > 0$. In addition, there exist positive constants μ_0 and κ_0 such that

$$(1 + x) H'(x) \exp\{-H(x)\} \leqslant \mu_0 (1 + x)^{-\kappa_0}.$$

Furthermore, there exists a constant $\rho_0 > 0$ such that

$$\sup_x \left\{ \frac{|H''(x)| + |H^{(3)}(x)| + |H^{(4)}(x)|}{H'(x) (1 + x)^{\rho_0}} \right\} < \infty,$$

where $H^{(3)}$ and $H^{(4)}$ are the third and fourth derivatives.

(A6) For some $t \in [0, \tau]$, if there exist a deterministic function $c(t)$ and v such that $c(t) + v^T X(t) = 0$ with probability 1, then $c(t) = 0$ and $v = 0$.

(A7) With some positive probability, $\boldsymbol{Z}_1^T \boldsymbol{Z}_1$ has full rank, where $\boldsymbol{Z}_1$ denotes a matrix with each row equal to the observed covariate $Z_1(t)^T$ at the time of each measurement.

(A8) The potential observation process of $Y(t)$ has a continuous intensity over $[0, \tau]$.

(A9) For some fixed integer $r > 4$, $\alpha_0(t)$ lies in $W^{r,\infty}(R)$, where $W^{r,\infty}(R)$ is a Sobolev space consisting of the functions with bounded rth derivatives.

(A10) For fixed constant r_0 such that $1/(4r) < r_0 < 1/7$, there exist m and M_n satisfying

$$m = O(n^{r_0}), \text{ and } M_n = O(\log \log n).$$

Conditions (A1) - (A3) are the standard assumptions in survival analysis. Condition (A4) is necessary to prove the existence of the NPMLEs. It can be easily verified that Condition (A5) holds for all transformations commonly used, including the logarithmic transformations described in Section 5.2. Conditions (A6) - (A7) entail the linear independence of design matrices of covariates for the fixed and random effects. Condition (A8) prescribes that some subjects have sufficient repeated measures. Finally, Condition (A9) grants sufficient smoothness of α_0, and the size of the sieve space S_n is determined by Condition (A10).

Under the above conditions, the following theorem shows the consistency of the MLEs $(\hat{\theta}, \hat{\alpha}, \hat{\Lambda})$.

Theorem 5.1. *Under Conditions (A1) - (A10),*

$$|\hat{\theta} - \theta_0| \to 0, \quad \|\hat{\alpha}(t) - \alpha_0(t)\|_{W^{1,\infty}(\mathbb{T})} \to 0, \quad \|\hat{\Lambda}(t) - \Lambda_0(t)\|_{L_\infty(\mathbb{T})} \to 0, \quad a.s.,$$

where $\|\cdot\|_{W^{1,\infty}(\mathbb{T})}$ *is the Sobolev norm on* $\mathbb{T}$ *and* $\|\cdot\|_{L_\infty(\mathbb{T})}$ *is the supremum norm on* $\mathbb{T}$.

Now, we need to obtain a tighter bound for the convergence rate of the estimates, which is stated in Theorem 5.2.

Theorem 5.2. *Under Conditions (A1) - (A10),*

$$\|\hat{\alpha}(t) - \alpha_0(t)\|^2_{L_2(P)} + \|\hat{\Lambda}(t) - \Lambda_0(t)\|^2_{L_2(P)} \leqslant O_p(m^{-2r}) + o_p(n^{-1/2}),$$

where $\|\cdot\|_{L_2(P)}$ *is the* L_2*-norm.*

Theorems 5.1 - 5.2 then lead to the following results on the asymptotic normality and semiparametric efficiency of $\hat{\theta}$.

Theorem 5.3. *Under Conditions (A1) - (A10),* $\sqrt{n}\,(\hat{\theta} - \theta_0)$ *weakly converges to a zero-mean Gaussian process in* R^{d_θ}*, where* d_θ *is the dimension of* θ*. Furthermore, the asymptotic covariance matrix of* $\sqrt{n}\,(\hat{\theta} - \theta_0)$ *achieves the semiparametric efficiency bound.*

Furthermore, in Section 5.8, we show that the inverse of the observed information matrix is a consistent estimator of the asymptotic covariance matrix of $\sqrt{n}\,\hat{\theta}$. This result allows us to make inference for any functional of θ. The detailed proofs are provided in Section 5.8.

5.5 Simulation Studies

In this section, we conduct extensive simulation studies with small sample sizes to assess the validity of the proposed method. The longitudinal data are generated from

$$Y(t \,|\, x_1, x_2, b) = \alpha(t) + x_1 - 0.5x_2 + b + \epsilon(t),$$

and the informative censoring times are generated from the following transformed survival model

$$\Lambda(t \,|\, x_1, x_2, b) = H(\exp\{x_1 - 0.5x_2 + \phi b\}\,\Lambda(t)),$$

where x_1 is a dichotomous covariate taking the value of 0 or 1 with the equal probability of 0.5, x_2 is a continuous covariate generated from a uniform distribution on [-1, 1], and $\epsilon(t) \sim N(0, \sigma_e^2)$ is assumed with $\sigma_e^2 = 1$. The true cumulative hazard function is set to be $\Lambda(t) = t$. Two types of baseline function $\alpha(t)$ are considered to represent non-linear trends in $Y(t)$: $\alpha(t) = \sin(\pi t)\, e^{\frac{t}{2}}/(1 + e^{\frac{t}{2}})$ and $\alpha(t) = (t - 0.8)^2$.

For each subject, the correlation within repeated measures is reflected by the

subject-specific random intercept $b \sim N(0, \sigma_b^2)$ with $\sigma_b^2 = 0.5$, and the negative, no, and positive dependences between the longitudinal measures and the informative censoring rate are simulated through different ϕ values of -0.3, 0, and 0.3, respectively. Also, we consider the transformations for $H(\cdot)$ representing the proportional hazards model $H(x) = x$ (when $\eta = 0$) and the proportional odds model $H(x) = \log(1 + x)$ (when $\eta = 1$) in the survival data.

We generate the non-informative censoring time C_i from a uniform distribution [0.5, 2.5] and set longitudinal responses to be repeatedly measured about 6 times, on average, at any times before censoring to design a 60$\sim$75% of subjects who are informatively censored, depending on the chosen transformation.

We also investigate the effect of the number of knots on $\alpha(t)$ estimation by comparing estimates from three different sets of the control points used for the B-spline approximation. The considered sequences of the control points are three sets of the pth percentiles of observation times $\{q_p\}$, specifically, $\{q_{25}, q_{50}, q_{75}\}$, $\{q_{15}, q_{30}, \ldots, q_{90}\}$, and $\{q_{10}, q_{20}, \ldots, q_{90}\}$, which are denoted by $m = 3$, $m = 6$, and $m = 9$, respectively, in Tables 5.1 - 5.4.

The simulation results based on 1000 replications are presented in Tables 5.1 - 5.4 for n=200 and n=400 when $\alpha(t) = \sin(\pi t)\, e^{\frac{t}{2}}/(1 + e^{\frac{t}{2}})$. Table 5.1 and Table 5.3 include the average of the differences between the true parameter and the estimates (Bias), the sample standard deviation of the parameter estimators (SE), and the average of the standard error estimators (SEE), and the coverage probability of 95% confidence intervals (CP). The confidence intervals for σ_e^2 and σ_b^2 are constructed based on the the Satterthwaite approximation. To summarize the performance of the proposed $\hat{\alpha}(t)$, Bias, SE, the mean square error (MSE), and the ratio of the MSE for $\alpha(t)$ estimates in the joint model to the counterpart in the marginal model (MSER) are provided in Table 5.2 and Table 5.4.

Table 5.1 shows that the NPMLEs under $H(x) = x$ are noticeably unbiased, the standard error estimators calculated via the Louis formula well reflect the true variations of the proposed estimators, and the coverage probabilities are in a reasonable range, even with a small sample size of 200. As the sample size increases to 400, we may see that the biases slightly increase for some estimates; however, they are still very small comparing to the sizes of true parameter values and the variations of the parameter estimators become smaller, and hence the coverage probabilities still lie in a reasonable range.

Table 5.2 shows that the variations among the estimates and the MSE in estimating $\alpha(t)$ become smaller as the sample size increases, whereas they become larger as the number of knots increases. These results appear rather general since the biases are negligible. However, in the comparison of the biasedness, there exist no general rules by the sample size or the number of knots; use of 6 control points yields the smallest bias with n=200, in contrast to 9 control points with n=400. When we compare the estimates from the joint model to the marginal model, the use of the joint model clearly produces the more accurate and efficient estimation in that the approach using the joint model reduces the Bias and MSE with the smaller variations in the estimates.

The simulation results shown in Tables 5.1 - 5.2 are similar to those for Tables 5.3 - 5.4, indicating that the proposed method seems to work well for the other transformation models $H(x) = \log(1+x)$. Figure 1 also compares the performance of the number of knots and transformations visually, and it supports that the proposed estimators of $\alpha(t)$ behave well in both transformation models when $\alpha(t) = \sin(\pi t)\, e^{\frac{t}{2}}/(1+e^{\frac{t}{2}})$ as well as $\alpha(t) = (t-0.8)^2$. We have also studied the performance with $\alpha(t) = (t-0.8)^2$, and the results are similar to those for $\alpha(t) = \sin(\pi t)\, e^{\frac{t}{2}}/(1+e^{\frac{t}{2}})$ and hence omitted here.

5.6 Data Application

We illustrate the application of the proposed method through medical cost data for chronic heart failure patients, which are obtained from the clinical data repository database in the University of Virginia (UVa) Health System. The data were collected from a total of 1475 patients who were at least 60 years old and first diagnosed and treated with heart failure in 2004 until their death or last hospital admission (up to July 31, 2006). The cohort consists of 55% males and 73% whites with an average age of 72.

A main focus of the study is on adjusting the underlying trend of the medical cost which is defined by the actual monetary expense of the hospital at each visit. Preliminary studies, however, showed that the medical costs were subject to the time of the patient's death, and hence we believe that censoring by death can be a crucial factor in describing the trend of medical costs. Thus, we propose to model the medical cost data jointly with the informative censorship (death). In this study, medical costs of each patient were measured about 11 times on average, up to 45 times, with a median follow-up time of 21 months. Among the cohort members, 297 patients (about 20%) were informatively censored by death.

We model the log-transformed costs and the censoring time with gender, race, and age (up to a quadratic term) as covariates. The age variable is centered at mean 0 and rescaled to represent every 10-year unit change. Visiting time (in years), as a nonparametric function $\alpha(t)$, is also included into the longitudinal model component. To construct B-spline curves for $\alpha(t)$, as discussed in Section 5.3, we use m percentile points of visiting times as the sequence of control points. In addition, a subject-specific random intercept is included in both longitudinal and survival models to account for the dependence of medical costs on informative censoring. We notice that the coefficient of shared random intercept can also quantify their dependence,

conditional on covariates.

For the survival component, we apply transformation models $H(x) = \log(1+\eta)/\eta$ by varying values of η in [0, 1]. This class of transformation allows us to explore the possibility of the proportional hazards and the proportional odds structures in survival data. Under each set of (m, η), the MLEs for the regression coefficients are computed using the proposed method in Section 5.3, using the least square estimates as initial values in the EM algorithm. To select the number of control points m, we can adopt model selection approaches such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), and 'leave-one-subject-out' cross-validation, although we let the BIC choose the best set of the number of control points and transformation. In Figure 5.3, the smallest BIC value corresponding to $(m, \eta) = (5, 0.5)$ indicates that use of 5 control points and the transformation in the middle of proportional hazards and odds models, (i.e. $H(x) = 2\log(1 + 0.5x)$) produces the best fit to the data.

Under the selected best model, Table 5.5 summarizes the estimation results. The results show that nonwhite male patients tend to bear a higher cost incurred by chronic heart failure at each hospital visit. The squared age is also a significant factor to explain the trend of medical costs. In more detail, patients 69 years old meet the highest medical cost, and the younger or older patients than 69 years of age spend less money at each hospital visit. In the survival model, nonwhite male patients have a higher risk of death, and the risk grows with aging. For the random effect b, through the significant variance component $\hat{\sigma_b}^2$, we note that heterogeneity between patients exists in repeated measures. We also notice that the highly significant $\hat{\phi}$ suggests that patients at the higher risk of death are likely to meet a higher medical cost by chronic heart failure. This result is consistent what is shown in Figure 5.4. By comparing the estimated underlying trajectory of cost in the joint model fit to the marginal model fit along with residual means of $\{Y(t) - \hat{\beta}^T X_1(t)\}$ in Figure 5.4,

we can see that the marginal model appears to overestimate the underlying medical costs in the first half of the curve, where most of patients are informatively censored by death, and appears to underestimate in the second half of the curve, where most of patients are randomly censored. We can thus conclude that the joint model adjusts well the bias caused by dying patients' high medical costs.

5.7 Concluding Remarks

We have proposed the partially linear model for longitudinal data with informative censoring by modeling the longitudinal data simultaneously with the transformed survival model to adjust the dependence on informative censoring. For estimation, we have used the MLEs through B-spline approximation of the baseline function in the longitudinal data and through step functions of the baseline cumulative hazard in the censoring mechanism. In addition, the EM algorithm has been presented for implementation, which is shown to be computationally efficient. The resulting MLEs are theoretically justified, and the proposed joint approach has clearly shown the potential to correct biases induced by ignoring informative censoring using the simulated data and a real example.

In this paper, we assumed that the underlying counting process of measurement times are independent of the pattern of longitudinal data. Our partially linear model can be extended in a way to account for the informative observation process. The BIC was used to determine both the best transformation and the selection of the number of knots, but we can explore and compare the validity of other model selection criteria such as the AIC and cross-validation in the future. Methods on model checking would be useful for the practical application of the joint models.

5.8 Proof of Asymptotic Properties

This section proves Theorems 5.1 - 5.3 stated in Section 5.4 by using techniques from the empirical process theory.

Proof of Theorem 5.1

Proof. The whole proof can be divided into three steps: first, we construct some functions in the sieve space, which approximate the true parameters; then by using empirical process theory, we obtain one key inequality; finally, this inequality is used to obtain the consistency.

Step 1. We construct some functions in $S_n(p, m, M_n)$ to approximate the true parameters. From the properties of B-spline functions, we can define a linear operator $\mathcal{Q}$ mapping $W^{r,\infty}(\mathbb{T})$,

$$\mathcal{Q}[g] = \sum_{k=1}^{m+p+1} \Psi_k[g] B_k^p(t),$$

where Ψ_k are linear functionals in $L_\infty(\mathbb{T})$ and $(\mathbb{T}) = [0, \tau]$. Moreover,

$$\sum_{k=1}^{m+p+1} |\Psi_k[g]| \leqslant (2(p+1)+1)9^p \|g\|_{L_\infty(\mathbb{T})},$$

and according to Theorem 12.7 of Schumaker (2007),

$$\|\mathcal{Q}[g] - g\|_{L_\infty(\mathbb{T})} \leqslant \frac{C(p+1)}{m^r} \|g\|_{L_\infty(\mathbb{T})}.$$

Thus, we define $\alpha_n(t) = \mathcal{Q}[\alpha_0]$. As a result of the fact that $\sum_{k=1}^{m+p+1} B_k^p(t) = 1$, $\alpha_0(t)$ lies in the sieve space $S_n(p, m, M_n)$ and moreover, the following boundness hold

$$\|\alpha_n - \alpha_0\|_{L_\infty(\mathbb{T})} \leqslant O(m^{-r}).$$

Step 2. We obtain a key inequality based on empirical process theory. Let $\mathbb{P}_n$ be the empirical measure determined by n *iid* subjects, let $\boldsymbol{P}$ be its expectation, and let $\mathbb{G}_n$ be the empirical process given by $\sqrt{n}(\mathbb{P}_n - \boldsymbol{P})$. For simplicity of notation, we denote $G(\alpha, \theta, \Lambda)$ as the likelihood function from one single observation. Since $(\hat{\alpha}, \hat{\theta}, \hat{\Lambda})$ maximizes $\mathbb{P}_n[\log G(\alpha, \theta, \Lambda)]$ over the sieve space, it follows that

$$\mathbb{P}_n[\log G(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})] \geqslant \mathbb{P}_n[\log G(\alpha_n, \theta_0, \tilde{\Lambda})],$$

where $\tilde{\Lambda}$ is a step function with jumps only at the observed failure times and it uniformly converges to Λ_0 with convergence rate $n^{-1/2}$. Equivalently,

$$n^{-1/2}\, \mathbb{G}_n \left[\log \frac{G(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})}{G(\alpha_n, \theta_0, \tilde{\Lambda})} \right] \geqslant \boldsymbol{P} \left[\log \frac{G(\alpha_n, \theta_0, \tilde{\Lambda})}{G(\alpha_0, \theta_0, \Lambda_0)} \right] + \boldsymbol{P} \left[\log \frac{G(\alpha_0, \theta_0, \Lambda_0)}{G(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})} \right]. \quad (5.8)$$

First, we can show that the left-hand side of (5.8) is bounded, using empirical process theory. Consider a class of functions $\mathcal{L}_n$ defined by

$$\mathcal{L}_n = \left\{ \log \frac{G(\tilde{\alpha}_n, \theta_0, \tilde{\Lambda})}{G(\alpha_n, \theta_0, \tilde{\Lambda})};\ \tilde{\alpha}_n \in S_n(p, m, M_n) \right\}.$$

Since $\|B_k^p(\cdot)\|_{L_\infty} = 1$, any function of $\tilde{\alpha}_n$ given in $\mathcal{L}_n$ is bounded by $O(e^{2M_n})$. By assumptions (A1)-(A6), $G(\alpha_n, \theta_0, \tilde{\Lambda})$ is bounded away from 0, and hence the class $\mathcal{L}_n$ has an upper bound $O_p(M_n)$. After tedious calulation, we can show that the function in $\mathcal{L}_n$ is Lipschitz continuous with respect to α, and the Lipschitz constant is bounded by $O_p(e^{c_1 M_n})$, for a fixed constant c_1. By computing the bracket number of $\mathcal{L}_n$ and applying Theorem 19.35, van der Vaart (1998), in probability we have

$$\begin{aligned} \sqrt{n}\, E_p^* \|\mathbb{P}_n - \boldsymbol{P}\|_{\mathcal{L}_n} &\leqslant O_p(1) \int_0^{O(M_n)} \sqrt{\log \left(\frac{M_n e^{c_1 M_n} (m + p + 1)}{\epsilon} \right)^{(m+p+1)}}\, d\epsilon \\ &\leqslant O_p(1)\, m^{1/2} (\log M_n) M_n^2, \end{aligned}$$

so, the left-hand side of (5.8) is bounded by $O_p(M_n^2 m^{1/2}(\log M_n)/\sqrt{n})$ from above. Second, since the functional $G(\cdot)$ is Lipschitz continuous with α and $\tilde{\Lambda}$ uniformly converges to Λ_0, we can show that the first term of the right side of (5.8) is

$$\begin{aligned} \boldsymbol{P}\left[\log \frac{G(\alpha_n, \theta_0, \tilde{\Lambda})}{G(\alpha_0, \theta_0, \Lambda_0)}\right] &\geqslant -O_p(1)\left\{\|\alpha_n - \alpha_0\|_{L_\infty} + \|\tilde{\Lambda} - \Lambda_0\|_{L_\infty}\right\} \\ &\geqslant -O_p(1)\left\{\frac{1}{m^r} + O_p(n^{-1/2})\right\}. \end{aligned}$$

Third, since the second term of the right side of (5.8) is the Kullback-Leibler information, and by linearizing, we obtain

$$\boldsymbol{P}\left[\log \frac{G(\alpha_0, \theta_0, \Lambda_0)}{G(\alpha_n, \theta_0, \tilde{\Lambda})}\right] \geqslant O(e^{c_1 M_n/2})\|G(\alpha_0, \theta_0, \Lambda_0) - G(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})\|_{L_2(P)}^2.$$

Thus, combining the above results, we can show that

$$\|G(\alpha_0, \theta_0, \Lambda_0) - G(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})\|_{L_2(P)}^2 \leqslant O_p\left(\frac{e^{c_1 M_n/2}}{m^r} + \frac{e^{c_1 M_n/2} M_n^2 m^{1/2} \log M_n}{\sqrt{n}}\right) \quad (5.9)$$

Step 3. We obtain the L_2-convergence of the estimators. Suppose we select m and M_n satisfying Assumption (A10), then we can obtain from (5.9)

$$\|G(\alpha_0, \theta_0, \Lambda_0) - G(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})\|_{L_2(P)}^2 \leqslant O_p(1)\, D_n(m, M_n),$$

where $D_n(m, M_n) = e^{c_1 M_n/2} m^{-r} + e^{c_1 M_n} m^{1/2} \log(M_n)/\sqrt{n}$. From the above inequality and identifiability conditions of the parameters, we can obtain that $\|\hat{\alpha} - \alpha_0\|_{L_2(\mathbb{T})}^2$ is bouned by $O_p(1)\, D_n(m, M_n)$ from above. Moreover, $D_n(m, M_n)^{1/2}$ is the convergence rate of $\hat{\alpha}$.

To obtain the convergence of $\hat{\alpha}$ in $W^{1,\infty}$-space, we notice from Theorem 4.22 of Schumaker (1981) that the $W^{r,\infty}$-norm of $\hat{\alpha}$ is bounded by $O(e^{c_2 M_n} m^r)$ from above for

some constant c_2. Hence, according to the Sobolev interpolation inequality (Adams and Fournier, 1975), we obtain

$$\|\hat{\alpha}(t) - \alpha_0(t)\|_{W^{1,\infty}(\mathbb{T})} \leqslant O(1)\, e^{c_2 \tau_1 M_n} m^{\tau_1 r} D_n(m, M_n)^{(1-\tau_1)/2}, \tag{5.10}$$

where $\tau_1 = 3/(2r)$. By Assumption (A10), the right side of (5.10) converges to zero. Thus, Theorem 5.1 holds. □

Proof of Theorem 5.2

Proof. We use the results of Theorem 5.1. Since $\hat{\alpha}$ is within a $W^{1,\infty}$-neighborhood of α_0, now $\mathcal{L}_n$ has a bound covering function and the integration of the entropy for the class $\mathcal{L}_n$ is finite. Moreover, the function in the left side of (5.8) uniformly converges to zero. Thus, we can apply Theorem 2.11.23 of van der Vaart and Wellner (1996), to obtain that the left side of (5.8) is bounded by $o_p(1/\sqrt{n})$. By Taylor expansion of the right side of (5.8) at the true parameters and Theorem 5.1, the right side of (5.8) is bounded from below by

$$\begin{aligned} &- O_p(1) \left\{ \|\alpha_n - \alpha_0\|^2_{L_2(\mathbb{T})} + \|\tilde{\Lambda} - \Lambda_0\|^2_{L_\infty(\mathbb{T})} \right\} \\ &+ O_p(1) \|G(\alpha_0, \theta_0, \Lambda_0) - G(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})\|^2_{L_2(P)}. \end{aligned}$$

Providing Theorem 5.1, we obtain that

$$\|G(\alpha_0, \theta_0, \Lambda_0) - G(\hat{\alpha},\, \hat{\theta},\, \hat{\Lambda})\|^2_{L_2(P)} \leqslant \frac{o_p(1)}{\sqrt{n}} + \frac{O_p(1)}{m^{2r}}.$$

From the parameter identifiability condition, the left-hand side can be further bounded from below by the $L_2(P)$-norm of $|\hat{\Lambda} - \Lambda_0|$ and $|\hat{\alpha} - \alpha_0|$. Thus, we conclude Theorem 5.2. □

Proof of Theorem 5.3

Proof. We will prove Theorem 5.3 by writing $\sqrt{n}\,(\hat{\theta} - \theta_0)$ as a linear functional of the empirical process $\mathbb{G}_n$. Let $\ell(\alpha,\, \Lambda,\, \theta)$ be the log-likelihood function from a single subject, and let $\ell_0 = \ell(\alpha_0, \Lambda_0, \theta_0)$.

Step 1. We define a least favorable direction for θ_0. We treat $\psi = (\alpha,\, \Lambda)$ as the vector of nuisance parameters with $\psi_0 = (\alpha_0,\, \Lambda_0)$, and then the tangent space for ψ is given by $\mathcal{H} = \{h(t) = (h_1(t),\, h_2(t));\ h(t) \in L_2(\mathbb{T}^2)\}$. Let $\ell_\psi(\psi_0, \theta_0)[h]$ be the derivative of ℓ_0 with respect to ψ along with the direction h_1 for α and the direction h_2 for Λ, and let $\ell_\theta(\psi_0, \theta_0)$ be the derivative of ℓ_0 with respect to θ. Then, a least favorable direction for θ_0 is defined as a tangent function $h(t) \in \mathcal{H}$ for ψ that satisfies

$$\ell^*_\psi(\psi_0, \theta_0)\ell_\psi(\psi_0, \theta_0)[h] = \ell^*_\psi(\psi_0, \theta_0)\ell_\theta(\psi_0, \theta_0) \quad a.s.,$$

where $\ell^*_\psi(\psi_0, \theta_0)$ is the adjoint operator of $\ell_\psi(\psi_0, \theta_0)$ in the Hilbert space $L_2(P)$.

Step 2. We prove the existence and smoothness of the least favorable direction. The existence can be shown by proving the operator $\ell^*_\psi(\psi_0, \theta_0)\ell_\psi(\psi_0, \theta_0)$ is invertible based on the Lax-Milgram theorem. The details of proofs are the same as in Zeng (2005).

Step 3. We construct the projection of $h_1(t)$ on the tangent space of the sieve space. The tangent function for ψ at $\hat{\psi} = (\hat{\alpha}, \hat{\Lambda})$ in the sieve space can be chosen by $h_n = (h_{1n}(t), h_2 d\hat{\Lambda})$ in $L_2(\mathbb{T}^2)$ such that

$$\|h_{1n} - h_1\|^2_{L_2(P)} \leqslant O(m^{-2r}) + o_p(n^{-1/2}).$$

Step 4. We derive the empirical process for $\sqrt{n}\,(\hat{\theta} - \theta_0)$. Since $(\hat{\psi},\, \hat{\theta})$ maximizes the log-likelihood over the sieve space, the score along the path $(\hat{\psi} + \nu h_n,\, \hat{\theta} + \nu)$ is zero when $\nu = 0$. Thus, it holds that

$$\mathbb{G}_n\{\ell_\psi(\hat{\psi}, \hat{\theta})[h_n] + \ell_\theta(\hat{\psi}, \hat{\theta})\} = -\sqrt{n}\, \boldsymbol{P}\{\ell_\psi(\hat{\psi}, \hat{\theta})[h_n] + \ell_\theta(\hat{\psi}, \hat{\theta})\}. \tag{5.11}$$

Since the function in the left side of (5.11), indexed by both $(\hat{\psi}, h_n) \in W_{1,\infty}$ and $\hat{\theta} \in \Theta$, belongs to P-Donsker class, we apply Theorem 2.11.23 of van der Vaart and Wellner (1996). By linearizing the right side of (5.11) at the true parameters and approximating h_n to h, we obtain that

$$\begin{aligned} -\, \boldsymbol{P}\{\ell_{\psi\theta}(\psi_0, \theta_0)[h] + \ell_{\theta\theta}(\psi_0, \theta_0)\}\sqrt{n}\,(\hat{\theta} - \theta_0) & \\ = \mathbb{G}_n\{\ell_\psi(\psi_0, \theta_0)[h] + \ell_\theta(\psi_0, \theta_0)\} & \\ + \sqrt{n}\, O_p(\|\hat{\psi} - \psi_0\|^2_{L_2(P)} + \|h_n - h\|^2_{L_2(P)} + |\hat{\theta} - \theta_0|^2). & \end{aligned}$$

Since the second term in the right side of the above equation is $o_p(1)$ by Theorem 5.2 and (A10) and $-\boldsymbol{P}\{\ell_{\psi\theta}(\psi_0, \theta_0)[h] + \ell_{\theta\theta}(\psi_0, \theta_0)\} > 0$, the asymptotic normality of $\sqrt{n}\,(\hat{\theta} - \theta_0)$ holds. Moreover, the influence function of $\hat{\theta}$ is given by

$$[-\boldsymbol{P}\{\ell_{\psi\theta}(\psi_0, \theta_0)[h] + \ell_{\theta\theta}(\psi_0, \theta_0)\}]^{-1}\{\ell_\psi(\psi_0, \theta_0)[h] + \ell_\theta(\psi_0, \theta_0)\}.$$

Clearly, the above influence function is contained in the tangent space, therefore, we conclude that $\hat{\theta}$ is semiparametrically efficient. □

Table 5.1: Simulation results for $H(x) = x$ and $\alpha(t) = \sin(\pi t)\ \exp(t/2)/\{1+\exp(t/2)\}$ based on m=6 control points of B-spline curves. τ_p represents $p\%$ of τ (study duration).

			$n = 200$				$n = 400$			
ϕ		*True*	*Bias*	*SE*	*SEE*	*CP*	*Bias*	*SE*	*SEE*	*CP*
-0.3	β_1	1.0	0.004	0.125	0.129	0.956	0.001	0.088	0.090	0.953
	β_2	-0.5	-0.000	0.107	0.108	0.945	0.001	0.075	0.076	0.953
	σ_e^2	1.0	-0.007	0.039	0.041	0.963	-0.003	0.028	0.029	0.953
	σ_b^2	0.5	-0.006	0.073	0.073	0.947	-0.006	0.049	0.051	0.956
	γ_1	1.0	-0.008	0.180	0.202	0.968	-0.013	0.126	0.135	0.968
	γ_2	-0.5	0.003	0.151	0.152	0.957	-0.005	0.107	0.106	0.946
	ϕ	-0.3	0.005	0.154	0.160	0.972	0.006	0.105	0.109	0.956
	$\Lambda(\tau_{20})$	0.3	0.003	0.048	0.052	0.963	0.004	0.033	0.036	0.968
	$\Lambda(\tau_{40})$	0.6	0.008	0.085	0.092	0.962	0.008	0.060	0.063	0.952
	$\Lambda(\tau_{60})$	0.8	0.012	0.127	0.132	0.954	0.014	0.089	0.090	0.947
	$\Lambda(\tau_{80})$	1.1	0.024	0.182	0.179	0.962	0.019	0.123	0.122	0.948
0	β_1	1.0	-0.007	0.128	0.125	0.940	-0.002	0.088	0.088	0.952
	β_2	-0.5	-0.002	0.108	0.107	0.951	0.002	0.075	0.075	0.953
	σ_e^2	1.0	-0.008	0.043	0.041	0.935	-0.004	0.029	0.029	0.946
	σ_b^2	0.5	-0.010	0.075	0.072	0.936	-0.008	0.052	0.051	0.951
	γ_1	1.0	-0.002	0.174	0.178	0.949	-0.016	0.122	0.123	0.949
	γ_2	-0.5	-0.005	0.147	0.148	0.954	0.001	0.103	0.103	0.955
	ϕ	0.0	0.002	0.148	0.151	0.955	0.001	0.104	0.105	0.953
	$\Lambda(\tau_{20})$	0.3	0.003	0.046	0.048	0.962	0.005	0.034	0.033	0.946
	$\Lambda(\tau_{40})$	0.6	0.006	0.081	0.084	0.961	0.011	0.058	0.059	0.948
	$\Lambda(\tau_{60})$	0.8	0.012	0.120	0.122	0.961	0.013	0.084	0.085	0.949
	$\Lambda(\tau_{80})$	1.1	0.021	0.168	0.169	0.964	0.024	0.116	0.118	0.959
0.3	β_1	1.0	-0.004	0.121	0.129	0.961	-0.006	0.085	0.090	0.962
	β_2	-0.5	-0.003	0.107	0.107	0.947	-0.001	0.076	0.076	0.947
	σ_e^2	1.0	-0.006	0.042	0.041	0.943	-0.005	0.030	0.029	0.940
	σ_b^2	0.5	-0.009	0.072	0.072	0.954	-0.004	0.053	0.051	0.951
	γ_1	1.0	-0.010	0.184	0.201	0.962	-0.024	0.131	0.136	0.958
	γ_2	-0.5	0.000	0.153	0.152	0.941	0.000	0.102	0.106	0.951
	ϕ	0.3	-0.006	0.154	0.160	0.966	0.003	0.104	0.109	0.969
	$\Lambda(\tau_{20})$	0.3	0.003	0.048	0.052	0.956	0.004	0.034	0.036	0.955
	$\Lambda(\tau_{40})$	0.6	0.006	0.085	0.091	0.957	0.010	0.060	0.063	0.954
	$\Lambda(\tau_{60})$	0.8	0.014	0.126	0.131	0.957	0.017	0.088	0.091	0.947
	$\Lambda(\tau_{80})$	1.1	0.022	0.175	0.178	0.956	0.025	0.121	0.123	0.948

Table 5.2: Simulation results for $H(x) = x$ and $\alpha(t) = \sin(\pi t)\ \exp(t/2)/\{1+\exp(t/2)\}$ based on m control points of B-spline curves. τ_p represents $p\%$ of τ (study duration).

			Joint Model			**Marginal Model**			
m		*True*	*Bias*	*SE*	*MSE*	*Bias*	*SE*	*MSE*	*MSER*
					$n = 200$				
3	$\alpha(\tau_{20})$	0.52	0.009	0.099	0.010	-0.029	0.104	0.012	0.835
	$\alpha(\tau_{40})$	0.35	-0.020	0.106	0.012	-0.116	0.122	0.028	0.411
	$\alpha(\tau_{60})$	-0.38	0.027	0.117	0.014	-0.115	0.143	0.033	0.430
	$\alpha(\tau_{80})$	-0.66	-0.033	0.146	0.022	-0.213	0.203	0.086	0.260
6	$\alpha(\tau_{20})$	0.52	-0.001	0.116	0.013	-0.040	0.126	0.017	0.770
	$\alpha(\tau_{40})$	0.35	-0.001	0.115	0.013	-0.099	0.133	0.028	0.476
	$\alpha(\tau_{60})$	-0.38	-0.004	0.133	0.018	-0.142	0.161	0.046	0.383
	$\alpha(\tau_{80})$	-0.66	0.007	0.171	0.029	-0.176	0.233	0.085	0.343
9	$\alpha(\tau_{20})$	0.52	-0.004	0.122	0.015	-0.041	0.134	0.020	0.759
	$\alpha(\tau_{40})$	0.35	-0.006	0.124	0.015	-0.104	0.145	0.032	0.484
	$\alpha(\tau_{60})$	-0.38	-0.005	0.135	0.018	-0.143	0.163	0.047	0.387
	$\alpha(\tau_{80})$	-0.66	0.009	0.172	0.029	-0.175	0.234	0.085	0.347
					$n = 400$				
3	$\alpha(\tau_{20})$	0.52	0.013	0.068	0.005	-0.029	0.072	0.006	0.781
	$\alpha(\tau_{40})$	0.35	-0.014	0.073	0.006	-0.115	0.083	0.020	0.276
	$\alpha(\tau_{60})$	-0.38	0.026	0.078	0.007	-0.121	0.099	0.024	0.278
	$\alpha(\tau_{80})$	-0.66	-0.038	0.100	0.011	-0.220	0.145	0.069	0.164
6	$\alpha(\tau_{20})$	0.52	0.004	0.079	0.006	-0.038	0.087	0.009	0.700
	$\alpha(\tau_{40})$	0.35	0.005	0.080	0.006	-0.096	0.091	0.017	0.374
	$\alpha(\tau_{60})$	-0.38	-0.004	0.089	0.008	-0.152	0.112	0.036	0.221
	$\alpha(\tau_{80})$	-0.66	-0.000	0.116	0.014	-0.181	0.163	0.059	0.228
9	$\alpha(\tau_{20})$	0.52	0.003	0.083	0.007	-0.039	0.091	0.010	0.704
	$\alpha(\tau_{40})$	0.35	0.002	0.088	0.008	-0.100	0.099	0.020	0.392
	$\alpha(\tau_{60})$	-0.38	-0.006	0.090	0.008	-0.153	0.114	0.036	0.224
	$\alpha(\tau_{80})$	-0.66	0.001	0.117	0.014	-0.179	0.164	0.059	0.231

Table 5.3: Simulation results for $H(x) = \log(1+x)$ and $\alpha(t) = \sin(\pi t)\exp(t/2)/\{1+\exp(t/2)\}$ based on m=6 control points of B-spline curves. τ_p represents $p\%$ of τ (study duration).

			$n = 200$				$n = 400$			
ϕ		*True*	*Bias*	*SE*	*SEE*	*CP*	*Bias*	*SE*	*SEE*	*CP*
-0.3	β_1	1.0	0.005	0.124	0.121	0.947	0.001	0.082	0.085	0.956
	β_2	-0.5	0.002	0.106	0.104	0.950	0.001	0.074	0.074	0.944
	σ_e^2	1.0	-0.005	0.036	0.037	0.944	-0.003	0.026	0.026	0.951
	σ_b^2	0.5	-0.006	0.071	0.069	0.949	-0.005	0.049	0.049	0.945
	γ_1	1.0	-0.054	0.264	0.278	0.956	-0.056	0.193	0.193	0.938
	γ_2	-0.5	-0.005	0.228	0.232	0.954	0.006	0.159	0.162	0.954
	ϕ	-0.3	0.002	0.264	0.258	0.950	0.005	0.185	0.180	0.945
	$\Lambda(\tau_{20})$	0.3	0.011	0.063	0.065	0.953	0.010	0.045	0.045	0.956
	$\Lambda(\tau_{40})$	0.6	0.026	0.121	0.122	0.945	0.020	0.083	0.084	0.949
	$\Lambda(\tau_{60})$	0.8	0.041	0.185	0.185	0.943	0.035	0.128	0.128	0.952
	$\Lambda(\tau_{80})$	1.1	0.055	0.258	0.258	0.956	0.048	0.183	0.179	0.951
0	β_1	1.0	-0.003	0.118	0.121	0.960	0.001	0.087	0.085	0.945
	β_2	-0.5	0.002	0.107	0.104	0.943	0.001	0.074	0.074	0.947
	σ_e^2	1.0	-0.005	0.038	0.037	0.935	-0.004	0.025	0.026	0.958
	σ_b^2	0.5	-0.009	0.069	0.068	0.953	-0.005	0.049	0.049	0.953
	γ_1	1.0	-0.047	0.262	0.271	0.945	-0.058	0.187	0.189	0.937
	γ_2	-0.5	-0.006	0.227	0.229	0.946	0.006	0.156	0.161	0.962
	ϕ	0.0	0.000	0.248	0.256	0.959	0.002	0.180	0.178	0.952
	$\Lambda(\tau_{20})$	0.3	0.008	0.062	0.063	0.956	0.013	0.046	0.045	0.945
	$\Lambda(\tau_{40})$	0.6	0.022	0.119	0.119	0.952	0.025	0.085	0.084	0.944
	$\Lambda(\tau_{60})$	0.8	0.036	0.188	0.181	0.946	0.037	0.129	0.127	0.942
	$\Lambda(\tau_{80})$	1.1	0.054	0.264	0.255	0.945	0.052	0.183	0.178	0.940
0.3	β_1	1.0	-0.005	0.122	0.121	0.943	-0.005	0.085	0.085	0.951
	β_2	-0.5	0.003	0.107	0.104	0.944	-0.005	0.074	0.074	0.935
	σ_e^2	1.0	-0.005	0.039	0.037	0.937	-0.003	0.027	0.026	0.952
	σ_b^2	0.5	-0.011	0.071	0.068	0.950	-0.005	0.051	0.049	0.943
	γ_1	1.0	-0.052	0.278	0.278	0.941	-0.044	0.187	0.193	0.952
	γ_2	-0.5	0.003	0.230	0.232	0.957	0.002	0.162	0.162	0.948
	ϕ	0.3	0.007	0.257	0.259	0.955	-0.007	0.176	0.180	0.954
	$\Lambda(\tau_{20})$	0.3	0.009	0.063	0.064	0.962	0.008	0.047	0.045	0.936
	$\Lambda(\tau_{40})$	0.6	0.022	0.115	0.121	0.966	0.017	0.086	0.084	0.935
	$\Lambda(\tau_{60})$	0.8	0.037	0.178	0.183	0.956	0.029	0.128	0.127	0.949
	$\Lambda(\tau_{80})$	1.1	0.057	0.267	0.258	0.955	0.037	0.181	0.178	0.942

Table 5.4: Simulation results for $H(x) = \log(1+x)$ and $\alpha(t) = \sin(\pi t)\exp(t/2)/\{1+\exp(t/2)\}$ based on m control points of B-spline curves. τ_p represents $p\%$ of τ (study duration).

			Joint Model			**Marginal Model**			
m		*True*	*Bias*	*SE*	*MSE*	*Bias*	*SE*	*MSE*	*MSER*
					$n = 200$				
3	$\alpha(\tau_{20})$	0.52	0.003	0.095	0.009	-0.035	0.100	0.011	0.806
	$\alpha(\tau_{40})$	0.35	-0.012	0.099	0.010	-0.069	0.107	0.016	0.612
	$\alpha(\tau_{60})$	-0.38	0.024	0.107	0.012	-0.049	0.128	0.019	0.642
	$\alpha(\tau_{80})$	-0.66	-0.014	0.126	0.016	-0.097	0.168	0.038	0.426
6	$\alpha(\tau_{20})$	0.52	0.000	0.100	0.010	-0.037	0.106	0.013	0.784
	$\alpha(\tau_{40})$	0.35	0.003	0.108	0.012	-0.054	0.117	0.017	0.701
	$\alpha(\tau_{60})$	-0.38	0.005	0.112	0.013	-0.068	0.136	0.023	0.547
	$\alpha(\tau_{80})$	-0.66	0.009	0.133	0.018	-0.073	0.174	0.036	0.499
9	$\alpha(\tau_{20})$	0.52	0.000	0.115	0.013	-0.037	0.122	0.016	0.804
	$\alpha(\tau_{40})$	0.35	0.006	0.115	0.013	-0.050	0.127	0.018	0.714
	$\alpha(\tau_{60})$	-0.38	0.004	0.122	0.015	-0.069	0.146	0.026	0.567
	$\alpha(\tau_{80})$	-0.66	0.009	0.135	0.018	-0.073	0.176	0.036	0.504
					$n = 400$				
3	$\alpha(\tau_{20})$	0.52	0.005	0.065	0.004	-0.030	0.071	0.006	0.729
	$\alpha(\tau_{40})$	0.35	-0.013	0.065	0.004	-0.072	0.076	0.011	0.405
	$\alpha(\tau_{60})$	-0.38	0.021	0.071	0.005	-0.052	0.088	0.010	0.526
	$\alpha(\tau_{80})$	-0.66	-0.018	0.082	0.007	-0.100	0.110	0.022	0.319
6	$\alpha(\tau_{20})$	0.52	0.002	0.069	0.005	-0.033	0.075	0.007	0.709
	$\alpha(\tau_{40})$	0.35	0.003	0.072	0.005	-0.056	0.083	0.010	0.511
	$\alpha(\tau_{60})$	-0.38	0.001	0.074	0.006	-0.071	0.093	0.014	0.404
	$\alpha(\tau_{80})$	-0.66	0.006	0.087	0.008	-0.076	0.117	0.019	0.390
9	$\alpha(\tau_{20})$	0.52	0.000	0.082	0.007	-0.033	0.090	0.009	0.720
	$\alpha(\tau_{40})$	0.35	0.004	0.075	0.006	-0.053	0.087	0.010	0.532
	$\alpha(\tau_{60})$	-0.38	-0.001	0.080	0.006	-0.074	0.100	0.015	0.416
	$\alpha(\tau_{80})$	-0.66	0.007	0.087	0.008	-0.075	0.117	0.019	0.396

Table 5.5: Joint analysis results of the medical costs data. The 50:50 mixture of χ^2 distributions is used for testing variances.

Effect	*Estimate*	*Std.Error*	*p-value*
	Longitudinal medical cost		
Age	-0.116	0.042	0.0054
Age^2	-0.184	0.048	0.0001
Male (vs. Female)	0.112	0.061	0.0691
White (vs. Nonwhite)	-0.248	0.069	0.0003
σ_e^2	2.336	0.027	$< .0001$
σ_b^2	1.034	0.051	$< .0001$
	Death (informative censoring)		
Age	0.524	0.088	$< .0001$
Male (vs. Female)	0.246	0.127	0.0535
White (vs. Nonwhite)	-0.272	0.127	0.0319
ϕ	0.867	0.078	$< .0001$

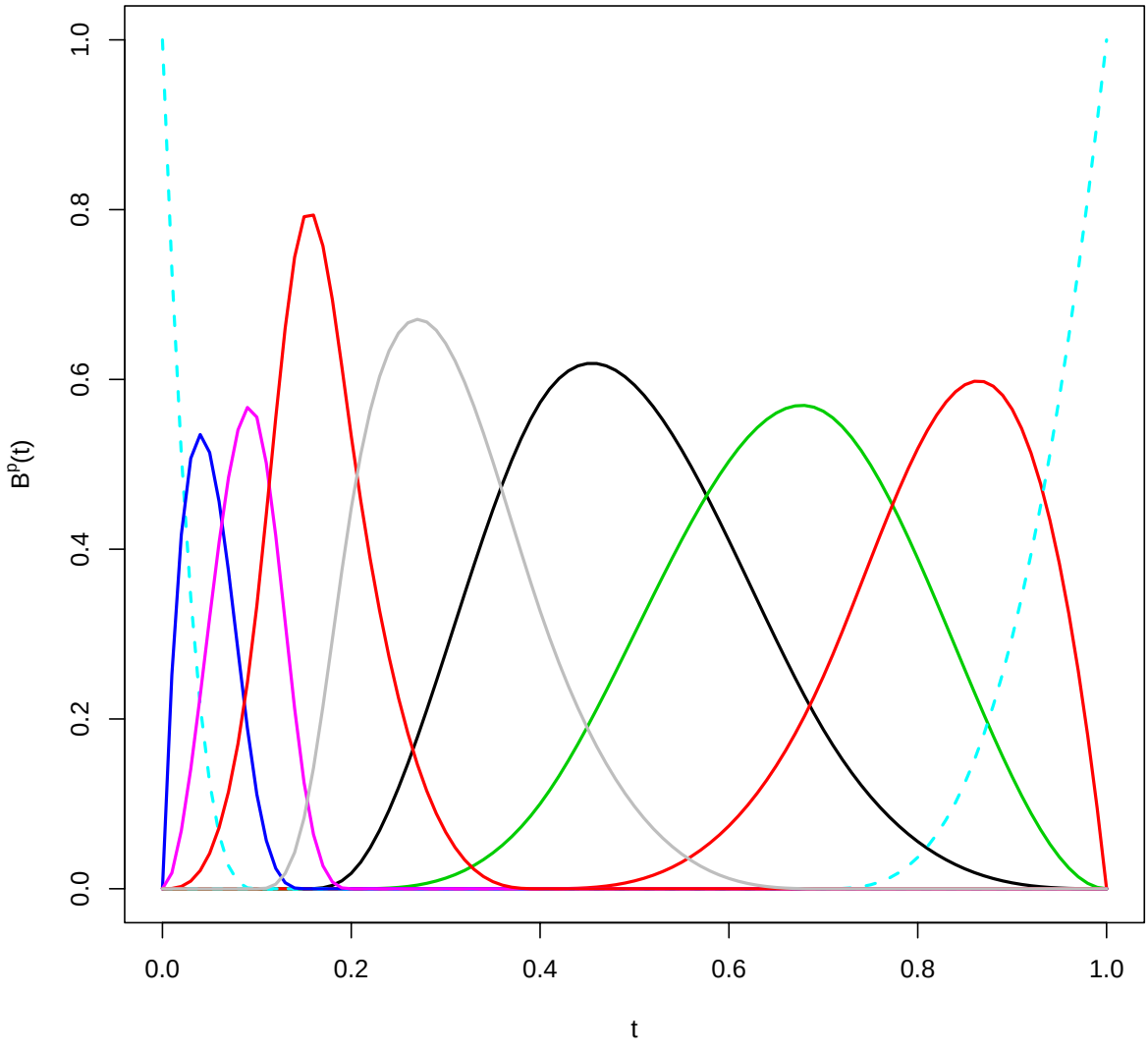

Figure 5.1: Example of basis functions (cubic B-spline) for time t in [0, 1] under 5 control points $\{0.1, 0.15, 0.2, 0.4, 0.7\}$.

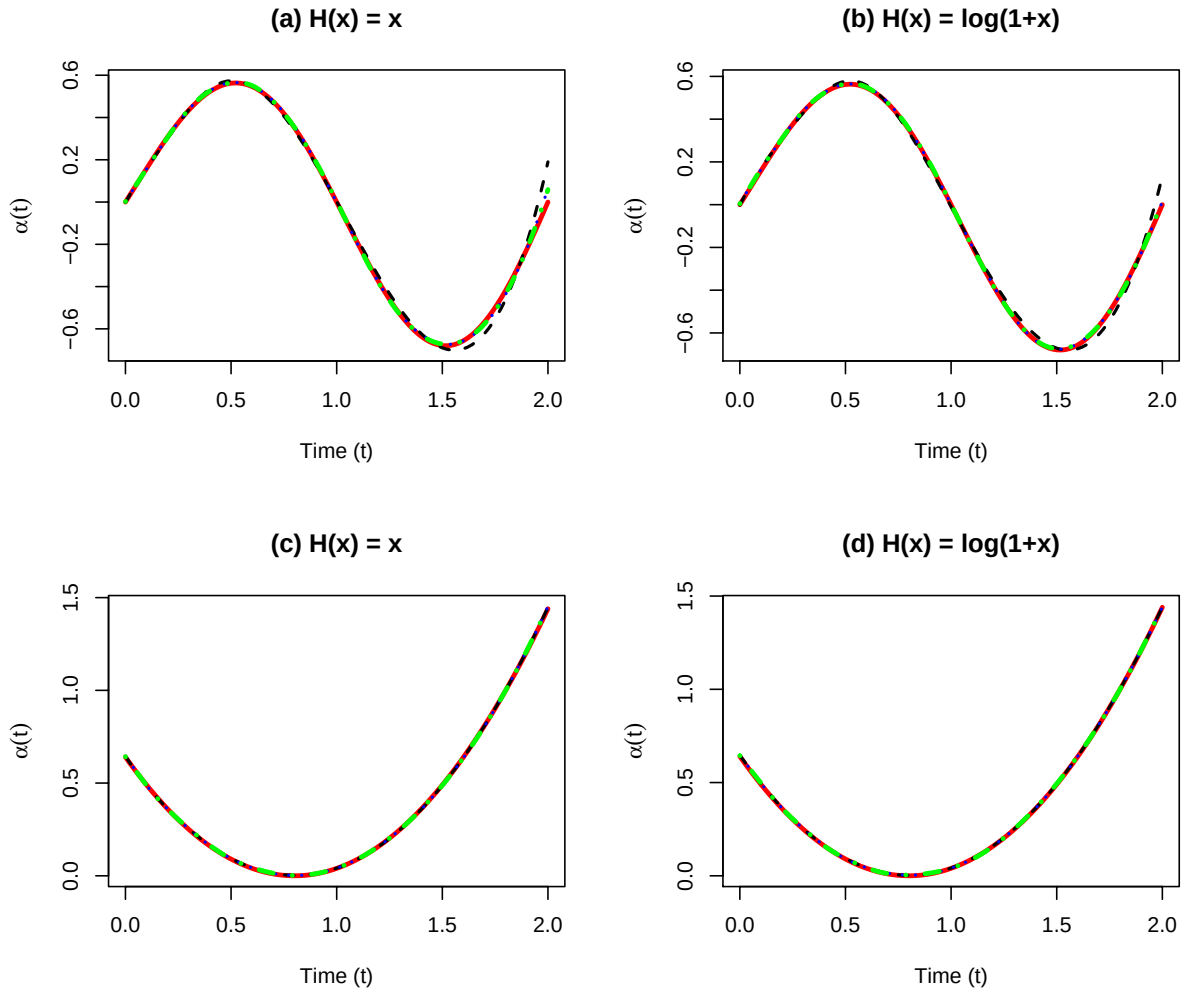

Figure 5.2: Simulation results for the baseline coefficient function by (a) $H(x) = x$ and $\alpha(t) = \sin(\pi t)\, e^{\frac{t}{2}}/(1+e^{\frac{t}{2}})$; (b) $H(x) = \log(1+x)$ and $\alpha(t) = \sin(\pi t)\, e^{\frac{t}{2}}/(1+e^{\frac{t}{2}})$; (c) $H(x) = x$ and $\alpha(t) = (t-0.8)^2$; and (d) $H(x) = \log(1+x)$ and $\alpha(t) = (t-0.8)^2$. The solid curves are true values, the dashed curves are estimates under m=3, the dash-dotted curves are estimates under m=6, and the dotted curves are estimates under m=9.

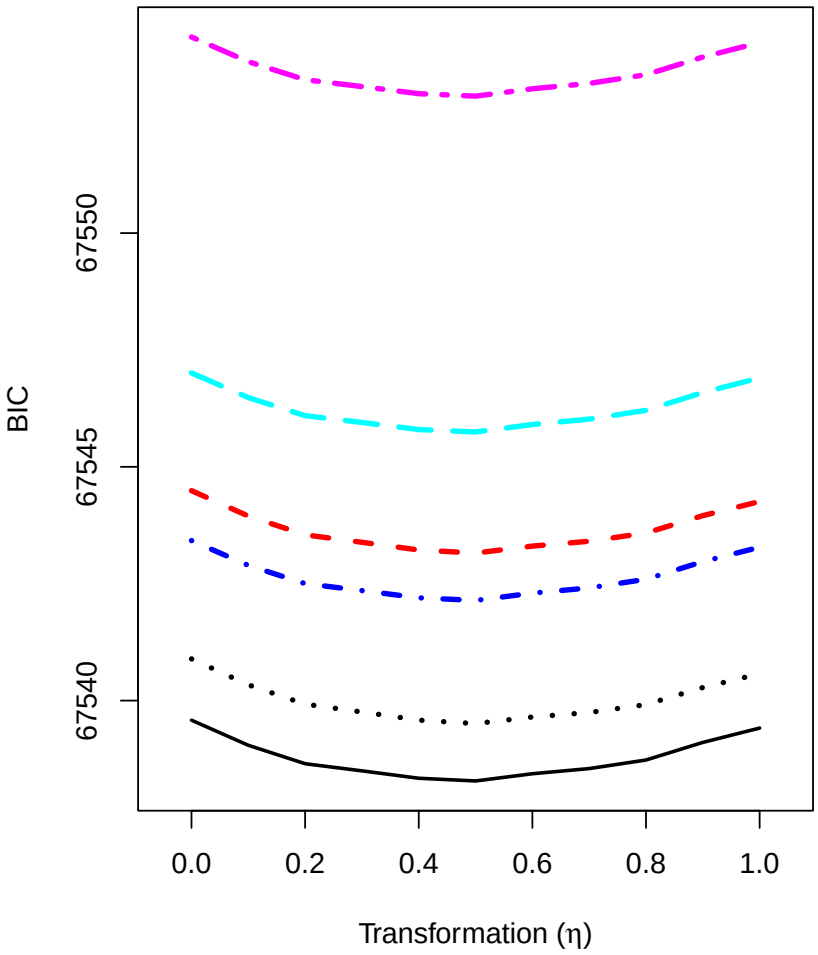

Figure 5.3: Bayesian information criterion (BIC) for the transformation $H(x) = \log(1 + \eta x)/\eta$ and the number of control knots (m). From top to bottom, the dot-long-dashed curve is for m=8, the long-dashed curve is for m=7, the short-dashed curve is for m=4, the dot-short-dashed curve is for m=6, the dotted curve is for m=3, and the solid curve is for m=5.

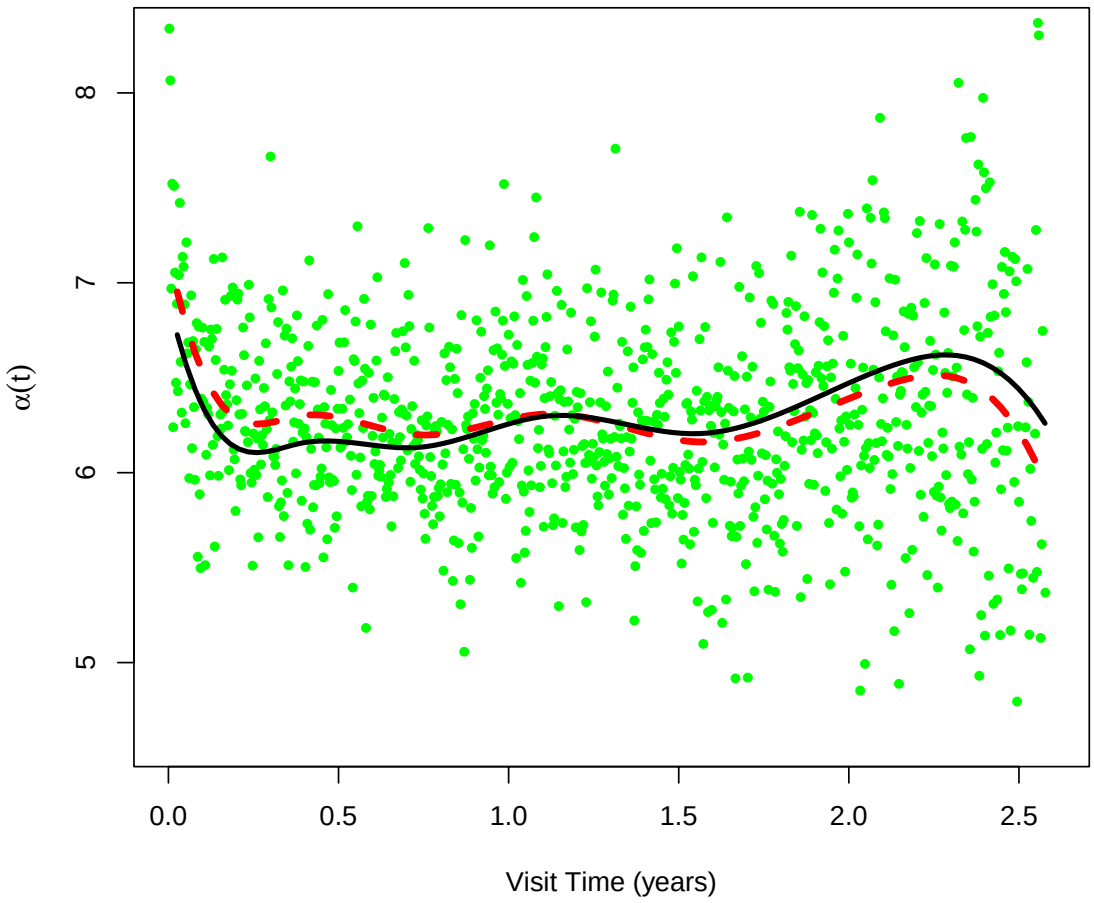

Figure 5.4: Baseline coefficient function of hospital visit time in the medical cost data under the best fit of transformation $H(x) = 2\log(1 + 0.5x)$ and 5 control points. The solid curves are estimates from the joint model, the dashed curves are estimates from the marginal model, and the dots are residual means of $\{Y(t) - \hat{\beta}^T X_1(t)\}$.

Chapter 6

Summary and Future Research

In this dissertation, we have studied semiparametric models for joint analysis of longitudinal data and counting processes, where the models of each component were connected through the shared random effects. Particularly, in Chapter 3, we developed joint models of longitudinal data via the linear mixed effects model and recurrent and terminal events via transformation models. The proposed joint models captured latent relationships among the longitudinal responses and two events. In Chapter 4, we studied the joint analysis of longitudinal data and cure-survival data, using the linear mixed effects model and the transformed promotion time cure model. We found that the proposed joint models corrected biases in the regression coefficient estimates, induced by ignoring the true cure proportion in survival data and the correlation between longitudinal and cure-survival data. In Chapter 5, we modeled nonlinear trajectories of longitudinal data with informative censoring. We relaxed usual parametric model specification of the longitudinal data by using the partially linear model, and we adopted transformed survival models to account for informative censoring. We showed that the proposed joint modeling approach can reduce biases in the estimation of the parametric and nonparametric coefficients.

In all of the methods, the maximum likelihood approach was used under proper

conditions on infinite-dimensional parameters. We assumed the baseline cumulative intensity or hazards functions to be step functions, and the underlying trajectory of longitudinal responses were assumed to be smooth enough for B-spline approximation. By treating the shared random effects as missing data, simple EM algorithms were provided to compute the MLEs.

The asymptotic properties of the MLEs were studied and were shown to provide desirable properties, consistency, normality and semiparametric efficiency. Most of the proofs relied on modern empirical process theory. We also investigated the finite sample properties of the proposed methods via extensive simulation studies. Simulation results based on various scenarios confirmed that the proposed methods worked properly under reasonably finite sample sizes. The proposed methods were also applied to real data examples for illustration; specifically, the ARIC data was used to analyze the longitudinal SBP, recurrent CHD events, and death in Chapter 3 and to analyze the longitudinal SBP with MI or fatal CHD event in Chapter 4. In Chapter 5, we analyzed the medical costs of chronic heart failure patients, while accounting for death as the informative censoring event.

The proposed methods in this dissertation research can be extended in several directions. In Chapters 4 - 5, measurement times of longitudinal data were assumed to be non-informative. In practice, they may contain important information if observed, for example, at emergency admissions. A natural way of adjusting the informative observation times is to combine another counting process with the proposed models. The proposed work in Chapter 5 can be continued to the context of time-varying coefficients models that accommodate more than one nonparametric functions of time. New methods for time-varying coefficients models with informatively censored data would be useful for analyzing clinical trial data where the effects of treatments may vary over time. We can also extend the joint modeling approach to develop methodol-

ogy for interval-censored data with longitudinal covariates. Lastly, further research on the model selection techniques to check the adequacy of the model assumptions and to select the best transformation would be worthwhile to continue for the practical applications of the research.

Bibliography

Adams, R. and Fournier, J. (1975). *Sobolev spaces*. New York: Academic Press.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2:273–277.

Berkson, J. and Gage, R. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47:501–515.

Brown, E. and Ibrahim, J. (2003a). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59:221–228.

Brown, E. and Ibrahim, J. (2003b). Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics*, 59:686–693.

Brown, E., Ibrahim, J., and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61:64–73.

Chambless, L., Folsom, A., Sharrett, A., Sorlie, P., Couper, D., Szklo, M., and Nieto, F. (2003). Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. *Journal of clinical epidemiology*, 56:880–890.

Chen, M., Ibrahim, J., and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94:909–919.

Chen, M., Ibrahim, J., and Sinha, D. (2004). A new joint model for longitudinal and survival data with a cure fraction. *Journal of Multivariate Analysis*, 91:18–34.

Cheng, S., Wei, L., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, 82:835–845.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, 34:187–220.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38.

Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press, USA.

Farewell, V. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38:1041–1046.

Farewell, V. (1986). Mixture models in survival analysis: Are they worth the risk? *The Canadian Journal of Statistics*, 14:257–262.

Green, P. and Silverman, B. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman & Hall/CRC.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and recurrent events. *Biostatistics*, 1:465–480.

Hogan, J., Lin, X., and Herman, B. (2004). Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics*, 60:854–864.

Hoover, D., Rice, J., Wu, C., and Yang, L. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85:809–822.

Hsieh, F., Tseng, Y., and Wang, J. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics*, 62:1037–1043.

Huang, C. and Wang, M. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, 99:1153–1165.

Huang, J., Wu, C., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89:111–128.

Huang, X. and Liu, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics*, 63:389–397.

Ibrahim, J., Chen, M., and Sinha, D. (2001). *Bayesian survival analysis*. New York: Springer-Verlag.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53:457–481.

Kuk, A. and Chen, C. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79:531–541.

Lancaster, T. and Intrator, O. (1998). Panel data with survival: Hospitalization of HIV-positive patients. *Journal of the American Statistical Association*, 93:46–53.

Laska, E. and Meisner, M. (1992). Nonparametric estimation and testing in a cure model. *Biometrics*, 48:1223–1234.

Law, N., Taylor, J., and Sandler, H. (2002). The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics*, 3:547–563.

Liang, Y., Lu, W., and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, 65:377–384.

Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96:103–126.

Lin, H., Scharfstein, D., and Rosenheck, R. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society, Series B*, 66:791–813.

Lin, X. and Carroll, R. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96:1045–1056.

Liu, L. and Huang, X. (2009). Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society, Series C*, 58:65–81.

Liu, L., Huang, X., and O'Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics*, 64:950–958.

Liu, L., Wolfe, R., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60:747–756.

Liu, L., Wolfe, R., and Kalbfleisch, J. (2007). A shared random effects model for censored medical costs and mortality. *Statistics in medicine*, 26:139–155.

Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44:226–233.

Lu, W. and Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika*, 91:331–343.

Moyeed, R. and Diggle, P. (1994). Rates of convergence in semi-parametric modelling of longitudinal data. *Australian & New Zealand Journal of Statistics*, 36:75–93.

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95:449–485.

Peng, Y. and Dear, K. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56:237–243.

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

Ratcliffe, S., Guo, W., and Ten Have, T. (2004). Joint modeling of longitudinal and survival data via a common frailty. *Biometrics*, 60:892–899.

Rice, J. and Wu, C. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57:253–259.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V., and Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*, 8:708–721.

Schumaker, L. (2007). *Spline functions: basic theory*. Cambridge Univ. Press.

Shen, X. (1997). On methods of sieves and penalization. *The Annals of Statistics*, 25:2555–2591.

Song, X., Davidian, M., and Tsiatis, A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58:742–753.

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

Sun, J., Park, D., Sun, L., and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association*, 100:882–889.

Sy, J. and Taylor, J. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, 56:227–236.

Taylor, J. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, 51:899–907.

Tsiatis, A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88:447–458.

Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14:809–834.

Tsiatis, A., DeGruttola, V., and Wulfsohn, M. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90:27–37.

Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, 54:1508–1516.

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge Univ. Press.

van der Vaart, A. and Wellner, J. (1996). *Weak convergence and empirical processes*. New York: Springer-Verlag.

Vonesh, E., Greene, T., and Schluchter, M. (2006). Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*, 25:143–163.

Wang, M., Qin, J., and Chiang, C. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, 96:1057–1065.

Wattanakit, K., Folsom, A., Chambless, L., and Nieto, F. (2005). Risk factors for cardiovascular event recurrence in the Atherosclerosis Risk in Communities (ARIC) study. *American Heart Journal*, 149:606–612.

Wulfsohn, M. and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:330–339.

Xu, J. and Zeger, S. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics*, 50:375–387.

Yakovlev, A., Tsodikov, A., and Asselain, B. (1996). *Stochastic models of tumor latency and their biostatistical applications*. New Jersey: World Scientific.

Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of "Permanent Employment" in Japan. *Journal of the American Statistical Association*, 87:284–292.

Ye, Y., Kalbfleisch, J., and Schaubel, D. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics*, 63:78–87.

Yu, M., Law, N., Taylor, J., and Sandler, H. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 14:835–862.

Yu, M., Taylor, J., and Sandler, H. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association*, 103:178–187.

Zeger, S. and Diggle, P. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, 50:689–699.

Zeng, D. (2005). Likelihood approach for marginal proportional hazards regression in the presence of dependent censoring. *Annals of statistics*, 33:501–521.

Zeng, D. and Cai, J. (2005a). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics*, 33:2132–2163.

Zeng, D. and Cai, J. (2005b). Simultaneous modelling of survival and longitudinal data with an application to repeated quality of life measures. *Lifetime Data Analysis*, 11:151–174.

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93:627–640.

Zeng, D. and Lin, D. (2007a). Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association*, 102:167–180.

Zeng, D. and Lin, D. Y. (2007b). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society, Series B*, 69:507–564.

Zeng, D. and Lin, D. Y. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics*, 65:746–752.

Zeng, D., Yin, G., and Ibrahim, J. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101:670–684.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 93:710–719.