University of California
Santa Barbara

Joint Modeling of Mixed Outcomes in Clinical Research

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Statistics and Applied Probability

by

Yuqi Chen

Committee in charge:

Professor Yuedong Wang, Chair
Professor Sreenivasa Rao Jammalamadaka
Professor Wendy Meiring

June 2016

The Dissertation of Yuqi Chen is approved.

Professor Sreenivasa Rao Jammalamadaka

Professor Wendy Meiring

Professor Yuedong Wang, Committee Chair

April 2016

Joint Modeling of Mixed Outcomes in Clinical Research

by

Yuqi Chen

To My Family

Acknowledgements

I wish to express my sincere gratitude to my advisor Dr. Yuedong Wang for his constant guidance, support and encouragement in my Ph.D. study. I really appreciate the enormous number of hours that he spent in answering my questions and revising the writings. This work would have been impossible without his great guidance and endless patience. I also thank him for providing me well-rounded experience consistent with my long-term career goal. During the past five years, I learned not only the knowledge and expertise, but also the enthusiasm, diligence and creative-thinking as a researcher through him, which I will benefit from for the rest of my life.

I am also very grateful to Dr. Meiring and Dr. Jammalamadaka for their willingness and precious time to serve as my committee members. I really appreciate their insights and comments on my dissertation.

It has been a wonderful experience studying and working in the Department of Statistics at University of California, Santa Barbara. I would like to thank all the faculty, staff and my fellow graduate students for providing such a friendly academic environment and for their various support during my graduate study.

Finally, I would like to thank my parents for their unconditional love and support though my life. It is their encouragement that makes what I am today. And my special thanks go to my husband Yingjie Hu, who is a great companion, a best friend, and my pillar of strength through all the ups and downs.

Curriculum Vitæ

Yuqi Chen

Education

2016	Ph.D. in Statistics (Expected), University of California, Santa Barbara.
2012	M.A. in Statistics, University of California, Santa Barbara.
2011	B.S. in Statistics, East China Normal University.

Working Experience

2015 Summer	Statistician, *Santa Barbara Actuaries Inc.*
2014 Summer	Biostatistician Intern, *Amgen Inc.*
2013 Summer	Statistical Consultant, *Fresenius Medical Care.*
2013-2015	Statistical Consultant, *StatLab at U.C. Santa Barbara.*
2011-2015	Teaching Assistant, *U.C. Santa Barbara.*
2011 Spring	Statistical Programmer Intern, *Tigermed Consulting Ltd.*

Publication

Yuqi Chen, Pang Du, Yuedong Wang, Variable Selection in Linear Models, *Wiley Interdisciplinary Reviews: Computational Statistics,* 2013

Yingjie Hu, Krzysztof Janowicz, **Yuqi Chen**, Task-oriented information value measurement based on space-time prisms, *International Journal of Geographical Information Science,* 2015

Presentation

Invited talk, *Ninth ICSA International Conference,* 2013, Hong Kong, China
Joint Modeling of Survival Times and Responses from Exponential Family

Contributed paper, *ENAR 2015 Meeting,* 2015, Miami, United States
Analysis of a Composite Endpoint Under Different Censoring Schemes For Component Events

Contributed paper, *JSM 2015 Meeting,* 2015, Seattle, United States
Mixed Effect Joint Model of Hospital Admission and Readmission via Shared Latent Random Variable

Abstract

Joint Modeling of Mixed Outcomes in Clinical Research

by

Yuqi Chen

Mixed types of multivariate outcomes are common in clinical investigations. Survival time is one of the primary goals in practice. In addition, hospitalization attracts increasing attention as it is a main contributor to the total cost of care, and the identification of related risk factors is of interest in many health economics studies. Meanwhile, we are also interested in the longitudinal path of important clinical measurements along the progress of disease. Joint modeling is often required as both hospitalization frequencies or longitudinal measurements can be informatively censored due to death. In this dissertation, we will propose three research projects which jointly model multiple aspects of the outcomes.

The first research project models survival time and hospitalization together through a latent subject-specific random frailty. B-spline bases are introduced for flexible forms of baseline hazard and the offset function. Computational methods to solve for the MLE and to select knots are developed. The proposed methods are applied to study the risk factors of hospitalization and survival time among end-stage-renal-disease (ESRD) patients.

The second part proposes a joint model of hospitalization and readmission. Number of hospitalizations is modeled as a Poisson random variable and number of readmissions is treated as a Binomial random variable with number of hospitalizations being the total number of trials. The proposed joint modeling framework is applied to evaluate the performance of an intervention program from Fresenius Medical Care in reducing number

of hospitalizations and readmissions.

The third research project jointly models survival time and multiple longitudinal observations. A penalized likelihood approach is described for variable selection. We design a Coordinate Descent Algorithm to solve for the penalized MLE and a two-stage estimation method to reduce the bias resulting from penalization. Simulation results demonstrate good selection and estimation property. We illustrate the practical usage of proposed method through an application to ESRD patients.

Contents

Chapter 1

Introduction

Mixed types of outcomes are very common in clinical research. Patients' survival time is often the primary interest. Other outcomes of interest include hospitalization, quality of life and important clinical measurements. The main interest is to investigate the relationship between mixed outcomes and predictors. The study may be cross-sectional where predictors are observed at baseline and outcomes are observed at follow-up time, or longitudinal where both outcomes and predictors are observed over time. Outcomes collected from the same patient are usually correlated since they reflect the patient's underlying health condition. Therefore fitting each outcome separately is less efficient compared with joint modeling, and may lead to bias. For example, Ibrahim et al. [32] and Henderson et al. [27] have shown that severe bias occurs for some parameters when latent association between longitudinal measures and event time data is ignored. Liu et al. [40] have also shown that ignoring the dependence between the terminal and recurrent events can result in significant biases. Statistical techniques of joint modeling are required for valid data analysis and inference.

Our research is motivated by the need for improvement in care for patients who have end-stage renal disease (ESRD). Chronic kidney disease (CKD) is a major public health

problem affecting 11% of the US population [8]. ESRD is the last stage of CKD when kidneys can no longer support the body's needs. ESRD patients need kidney transplants or rely on dialysis to remove extra salt, water, and waste products. ESRD is a complex condition, the failure of kidney function is accompanied by numerous metabolic changes which affect almost all organ systems of the human body. In general, ESRD patients suffer from multiple comorbidities, such as diabetes and cardiovascular diseases, resulting in frequent hospitalizations and substantial mortality. Many scientific questions arise in practice in the care for ESRD patients that require novel statistical methodologies. This dissertation aims to develop efficient and valid statistical methodologies to address these scientific questions.

We will develop statistical methods for three situations with mixed outcomes and apply them to investigate the relationships between mortality, hospitalization and predictors. We introduce these three situations in the next three sections.

1.1 Joint Modeling of Mortality and Hospitalization with Cross-Sectional Data

Hospitalization, an important marker of disease severity, and a substantial contributor to medical cost, has attracted much attention in the literature. The aggregated cost for all hospital stays was $387.3 billion in 2011, nearly one-third of all health care expenses in the United States [44]. Specifically, the aggregate cost for stays with acute and unspecified renal failure jumped from $1.0 billion in 1997 to $4.7 billion in 2011.

The high stake of hospitalization motivates health care professionals and policy makers to investigate associated risk factors for better intervention. A large number of studies have studied the risk factors for various diseases. For example, Moss et al. [43] identified

risk factors for hospitalization in diabetic patients, Eisner et al. [13] studied the risk factors for hospitalization among adult asthma patients, Dennehy et al. [11] investigated risk factors for hospitalization in pediatric rotavirus gastroenteritis patients. Inrig et al. [34] and Brotman et al. [3] examined the effects of blood pressure changes and heart rate variability on hospitalization in ESRD patients respectively.

One challenge in modeling hospitalization is that it is informatively censored by a terminal event, namely death. Therefore it is desirable to model the two outcomes (death and hospitalization) jointly since they are correlated. To investigate risk factors associated with mortality and hospitalization together, in Chapter 2 I will propose a semi-parametric joint model for survival time and a random variable from exponential family. Hospitalization can be easily fitted into this framework. For example, the number of hospitalizations can be treated as Poisson random variables and the total length of stay, another important measure of hospitalization, can be modeled by Gamma distribution. Shared frailty is introduced to account for the correlation between the two outcomes. An offset function is incorporated in the hospitalization sub-model to accommodate the fact that hospitalization can only be observed prior to death. To allow flexible form of baseline hazard and the offset set, I assume each of these is a smooth function and model them using B-splines with the number of knots selected by the AIC criterion. A method for computing the maximum likelihood estimates will be developed and implemented using SAS Proc NLMIXED. Simulations will be conducted to evaluate the proposed method. I apply the proposed method to study the risk factors for hospitalization and mortality of ESRD patients who are on hemodialysis. I analyze two applications on ESRD patients: joint model of frequency of hospitalization and survival time, and joint model of total length of stays and survival time.

1.2 Joint Modeling of Hospitalization and Readmission

Readmission is defined as a patient being admitted to a hospital within a certain time period from an initial admission. A 30-days rule is usually used by Medicare. A study of Medicare claims data from 2003 to 2004 found that almost one fifth (19.6%) of Medicare beneficiaries who had been discharged from a hospital were rehospitalized within 30 days [35]. The estimated cost to Medicare of unplanned rehospitalization in 2004 was as high as $17.4 billion.

If a hospital has a high proportion of patients readmitted within a short time frame, it may be an indication of inadequate quality of care in the hospital or a lack of appropriate coordination of post-discharge care. Preventable hospital readmissions result from several factors: lack of discharge instruction, poor quality post-hospitalization care or poor transition of patients among providers. Many efforts have been made to alleviate these problems [23].

Hospitalization cost and readmission rates are high in ESRD patients. United States Renal Data System reports that in 2015, hospitalization accounts for approximately 40% of total Medicare expenditures for dialysis patients and about 30% of ESRD patients have an unplanned rehospitalization within the 30 days following discharge [50].

Targeting on ESRD patients on hemodialysis, Fresenius Medical Care initiated an intervention program in 2013 in order to reduce the number of hospitalizations and readmissions for ESRD patients on hemodialysis. Details about this study will be discussed in Chapter 3.

Motivated by the need of evaluation for this intervention program, I propose a joint model for hospitalization and readmission in Chapter 3. The number of hospitalizations is modeled using a Poisson distribution, and conditional on the number of hospitalizations,

the number of readmissions is modeled using a Binomial distribution where the number of hospitalizations serves as the total number of trials. A shared frailty term is introduced and an extra function is incorporated to explain the possible non-linear association between the two responses. I apply the proposed method to evaluate the effectiveness of the intervention program in reducing patients hospitalizations and readmissions.

1.3 Variable Selection in Joint Modeling of Multivariate Longitudinal Covariates and Survival Time

Methods in the previous two sections are intended for cross sectional data, where a baseline period and a follow-up period are established. This section considers longitudinal studies. The important clinical measurements are recorded repeatedly over time and the longitudinal patterns provide insights of disease progression. In some situations, the course of a disease is determined by one underlying clinical measurement which is referred as a biomarker. For example, for HIV patients, CD4 cell counts is a well known biomarker and is often monitored longitudinally together with patients' survival time. In some other complicated diseases, no one golden biomarker exists and therefore multiple longitudinal processes need to be recorded. For example in ESRD patients, the impaired kidney function is accompanied by numerous biological changes in the body and there is no one simple clinical measurement that fully captures the disease progression.

Since the longitudinal processes are correlated and terminated by patients' death, a large body of literatures have explored joint modeling of longitudinal processes and survival time [62, 9, 17, 56, 31]. Mixed effects models are commonly used to model the longitudinal covariate and the Cox proportional hazard model is a classical choice for survival time. Shared random effects are often introduced to account for the correlation

between the two outcomes. Joint modeling of multiple longitudinal processes and survival time has also been proposed from a Bayesian's perspective [4]. However variable selection, an essential part in statistical theory and practice, is rarely discussed in the literature of joint modeling framework. Variable selection is especially important when there are multiple longitudinal outcomes.

To fill this gap, I will propose a joint model for multiple longitudinal outcomes and survival time in Chapter 4. Longitudinal outcomes are modeled by multivariate linear mixed effect models. The random intercepts and slopes, as trajectories of longitudinal measurements, are then incorporated in the sub-model of survival time. A penalized likelihood approach is proposed to perform variable selection of these random effects in a Cox model. I will develop a coordinate descent algorithm for the optimization of penalized likelihood. A two-stage estimation method will be adopted to reduce the estimation biases resulting from penalization. The proposed procedure will be applied to an ESRD study to identify important predictors for mortality.

Chapter 2

Joint Modeling of Mortality and Hospitalization with Cross-Sectional Data

2.1 Introduction and Related Work

Hospitalization is a main contributor to the total cost of care, and identification of the related risk factors is of interest in many health care studies. The main difficulty in modeling hospitalization data is due to the fact that the frequency of hospitalization and the total length of hospital stays are functions of follow-up time that can be informatively censored due to death. Since both the hospitalization outcome and time-to-death are related to the underlying health, it is desirable to jointly model them as bivariate outcomes. Mixed types of multivariate outcomes are common in many fields of science and social science. Various statistical models and methods have been proposed to deal with different types of mixed outcomes [10]. For example, Fitzmaurice and Laird proposed regression models for continuous and binary outcomes [18], Sammel et al. pro-

posed latent variable models for mixed discrete and continuous outcomes [49], Catalano proposed a latent variable model for continuous and ordinal outcomes [6], and Dunson and Herring proposed Bayesian latent variable models for mixed outcomes [12]. These methods can not handle censored data which is needed for joint modeling of survival time and hospitalization in health studies.

A large number of models have been developed aiming at the joint modeling of survival hazard function and hospitalization rate simultaneously. For example, Lancaster and Intrator 1998 [37], Wang, Qin and Chiang 2001 [60], Huang and Wang 2004 [30], Liu, Wolfe and Huang 2004 [40], Huang, Qin and Wang 2010 [29] and the references therein. These researches treat hospitalization as a recurrent event and death as a terminal event. And they are interested in modeling the intensity function of the recurrent process. In this research, we are interest in modeling the expected number of hospital admissions and hospital stay.

Our research is motivated by the need for improvement in care for end-stage renal disease (ESRD) patients. Hemodialysis (HD) is the most frequently used treatment modality for ESRD patients. In general, HD patients suffer from multiple comorbidities, such as diabetes and cardiovascular diseases, resulting in frequent hospitalizations and substantial mortality. In spite of improvements over the years, hospitalization and mortality rates of ESRD patients on HD remain much higher than those of the general population [7]. In this chapter I am interested in identifying risk factors for hospitalization and mortality. The data come from an observational study of patients on HD in Fresenius Medical Care. Covariates at baseline and outcomes including survival time, hospital admissions and total length of hospital stay at follow-up were collected. Approximately 20% of patients died during the follow-up period and observational times for hospitalization outcomes of these patients are censored due to death. Since both survival time and hospitalization are associated with the underlying health condition, it is likely

that these outcomes from the same subject are correlated. Therefore, it is necessary to develop a joint model for survival time and hospitalization. Details of the data are given in Section 2.5.

In this chapter I propose a semi-parametric latent variable model for joint modeling of a survival time and an outcome from exponential family. The survival time is modeled by a semi-parametric proportional hazard model with a subject-specific random effect. The hospitalization related endpoint, such as the number of admissions, length of stay or whether a subject has ever been hospitalized, can be model by a generalized linear mixed effects model. Since the hospitalization outcome may only be observed before death, an offset function will be included in the generalized linear model to take into account the follow-up time. To allow a flexible relationship between the hospitalization endpoint and the follow-up time, I introduce a nonparametric smooth offset function that includes parametric functions, such as logarithm, as special cases. When the offset function is parametric, these models reduce to the standard generalized mixed effects models and parameters of interest may be interpreted in terms of the constant conditional means such as incident rate, mean duration and average probability. The smooth offset function allows deviation from this rigid assumption. The forms of the baseline hazard function and the offset function are usually unknown. They will be modeled non-parametrically using spline functions with non-negative and, when appropriate, monotone constraints. A latent random variable will be used to model potential correlation between survival time and hospitalization outcome from the same subject [42]. I will further discuss the estimation procedures which can be conveniently carried out using existing softwares.

The rest of this chapter is organized as follows. Section 2.2 introduces the semi-parametric latent variable model. Section 2.3 provides details about our estimation procedure. Sections 2.4 and 2.5 present simulation results and applications to patients on HD. The chapter ends with a discussion in Section 2.6.

2.2 The Semi-parametric Latent Variable Model

2.2.1 The Overall Model

For subject i, I denote D_i as the death time, C_i as the censoring time, $T_i = \min\{C_i, D_i\}$ as the observed time, $\Delta_i = I(D_i < C_i)$ as the event indicator and $h_i(t)$ as the hazard function. Let Y_i be another outcome variable from exponential family. Let $\mathbf{Z}_i^D$ and $\mathbf{Z}_i^Y$ be covariates associated with the outcomes D_i and Y_i respectively. Note that $\mathbf{Z}_i^Y$ may include Δ_i. We will consider the following joint model:

$$\begin{aligned} h_i(t) &= h_0(t)\exp(\beta'\mathbf{Z}_i^D + \nu_i), \\ g\left(\mathrm{E}(Y_i|T_i,\nu_i)\right) &= w(T_i) + \alpha'\mathbf{Z}_i^Y + \eta\nu_i, \end{aligned} \tag{2.1}$$

where $h_0(t)$ is the baseline hazard, g is the link function, $\nu_i \overset{iid}{\sim} \mathrm{N}(0,\sigma^2)$ is a shared frailty for subject i, and w is an offset function. The first equation in (2.2) is a Cox proportional hazard model for survival time while the second equation in (2.2) is a generalized linear model for Y_i. The shared frailty is introduced to model heterogeneity among subjects and correlation between D_i and Y_i within a subject. The distribution of the shared frailty is assumed to be normal for simplicity. Extensions to other distributions are straightforward. The offset term $w(T_i)$ is introduced to account for the fact that Y_i is only observed prior to time T_i.

2.2.2 A Spline Model for the Baseline Hazard

The form of the baseline hazard function $h_0(t)$ is generally unknown in practice. We will assume that $h_0(t)$ is a smooth function and model it using B-spline basis functions:

$$h_0(t) = \sum_{k=1}^{K+1+L_h} d_k B_k(t|K_1, \tau_h),$$

where $B_k(t|K, \tau_h)$ denotes the evaluation at t of the K-degree B-spline basis functions generated with interior knots $\tau_h = \{t_{h1}, t_{h2}, \cdots, t_{hL_h}\}$. We will use the constraints $d_k \geq 0$ to enforce the non-negativity constraint of the function $h_0(t)$. The function $h_0(t)$ is decided by coefficients d_k as well as the number and locations of knots. The estimation of coefficients and the selection of knots will be discussed in Section 2.3.

2.2.3 A Spline or Monotone Spline Model for the Offset Function

When Y_i represents counts such as hospital admissions, one possible assumption is that Y_i is generated from a homogeneous Poisson process. Under this assumption and canonical link for Poisson data, the offset function $w(t) = \log(t)$. However in practice Y_i may be generated from a non-homogeneous Poisson process [59]. It is therefore desirable to leave the functional form of w unspecified. Again I model w nonparametrically using B-spline basis functions:

$$w(t) = \sum_{k=1}^{K+1+L_w} c_k B_k(t|K, \tau_w),$$

where $B_k(t|K, \tau_w)$ denotes the evaluation at t of the K-degree B-spline basis functions generated with internal knots $\tau_w = \{t_{w1}, t_{w2}, \cdots, t_{wL_w}\}$.

For Poisson data, it is natural to assume that the expectation of Y_i increases with the observational time T_i. In this case I assume that $w(t)$ is a smooth non-decreasing

function. Ramsay used integrated M-splines to fit a monotone spline [47]. We will adopt a similar approach using integrated B-splines. Specifically, denote integrated B-splines as $I_k(t|K,\tau) = \int_0^t B_k(u|K,\tau)du$ for $k = 1,\ldots,K$. Since B_k's are non-negative, then I_k's provide a set of non-decreasing basis functions. We model w using integrated B-spline basis functions:

$$w(t) = \sum_{k=1}^{K+1+L_w} c_k I_k(t|K,\tau_w) + c,$$

where c is an unknown constant and c_k's are coefficients with constraints $c_k \geq 0$.

2.3 Estimation Method

The full likelihood is

$$L = \prod_{i=1}^{n} \int f(Y_i|T_i,\Delta_i,\nu_i) l_i(T_i,\Delta_i|\nu_i) f_\nu(\nu_i) d\nu_i, \tag{2.2}$$

where n is the total number of subject, $f(Y_i|T_i,\Delta_i,\nu_i)$ is the conditional density of Y_i in the exponential family, $f_\nu(\nu_i)$ is the density function of the latent random variable ν, and

$$l_i(T_i,\Delta_i|\nu_i) = \left\{h_0(T_i)\exp(\beta'\mathbf{Z}_i^D + \nu_i)\right\}^{\Delta_i} \exp\left\{-\int_0^{T_i} h_0(t)\exp(\beta'\mathbf{Z}_i^D + \nu_i)dt\right\}.$$

Our goal is then to obtain parameter estimates by maximizing the likelihood. Since there is no closed form solution, I apply the Newton-Raphson methods to compute parameter estimates numerically. For stability, I apply the Newton-Raphson ridge optimization where a pure Newton step is used when the Hessian is positive definite and when the Newton step successfully increases the value of the likelihood, otherwise a multiple of the identity matrix is added to the Hessian matrix [38]. To calculate the gradient and Hessian matrix, I need to evaluate integrals derived from the likelihood function. The Gaussian

quadrature method is used to approximate these integrals. We estimate random effects ν_i by their empirical Bayes estimators $\hat{\nu}_i$ that maximize $f(y_i|T_i, \Delta_i, \nu_i) l_i(T_i, \Delta_i|\nu_i) f_\nu(\nu_i)$.

Numerically stable implementations of these methods can be obtained from a variety of publicly available softwares [46]. In our simulation and example, I employed SAS procedure `Proc NLMIXED` to perform the computation. `Proc NLMIXED` has an appealing feature which allows a user-specified log likelihood functions with respect to the random effects. See [38] for details on this procedure.

The number and location of knots are fixed in the above discussion. While increasing the number of knots has the capability to model a more flexible function, having too many knots will increase the complexity of the model and result in over-fitting. A data-driven procedure for the selection of number and location of knots is desirable. We allow $h_0(t)$ and $w(t)$ to have different numbers and locations of knots. In practice one may place knots evenly in a range or at equally spaced quantiles of data. We select the numbers of knots by minimizing the following AIC [1]:

$$\mathrm{AIC}(L_h, L_w) = -2 \log L + 2(L_h + L_w + 2K + 2), \tag{2.3}$$

where L_h and L_w are the number of knots for baseline hazard function and offset function respectively.

2.4 Simulations

We generate simulation samples from the following model

$$\begin{aligned} h_i(t|\nu_i) &= h_0(t) \exp(\beta Z_i + \nu_i), \\ \log(\mathrm{E}(Y_i|T_i, \nu_i)) &= w(T_i) + \alpha Z_i + \eta \nu_i, \end{aligned} \tag{2.4}$$

where Z_i's are iid random variables with $P(Z_i = 0) = P(Z_i = 1) = 0.5$, $\nu_i \overset{iid}{\sim} \mathrm{N}(0, 0.5)$, and conditional on T_i and ν_i, Y_i follows a Poisson distribution with mean $\exp(w(T_i)+\alpha Z_i+\eta\nu_i)$. The censoring time $C_i = \min\{E_i, 4\}$ where $E_i \overset{iid}{\sim} \mathrm{Exp}(0.1)$. The true parameters are set to be $(\alpha, \beta, \eta) = (0.5, 0.5, 1)$. We consider two baseline hazard functions, Exponential baseline $h_0(t) = 1/2$ and Weibull baseline $h_0(t) = t/2$, and two offset functions, linear function $w(t) = t/2$ and log function $w(t) = \log(t)$.

The baseline hazard $h_0(t)$ is estimated using cubic B-spline basis functions. The offset function $w(t)$ is estimated using cubic integrated B-spline basis functions under the monotone constraint. Interior knots are equally spaced within the time period $(0, 4]$, and the number of knots for $h_0(t)$ and $w(t)$ range from 2 to 4 respectively. The optimal combination of number of knots is selected by minimizing the AIC (2.3).

Simulation under each setting is repeated 500 times. For the estimation of parameters, I compute bias, mean squared error (MSE) and coverage probability of 95% confidence intervals (CP). The 95% confidence interval is constructed as the MLE plus-minus 1.96 times the standard errors obtained from the variance-covariance matrix. For the estimation of functions $h_0(t)$ and $w(t)$, I compute the integrated mean square error (IMSE)

$$\mathrm{IMSE}(\hat{f}) = \int_0^4 (\hat{f}(t) - f(t))^2 dt$$

for each replicate, where f is either h_0 or w.

Tables 2.1, 2.2, 2.3, 2.4, 2.5 summarize performances of parameter and function estimates under four simulation settings. Overall the proposed estimation procedure performs well: bias and MSE are small, and the coverages of 95% confidence intervals are close to the nominal value. The performances improve as sample size increases.

As an illustration, Figure 2.1 shows the $5th$, $25th$, $50th$, $75th$ and $95th$ best estimates of $\hat{h}_0(t)$ and $\hat{w}(t)$ ordered by the IMSE under the simulation setting when $h_0(t) = t/2$,

Table 2.1: Bias, mean squared error (MSE) and coverage probability of 95% confidence intervals (CP) based on the joint model when $h_0(t) = 1/2$ and $w(t) = t/2$.

$h_0(t) = 1/2$	$w(t) = t/2$	α	β	η	σ^2
$n = 300$	Bias	0.007	0.045	-0.064	0.337
	MSE	0.017	0.037	0.066	0.65
	CP	0.938	0.981	0.809	0.965
$n = 500$	Bias	0.002	0.014	-0.008	0.149
	MSE	0.010	0.022	0.871	0.936
	CP	0.946	0.946	0.871	0.936
$n = 1000$	Bias	0.002	0.008	0.003	0.063
	MSE	0.005	0.01	0.031	0.062
	CP	0.94	0.948	0.916	0.94

Table 2.2: Bias, mean squared error (MSE) and coverage probability of 95% confidence intervals (CP) based on the joint model when $h_0(t) = 1/2$ and $w(t) = \log(t)$.

$h_0(t) = 1/2$	$w(t) = log(t)$	α	β	η	σ^2
$n = 300$	Bias	0.033	0.084	-0.106	0.779
	MSE	0.025	0.064	0.109	2.833
	CP	0.966	0.968	0.774	0.957
$n = 500$	Bias	0.016	0.046	-0.03	0.381
	MSE	0.016	0.030	0.088	0.912
	CP	0.955	0.973	0.842	0.953
$n = 1000$	Bias	0.004	0.017	0.005	0.127
	MSE	0.007	0.011	0.053	0.156
	CP	0.947	0.966	0.890	0.951

Table 2.3: Bias, mean squared error (MSE) and coverage probability of 95% confidence intervals (CP) based on the joint model when $h_0(t) = t/2$ and $w(t) = t/2$.

$h_0(t) = t/2$	$w(t) = t/2$	α	β	η	σ^2
$n = 300$	Bias	-0.003	0.008	0.016	0.056
	MSE	0.011	0.025	0.06	0.075
	CP	0.968	0.963	0.925	0.951
$n = 500$	Bias	-0.006	0.008	0.007	0.044
	MSE	0.007	0.015	0.038	0.052
	CP	0.944	0.962	0.912	0.930
$n = 1000$	Bias	-0.002	0.003	0.011	0.017
	MSE	0.003	0.008	0.022	0.025
	CP	0.950	0.946	0.942	0.928

Table 2.4: Bias, mean squared error (MSE) and coverage probability of 95% confidence intervals (CP) based on the joint model when $h_0(t) = t/2$ and $w(t) = \log(t)$.

$h_0(t) = t/2$	$w(t) = log(t)$	α	β	η	σ^2
$n = 300$	Bias	0.033	0.064	-0.025	0.346
	MSE	0.020	0.064	-0.025	0.346
	CP	0.958	0.973	0.859	0.936
$n = 500$	Bias	0.014	0.036	-0.014	0.227
	MSE	0.013	0.027	0.070	0.386
	CP	0.945	0.955	0.850	0.951
$n = 1000$	Bias	0.009	0.015	-0.009	0.117
	MSE	0.006	0.011	0.040	0.100
	CP	0.954	0.950	0.892	0.942

Table 2.5: Integrated Mean Square Error (IMSE) of the baseline hazard $h_0(t)$ and offset function $w(t)$ fitted by the joint model.

		$h_0(t)$	$w(t)$
$h_0(t) = 1/2$	$n = 300$	0.078	0.079
$w(t) = t/2$	$n = 500$	0.050	0.052
	$n = 1000$	0.027	0.027
$h_0(t) = 1/2$	$n = 300$	0.109	0.151
$w(t) = \log(t)$	$n = 500$	0.063	0.097
	$n = 1000$	0.033	0.052
$h_0(t) = t/2$	$n = 300$	0.665	0.066
$w(t) = t/2$	$n = 500$	0.456	0.043
	$n = 1000$	0.230	0.025
$h_0(t) = t/2$	$n = 300$	0.856	0.165
$w(t) = \log(t)$	$n = 500$	0.662	0.114
	$n = 1000$	0.340	0.057

$w(t) = \log(t)$ and $n = 500$. Overall, the estimates are close to the true functions except for the baseline hazard with large t. The poor estimation of the baseline hazard with large t is likely caused by censoring. Estimation performance of $\hat{h}_0(t)$ and $\hat{w}(t)$ under other scenarios are provided in Appendix A.

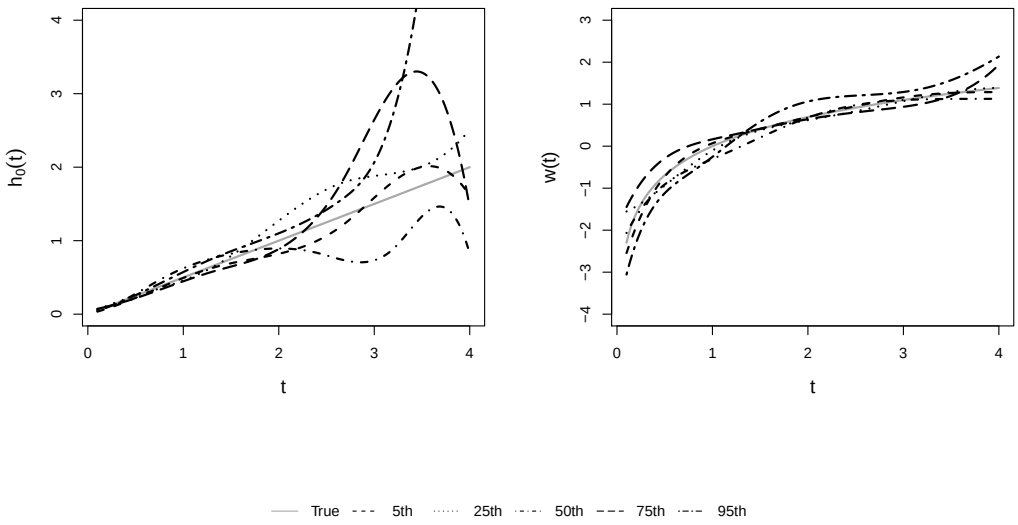


Figure 2.1: True function (solid lines) and estimates (dashed lines) of $h_0(t) = t/2$ (left) and $w(t) = \log(t)$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = t/2$, $w(t) = \log(t)$ and $n = 500$.

2.5 Application

We now apply the proposed method to model mortality and hospitalization outcomes for patients on HD. Baseline covariates are collected from 1999 HD patients from January 1, 2007 to December 31, 2007. Survival time, the number of hospital admissions and total length of stay of these patients during the period of January 1, 2008 and December 31, 2009 are collected. 1078 (53.93%) patients are male. 984 (49.22%) patients are black, 834 (41.72%) patients are white, the rest are from other races.

Table 2.6: Summary statistics of covariates

	(Min, Max)	Mean (Std)
Age (year)	(1.00, 96.62)	62.39 (14.84)
BMI (kg/m^2)	(13.75, 49.51)	27.65 (6.46)
Albumin (g/dL)	(1.60, 4.74)	3.84 (0.37)
IDWG (%)	(0.41, 7.99)	3.48 (1.05)
PreSBP (mmHg)	(81.88, 219.29)	149.38 (18.86)
eKt/V	(0.68, 3.77)	1.46 (0.26)
NLR	(0.51, 31.18)	3.70 (2.32)
Vintage (year)	(0.08, 7.90)	2.56 (1.92)

In previous studies, albumin and systolic blood pressure prior to dialysis (PreSBP) have been found as significant risk factors for mortality [45, 25, 28]. Albumin serves as an indication of nutrition levels. Erdem et al. observed that HD patients with high neutrophil-to-lymphocyte ratio (NLR) levels have increased risk of short term mortality as it has been discovered as an important marker of inflammation in ESRD patients [14]. Our preliminary data analytical results indicate that vintage, inter-dialytic weight gain (IDWG) and eKt/V also have significant effect on mortality. Vintage is recorded as the time in years since the patient initiated dialysis. IDWG is measured at the beginning of dialysis and excessive IDWG is usually related to the overload of sodium and water in the body and indicates a poor residual kidney function. eKt/V is a quantity comparing the level of urea in the blood before and after dialysis, and serves as a measurement of dialysis adequacy. Larger level of eKt/V implies that more waste is cleared from the body through dialysis. Each time-varying covariates is summarized by its mean in baseline period for each patient. In addition, I will include gender, race and BMI as potential risk factors. The summary statistics for these covariates are listed in Table 2.6.

In modeling the hospitalization, the number of hospital admissions is usually the primary outcome which will be studied in Section 2.5.1 using a Poisson model. We are sometimes also interested in whether a patient has ever been hospitalized as a binary

outcome. Since the probability of ever been hospitalized can be derived from the Poisson model, I omit the details of modeling the binary outcome in this chapter. Given the subject has been hospitalized, a further goal is to identify the risk factors that lead to longer total length of stay which will be studied in Section 2.5.2 using a Gamma model. For simplicity I will consider the same set of covariates for all models.

2.5.1 Joint Analysis of Mortality and Hospital Admission

359 (17.96%) patients died during the follow-up period. The number of hospital admissions in the data ranges from 0 to 37 with mean 2.53. We consider the following joint model:

$$\begin{aligned} h_i(t|\nu_i) &= h_0(t)\exp\{\beta_1 * Age_i + \beta_2 * Albumin_i + \beta_3 * PreSBP_i + \beta_4 * NLR_i \\ &\quad + \beta_5 * BMI_i + \beta_6 * Male_i + \beta_7 * IDWG_i + \beta_8 * eKt/V_i \\ &\quad + \beta_9 * Vintage_i + \beta_{10} * RaceWhite_i + \beta_{11} * RaceBlack_i + \nu_i\}, \\ g(E(Y_i|T_i,\nu_i)) &= w(T_i) + \alpha_1 * Age_i + \alpha_2 * Albumin_i + \alpha_3 * PreSBP_i + \alpha_4 * NLR_i \\ &\quad + \alpha_5 * BMI_i + \alpha_6 * Male_i + \alpha_7 * IDWG_i + \alpha_8 * eKt/V_i \\ &\quad + \alpha_9 * Vintage_i + \alpha_{10} * RaceWhite_i + \alpha_{11} * RaceBlack_i + \eta\nu_i, \end{aligned} \tag{2.5}$$

where Y_i represents the number of hospital admissions of patient i and is assumed to follow a Poisson distribution, and $\nu_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$. We use log link $g(.) = log(.)$.

As in the previous section I set the interior knots for baseline hazard and offset function equally spaced within the time period. The number of interior knots ranges from 2 to 4. Among all the combinations, the AIC selects 2 knots for the baseline hazard and 2 knots for the offset function.

I summarize the estimation results in Table 2.7. All covariates except BMI have sig-

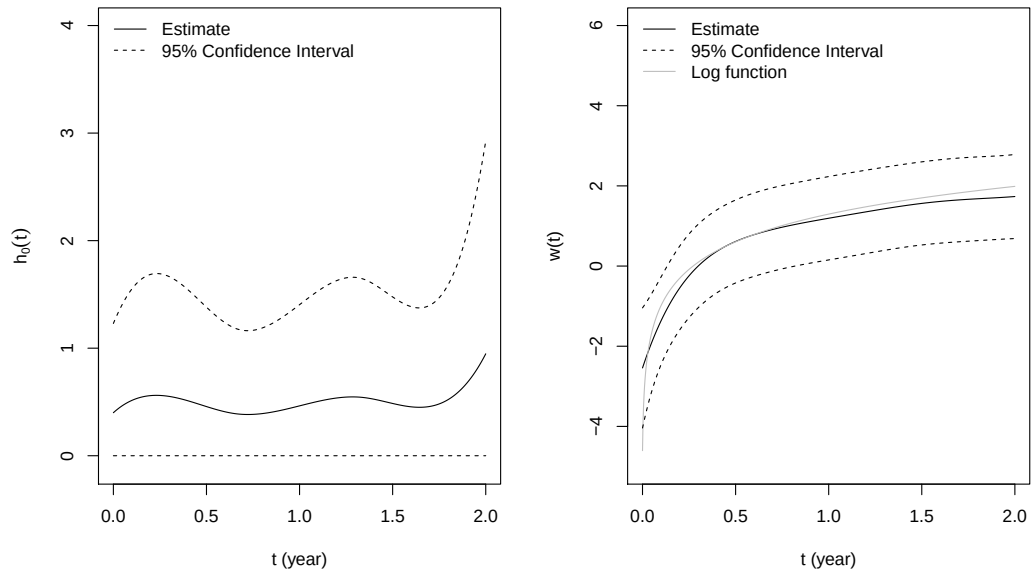


Figure 2.2: The estimated baseline function $h_0(t)$ and offset function $w(t)$ for the joint model of mortality and number of hospitalization.

nificant effect on the expected number of hospital admissions, while age, albumin, NLR, eKt/V and vintage have significant effect on the hazard function. Overall age, NLR and vintage are positively associated with both survival hazard and the number of hospital admissions, while albumin and eKt/V are negatively associated with the outcomes. Furthermore, pre-dialysis SBP and IDWG are positively associated with the number of hospital admissions, and female patients tend to have more hospital admissions. The directions of associations are as expected.

The latent random variable is significant ($\hat{\sigma}^2 = 0.6008$, $p = 0.0057$), which supports the model with random effect. Furthermore $\hat{\eta}$ is significantly larger than 0 ($p < 0.0001$). It implies that the survival time and the number of hospital admissions are positive correlated. The estimated baseline function $h_0(t)$ and offset function $w(t)$ are shown in Figure 3.2. While our model allows for inhomogeneous Poison model, the logarithm function is close to the estimated offset function and well within the 95% confidence

Table 2.7: Analysis of ESRD data

	Covariates	Estimate	SE	p-value
Mortality	Age	0.0355	0.0048	< 0.0001
	Albumin	-1.2736	0.1681	< 0.0001
	PreSBP	-0.0004	0.0031	0.8990
	NLR	0.1061	0.0217	< 0.0001
	BMI	-0.0198	0.0106	0.0619
	Male	0.0748	0.1210	0.5365
	IDWG	0.0684	0.0625	0.2737
	eKt/V	-0.5936	0.2474	0.0165
	Vintage	0.1244	0.0307	< 0.0001
	Race(White)	0.1341	0.2151	0.5329
	Race(Black)	-0.2446	0.2184	0.2629
Hospitalization	Age	0.0089	0.0022	< 0.0001
	Albumin	-0.8126	0.0856	< 0.0001
	PreSBP	0.0072	0.0015	< 0.0001
	NLR	0.0776	0.0129	< 0.0001
	BMI	-0.0026	0.0049	0.5974
	Male	-0.1612	0.0600	0.0073
	IDWG	0.1018	0.0307	0.0009
	eKt/V	-0.2360	0.1170	0.0437
	Vintage	0.0386	0.0157	0.0140
	Race(White)	0.2634	0.1094	0.0162
	Race(Black)	0.3130	0.1069	0.0035
	σ^2	0.6008	0.2172	0.0057
	η	1.2225	0.2039	< 0.0001

intervals, suggesting that it is reasonable to model the offset function by the logarithm function in this case.

2.5.2 Joint Analysis of Mortality and Total Length of Stay

To further investigate the features of patients with hospitalizations, another interesting application is to model mortality and total length of hospital stay. We will focus on the patients who had non-zero length of stays (1396 patients). The total length of stay

ranges from 1 to 368 with mean 26.13. We consider the following joint model:

$$\begin{aligned} h_i(t|\nu_i) &= h_0(t)\exp\{\beta_1 * Age_i + \beta_2 * Albumin_i + \beta_3 * PreSBP_i + \beta_4 * NLR_i \\ &\quad + \beta_5 * BMI_i + \beta_6 * Male_i + \beta_7 * IDWG_i + \beta_8 * eKt/V_i \\ &\quad + \beta_9 * Vintage_i + \beta_{10} * RaceWhite_i + \beta_{11} * RaceBlack_i + \nu_i\}, \\ g(E(Y_i|T_i,\nu_i)) &= w(T_i) + \alpha_1 * Age_i + \alpha_2 * Albumin_i + \alpha_3 * PreSBP_i + \alpha_4 * NLR_i \\ &\quad + \alpha_5 * BMI_i + \alpha_6 * Male_i + \alpha_7 * IDWG_i + \alpha_8 * eKt/V_i \\ &\quad + \alpha_9 * Vintage_i + \alpha_{10} * RaceWhite_i + \alpha_{11} * RaceBlack_i + \eta\nu_i, \end{aligned} \tag{2.6}$$

where Y_i represents the total length of stay of patient i and is assumed to follow a Gamma distribution, and $\nu_i \overset{iid}{\sim} \mathrm{N}(0,\sigma^2)$. We use natural log link function $g(.) = log(.)$.

Similar process for knots selection applies, which results in 2 knots for the baseline hazard and 2 knots for the offset function. The estimation results are summarized in Table 2.8. All covariates except BMI and vintage have significant effect on the expectation of total length of stay, while age, albumin, NLR and vintage have significant effect on the hazard function. We note that the results of this subsection are consistent with those in the previous subsection. The latent random variable is borderline significant ($\hat{\sigma}^2 = 0.2108$, $p = 0.0542$). The estimated baseline function $h_0(t)$ and offset function $w(t)$ are shown in Figure 2.3.

2.6 Discussion

In this chapter, we propose a semi-parametric joint model for survival time and hospitalization. In particular, we consider the number of hospital admissions and total length of stay as hospitalization outcomes. A shared random effect is introduced to accommodate heterogeneity among subjects. The baseline hazard and offset functions

Table 2.8: Analysis of ESRD data

	Covariates	Estimate	SE	p-value
Mortality	Age	0.0302	0.0053	< 0.0001
	Albumin	-1.0237	0.1730	< 0.0001
	PreSBP	-0.0040	0.0035	0.2472
	NLR	0.0751	0.0228	0.0010
	BMI	-0.0204	0.0118	0.0843
	Male	0.0830	0.1340	0.5359
	IDWG	0.0939	0.0704	0.1826
	eKt/V	-0.4920	0.0704	0.0812
	Vintage	0.0944	0.0332	0.0045
	Race(White)	-0.0361	0.2303	0.8754
	Race(Black)	-0.3373	0.2335	0.1489
Length of Stay	Age	0.0070	0.0022	0.0017
	Albumin	-0.5335	0.0870	< 0.0001
	PreSBP	0.0047	0.0016	0.0036
	NLR	0.0484	0.0137	0.0004
	BMI	-0.0045	0.0051	0.3788
	Male	-0.1269	0.0626	0.0430
	IDWG	0.0740	0.0319	0.0205
	eKt/V	-0.2495	0.1234	0.0433
	Vintage	0.0275	0.0165	0.0955
	Race(White)	0.1338	0.1123	0.2336
	Race(Black)	0.2933	0.1110	0.0084
	σ^2	0.2108	0.1094	0.0542
	η	1.8883	0.5550	0.0007

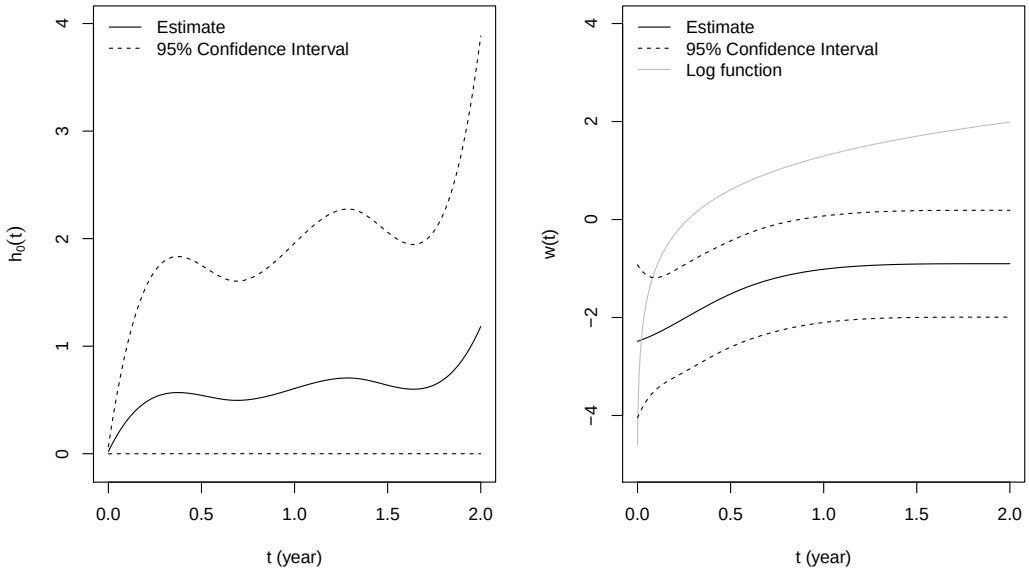


Figure 2.3: The estimated baseline function $h_0(t)$ and offset function $w(t)$ for the joint model of mortality and total length of stay.

are modeled non-parametrically through B-spline or monotone B-spline bases in order to gain flexibility. With fixed number of knots, the techniques to numerically obtain maximum likelihood estimation are presented. We have also discussed the AIC method for selecting the number of knots. Standard large sample properties of maximum likelihood estimation apply when knots are fixed. Simulation results indicate that the proposed estimation method performs well.

Throughout this chapter, we assume Normal distribution for the random effect. Our method can be easily generalized to other parametric distributions for the random effect. We have analyzed the different aspects of the hospitalization separately. One future research is to build a joint model for survival time, hospital admission and length of stay. Zero-inflated Poisson model can also be incorporated to account for the fact that a large number of zeros. It is worth noting here that even though we focus on the joint modeling of survival time and hospitalization in this research, the model and estimation method we

proposed is in fact general in theory. It can accommodate the joint modeling of survival time and any outcome which can be reasonably assumes to follow a distribution from Exponential family.

Chapter 3

Joint Modeling of Hospitalization and Readmission

3.1 Introduction and Related Work

Hospitalization and readmission are increasing focuses for medical researchers and policy makers. Readmission refers to a patient being admitted to a hospital within a certain time period from an initial admission. Medicare uses 30 days as the time period. If a hospital has a high proportion of patients readmitted within a short time frame, it may be an indication of inadequate quality of care in the hospital or a lack of appropriate coordination of postdischarge care.

Hospital admissions and readmissions are expensive. A variety of intervention programs have been implemented to reduce hospital readmission rates. Hansen et al. provided an excellent systematic review of 43 reports and articles of such intervention programs published between January 1975 and January 2011 [23].

Objective methods are required to evaluate the effectiveness of intervention programs. Traditional evaluation methods focus on the readmission after one initial hospitalization

during a specified observational period. Logistic regression or Cox proportional hazard models are commonly used to model the rate and time of readmission [54, 24, 53]. However, one patient may experience multiple initial hospitalizations and/or multiple readmissions. Focusing only on one initial occurrence of hospitalization and/or one readmission does not take full advantage of whole data.

Considering all hospitalizations during the period brings two challenges: 1) outcomes are correlated among the same subject; and 2) number of readmissions is naturally bounded by the number of total hospitalizations. A joint modeling approach is needed to consider these two outcomes simultaneously.

This research is motivated by the need of evaluation of an intervention program implemented by Fresenius Medical Care for ESRD patients who are on hemodialysis (HD). The intervention program is a combination of targeted interventions. This program involves a post-hospital care management system, including pre- and post-hospital checking lists, regular follow-up on patients' post-hospital needs and adherence to the discharge instructions, and a centralized approach for transfer of patient among providers. It is of interest to evaluate the program performance in reducing number of hospitalizations and readmissions.

The aim of this research is to develop a joint modeling approach for the analysis of hospitalization and readmission. The rest of this chapter is organized as following: Section 3.2 introduces the statistical model for the joint analysis. Section 3.3 discusses the estimation procedure using SAS Proc NLMIXED and the selection of number of knots. Systematic simulation results are demonstrated in Section 3.4. The proposed joint modeling framework is applied to evaluate the performance of the intervention program from Fresenius Medical Care in reducing number of hospitalizations and readmissions. The results are presented in Section 3.5. Finally this chapter ends with a conclusion in Section 3.6.

3.2 Model Formulation

For each patient, we observe numbers of hospital admissions and readmissions as two outcomes. Let Y_{ij}, Z_{ij} and $\mathbf{X}_{ij}$ be the number of hospitalizations, the number of hospital readmissions and a vector of covariates for patient i at time period j, $i = 1, \cdots, n$, $j = 1, \cdots, n_i$.

We begin by modeling the number of hospitalizations with log linear model with random intercept. We assume that given covariates $\mathbf{X}_{ij}^Y$, Y_{ij} follows Poisson distribution with mean λ_{ij} and

$$\begin{aligned} Y_{ij}|b_i &\sim Poisson(\lambda_{ij}), \\ \log(\lambda_{ij}) &= \log(T_{ij}) + \alpha_0 + \boldsymbol{\alpha}\mathbf{X}_{ij}^Y + b_i, \end{aligned} \tag{3.1}$$

where T_{ij} is the exposure time for patient i in period j and b_i is the random intercept. We assume that $b_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$.

We then treat number of hospital readmission Z_{ij} as a Binomial random variable with total number of trials Y_{ij} and probability p_{ij}, and model p_{ij} with a semi-parametric logistic regression model with random intercept. Specifically, we assume that

$$\begin{aligned} Z_{ij}|Y_{ij}, \mathbf{X}_{ij}^Z &\sim Binom(Y_{ij}, p_{ij}), \\ logit(p_{ij}) &= f(Y_{ij}) + \boldsymbol{\gamma}\mathbf{X}_{ij}^Z + \eta b_i. \end{aligned} \tag{3.2}$$

The random effect b_i is shared by (3.1) and (3.2) to accommodate the correlation between the observations from the same subjuect i. η is a scale parameter, which allows the random effect b_i to have different impact on each of the two responses. $\boldsymbol{X}_{ij}^Y$ and $\boldsymbol{X}_{ij}^Z$ are part of the vector $\boldsymbol{X}_{ij}$ and they do not need to be the same.

The probability p_{ij} may depend on Y_{ij}. We introduce function $f(Y_{ij})$ into the model to account for the potential association. The form of $f(Y_{ij})$ is often unknown in practice and we will model it non-parametrically.

We assume that $f(x)$ is a smooth function and propose to use a natural cubic spline to model it. Natural cubic spline is piecewise cubic polynomials joined at pre-determined knots where first-order and second-order derivatives of the function are constrained to be continuous. In addition it requires the evaluations of second order derivative at boundary points to equal 0. The spline function is then modeled by a linear combination of B-spline basis functions

$$f(x) = \sum_{k=1}^{h} d_k B_k(x|\tau),$$

where $B_k(x|\tau)$ denote the evaluation at x of the kth B-spline basis functions, generated with internal knots $\tau = \{t_1, t_2, \cdots, t_h\}$. The function $f(x)$ is decided by coefficients d_k as well as the number and locations of knots.

3.3 Estimation

For a fixed number and location of knots, the likelihood function can be obtained by integrating out the latent random variable with respect to its density. We define $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{in_i})$ and $\mathbf{Z}_i = (Z_{i1}, \cdots, Z_{in_i})$, the likelihood function for subject i is

$$l_i(\Theta|\mathbf{Y}_i, \mathbf{Z}_i) = \int f(\mathbf{Y}_i, \mathbf{Z}_i|b_i) f_b(b_i) \mathtt{d}b_i,$$

where $\Theta = (\alpha_0, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \eta, \sigma^2, \mathbf{d})$ is the parameter vector and $f_b(b_i)$ is the density function for the latent random variable b_i.

We can further express $f(\mathbf{Y}_i, \mathbf{Z}_i|b_i)$ as the product of conditional densities:

$$\begin{aligned} f(\mathbf{Y}_i, \mathbf{Z}_i|b_i) &= f_Y(\mathbf{Y}_i|b_i) f_Z(\mathbf{Z}_i|\mathbf{Y}_i, b_i) \\ &\propto \prod_{j=1}^{n_i} \lambda_{ij}^{Y_{ij}} \exp(-\lambda_{ij}) \prod_{j=1}^{n_i} p_{ij}^{Z_{ij}} (1-p_{ij})^{Y_{ij}-Z_{ij}}. \end{aligned}$$

We defind $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_n)$, $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_n)$, the full log likelihood is:

$$l(\Theta|\mathbf{Y}, \mathbf{Z}) \propto \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} \lambda_{ij}^{Y_{ij}} \exp(-\lambda_{ij}) \prod_{j=1}^{n_i} p_{ij}^{Z_{ij}} (1-p_{ij})^{Y_{ij}-Z_{ij}} f_b(b_i) \mathrm{d}b_i. \tag{3.3}$$

Then it is our goal to obtain parameter estimates by maximizing the likelihood. I adopt a similar computational approach as in Chapter 2 using SAS Proc NLMIXED for its flexibility to provide a user-specific likelihood function with respect to random effects. This procedure approximates the integral numerically using Gaussian quadrature method and finds maximizers using the Newton-Raphson methods. A Newton-Raphson ridge optimization is employed here for stable results [38]. Random effect $\hat{\nu}_i$ is estimated by the empirical Bayes estimators $\hat{\nu}_i$ that maximize $f(\mathbf{Y}_i, \mathbf{Z}_i|b_i) f_b(b_i)$.

To select the number and location of knots, we propose a data-driven method to strike the balance between goodness-of-fit and model complexity. Suppose we have h knots $\tau_1, \tau_2, \cdots, \tau_h$, where τ_i is placed at the $(i-1)/(h-1)$th percentile of the observed Y_{ij}. τ_1 and τ_h are the minimum and maximum of Y_{ij}. We then select the optimal h by minimizing the AIC [1]:

$$\mathrm{AIC}(h) = -2 \log L + 2h.$$

3.4 Simulations

We set $n_i = 1$ or 2. We randomly assigned subjects such that they are observed only in 1 period with probability 0.4 and in 2 periods with probability 0.6. Subjects are randomly assigned to two treatment groups with same probability.

We generate the simulation sample from the following model:

$$\begin{aligned} Y_{ij}|b_i &\sim Poisson(\lambda_{ij}), \\ \log(\lambda_{ij}) &= \alpha_0 + \alpha_1 * group_i + \alpha_2 * period_{ij} + b_i, \\ Z_{ij}|Y_{ij}, b_i &\sim Binom(Y_{ij}, p_{ij}), \\ logit(p_{ij}) &= f(Y_{ij}) + \gamma_1 * group_i + \gamma_2 * period_{ij} + \eta b_i, \\ b_i &\overset{i.i.d.}{\sim} N(0, \sigma^2). \end{aligned} \tag{3.4}$$

We consider two situations for the function $f(x)$: linear and quadratic. Knots are placed at percentiles of the values of Y_{ij}'s in the sample. The number of interior knots are selected as the minimizer of AIC among three values: 3, 4 and 5. Simulation results are based on 500 replicates in each scenario.

As a comparison, we have also conducted the simulation by separate modeling (SM) where there is no shared random variable b_i, and Y_{ij} and Z_{ij} are modeled marginally with a log-linear model and a logistic regression respectively. The estimation results for parameters are listed in Tables 3.1 and 3.2. Our joint model (JM) has small bias and MSE, and the coverage probability is closed to nominal value. Moreover, the performance of estimation improves when we increase sample size. The estimates from SM for some parameters are severely biased, especially $\alpha_0, \gamma_1, \gamma_2$. The coverage probabilities generally are poor for SM and increasing sample size does not improve the performance. The simulation results verifies the necessity of modeling two responses jointly when they are correlated.

We investigate the performance of curve fitting by their sum weighted mean square error (SWMSE) defined as

$$\text{SWMSE}(\hat{f}) = \sum_{i=1}^{N} w_i(\hat{f}(x_i) - f(x_i))^2, \tag{3.5}$$

Table 3.1: Mean squared error (MSE), bias and coverage of 95% confidence intervals (CP) based on the joint model (JM) and the separate model (SM).

			JM			SM		
$f(x) = 0.5 + 0.1x$	Parameter	True	Bias	MSE	CP	Bias	MSE	CP
$n = 400$	α_0	0.5	.000	.004	.952	.123	.019	.338
	α_1	-0.2	-.004	.006	.966	-.004	.006	.896
	α_2	-0.2	-.002	.004	.964	-.002	.005	.952
	γ_1	-0.2	.000	.026	.942	.064	.027	.906
	γ_2	-0.2	-.005	.023	.954	.058	.025	.942
	η	1	-.012	.130	.966	-	-	-
	σ^2	0.25	-.004	.003	.934	-	-	-
$n = 600$	α_0	0.5	.000	.003	.938	.124	.018	.174
	α_1	-0.2	-.003	.004	.960	-.003	.004	.890
	α_2	-0.2	-.002	.003	.950	-.002	.003	.940
	γ_1	-0.2	-.002	.017	.946	.064	.019	.910
	γ_2	-0.2	-.003	.018	.930	.062	.021	.896
	η	1	.005	.095	.960	-	-	-
	σ^2	0.25	-.001	.002	.946	-	-	-
$n = 800$	α_0	0.5	.000	.002	.948	.126	.018	.086
	α_1	-0.2	-.003	.003	.952	-.003	.003	.864
	α_2	-0.2	-.001	.003	.942	-.001	.003	.928
	γ_1	-0.2	-.003	.013	.948	.063	.015	.900
	γ_2	-0.2	-.004	.012	.952	.059	.014	.928
	η	1	-.007	.073	.948	-	-	-
	σ^2	0.25	.001	.001	.938	-	-	-

where $x_1, \ldots, x_N$ are integers from 0 to the maximum value of Y_{ij} in the sample.

Here we use weight w_i in Equation (3.5) proportional to the observed frequency of x_i in the data. Figure 3.4 provides the $5th$, $25th$, $50th$, $75th$ and $95th$ best estimates of $\hat{f}(x)$ ordered by SWMSE when the true function is linear and quadratic respectively. Overall, the estimates are close to the true functions when sample size is median or large.

Table 3.2: Mean squared error (MSE), bias and coverage of 95% confidence intervals (CP) based on the joint model.

			JM			SM		
$f(x) = 0.02(x-5)^2$	Parameter	True	Bias	MSE	CP	Bias	MSE	CP
$n = 400$	α_0	0.5	.002	.004	.942	.126	.020	.290
	α_1	-0.2	-.004	.007	.954	-.004	.007	.854
	α_2	-0.2	-.001	.071	.944	-.002	.005	.934
	γ_1	-0.2	.002	.023	.936	.068	.025	.884
	γ_2	-0.2	.001	.021	.954	.065	.023	.934
	η	1	-.040	.134	.964	-	-	-
	σ^2	0.25	-.004	.002	.936	-	-	-
$n = 600$	α_0	0.50	.001	.003	.938	.124	.018	.174
	α_1	-0.20	-.001	.005	.936	-.001	.005	.840
	α_2	-0.20	.001	.003	.964	.001	.003	.952
	γ_1	-0.20	.000	.016	.934	.067	.019	.866
	γ_2	-0.20	-.003	.014	.952	.062	.017	.920
	η	1	-.025	.082	.950	-	-	-
	σ^2	0.25	-.004	.001	.946	-	-	-
$n = 800$	α_0	0.50	.000	.002	.940	.124	.017	.104
	α_1	-0.20	.002	.003	.964	.001	.004	.868
	α_2	-0.20	.000	.002	.972	.000	.002	.954
	γ_1	-0.20	.007	.012	.934	.074	.015	.848
	γ_2	-0.20	-.003	.010	.950	.063	.013	.884
	η	1	-.017	.058	.966	-	-	-
	σ^2	0.25	-.002	.001	.962	-	-	-

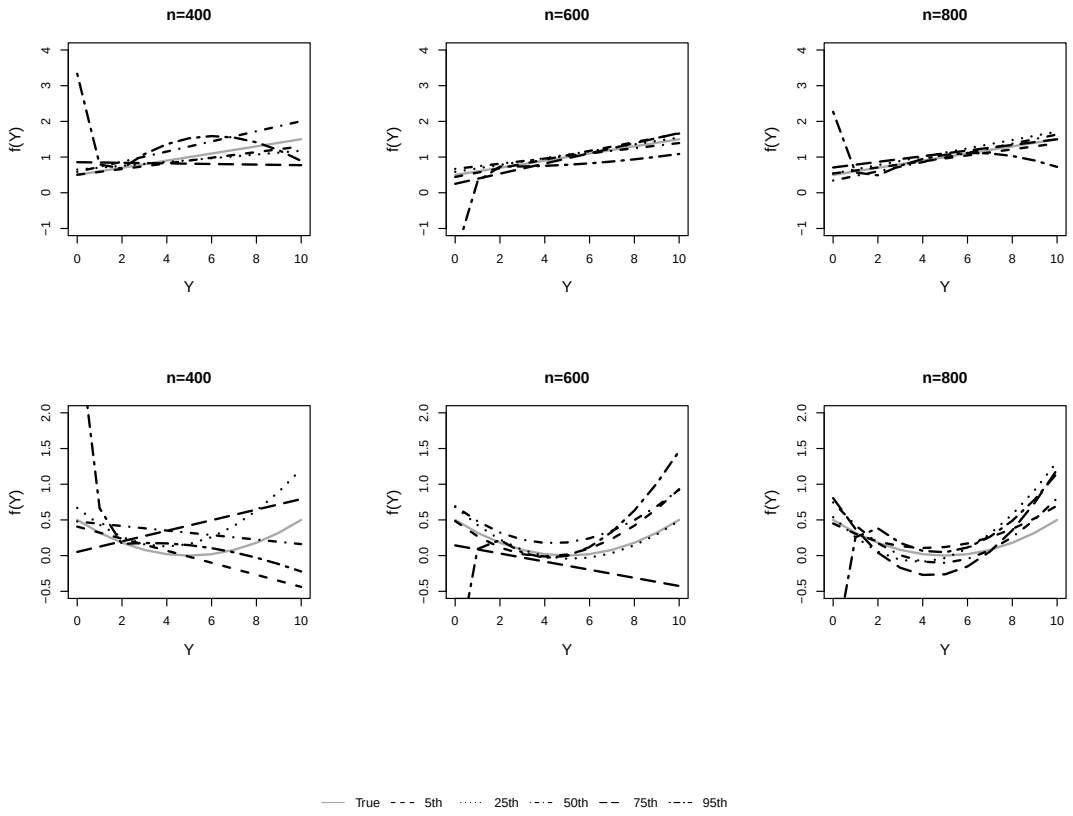


Figure 3.1: True function (solid lines) and estimates (dashed lines) of $f(Y)$ (left) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the SWMSE when $f(Y)$ is linear (top) or quadratic (bottom).

3.5 Application

We apply our method to an observational study of ESRD patients. The data come from Fresenius Medical Care. There are two treatment groups: intervention group and control group. The intervention group includes 26 clinics in West Virginia, and the control group includes 18 clinics from other places. Clinics in the control group were matched to those in the intervention group based on Urban vs. rural, clinic size, hospital admission rate and percent of readmissions within 30 days. We have 1730 and 1400 patients in each group respectively. The intervention was applied to patients in 2013. Each patient is expected to be observed in two periods: 1) Before treatment: 12 months ending 12/31/2012 and 2) After treatment: 12 months ending 12/31/2014. However some patients died or lost-to-follow in the After period. In the intervention group, there are 889 patients who have both observations and 841 patients who only have the observation in Before period. In the control group, there are 749 patients who have both observations and 651 patients who only have the observation in Before period. The number of hospitalization admissions and readmissions are recorded. Summaries are provided in Table 3.3. Other covariates include age, vintage, gender, and comorbidities diabetes, congestive heart failure (CHF) and chronic obstructive pulmonary disease (COPD). Ages of the patients in our data range from 2.4 to 98.8 years with mean 61.4. Vintage is the number of years since the patient started dialysis, it ranges from 0.002 to 28.613 years with mean 2.497. 56.26% of patients are male. The proportions of diabetes, CHF and COPD patients in our data are 68.56%, 18.18% and 9.11% respectively.

The motivation of this analysis is to investigate the effect of intervention program on both the hospitalization and readmission rate. Since hospitalization will naturally change over time with patients' disease progression, as illustrated in Table 3.3, our main goal is to test whether intervention changes the difference between the outcomes in Before

Table 3.3: Mean (SE) of the number of hospitalizations and readmissions.

		Before	After
Hospitalization	Intervention	1.82 (2.23)	1.50 (2.12)
	Control	1.57 (2.20)	1.46 (1.98)
Readmission	Intervention	0.65 (1.41)	0.49 (1.32)
	Control	0.55 (1.45)	0.45 (1.20)

period and After period. The model is fitted as

$$
\begin{aligned}
Y_{ij}|\mathbf{X}_{ij} &\sim Poisson(\lambda_{ij}),\\
\log(\lambda_{ij}) &= \log(Exp_{ij}) + \alpha_0 + \alpha_1 Trt_i + \alpha_2 Prd_{ij} + \alpha_3(Trt_i \times Prd_{ij}) + \alpha_4 Age_{ij}\\
&\quad + \alpha_5 Vintage_{ij} + \alpha_6 Male_i + \alpha_7 Diabetic_i + \alpha_8 CHF_i + \alpha_9 COPD_i + b_i,\\
Z_{ij}|Y_{ij}, \mathbf{X}_{ij} &\sim Binom(Y_{ij}, p_{ij}),\\
logit(p_{ij}) &= f(Y_{ij}) + \gamma_1 Trt_i + \gamma_2 Prd_{ij} + \gamma_3(Trt_i \times Prd_{ij}) + \gamma_4 Age_{ij}\\
&\quad + \gamma_5 Vintage_{ij} + \gamma_6 Male_i + \gamma_7 Diabetic_i + \gamma_8 CHF_i + \gamma_9 COPD_i + \eta b_i,\\
b_i &\overset{i.i.d.}{\sim} N(0, \sigma^2),
\end{aligned}
$$

where we define $Trt_i = 1$ if patient i is assigned to treatment group and 0 otherwise. $Prd_{ij} = 0$ if patient i is observed in Before period, and $Prd_{ij} = 1$ if in After period, $j = 1$ or 2 respectively. Exp_{ij} is the exposure time of patient i in period j. To test the impact of intervention on hospitalization and readmission, it is equivalent to test $H_0 : \alpha_3 = 0$ and $H_0 : \gamma_3 = 0$ respectively. Parameter estimates are listed in Table 2.7. We select the number of knots as the minimizer of AIC among three values, 2, 3 and 4.

From Table 3.4, we can conclude that intervention has significantly reduced the number of hospitalizations ($\hat{\alpha}_3 = -0.218$, p-value $= 0.016$), while its impact on readmission rate is not significant. Age, diabetes, COPD are positively associated with number of hospitalizations. Vintage, male are negatively associated with the number of hospital-

Table 3.4: Parameter estimation for the ESRD data.

	Hospitalization			Readmission		
	Estimation	Error	p-value	Estimation	Error	p-value
Intercept	-0.053	0.115	0.647			
Treatment	0.177	0.053	0.001	0.058	0.078	0.458
Period	-0.216	0.068	0.002	-0.136	0.107	0.205
Trt*Prd	-0.218	0.090	0.016	-0.010	0.140	0.942
Age	0.004	0.002	0.003	0.006	0.002	0.012
Vintage	-0.020	0.007	0.007	-0.015	0.011	0.188
Male	-0.094	0.042	0.027	0.028	0.064	0.698
Diabetic	0.283	0.048	< 0.001	0.058	0.076	0.443
CHF	0.004	0.053	0.937	-0.126	0.082	0.125
COPD	0.192	0.068	0.005	0.081	0.101	0.425
σ^2	1.128	0.047	< 0.001			
η	0.812	0.063	< 0.001			

izations. Only age is positively associated with readmission rate.

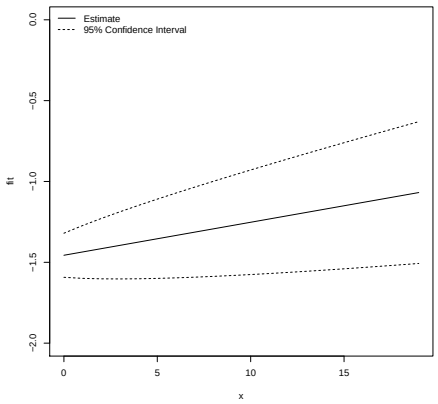
The estimates of both $\hat{\sigma}^2$ and $\hat{\eta}$ are significantly different from 0 at significance level 0.05, which suggests that the correlation among outcomes from the same patient is non-ignorable. Figure 3.2 shows the fitted curve $\hat{f}(x)$ and its 95% pointwise confidence interval. We can observe that it is an increasing linear function.

3.6 Conclusion

The methods proposed in this chapter fill the gap of statistical methodologies on evaluating intervention effect on both hospitalization and readmission rate. A joint modeling approach is proposed treating the number of hospitalizations as observations from Poisson distribution and number of readmissions as observations from Binomial distribution with the total number of trials being the number of hospitalizations. The means of the Poisson distribution and Binomial distribution are linked by shared random effects.

The estimation procedure can be conveniently carried out by SAS Proc NLMIXED.

Figure 3.2: Fitted $\hat{f}(x)$ in ESRD data and its 95% pointwise confidence interval.



The simulation demonstrates good estimation properties. Furthermore, by comparison of the estimation between joint analysis and separate modeling, we found that separate modeling neglecting the possible correlation introduces biases on the treatment effects.

The proposed joint modeling framework is applied to a real world example to examine the effect of an intervention program on reducing number of hospitalizations and admissions. Data analytical results show that this program effectively reduces the number of hospitalizations, however its effect on reducing rate of readmission is not significant comparing with the control group.

It may be noticed that in the application, only a subset of patients are observed in both before intervention and after intervention period. This is caused by the death or censoring of patients during the time period in-between. As mortality is another important measure of treatment effect and is underlying associated with patients' health condition, it is desirable to include it as one of the outcomes and model it with hospitalization and readmission jointly. This is one potential direction of our future research.

We considered all-cause readmission. Planned and unplanned readmissions are mixedly observed and unbraiding the two provides different interpretation of statistical analysis. However it is beyond the scope of this research. See Kossovsky et al. about distinctions

between planned and unplanned readmissions [36].

Chapter 4

Variable Selection in Joint Modeling of Multivariate Longitudinal and Survival Data

4.1 Introduction and Literature Review

Joint modeling of longitudinal measurements and survival data has attracted a great deal of attention. Both longitudinal measurements and survival time may be driven by an underlying biological process which governs the health condition of patients. The complicated features in the data structure brings challenges in statistical modeling. Previous studies have shown that a naive combination of longitudinal and survival time data analyses neglecting the dependence structure will lead to bias [40, 62]. Modeling the longitudinal measurements and survival time jointly reduces biases and provides improvements of efficiency in the estimates [32].

The joint modeling framework links a submodel for longitudinal processes and a submodel for survival time. Cox proportional hazard model for the survival time has

been widely adopted in literature. Standard approaches for the longitudinal process are mixed effect models [62, 9, 17, 56, 31]. These models assume that true longitudinal process (without measurement error) changes smoothly, possibly depending on time-independent covariates and subject-specific random effects. Excellent reviews were given by Tsiatis and Davidian [58] and Rizopoulos [48]. Some other authors introduced a stochastic process as time varying frailty which can be regarded as biological fluctuations in the process as disease progresses (eg, Wang and Taylor discussed integrated OU process [61], Henderson et al. suggested stationary Gaussian process [27]). To allow for more flexibility in modeling of longitudinal measurements, Tsiatis and Davidian specified nonparametric distributions for the random effects [57] while Brown, Ibrahim and DeGruttola proposed a multivariate B-Spline model for the longitudinal biomarkers [4].

Variable selection is an essential topic in statistical modeling, for improved inference and interpretation. Many classical methods such as Akaike information criterion (AIC) [1], Bayesian information criterion (BIC) [51], Mallow's C_p [41] and risk inflation criterion (RIC) [19] have been developed to compare different models. These methods are computationally intensive when the number of models is large. Regularization methods are widely used for variable selection for high dimensional data [5]. Tibshirani proposed the famous LASSO method for fixed-effect selection in linear models [55]. Other types of penalties, including SCAD [16], elastic-net [65] and adaptive LASSO [64], are well studied. Bondell et al. [2] and Ibrahim et al. [33] studied the selection of fixed and random effects in mixed effect models. Fan et al. provided a systematic review on the variable selection methods for survival analysis and proposed a unified nonconcave penalized likelihood approach [15]. However research of variable selection in the joint modeling framework is still limited.

Recently He et al. [26] proposed a variable selection method for the fixed effects and random effects in the joint modeling of one longitudinal biomarker and the survival time

. In many situations there does not exist a single biomarker and multiple variables are observed longitudinally. For example, ESRD patients suffer from many comorbidities. The reduced kidney function is accompanied by numerous metabolic changes which affect almost all organ systems of the human boby. Therefore multiple biological changes of the patients need to be monitored.

The goal of this research is to develop a joint modeling framework linking multivariate longitudinal processes and survival time. A penalized likelihood approach is proposed to perform variable selection and estimation simultaneously. The rest of this chapter is organized as follows. Section 4.2 introduces the joint model. Section 4.3 proposes the penalized likelihood estimation method. Section 4.4 discusses details about our estimation procedure. Sections 4.5 and 4.6 present simulation results and applications to patients on HD. The chapter ends with a discussion in Section 4.7.

4.2 Model Formulation

For each subject i, $i = 1, \cdots, n$, we observe J longitudinal outcomes $\boldsymbol{y}_{ij}$, $j = 1, \cdots, J$, at time points t_{ijk}, $k = 1, \cdots, K_{ij}$. The observation time points may be different for each longitudinal outcome and for each subject. Denote D_i as the death time of subject i which is subject to right censoring at censoring time C_i. We denote $T_i = \min\{C_i, D_i\}$ as the observed time, and $\Delta_i = I(D_i < C_i)$ as the death event indicator.

For longitudinal outcomes, consider the following multivariate linear mixed effect model:

$$y_{ij}(t_{ijk}) = \boldsymbol{X}_i'\boldsymbol{\alpha}_j + \beta_{j1} + b_{ij1} + (\beta_{j2} + b_{ij2})t_{ijk} + \epsilon_{ijk}, \tag{4.1}$$

where $\boldsymbol{X}_i'$ is the vector of fixed covariates for the longitudinal outcomes of subject i, β_{j1} and β_{j2} are population mean intercept and slope for the jth longitudinal outcome

respectively, b_{ij1} and b_{ij2} are the subject-specific random intercept and slope for subject i respectively and ϵ_{ijk} are random errors. We denote $\boldsymbol{b}_i = (b_{i11}, b_{i12}, b_{i21}, b_{i22}, \cdots, b_{iJ1}, b_{iJ2})'$ as the vector of random effects of subject i and assume $\boldsymbol{b}_i$ follows a $2J$-dimension multivariate Normal distribution, $\boldsymbol{b}_i \overset{iid}{\sim} \mathrm{N}(\mathbf{0}, \Sigma)$. We assume that $\epsilon_{ijk} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$ and furthermore they are independent of subject-specific random effects $\boldsymbol{b}_i$ for all i, j, k.

For the survival outcome, we consider the Cox proportional hazard model:

$$h_i(t) = h_0(t) \exp\{\boldsymbol{X}_i'\boldsymbol{\gamma}_0 + \boldsymbol{b}_i'\boldsymbol{\gamma}\}, \tag{4.2}$$

where $h_0(t)$ is the baseline hazard function and $\boldsymbol{X}_i'$ is the vector of fixed covariates. Subject-specific random intercepts and slopes for longitudinal covariates $\boldsymbol{b}_i$ are also included as features of the longitudinal trajectories. Therefore $\boldsymbol{b}_i$ is the key component in our joint modeling framework since it is responsible for not only the correlation among multiple longitudinal covariates but also the correlation between survival time and longitudinal outcome. Since $\boldsymbol{b}_i$'s are the ith subject's deviation from population mean intercept and population mean slope, $\exp(\boldsymbol{\gamma})$ is interpreted as the hazard ratio of increasing one unit of subject intercept or slope away from the population mean. The idea of introducing random intercept and slope of longitudinal processes in the Cox model is not new. A similar modeling technique has been considered by Liu and Huang [39] in 2009.

For simplicity of notation, we consider the situation where all J longitudinal outcomes and the survival outcome shares the same set of fixed covariates. Our model formulation can be easily generalized to the situation where the components of fixed covariates are different in different sub-models.

4.3 Variable Selection Through Penalized Likelihood

We denote $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}', \boldsymbol{y}_{i2}', \cdots, \boldsymbol{y}_{iJ}')'$ as the stacked longitudinal outcomes for subject i, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_J)'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J)'$. Let $\boldsymbol{\theta}$ be the collection of all unknown parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}, \sigma^2, \Sigma)$ in the model. The log likelihood can be written as

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \int f_y(\boldsymbol{y}_i | \boldsymbol{X}_i, \boldsymbol{b}_i; \boldsymbol{\theta}) f_T(T_i, \Delta_i | \boldsymbol{X}_i, \boldsymbol{b}_i, h_0(T_i); \boldsymbol{\theta}) f_b(\boldsymbol{b}_i) \mathtt{d}\boldsymbol{b}_i, \tag{4.3}$$

where $f_y(\boldsymbol{y}_i | \boldsymbol{X}_i, \boldsymbol{b}_i; \boldsymbol{\theta})$ is the density function of the multivariate Normal random variables $\boldsymbol{y}_i$ conditional on $\boldsymbol{b}_i$, $f_b(\boldsymbol{b}_i)$ is the density function of random effects $\boldsymbol{b}_i$, and $f_T(T_i, \Delta_i | \boldsymbol{X}_i, \boldsymbol{b}_i, h_0(T_i); \boldsymbol{\theta})$ is the likelihood function of survival outcome (T_i, Δ_i) conditional on $\boldsymbol{b}_i$:

$$\begin{aligned} & f_T(T_i, \Delta_i | \boldsymbol{X}_i, \boldsymbol{b}_i, h_0(T_i); \boldsymbol{\theta}) \\ = & \{h_0(T_i) \exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_0 + \boldsymbol{b}_i' \boldsymbol{\gamma})\}^{\Delta_i} \exp\left\{ - \int_0^{T_i} h_0(t) \exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_0 + \boldsymbol{b}_i' \boldsymbol{\gamma}) dt \right\}. \end{aligned} \tag{4.4}$$

The likelihood function depends on $h_0(T_i)$. It can be modeled parametrically (e.g. using an Exponential or Weibull distribution) or non-parametrically using a piece-wise constant function or B-spline bases. Throughout this chapter, I present an Exponential baseline, that is $h_0(T_i)$ equals a constant h_0. Our methods can be easily adapted to other models for the baseline function.

The aim of our research is to perform variable selection of the features of longitudinal outcomes in the survival sub-model. In order to select the random effects $\boldsymbol{b}_i$, we consider the following negative penalized likelihood

$$pl(\boldsymbol{\theta}) = -l(\boldsymbol{\theta}) + p_1(\boldsymbol{\gamma}) + p_2(\Sigma), \tag{4.5}$$

where $p_1(\boldsymbol{\gamma})$ controls the sparsity of the coefficients of random intercepts and slopes in the survival sub-model $\boldsymbol{\gamma}$, and $p_2(\Sigma)$ controls the sparsity of the covariance matrix of $\boldsymbol{b}_i$.

For $p_1(\boldsymbol{\gamma})$, we consider LASSO penalty, defined as $p_1(\boldsymbol{\gamma}) = n\lambda_1||\boldsymbol{\gamma}||_1 = n\lambda_1 \sum_{i=1}^{2J} |\gamma_i|$. The LASSO penalty was first introduced by Tibshirani [55] in 1996 for fixed-effect selection in linear models. Since L_1 penalty shrinks small estimates all the way towards zero, it is a useful tool for simultaneous variable selection and parameter estimation. Therefore the random intercepts and slopes of longitudinal covariates are selected in the survival sub-model. As the dimension of longitudinal outcomes may be high, it is reasonable to assume that the covariance matrix of random effects $\boldsymbol{b}_i$ is sparse. Therefore, we also consider L_1 penalty for $p_2(\Sigma)$. We define $p_2(\Sigma) = (n\lambda_2/2) \sum_{i \neq j} |\sigma_{ij}|$, where σ_{ij} is the element in Σ at ith row and jth column. Penalties are only imposed on off-diagonal elements to avoid the non-identifiability problem in our joint model. This off-diagonal penalty was considered on the precision matrix in graphical lasso problems by Zhang and Zou [63]. λ_1 and λ_2 are tuning parameters which control the sparsity of coefficients vector and covariance matrix respectively. It is worth mentioning that we do not penalize the variances of random intercepts and slopes in the longitudinal sub-models as they are essential features of trajectories, this is different than the penalty proposed by He et al. [26]. Their research aims at the selection of random covariates and therefore imposes penalties on the diagonal elements of covariances matrix as well.

We note that in this work we do not impose penalties on other coefficients since our main interest is to select important longitudinal covariates in the survival sub-model. However, our procedure can been easily generalized to penalize other coefficients, for example, $\boldsymbol{\alpha}$ or $\boldsymbol{\gamma}_0$, in order to perform variable selection of fixed covariates in longitudinal or survival sub-models respectively.

4.4 Computational Method

The negative penalized likelihood involves a complicated integral which does not have a closed form, therefore solving the whole solution path is difficult. In this section, I will propose a numeric estimation method based on coordinate descent algorithm.

4.4.1 Laplace Approximation

First, the integral in the negative penalized likelihood function is approximated by a Laplace approximation. We start by writing the non-penalized log likelihood function for subject i as:

$$\begin{aligned} & l_{0i}(\boldsymbol{\theta}) \\ & = \log \int \exp\{\log f_y(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{b}_i; \boldsymbol{\theta}) + \log f_T(T_i, \Delta_i|\boldsymbol{X}_i, \boldsymbol{b}_i, h_0(T_i); \boldsymbol{\theta}) + \log f_b(\boldsymbol{b}_i; \boldsymbol{\theta})\} \mathtt{d}\boldsymbol{b}_i \\ & = \log \int \exp\{\kappa(\boldsymbol{b}_i)\} \mathtt{d}\boldsymbol{b}_i. \end{aligned} \tag{4.6}$$

By second order Taylor expansion of $\kappa(\boldsymbol{b}_i)$, we get the following approximation:

$$\kappa(\boldsymbol{b}_i) \approx \kappa(\tilde{\boldsymbol{b}}_i) + \frac{1}{2}(\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i)' H_{\tilde{\boldsymbol{b}}_i} (\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i),$$

where $\tilde{\boldsymbol{b}}_i = \mathsf{argmax}_{\boldsymbol{b}_i} \kappa(\boldsymbol{b}_i)$ and $H_{\tilde{\boldsymbol{b}}_i}$ is the Hessian evaluated at $\tilde{\boldsymbol{b}}_i$.

Substituting the above approximation in (4.6) yields approximate log likelihood function for subject i:

$$l_{0i}(\boldsymbol{\theta}) \approx \kappa(\tilde{\boldsymbol{b}}_i) + \log \int \exp\left\{\frac{1}{2}(\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i)' H_{\tilde{\boldsymbol{b}}_i} (\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i)\right\} \mathtt{d}\boldsymbol{b}_i.$$

We note that the latter term is a multivariate Gaussian integral up to a constant and

therefore

$$l_{0i}(\boldsymbol{\theta}) \approx \kappa(\tilde{\boldsymbol{b}}_i) - \frac{1}{2}\log|-H_{\tilde{\boldsymbol{b}}_i}| + J\log(2\pi). \tag{4.7}$$

4.4.2 Coordinate Descent Algorithm

Coordinate descent is an algorithm designed to optimize a multivariate objective function. The idea is that the multivariate objective function can be optimized by optimizing over each direction in a loop. In each step, one focuses on one parameter alone while holding other parameters fixed, therefore the task of this algorithm becomes the optimization of a sequence of univariate objective functions. This algorithm is easy to implement since optimization of univariate objective functions are generally much easier than that of multivariate objective functions.

Coordinate descent and its extensions has been proposed for optimizing objective functions with L_1 regularization a number of times, for example in linear regression models [22, 20] or generalized linear models [52, 21]. In this section, we describe a coordinate descent algorithm to optimize our negative penalized likelihood.

With the Laplace approximation in (4.7), the approximate penalized log-likelihood can be written as

$$pl(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left[-\kappa(\tilde{\boldsymbol{b}}_i)\right] + \sum_{i=1}^{n}\frac{1}{2}\log|-H_{\tilde{\boldsymbol{b}}_i}| + n\lambda_1||\boldsymbol{\gamma}||_1 + (n\lambda_2/2)\sum_{j\neq k}|\sigma_{jk}|, \tag{4.8}$$

where σ_{jk} is the element of Σ at jth row and kth column, n is the number of subjects, and

$$\begin{aligned}-\kappa(\tilde{\boldsymbol{b}}_i) =& -\log f_y(\boldsymbol{y}_i|\boldsymbol{b}_i) - \log f_T(T_i, \Delta_i|\boldsymbol{b}_i) - \log f_b(\boldsymbol{b}_i)\\ =& \frac{n_i}{2}\log\sigma^2 + \frac{(\boldsymbol{y}_i-\boldsymbol{\mu}_i)'(\boldsymbol{y}_i-\boldsymbol{\mu}_i)}{2\sigma^2} - \Delta_i \log h_0 - \Delta_i(\boldsymbol{X}_i'\boldsymbol{\gamma}_0 + \boldsymbol{\gamma}'\tilde{\boldsymbol{b}}_i)\\ &+ h_0 T_i \exp(\boldsymbol{X}_i'\gamma_0 + \boldsymbol{\gamma}'\tilde{\boldsymbol{b}}_i) + \frac{1}{2}\log|\Sigma| + \frac{1}{2}\tilde{\boldsymbol{b}}_i'\Sigma^{-1}\tilde{\boldsymbol{b}}_i,\end{aligned}$$

$$H_i \triangleq H_{\tilde{\boldsymbol{b}}_i} = \left.\frac{\partial\kappa(\boldsymbol{b}_i)}{\partial\boldsymbol{b}_i'\partial\boldsymbol{b}_i}\right|_{\boldsymbol{b}_i=\tilde{\boldsymbol{b}}_i} = -\frac{\boldsymbol{Z}_i\boldsymbol{Z}_i'}{\sigma^2} - h_0 T_i \exp(\boldsymbol{X}_i'\boldsymbol{\gamma}_0 + \boldsymbol{\gamma}'\tilde{\boldsymbol{b}}_i)\boldsymbol{\gamma}\boldsymbol{\gamma}' - \Sigma^{-1},$$

where n_i is the total number of longitudinal observations for subject i, $\boldsymbol{\mu}_i$ is the mean vector of $\boldsymbol{y}_i$, $\boldsymbol{Z}_i$ is the design matrix for $\boldsymbol{b}_i$.

The last part in Equation (4.7) is a constant and therefore is dropped.

At the beginning of the algorithm, we initialize our parameter vector $\hat{\boldsymbol{\theta}}^{(0)}$ by a naive combination of longitudinal and survival outcomes. We fit a linear mixed effects model first on the longitudinal outcomes only to obtain estimates $\hat{\boldsymbol{\alpha}}^{(0)}$ and $\hat{\boldsymbol{\beta}}^{(0)}$. We then treat the estimated random effects as known covariates and fit the Cox proportional hazard model to get estimates $\hat{\boldsymbol{\gamma}}_0^{(0)}$ and $\hat{\boldsymbol{\gamma}}^{(0)}$.

Then in each iteration, we first calculate $\tilde{\boldsymbol{b}}_i \triangleq \mathsf{argmax}_{b_i}\kappa(\boldsymbol{b}_i)$ with the current estimates of parameters. We use a Newton method to solve $\tilde{\boldsymbol{b}}_i$ by updating equation $\tilde{\boldsymbol{b}}_i^{new} = \tilde{\boldsymbol{b}}_i^{old} - \kappa'(\boldsymbol{b}_i^{old})H_{\boldsymbol{b}_i^{old}}$, where

$$\kappa'(\boldsymbol{b}_i) = \frac{(\boldsymbol{y}_i-\boldsymbol{\mu}_i)'\boldsymbol{Z}_i}{\sigma^2} + \Delta_i\boldsymbol{\gamma}' - h_0 T_i \exp(\boldsymbol{X}_i'\boldsymbol{\gamma}_0 + \boldsymbol{\gamma}'\boldsymbol{b}_i) - \boldsymbol{b}_i'\Sigma^{-1},$$

$$H_{\boldsymbol{b}_i} = -\frac{\boldsymbol{Z}_i\boldsymbol{Z}_i'}{\sigma^2} - h_0 T_i \exp(\boldsymbol{X}_i'\boldsymbol{\gamma}_0 + \boldsymbol{\gamma}'\boldsymbol{b}_i)\boldsymbol{\gamma}\boldsymbol{\gamma}' - \Sigma^{-1}.$$

In the next step, we update each parameter in $\boldsymbol{\theta}$ one at a time with other parameters fixed at the value obtained in the previous iteration. Specifically, to update parameter $\hat{\theta}_j^t \to \hat{\theta}_j^{t+1}$, we would compute the gradient $\partial pl(\cdots, \hat{\theta}_{j-1}^{(t)}, \theta_j, \hat{\theta}_{j+1}^{(t)}, \cdots)/\partial\theta_j$ at $\theta_j = \hat{\theta}_j^{(t)}$.

Note that the gradient only exists when $\hat{\theta}_j^{(t)} \neq 0$. We then update $\hat{\theta}_j^{(t)}$ by

$$\hat{\theta}_j^{(t+1)} \leftarrow \hat{\theta}_j^{(t)} - s_j^{(t)} \left. \frac{\partial pl(\cdots, \hat{\theta}_{j-1}^{(t)}, \theta_j, \hat{\theta}_{j+1}^{(t)}, \cdots)}{\partial \theta_j} \right|_{\theta_j = \hat{\theta}_j^{(t)}}. \tag{4.9}$$

$s_j^{(t)}$ in (4.9) is the step size. We use a backtracking line search scheme to select the appropriate $s_j^{(t)}$ to insure sufficient descent of objective function in each update. A sketch of the line search algorithm is shown in Algorithm 1.

Algorithm 1 Backtracking line search algorithm

Start at $s = 1$, choose $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$
repeat $s \leftarrow \beta s$
until $f(x - sf'(x)) < f(x) - \alpha s[f'(x)]^2$

Partial derivatives and Hessian matrix in the updating steps are as follows:

$$\frac{\partial pl(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = -\sum_{i=1}^{n} \frac{(\boldsymbol{y}_i - \boldsymbol{\mu}_i)' \boldsymbol{X}_i}{\sigma^2}, \quad \frac{\partial pl(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^{n} \frac{(\boldsymbol{y}_i - \boldsymbol{\mu}_i)' \boldsymbol{Z}_i}{\sigma^2},$$

$$\frac{\partial pl(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}_0} = \sum_{i=1}^{n} -\Delta_i \boldsymbol{X}_i' + h_0 T_i \exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}' \tilde{\boldsymbol{b}}_i) \boldsymbol{X}_i' - \frac{1}{2} tr(H_i^{-1} h_0 T_i \boldsymbol{\gamma}\boldsymbol{\gamma}' \exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}' \tilde{\boldsymbol{b}}_i) \boldsymbol{X}_i),$$

$$\frac{\partial pl(\boldsymbol{\theta})}{\partial h_0} = \sum_{i=1}^{n} -\frac{\Delta_i}{h_0} + T_i \exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}' \tilde{\boldsymbol{b}}_i) - \frac{1}{2} |tr(H_i^{-1} T_i \boldsymbol{\gamma}\boldsymbol{\gamma}' \exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}' \tilde{\boldsymbol{b}}_i)),$$

$$\begin{aligned} \left. \frac{\partial pl(\boldsymbol{\theta})}{\partial \gamma_j} \right|_{\gamma_j = \gamma_j^{(t)}} &= \sum_{i=1}^{n} -\Delta_i b_{ij} + h_0 T_i \exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}^{(t)'} \tilde{\boldsymbol{b}}_i) b_{ij} \\ &\quad - \sum_{i=1}^{n} tr(H_i^{-1} h_0 T_i \exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}^{(t)'} \tilde{\boldsymbol{b}}_i)(\boldsymbol{\gamma}^{(t)} \boldsymbol{\gamma}^{(t)'} b_{ij} + 2\boldsymbol{I}_j \boldsymbol{\gamma}^{(t)'})) + sign(\gamma_j) n \lambda_1 |\gamma_j^{(t)}|, \end{aligned}$$

where $\boldsymbol{I}_j$ is a vector of length $2J$, its jth element is 1 and others are 0.

$$\frac{\partial pl(\boldsymbol{\theta})}{\partial \sigma^2} = \sum_{i=1}^{n} \frac{n_i}{2\sigma^2} - \frac{(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'(\boldsymbol{y}_i - \boldsymbol{\mu}_i)}{2(\sigma^2)^2} + \frac{1}{2} tr\left(H_i^{-1} \frac{\boldsymbol{Z}_i \boldsymbol{Z}_i'}{(\sigma^2)^2} \right).$$

$$\left.\frac{\partial pl(\boldsymbol{\theta})}{\partial \sigma_{jk}}\right|_{\sigma_{jk}=\sigma_{jk}^{(t)}} = \sum_{i=1}^{n}\left[\frac{1}{2}tr(\Sigma^{-1}\boldsymbol{I}_{jk}) - \frac{1}{2}\tilde{\boldsymbol{b}}_i'\Sigma^{-1}\boldsymbol{I}_{jk}\Sigma^{-1}\tilde{\boldsymbol{b}}_i + \frac{1}{2}tr(H_i^{-1}\Sigma^{-1}\boldsymbol{I}_{jk}\Sigma^{-1})\right]$$
$$+ I_{(j=k)}sign(\sigma_{jk})\lambda_2 n|\sigma_{jk}|,$$

where $I_{(j=k)}$ is a scale indicator, it equals 1 if $j = k$ and 0 otherwise, $\boldsymbol{I}_{jk}$ is a matrix indicator of dimension $2J \times 2J$, the elements at its jth row, kth column and kth row, jth column are 1 and all the other elements are 0.

The estimation procedure is summarized in Algorithm 2. The algorithm iterates until convergence. Our convergence criteria is

$$\frac{|pl(\boldsymbol{\theta}^{(t+1)}) - pl(\boldsymbol{\theta}^{(t)})|}{|pl(\boldsymbol{\theta}^{(t)})|} < \delta,$$

where δ is a pre-specified number. We use $\delta = 10^{-6}$ throughout this chapter.

Algorithm 2 Coordinate Descent Algorithm

Initialize parameter vector $\boldsymbol{\theta}^{(0)}$ by fitting marginal models
repeat
  Obtain $\tilde{\boldsymbol{b}}_i$ which maximize $\kappa(\boldsymbol{b}_i)$ with current estimates $\boldsymbol{\theta}^{(t)}$
  for (each parameter θ_i in $\boldsymbol{\theta}$) **do**
    Select step size $s_j^{(t)}$;
    Update $\theta_i^{(t+1)} \leftarrow \theta_i^{(t)} - s_j^{(t)} \cdot \left(\partial pl(\cdots, \hat{\theta}_{j-1}^{(t)}, \theta_j, \hat{\theta}_{j+1}^{(t)}, \cdots)/\partial\theta_j\right)$;
  end for
  $t \rightarrow t+1$
until Converge

4.4.3 Tuning Parameters

We apply the above algorithm on a grid of values of tuning parameter λ_1 and λ_2. The optimal combination of λ_1 and λ_2 is selected by minimizing BIC:

$$BIC_{\lambda_1,\lambda_2} = -2l(\boldsymbol{\theta}) + \log(N) \cdot df_{\lambda_1,\lambda_2},$$

where $l(\boldsymbol{\theta})$ is the log likelihood function, N is the total sample size $N = \sum_{i=1}^{n} n_i + n$. df_{λ_1,λ_2} is defined as the number of effective (non-zero) estimates in the model.

4.4.4 Two-Stage Estimation

To reduce bias introduced by penalization, we use a two-stage estimation suggested by He et al. [26]. The first stage focuses on the variable selection procedure using penalized likelihood function. When the non-zero effects are identified, the second stage re-estimates the model with the non-zero variables only, using non-penalized likelihood. Our simulation results show that this two-stage estimation method reduces the biases successfully.

4.5 Simulations

The simulation data are generated from the following model:

$$\begin{aligned} y_{ij}(t_{ijk}) &= X_i\alpha_j + \beta_{j1} + b_{ij1} + (\beta_{j2} + b_{ij2})t_{ijk} + \epsilon_{ijk}, \quad i = 1, 2, \cdots, n, j = 1, \cdots, J \\ h_i(t) &= h_0 \exp\{X_i\gamma_0 + \boldsymbol{b}_i'\boldsymbol{\gamma}\}, \end{aligned}$$

where fixed effect X_i is a binary group indicator, $X_i \overset{i.i.d.}{\sim} Bin(0.5)$, $\boldsymbol{\gamma} = (1, 0, 0, 0, 0, 0, 1)$ which implies that only the random intercept of the first longitudinal outcome and the random slope of the third longitudinal outcome are non-zero, $(\alpha_1, \alpha_2, \alpha_3) = (1, -1, 1)$. $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \beta_{31}, \beta_{32}) = (1, -1, 1, 1, -1, 1)$, $h_0 = 0.2$, $\gamma_0 = 1$, and $\sigma^2 = 0.04$. Longitudinal observations are collected every 0.2 time unit from the time 0 to the death or

censored time T_i. The following covariance matrix is considered for $\boldsymbol{b}_i$:

$$\Sigma = \begin{pmatrix} 0.5 & 0.2 & 0 & 0 & 0 & 0 \\ 0.2 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.2 & 0 & 0 \\ 0 & 0 & 0.2 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.2 \\ 0 & 0 & 0 & 0 & 0.2 & 0.5 \end{pmatrix}$$

We consider the following four scenarios:

- Scenario I: Censoring rate 10%, $n = 500$ subjects.
- Scenario II: Censoring rate 30%, $n = 500$ subjects.
- Scenario III: Censoring rate 10%, $n = 1000$ subjects.
- Scenario IV: Censoring rate 30%, $n = 1000$ subjects.

On average for each longitudinal outcome of each subject, 18 observations are collected in scenario I and III, and 13 observations are collected in scenario II and IV. Our simulation under each scenario is repeated 100 times.

Tables 4.1 and 4.7 present the selection frequencies of $\boldsymbol{\gamma}$ and covariance matrix Σ respectively. Throughout the simulation, we use 0.05 as the cut-off value to determine zero effects. Our variable selection procedure demonstrates excellent selection properties. The selection frequencies of non-zero elements in $\boldsymbol{\gamma}$ are 100% in all scenarios and the mis-selection rates for zero effects are well controlled. And both true positive rates and true negative rates in the selection of components in Σ are closed to 100%. Simulation also shows that increasing sample size and reducing censoring rate improves both selection performance and estimation precision.

Table 4.1: Sel. Freq.(%) of random effects in survival submodel

		γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{31}	γ_{32}
	True	1	0	0	0	0	1
Senario I	Freq.	100	0	1	0	1	100
Senario II	Freq.	100	3	2	2	13	100
Senario III	Freq.	100	1	0	0	3	100
Senario IV	Freq.	100	0	0	0	10	100

Table 4.2: Sel. Freq.(%) of components in covariance matrix

	True Positive	True Negative
Scenario I	100	99.8
Scenario II	100	99.8
Scenario III	100	100
Scenario IV	100	100

Estimation results are shown in Tables 4.3, 4.4, 4.5 and 4.6. The bias of direct estimation of random effects ranges from 0% to 25%. After post-selection estimation, they are reduced to 0% to 6%. The estimates of fixed effects are almost the same in direct estimation and post-selection estimation. This is reasonable because we do not impose penalties on these parameters, and re-estimation with selected covariates using non-penalized likelihood will not have a large impact on them. In order to evaluate the precision of estimation of covariance matrix Σ, I use Frobenius norm to measure the difference matrix between Σ and $\hat{\Sigma}$. It is defined as $||\Sigma - \hat{\Sigma}||_F = (\sum(\sigma_{ij} - \hat{\sigma}_{ij})^2)^{1/2}$. Table 4.7 compares the direct estimation and post-selection estimation in all four scenarios. It can be observed that post-selection estimation increases the precisions of $\hat{\Sigma}$ by a large extent.

4.6 Application

To illustrate the model, we analyze the data from an observational study of ESRD patients on HD. The data are provided by Fresenius Medical Care. There are 872 patients in the study. They are enrolled in the study since they have been on HD for one year and are then observed for 1 subsequent year. The patient cohort includes 506 (58.0%) males. 561 (64.33%) patients have diabetes. The age of patients ranges from 21 to 99 with mean 64. Patients' survival time is monitored. Among 872 patients, 151 (17.3%) died within the observational time period.

We also collect patients' monthly measurements of systolic blood pressure prior to dialysis (PreSBP), neutrophil to lymphocyte ratio (NLR), albumin, inter-dialytic weight gain (IDWG) and eKt/V. Since NLR is heavily skewed, we consider $\log(NLR)$ in the modeling. PreSBP is divided by 100 to ensure that the scales of the five longitudinal covariates are approximately the same. The distribution of these variables are presented in Figure 4.1.

To illustrate the format of the ESRD data, we randomly select 10 patients who died during the observational period and 10 patients who were censored. The five longitudinal observations of them are plotted in Figure 4.2. Red lines represent patients who died and green lines represent patients who were censored. The figure illustrates that the longitudinal measurements are not always available in each month, and the observational times for different measurements are not always the same in the data. It is worth-mentioning that our model does not require any specific observational scheme and can handle this situation well.

For the five longitudinal covariates, we consider a multivariate linear mixed effect model. Each longitudinal process is adjusted by the fixed effects of gender, diabetic and

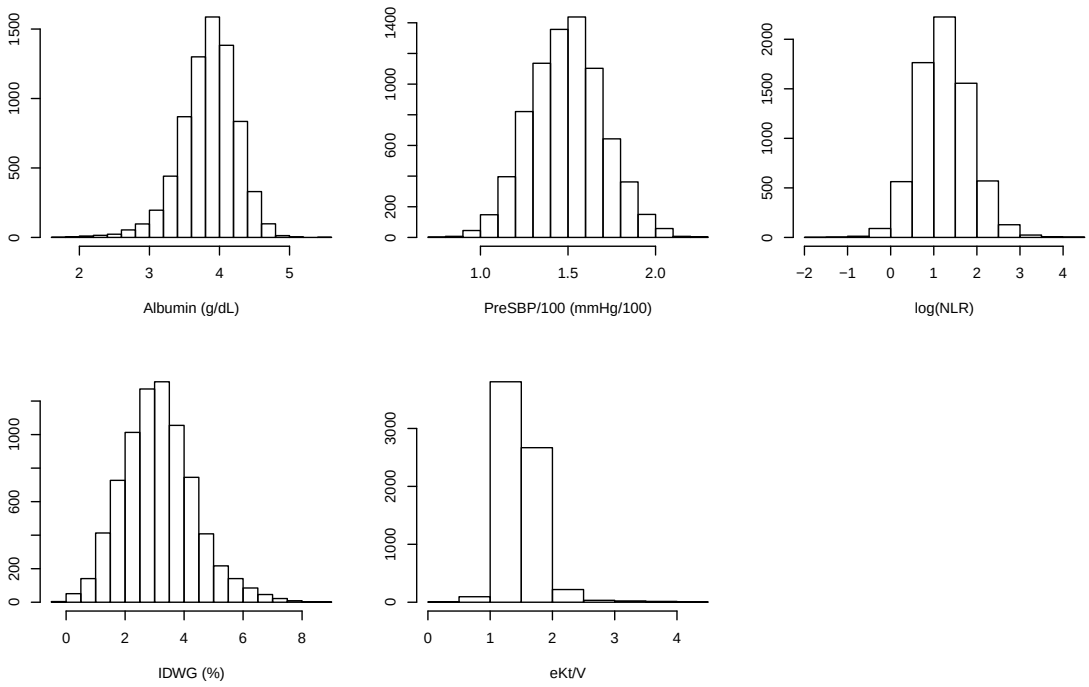


Figure 4.1: The histograms of longitudinal covariates: Albumin (g/dL), PreSBP/100 (mmHg/100), log(NLR), IDWG (%) and eKt/V in the data.

age:

$$
\begin{aligned}
Albumin_i(t_{i1k}) &= \alpha_{11}I_{Male_i} + \alpha_{12}I_{Diabetic_i} + \alpha_{13}Age_i + \beta_{11} + \beta_{12}t_{i1k} + b_{i11} + b_{i12}t_{i1k} + \epsilon_{i1k} \\
PreSBP_i(t_{i2k})/100 &= \alpha_{21}I_{Male_i} + \alpha_{22}I_{Diabetic_i} + \alpha_{23}Age_i + \beta_{21} + \beta_{22}t_{i1k} + b_{i21} + b_{i22}t_{i2k} + \epsilon_{i2k} \\
log(NLR)_i(t_{i3k}) &= \alpha_{31}I_{Male_i} + \alpha_{32}I_{Diabetic_i} + \alpha_{33}Age_i + \beta_{31} + \beta_{32}t_{i3k} + b_{i31} + b_{i32}t_{i3k} + \epsilon_{i3k} \\
IDWG_i(t_{i4k}) &= \alpha_{41}I_{Male_i} + \alpha_{42}I_{Diabetic_i} + \alpha_{43}Age_i + \beta_{41} + \beta_{42}t_{i4k} + b_{i41} + b_{i42}t_{i4k} + \epsilon_{i4k} \\
eKt/V_i(t_{i5k}) &= \alpha_{51}I_{Male_i} + \alpha_{52}I_{Diabetic_i} + \alpha_{53}Age_i + \beta_{51} + \beta_{52}t_{i1k} + b_{i51} + b_{i52}t_{i5k} + \epsilon_{i5k}
\end{aligned}
$$

For mortality, we use proportional hazard model with the fixed effects of gender,

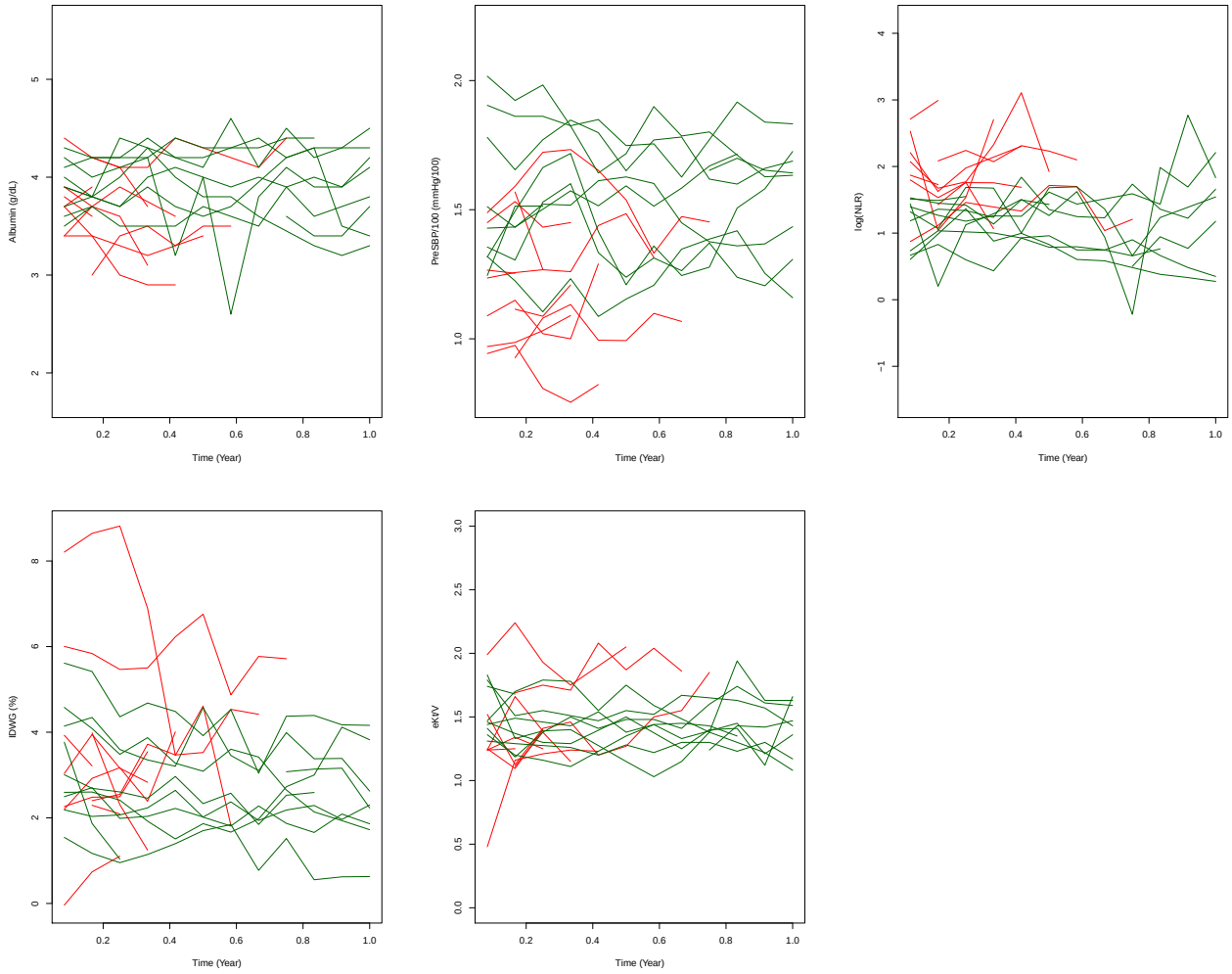


Figure 4.2: Longitudinal observations of albumin (g/dL), PreSBP/100 (mmHg/100), log(NLR), IDWG (%) and eKt/V from 20 random selected patients in the data. 10 patients died during the observational time, their observations are marked red. 10 other patients were censored, their observations are marked green.

diabetic, age, and random effects in the longitudinal sub-model:

$$h(t_i) = h_0(t_i) \exp\{\gamma_{01} I_{Male_i} + \gamma_{02} I_{Diabetic_i} + \gamma_{03} Age_i + \boldsymbol{\gamma} \boldsymbol{b}_i\},$$

where $\boldsymbol{b}_i = (b_{i11}, b_{i12}, \cdots, b_{i51}, b_{i52})$. We assume the survival time of ESRD patients follows an Exponential distribution, and therefore $h_0(t_i)$ is a constant h_0. We make this assumption mainly for the simplicity of illustration, however, it is not entirely unrealistic. In fact, a constant baseline hazard is suggested in the data analytic results in Chapter 2.

We apply our selection and estimation methods and present the selection and estimation results of random effects in Table 4.8. The standard errors are calculated based on the information matrix. Our procedure selects the random intercept of Albumin, the random slope of PreSBP/100, both the random intercept and slope of $\log(NLR)$, and the random intercept of IDWG as non-zero effects. Specifically, a high rate of survival hazard is associated with low level of albumin at the beginning, drop in PreSBP over time, high level of NLR at the beginning, increase of NLR over time and a high level of IDWG in the beginning. The estimation results of fixed effects are listed in the survival sub-model. Age is a significant risk factor. Senior patients tend to have higher survival hazard. Table 4.10 presents the estimation results of mean intercept, mean slope and the fixed effects in the longitudinal sub-model.

4.7 Conclusion

In this chapter, we propose a joint modeling framework for the joint analysis of multiple longitudinal outcomes and survival time. The two types of responses are correlated through shared random intercepts and slopes in the longitudinal sub-models. A penalized likelihood approach is introduced to perform variable selection and parameter estimation

simultaneously. A coordinate descent algorithm is proposed to calculate the penalized maximized likelihood estimates. Furthermore, a post-selection approach is introduced to reduce the estimation bias resulted from the penalization.

We demonstrate through simulation that our procedure carries excellent selection properties for both random effects and the elements of covariance matrix. The post-selection method effectively reduced the estimation bias. The performance of estimation procedure is further improved if we increase the sample size or decrease the censoring rate.

Our proposed method is applied to a study of ESRD patients as an illustration. The procedure successfully selects the non-zero random effects in the survival sub-model. They include the random intercept of albumin, random slope of PreSBP, random intercept of $\log(NLR)$, random intercept of IDWG and random slope of eKt/V. The signs of these estimated coefficients are as expected.

In this research, I assume the survival time is from an exponential distribution for simplicity. It can be easily generalized to other parametric distributions with corresponding forms of baseline hazard. To allow flexible forms of baseline hazard, piecewise-constant functions and B-splines can also be accommodated when the number and location of knots are specified. In addition, as the primary of interest of this research is the selection of non-zero longitudinal effects on the survival outcome, we only impose penalties on the coefficients of random trajectories of longitudinal outcome and their covariance matrix. Penalties on other parameters, for example fixed effects, could also be considered if of interest in future practical data analysis.

Table 4.3: Estimation of coefficients in survival submodel in scenario I

Survival Sub-model

		Random Effects						Fixed Effect
		γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{31}	γ_{32}	γ_0
True		1	0	0	0	0	1	1
Direct	Bias	0.130	-0.003	0.001	-0.001	-0.008	0.205	0.102
	MSE	0.024	0.000	0.000	0.000	0.000	0.005	0.023
Post-Selection	Bias	-0.006	0.000	0.000	0.000	0.000	0.047	0.015
	MSE	0.007	0.000	0.000	0.000	0.000	0.008	0.015

Longitudinal Sub-model

		Fixed Intercept and Slope					
		β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}
True		1	-1	1	1	-1	1
Direct	Bias	0.002	-0.007	0.001	0.007	0.002	0.099
	MSE	0.002	0.001	0.002	0.001	0.002	0.011
Post-Selection	Bias	0.002	-0.007	0.001	0.007	0.002	0.099
	MSE	0.002	0.001	0.002	0.001	0.002	0.010

		Fixed Effect		
		α_1	α_2	α_3
True		1	-1	1
Direct	Bias	-0.003	0.002	-0.019
	MSE	0.004	0.004	0.004
Post-Selection	Bias	-0.002	0.002	-0.019
	MSE	0.004	0.004	0.004

Table 4.4: Estimation of coefficients in survival submodel in scenario II

Survival Sub-model

		Random Effects						Fixed Effect
		γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{31}	γ_{32}	γ_0
True		1	0	0	0	0	1	1
Direct	Bias	0.179	-0.007	-0.001	0.002	-0.021	0.255	0.104
	MSE	0.036	0.000	0.000	0.000	0.002	0.073	0.023
Post-Selection	Bias	0.010	-0.006	-0.004	0.004	-0.026	0.065	0.011
	MSE	0.006	0.001	0.001	0.001	0.005	0.013	0.016

Longitudinal Sub-model

		Fixed Intercept and Slope					
		β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}
True		1	-1	1	1	-1	1
Direct	Bias	0.003	-0.007	0.007	-0.001	0.004	0.098
	MSE	0.002	0.001	0.002	0.001	0.002	0.011
Post-Selection	Bias	0.003	-0.007	0.007	-0.001	0.004	0.098
	MSE	0.002	0.001	0.002	0.001	0.002	0.011

		Fixed Effect		
		α_1	α_2	α_3
True		1	-1	1
Direct	Bias	-0.004	-0.013	-0.018
	MSE	0.004	0.003	0.004
Post-Selection	Bias	-0.002	-0.013	-0.017
	MSE	0.004	0.003	0.004

Table 4.5: Estimation of coefficients in survival submodel in scenario III

Survival Sub-model

		Random Effects						Fixed Effect
		γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{31}	γ_{32}	γ_0
True		1	0	0	0	0	1	1
Direct	Bias	0.122	0.000	0.000	0.000	-0.007	0.198	0.025
	MSE	0.018	0.000	0.000	0.000	0.000	0.042	0.011
Post-Selection	Bias	-0.002	-0.001	0.000	0.000	-0.003	0.049	0.000
	MSE	0.003	0.000	0.000	0.000	0.000	0.005	0.009

Longitudinal Sub-model

		Fixed Intercept and Slope					
		β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}
True		1	-1	1	1	-1	1
Direct	Bias	0.007	-0.007	-0.000	0.006	-0.002	0.103
	MSE	0.001	0.001	0.001	0.001	0.001	0.011
Post-Selection	Bias	0.007	-0.007	-0.000	0.006	-0.002	0.102
	MSE	0.001	0.001	0.001	0.001	0.001	0.011

		Fixed Effect		
		α_1	α_2	α_3
True		1	-1	1
Direct	Bias	-0.005	0.003	-0.016
	MSE	0.002	0.002	0.002
Post-Selection	Bias	-0.003	0.003	-0.016
	MSE	0.002	0.002	0.002

Table 4.6: Estimation of coefficients in survival submodel in scenario IV

Survival Sub-model

		Random Effects						Fixed Effect
		γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{31}	γ_{32}	γ_0
True		1	0	0	0	0	1	1
Direct	Bias	0.173	-0.001	-0.000	0.000	-0.016	0.246	0.089
	MSE	0.032	0.000	0.000	0.000	0.000	0.064	0.018
Post-Selection	Bias	0.014	0.000	0.000	0.000	-0.019	0.059	0.019
	MSE	0.003	0.000	0.000	0.000	0.003	0.008	0.010

Longitudinal Sub-model

		Fixed Intercept and Slope					
		β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}
True		1	-1	1	1	-1	1
Direct	Bias	-0.001	-0.007	0.005	0.001	0.003	0.098
	MSE	0.001	0.001	0.001	0.001	0.001	0.010
Post-Selection	Bias	-0.001	-0.006	0.005	0.001	0.003	0.098
	MSE	0.001	0.001	0.001	0.001	0.001	0.010

		Fixed Effect		
		α_1	α_2	α_3
True		1	-1	1
Direct	Bias	0.001	-0.005	-0.020
	MSE	0.002	0.002	0.002
Post-Selection	Bias	0.002	-0.005	-0.020
	MSE	0.002	0.002	0.002

Table 4.7: Frobenius norms of $(\Sigma - \hat{\Sigma})$ under all scenarios

	True Direct	Post-selection
Scenario I	0.161	0.104
Scenario II	0.162	0.107
Scenario III	0.161	0.074
Scenario IV	0.155	0.076

Table 4.8: Selection and Estimation (SE) of longitudinal covariates in survival sub-model

	Albumin	PreSBP/100	log(NLR)	IDWG	eKt/V
Intercept	-2.066(0.124)	-	1.034(0.151)	0.115(0.056)	-
Slope	-	-3.353(0.105)	0.795(0.156)	-	-

Table 4.9: Estimate (SE) of fixed effects in survival sub-model

Male	Diabetic	Age
0.199(0.104)	0.170(0.096)	0.004(0.001)

Table 4.10: Estimate (SE) of fixed covariates in longitudinal sub-model

	Albumin	PreSBP/100	log(NLR)	IDWG	EKTV
Intercept	4.175(0.004)	1.658(0.004)	0.784(0.004)	4.586(0.004)	1.344(0.004)
Slope	-0.079(0.006)	-0.023(0.006)	0.096(0.006)	-0.058(0.006)	-0.039(0.006)
Male	0.063(0.005)	-0.051(0.005)	0.010(0.005)	-0.052(0.005)	-0.090(0.005)
Diabetic	-0.083(0.005)	0.064(0.004)	-0.029(0.005)	-0.199(0.004)	0.037(0.005)
Age	-0.004(0.001)	-0.003(0.001)	0.007(0.001)	-0.020(0.001)	0.003(0.001)

Appendix A

Estimation Performance of $\hat{h}_0(t)$ and $\hat{w}(t)$ Under All Scenarios

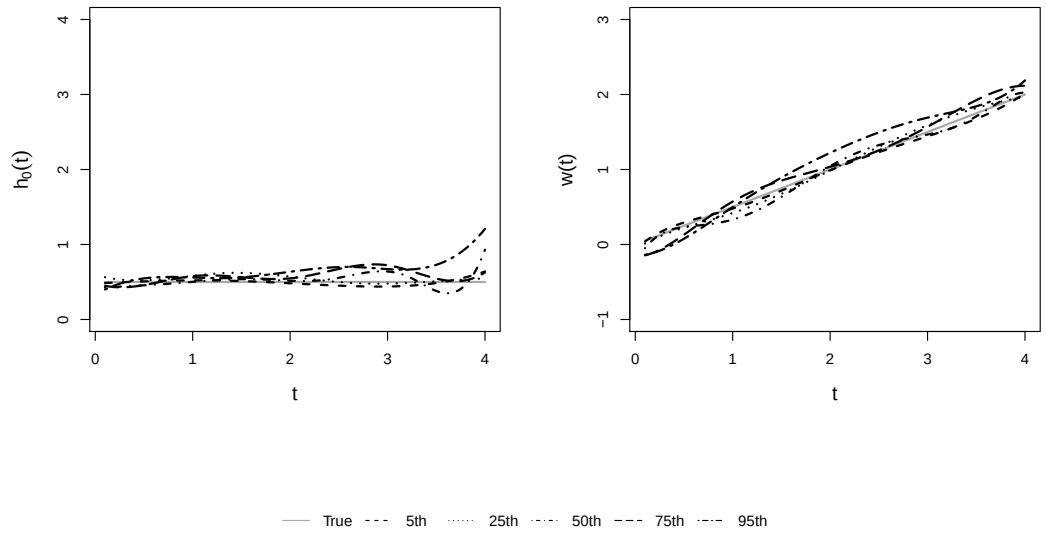


Figure A.1: True function (solid lines) and estimates (dashed lines) of $h_0(t) = 0.5$ (left) and $w(t) = t/2$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = 1/2$, $w(t) = \log(t)$ and $n = 1000$.

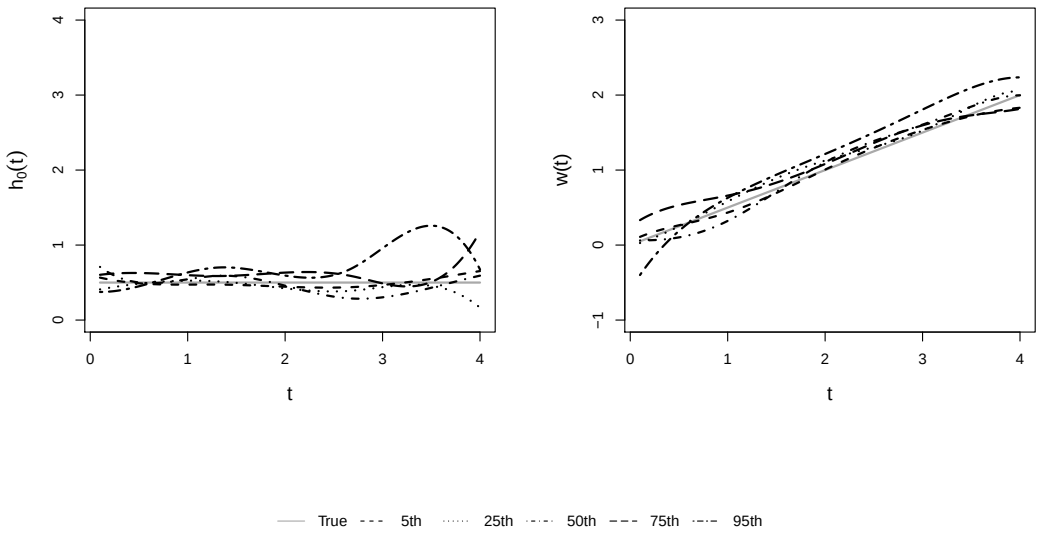


Figure A.2: True function (solid lines) and estimates (dashed lines) of $h_0(t) = 0.5$ (left) and $w(t) = t/2$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = 1/2$, $w(t) = \log(t)$ and $n = 500$.

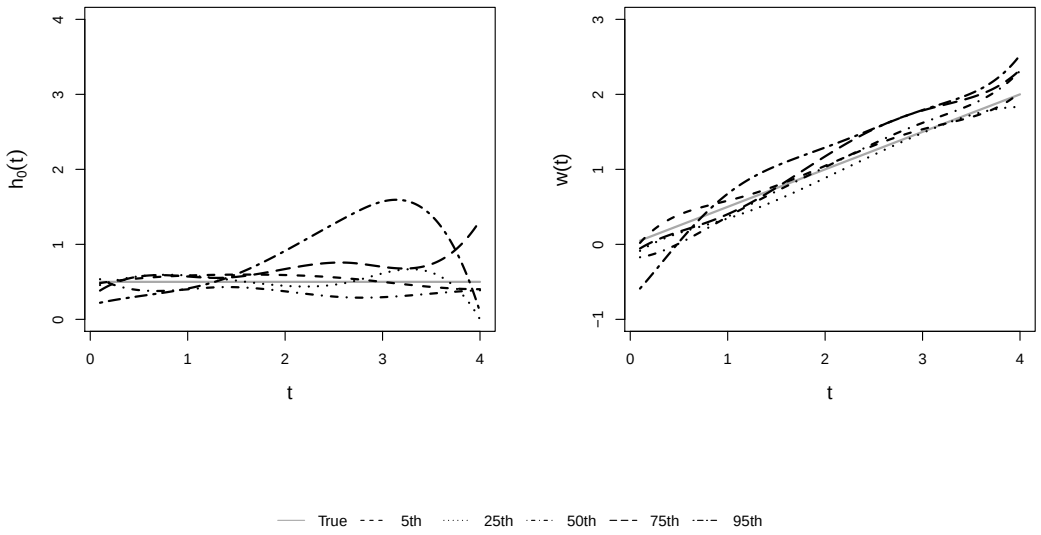


Figure A.3: True function (solid lines) and estimates (dashed lines) of $h_0(t) = 0.5$ (left) and $w(t) = t/2$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = 1/2$, $w(t) = \log(t)$ and $n = 300$.

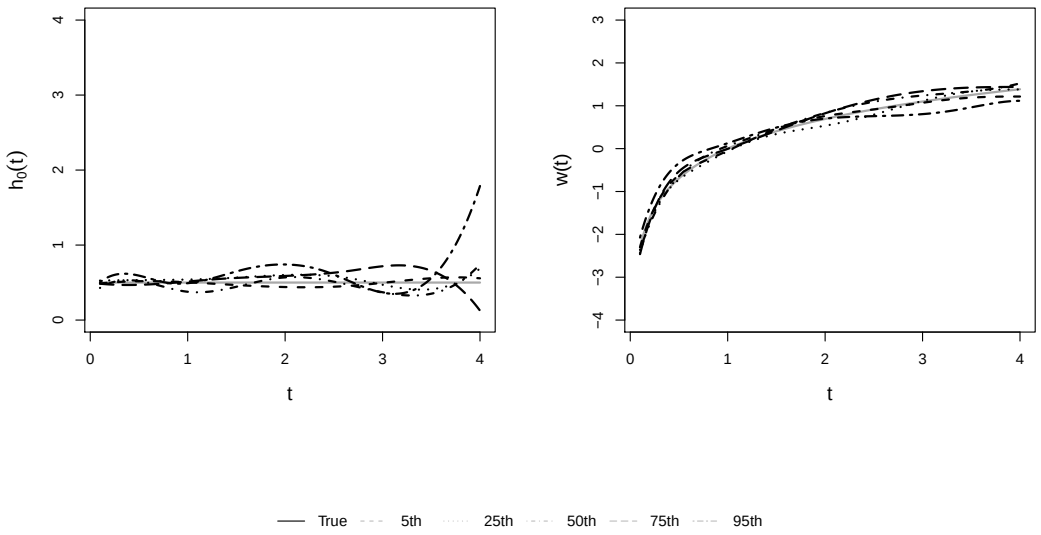


Figure A.4: True function (solid lines) and estimates (dashed lines) of $h_0(t) = 0.5$ (left) and $w(t) = log(t)$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = 1/2$, $w(t) = \log(t)$ and $n = 1000$.

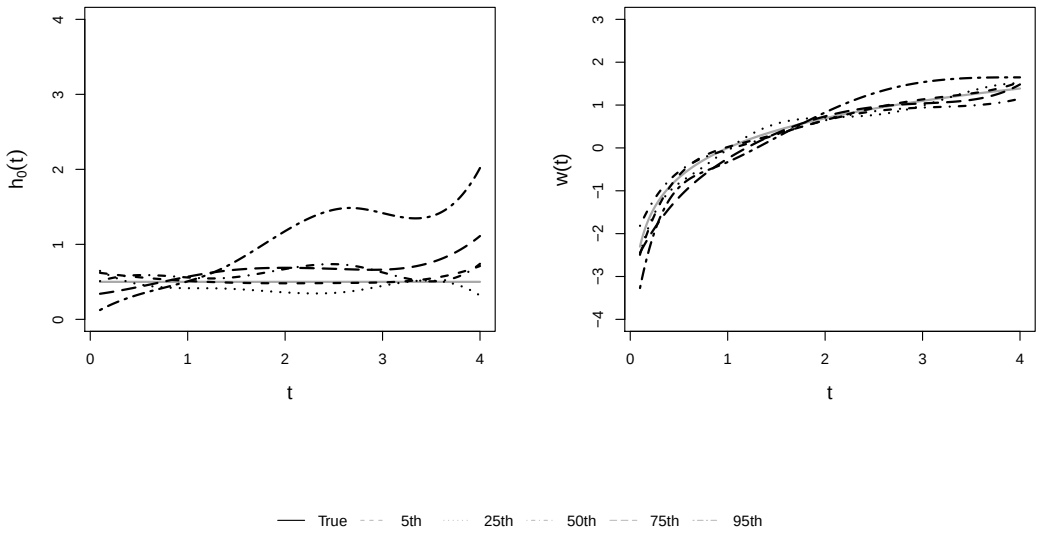


Figure A.5: True function (solid lines) and estimates (dashed lines) of $h_0(t) = 0.5$ (left) and $w(t) = log(t)$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = 1/2$, $w(t) = \log(t)$ and $n = 500$.

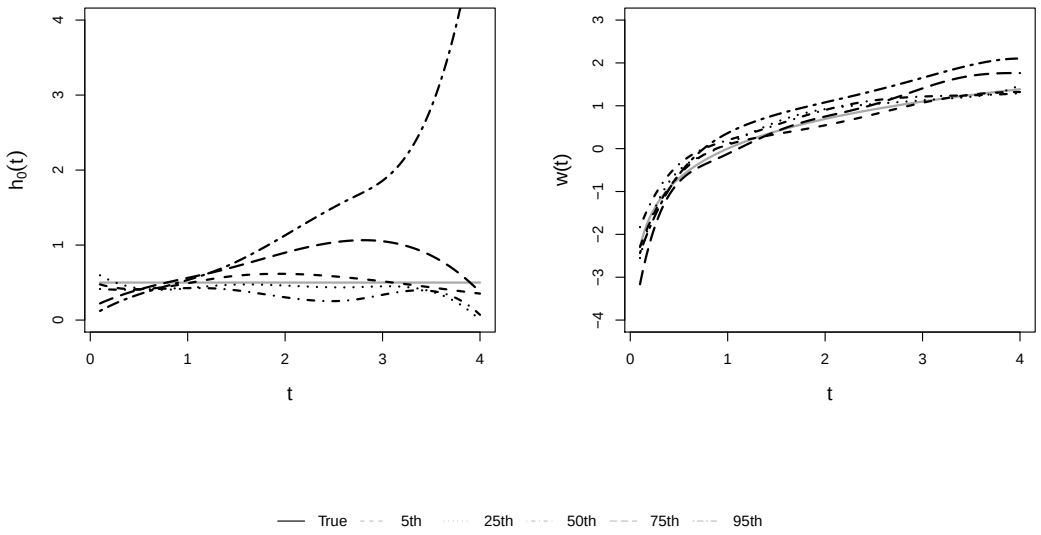


Figure A.6: True function (solid lines) and estimates (dashed lines) of $h_0(t) = 0.5$ (left) and $w(t) = log(t)$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = 1/2$, $w(t) = \log(t)$ and $n = 300$.

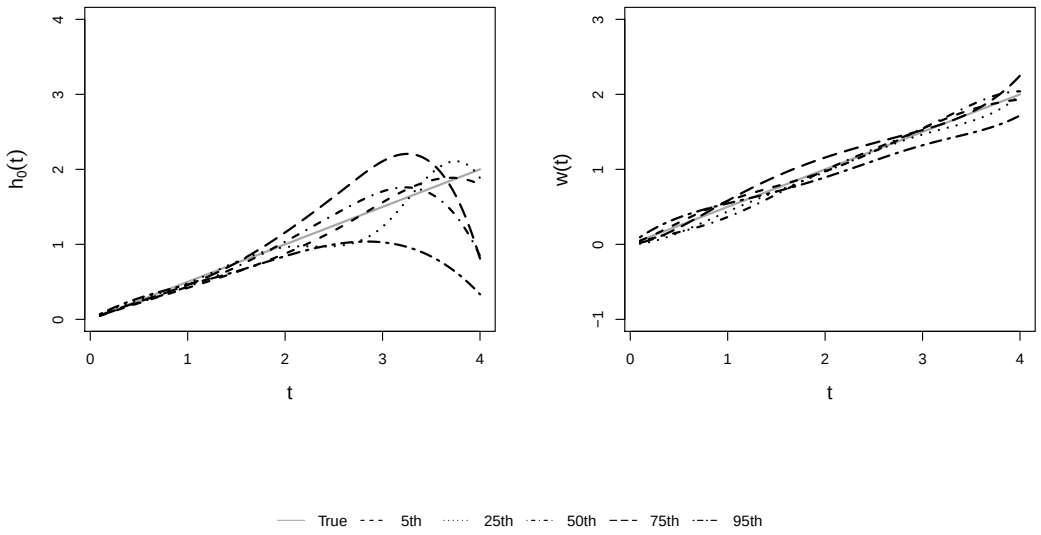


Figure A.7: True function (solid lines) and estimates (dashed lines) of $h_0(t) = t/2$ (left) and $w(t) = t/2$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = t/2$, $w(t) = t/2$ and $n = 1000$.

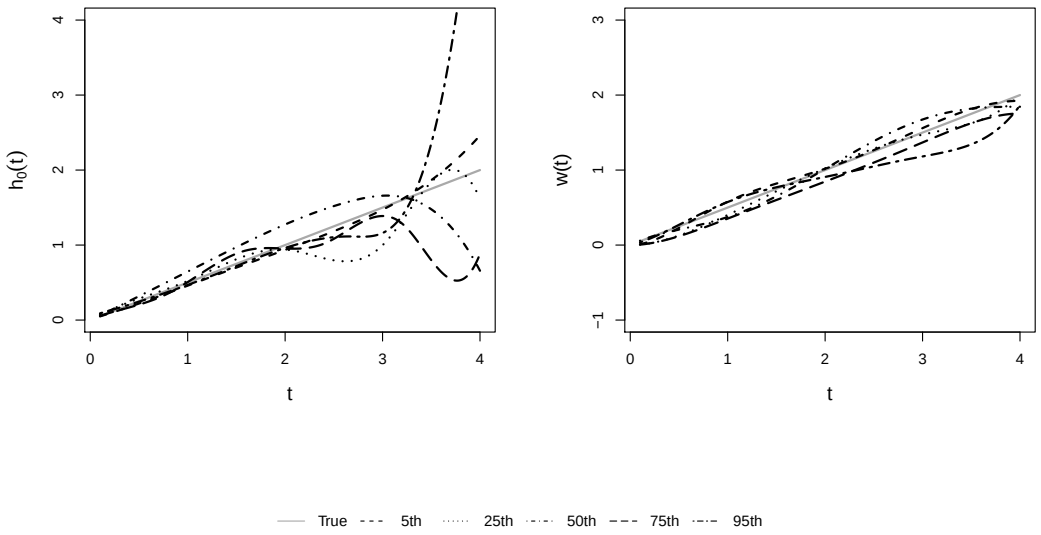


Figure A.8: True function (solid lines) and estimates (dashed lines) of $h_0(t) = t/2$ (left) and $w(t) = t/2$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = t/2$, $w(t) = t/2$ and $n = 500$.

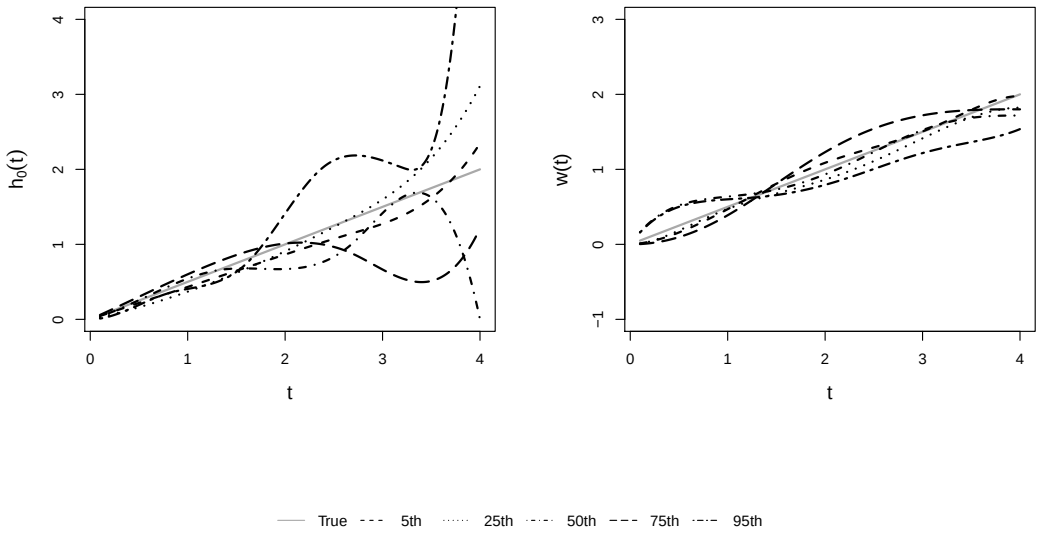


Figure A.9: True function (solid lines) and estimates (dashed lines) of $h_0(t) = t/2$ (left) and $w(t) = t/2$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = t/2$, $w(t) = t/2$ and $n = 300$.

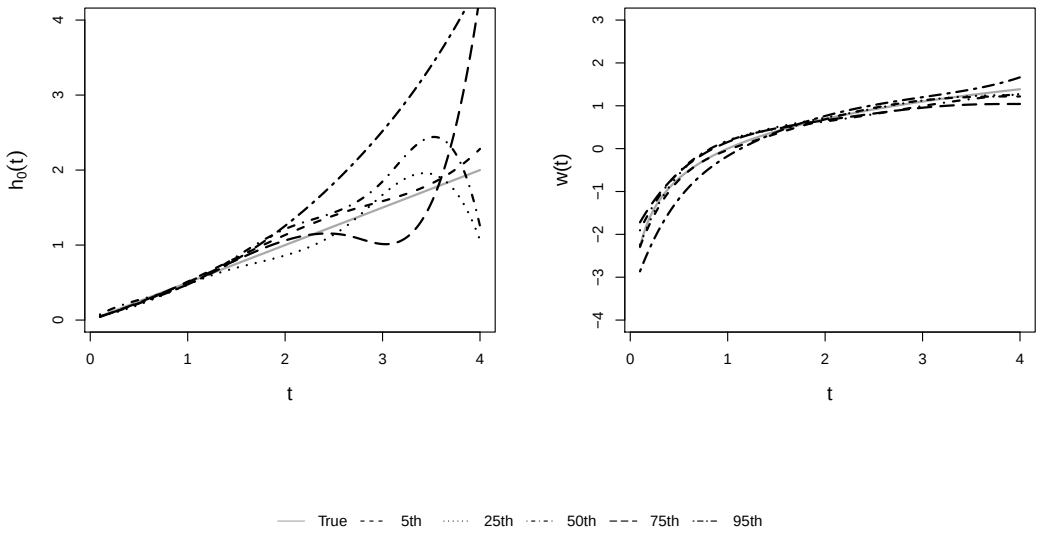


Figure A.10: True function (solid lines) and estimates (dashed lines) of $h_0(t) = t/2$ (left) and $w(t) = log(t)$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = t/2$, $w(t) = \log(t)$ and $n = 1000$.

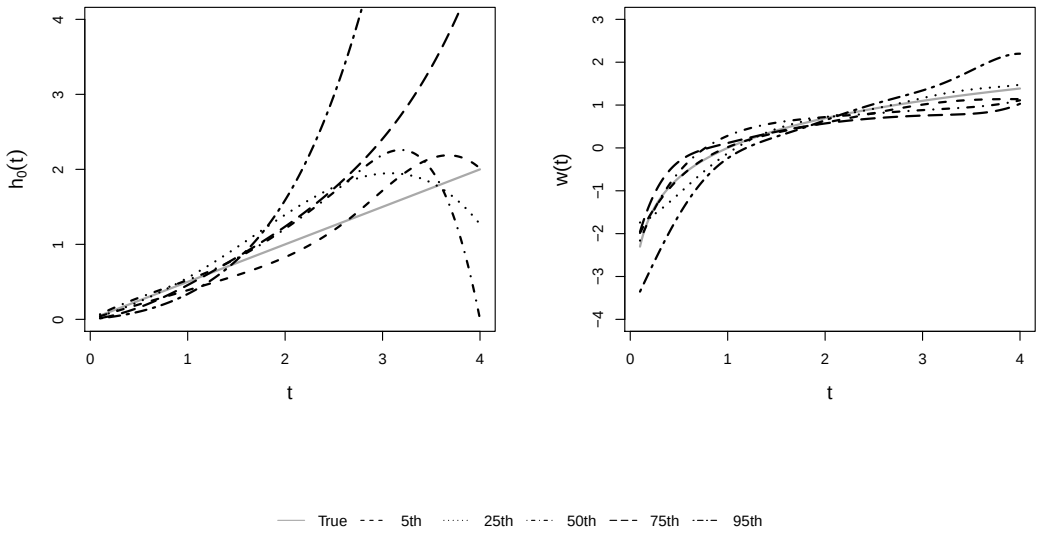


Figure A.11: True function (solid lines) and estimates (dashed lines) of $h_0(t) = t/2$ (left) and $w(t) = log(t)$ (right) correspond to the $5th$, $25th$, $50th$, $75th$ and $95th$ percentiles of the IMSE when $h_0(t) = t/2$, $w(t) = \log(t)$ and $n = 300$.

Appendix B

R and SAS Codes Used in Chapter 2

Here we provide the R and SAS codes for the data analysis in Chapter 2. The following R code is used for generating B-spline bases and integrated B-spline bases in R:

```
library(sas7bdat)
library(splines)

# read data of patients' survival time
train <- read.table("dat_time.csv", sep=",", header=T,
                    na.strings=c("NA",""))

# function to generate B-spline bases and integrated B-spline bases
# hk is number of knots for baseline
# wk is number of knots for offset function

bspline <-
  function(hk, wk, cut=730/100, n=dim(train)[1], K=1000, data=train)
  {
  dur <- function(i,j,knots){
    integrate(function(x)
      as.numeric(bs(x,knots=seq(0,cut,length=knots+2)[-c(1,knots+2)],
                 degree=3,intercept=T, Bound=c(-1/100,731/100))[,j]),
      0, tgrid[i], rel.tol=.Machine$double.eps^0.1)$value
  }

  t <- bs(data$time/100,
          knots=seq(0,cut,length=hk+2)[-c(1,hk+2)],intercept=T,
```

```
          Bound=c(-1/100,731/100))

  tgrid <- seq(0,cut,length=K)

  Tgrid <- mat.or.vec(K,hk+4)
  for (i in 1:K){
    for (j in 1:(hk+4)){
      Tgrid[i,j] <- dur(i,j,hk)
    }
  }
  d <- t(sapply(1:n, function(x)
                    Tgrid[which.min(abs(tgrid-data$time[x]/100)),]))

  Wgrid <- mat.or.vec(K,wk+4)
  for (i in 1:K){
    for (j in 1:(wk+4)){
      Wgrid[i,j] <- dur(i,j,wk)
    }
  }
  dw <- t(sapply(1:n, function(x)
                     Wgrid[which.min(abs(tgrid-data$time[x]/100)),]))

  data.frame(patient_id=data$patient_id,t=t,d=d,dw=dw)
}

time33 <- bspline(2,2)

write.table(time22,file="dat_time22.csv",sep=",",col.names=T,row.names=F)
```

The following sas code is for the joint model of survival time and number of hospitalization:

```
libname JM '~/JM2data';

proc import datafile = '~/JM2data/dat_time22.csv' dbms=csv replace
out = JM.dat_time22;
run;

data JM.dat_all22;
merge JM.dat_time22 JM.dat_all;
by patient_id;
```

```
run;

proc nlmixed data=JM.dat_all33 maxiter=4000 TECH=NRRIDG;
bounds sigma2 c1-c6 d1-d6>0;
parms  c0=-2 c1=0.3 c2=0.1 c3=0.5 c4=0.2 c5=1 c6=0.1 sigma2=0.5 eta=1
       a1=0.002  a2=-0.341 a3=0.003 a4=0.031 a5=-0.004 a6=-0.105
       a7=0.027 a8=-0.160 a10=0.018 a11=0.231 a12=0.444
       b1=0.028 b2=-0.906 b3=-0.004 b4=0.067 b5=-0.020 b6=0.095 b7=0.086
       b8=-0.387 b10=0.086 b11=0.002 b12=-0.281
       d1=0.1 d2=0.1 d3=0.1 d4=0.1 d5=0.1 d6=0.1;

lin_h = v + b1*age + b2*albumin_avg + b3*pre_sbp_avg + b4*nlr_avg
        + b5*post_bmi_avg + b6*male + b7*idwg_percent_avg + b8*ektv_avg
        + b9*vintage + b10*race_white +b11*race_black;

base_haz_d = d1*t_1 + d2*t_2 + d3*t_3 + d4*t_4 + d5*t_5 + d6*t_6;
cum_base_haz_d = d1*d_1 + d2*d_2 + d3*d_3 + d4*d_4 + d5*d_5 + d6*d_6;
h = base_haz_d * exp(lin_h);
S = exp(-cum_base_haz_d * exp(lin_h));

* loglik for death;
if (event=1) then loglikD = log(h) + log(S);
if (event=0) then loglikD = log(S);

* loglik for Y;
lin_Y = a1*age + a2*albumin_avg + a3*pre_sbp_avg + a4*nlr_avg
        + a5*post_bmi_avg + a6*male + a7*idwg_percent_avg + a8*ektv_avg
        + a9*vintage +a10*race_white +a11*race_black + eta*v;

w_T = c0 + c1*dw_1 + c2*dw_2 + c3*dw_3 + c4*dw_4 + c5*dw_5 + c6*dw_6;

lambda = exp(w_T + lin_Y);
loglikY = hosp_count*log(lambda) - lambda;

loglik = loglikD + loglikY;

model time ~ general(loglik);
random v ~ normal(0,sigma2) subject=patient_id out=JM.hosp22_nu;
ods output ParameterEstimates=JM.hosp22_est FitStatistics=JM.hosp22_fit;
run;
```

The following sas code is for the joint model of survival time and total length of hospital stay:

```
libname JM '~/JM2data';

data JM.dat_days22;
merge JM.dat_time22 JM.dat_all;
by patient_id;
run;

proc nlmixed data=JM.dat_days22 maxiter=4000 TECH=NRRIDG;
bounds sigma2 c1-c6 d1-d6 nu>0;
parms  c0=-2 c1=1 c2=1 c3=0.1 c4=0.3 c5=0.01 c6=0.1
       sigma2=1 eta=1 nu=2
       a1=0.002  a2=-0.341 a3=0.003 a4=0.031 a5=-0.004 a6=-0.105
       a7=0.027 a8=-0.160 a10=0.018 a11=0.231 a12=0.444
       b1=0.028 b2=-0.906 b3=-0.004 b4=0.067 b5=-0.020 b6=0.095
       b7=0.086 b8=-0.387 b10=0.086 b11=0.002 b12=-0.281
       d1=0.4 d2=0.8 d3=0.1 d4=0.8 d5=0.2 d6=1;

lin_h = v + b1*age + b2*albumin_avg + b3*pre_sbp_avg + b4*nlr_avg
        + b5*post_bmi_avg + b6*male + b7*idwg_percent_avg + b8*ektv_avg
        + b9*vintage + b10*race_white +b11*race_black;
*h = h0 * exp(lin_h);
*S = exp(-time*h/700);

base_haz_d = d1*t_1 + d2*t_2 + d3*t_3 + d4*t_4 + d5*t_5 + d6*t_6;
cum_base_haz_d = d1*d_1 + d2*d_2 + d3*d_3 + d4*d_4 + d5*d_5 + d6*d_6;
h = base_haz_d * exp(lin_h);
S = exp(-cum_base_haz_d * exp(lin_h));

* loglik for death;
if (event=1) then loglikD = log(h) + log(S);
if (event=0) then loglikD = log(S);

* loglik for Y;
lin_Y =  a1*age + a2*albumin_avg + a3*pre_sbp_avg + a4*nlr_avg
        + a5*post_bmi_avg + a6*male + a7*idwg_percent_avg + a8*ektv_avg
        + a9*vintage +a10*race_white +a11*race_black + eta*v;

*w_T = log(time/700);
```

```
w_T = c0 + c1*dw_1 + c2*dw_2 + c3*dw_3 + c4*dw_4 + c5*dw_5 + c6*dw_6;

mu = exp(w_T + lin_Y);
loglikY = nu*(-hosp_total_days/100/mu-log(mu))
         + nu*log(hosp_total_days/100)+nu*log(nu)-log(gamma(nu));

loglik = loglikD + loglikY;

model time ~ general(loglik);
random v ~ normal(0,sigma2) subject=patient_id out=JM.days22_nu;
ods output ParameterEstimates=JM.days22_est FitStatistics=JM.days22_fit;
run;
```

Appendix C

SAS Code Used in Chapter 3

The following SAS code is for the joint modeling of hospitalization and readmission in Chapter 3:

```
proc nlmixed data=rehosp.dat_oneyear maxiter=30000 tech=NRRIDG;
bounds sigma2 > 0 ;
parms a0=0.72 a1=0.15 a2=-0.17 a3=-0.14 a4=0.001
      a5=-0.03 a6=-0.10 a7=0.24 a8=-0.01 a9=0.19
          c1=0.04 c2=-0.13 c3=0.03 c4=-0.007
          c5=-0.02 c6=0.05 c7=-0.06 c8=-0.18 c9=0.11;

lambda_lin = exp_yrs_log + a0 + a1*txt + a2*prd + a3*(txt*prd) + a4*age
           + a5*vintage + a6*male + a7*DIABETIC + a8*comorbid_CHF
           + a9*comorbid_COPD + nv;
p_lin = b1*ns_1 + b2*ns_2 + c1*txt + c2*prd + c3*(txt*prd) + c4*age
      + c5*vintage + c6*male + c7*DIABETIC + c8*comorbid_CHF
      + c9*comorbid_COPD+ eta*nv;

loglikY = event_hosp*lambda_lin - exp(lambda_lin);
loglikZ = event_rehosp*p_lin - event_hosp*log(1+exp(p_lin));

loglik = loglikY + loglikZ;

model event_hosp ~ general(loglik);
random nv ~ normal(0,sigma2) subject=patient_id;
ods output ParameterEstimates=rehosp.dat_est FitStatistics=rehosp.dat_fit;
run;
```

Appendix D

R Code Used in Chapter 4

The following R code is for the variable selection in Chapter 4:

```
library(psych)
library(survival)
library(nlme)

# read data for longitudinal covariates
data.Y <- read.table("sample.y.csv", sep=",",header=TRUE)

# read data for survival time
data.T <- read.table("sample.t.csv", sep=",",header=TRUE)

# load initial values generated from marginal models
load("initial.rda")

varnames.fixed <- c("male","diabetic","age")

# main function to perform variable selection
lasso.JM <- function( maxiter=100, rtol=1e-6,
             lambda, lambda2, J=5, J1=3){
  res.alpha <- res.beta <- res.gamma <- res.gamma0 <- res.sigma.e2
    <- res.h0 <- res.Sigma <- res.negPL <- NULL
  idRange <- 1:nrow(data.T)

 # function to obtain b which maximize kappa(b)
 GS <- function() {
  b.out <- b
```

```
J <- ncol(b)/2
J1 <- length(alpha)/J

for (id in idRange){

  data <- data.Y[data.Y$id==id,]
  N <- nrow(data)
  X <- data[,colnames(data) %in% varnames.fixed]
  time <- data.T[data.T$id==id,]$time
  delta <- data.T[data.T$id==id,]$delta

  t <- data[data$y.name==1,]$y.t
  n <- length(t)
  Zi <- matrix(c(rep(1,n),t),n,2)
  for (j in 2:J) {
    t <- data[data$y.name==j,]$y.t
    n <- length(t)
    z <- matrix(c(rep(1,n),t),n,2)
    Zi <- superMatrix(list(Zi,z))
  }

  b.est <- rep(0,ncol(b))
  kbi.drv <- rep(1,ncol(b))

  GS.iter <- 0
  while (GS.iter < 10 & sum(kbi.drv^2) > 1e-10) {
    b.est.old <- b.est

    y.name.ind.all <- data$y.name
    alpha.sel.left.all <- J1*(y.name.ind.all-1)+1
    beta.sel.all <- 2*y.name.ind.all-1

    alpha.matrix <- matrix(rep(0,N*J1),nrow=N)
    for (i in 1:N)
      alpha.matrix[i,] <- alpha[(alpha.sel.left.all[i]):
                            (alpha.sel.left.all[i]+J1-1)]

    mu <- apply(X*alpha.matrix,1,sum) + beta[beta.sel.all]
       + beta[beta.sel.all+1]*data$y.t + Zi%*%b.est.old

    eta <- c(as.numeric(X[1,])%*%gamma0 + t(gamma)%*%b.est.old)
```

```
      kbi.drv <- t(data$y-mu)%*%Zi/sigma.e2 + delta*gamma
                 - h0*time*exp(eta)*gamma - t(b.est.old)%*%solve(Sigma)
      Hi <- -t(Zi)%*%Zi/sigma.e2 - c(h0*time*exp(eta))*gamma%*%t(gamma)
            - solve(Sigma)

      b.est <- b.est.old - solve(Hi)%*%t(kbi.drv)

      GS.iter <- GS.iter+1
    }

    b.out[rownames(b.out)==as.character(id),] <- b.est
  }
  b.out
}

# negative penalized likelihood

 negPL.fun <- function(alpha.est=alpha, beta.est=beta,
                gamma0.est=gamma0, gamma.est=gamma,
                h0.est=h0, sigma.e2.est=sigma.e2,
                Sigma.est=Sigma, b.est=b)
  {
  PL.neg <- 0
  for (id in 1:400) {
    data <- data.Y[data.Y$id==id,]
    N <- nrow(data)
    X <- data[,colnames(data) %in% varnames.fixed]
    time <- data.T[data.T$id==id,]$time
    delta <- data.T[data.T$id==id,]$delta

    t <- data[data$y.name==1,]$y.t
    n <- length(t)
    Zi <- matrix(c(rep(1,n),t),n,2)
    for (j in 2:J) {
      t <- data[data$y.name==j,]$y.t
      n <- length(t)
      z <- matrix(c(rep(1,n),t),n,2)
      Zi <- superMatrix(list(Zi,z))
    }
    mu <- sapply(1:N, function(i)
      c(as.numeric(X[i,])%*%alpha.est[(J1*(data$y.name[i]-1)+1)
      :(J1*data$y.name[i])]) + beta.est[2*data$y.name[i]-1]
```

```
      + beta.est[2*data$y.name[i]]*data$y.t[i]
      + b.est[rownames(b.est)==as.character(id),
      2*data$y.name[i]-1]
      + b.est[rownames(b.est)==as.character(id),
      2*data$y.name[i]]*data$y.t[i])
    eta <- c(as.numeric(X[1,])%*%gamma0.est)+c(gamma.est
           %*%b.est[rownames(b.est)==as.character(id),])
    kbi.neg <- N*log(sigma.e2.est)/2
      + c(t(data$y-mu)%*%(data$y-mu)/sigma.e2.est/2)
      - delta*log(h0.est) - delta*eta
      + h0.est*time*exp(eta) + log(det(Sigma.est)+0.0000001)/2
      + t(b.est[rownames(b.est)==as.character(id),])%*%
      solve(Sigma.est)%*%b.est[rownames(b.est)==as.character(id),]/2

    Hi <- -t(Zi)%*%Zi/sigma.e2.est - h0.est*time*exp(eta)*gamma.est
      %*%t(gamma.est) - solve(Sigma.est)
    PLi.neg <- kbi.neg+log(det(-Hi)+0.00000001)/2
    PL.neg <- PL.neg + PLi.neg
  }
  c(PL.neg) + lambda*sum(abs(gamma.est))*length(idRange) +
    lambda2*sum(abs(Sigma.est[upper.tri(Sigma.est)]))*length(idRange)
}

  data.X <- data.Y[!duplicated(data.Y$id),]
Z <- NULL
for (id in idRange){
  data <- data.Y[data.Y$id==id,]
  N <- nrow(data)

  t <- data[data$y.name==1,]$y.t
  n <- length(t)
  Zi <- matrix(c(rep(1,n),t),n,2)
  for (j in 2:J) {
    t <- data[data$y.name==j,]$y.t
    n <- length(t)
    z <- matrix(c(rep(1,n),t),n,2)
    Zi <- superMatrix(list(Zi,z))
  }
  Z <- rbind(Z, Zi)
}

alpha.hist <- alpha.old <- alpha <- alpha.ini
```

```
beta.hist <- beta.old <- beta <- beta.ini

sigma.e2.hist <- sigma.e2.old <- sigma.e2 <- sigma.e2.ini

Sigma <- Sigma.ini

gamma0.hist <- gamma0.old <- gamma0 <- gamma0.ini
gamma.hist <- gamma.old <- gamma <- gamma.ini

h0.hist <- h0.old <- h0 <- 1

b <- b.ini

negPL.hist <- negPL <- negPL.fun()

k <- 0
rdiff <- 1
scale <- 1/2
maxiter.opt <- 10

# Coordinate Decent Algorithm
while (k < maxiter & rdiff > rtol) {

  b <- GS()

  alpha.old <- alpha
  beta.old <- beta
  gamma0.old <- gamma0
  gamma.old <- gamma
  h0.old <- h0
  Sigma.old <- Sigma
  sigma.e2.old <- sigma.e2
  negPL.old <- negPL

  f <- f.new <- negPL.fun()

  # update h0
  drv1 <- 0

  drv1 <- sum(unlist(lapply(idRange, function(id) {
    data.Y.row.sel <- data.Y$id==id
    data <- data.Y[data.Y.row.sel,]
```

```
  N <- nrow(data)
  X <- data[,colnames(data) %in% varnames.fixed]

  data.T.row.sel <- data.T$id==id
  time <- data.T[data.T.row.sel,]$time
  delta <- data.T[data.T.row.sel,]$delta

  t <- data[data$y.name==1,]$y.t
  n <- length(t)
  Zi <- matrix(c(rep(1,n),t),n,2)
  for (j in 2:J) {
    t <- data[data$y.name==j,]$y.t
    n <- length(t)
    z <- matrix(c(rep(1,n),t),n,2)
    Zi <- superMatrix(list(Zi,z))
  }
  eta <- c(as.numeric(X[1,])%*%gamma0
    + gamma%*%b[rownames(b)==as.character(id),])
  H <- -t(Zi)%*%Zi/sigma.e2
    - h0*time*exp(eta)*gamma%*%t(gamma) - solve(Sigma)
  -tr(solve(H)%*%gamma%*%t(gamma))*h0*time*exp(eta)/2
})))

drv2 <- sum(-data.T$delta/h0 + data.T$time*
  exp(as.matrix(data.X[,colnames(data.X)%in%varnames.fixed])
  %*%gamma0+b%*%gamma))
drv <- drv1 + drv2

if (drv > 1e-6) {
  iter <- 0
  step <- 1*0.5^k
  h0.update <- h0 - step*drv
  while (h0.update < 0) {
    step <- scale*step
    iter <- iter + 1
    h0.update <- h0 - step*drv
  }
  if (abs(step*drv) >1e-5) {
    f.new <- negPL.fun(h0.est = h0.update)
    while ( abs(step*drv) >1e-5 & f - f.new  < step*drv^2/4) {
      step <- scale*step
      iter <- iter+1
```

```
      h0.update <- h0 - step*drv
      f.new <- negPL.fun(h0.est = h0.update)
    }
  }

  h0 <- ifelse(abs(step*drv)<1e-5, h0, h0 - step*drv)
  f <- ifelse(abs(step*drv)<1e-5, f, f.new)
}

print(paste("h0",h0,",","iter",iter))

# update gamma

drv2 <- -t(data.T$delta)%*%b + t(h0*data.T$time
  *exp(as.matrix(data.X[,colnames(data.X)%in%varnames.fixed])%*%gamma0
  + b%*%gamma))%*%b

for (i in 1:length(gamma)) {

  drv1 <- 0
  for (id in idRange){
    data <- data.Y[data.Y$id==id,]
    N <- nrow(data)
    X <- data[,colnames(data) %in% varnames.fixed]
    time <- data.T[data.T$id==id,]$time
    delta <- data.T[data.T$id==id,]$delta

    t <- data[data$y.name==1,]$y.t
    n <- length(t)
    Zi <- matrix(c(rep(1,n),t),n,2)
    for (j in 2:J) {
      t <- data[data$y.name==j,]$y.t
      n <- length(t)
      z <- matrix(c(rep(1,n),t),n,2)
      Zi <- superMatrix(list(Zi,z))
    }
    eta <- c(as.numeric(X[1,])%*%gamma0
      + gamma%*%b[rownames(b)==as.character(id),])
    H <- -t(Zi)%*%Zi/sigma.e2
      - h0*time*exp(eta)*gamma%*%t(gamma) - solve(Sigma)
    ind <- numeric(length(gamma))
    ind[i] <- 1
```

```
    drv1i <- -tr(solve(H)%*%(gamma%*%t(gamma)
      *c(b[rownames(b)==as.character(id),i])
      +2*ind%*%t(gamma)))*h0*time*exp(eta)/2
    drv1 <- drv1 + drv1i
  }

  drv <- drv1 + drv2[i] + sign(gamma[i])*lambda*length(idRange)

  if (abs(drv)>1e-5) {
    iter <- 0
    step <- 1*0.5^k
    gamma.update <- replace(gamma,i,gamma[i]-step*drv)
    f.new <- negPL.fun(gamma.est=gamma.update)
    while (is.na(f.new)) {
      step <- scale*step
      iter <- iter + 1
      gamma.update <- replace(gamma,i,gamma[i]-step*drv)
      f.new <- negPL.fun(gamma.est=gamma.update)
    }

    while ( abs(step*drv)>1e-5 & f - f.new < step*drv^2/4) {
      step <- scale*step
      iter <- iter+1
      f.new <- negPL.fun(gamma.est=replace(gamma,i,gamma[i]-step*drv))
    }
    gamma[i] <- ifelse(abs(step*drv)<1e-5, gamma[i], gamma[i]-step*drv)
    f <- ifelse(abs(step*drv)<1e-5, f, f.new)

  }
  print(paste("i",i))
}


print(paste("gamma",gamma,",","iter",iter))


# update alpha

mu <- sapply(1:nrow(data.Y), function(n){
  y.name.ind <- data.Y$y.name[n]
  alpha.sel <- (J1*(y.name.ind-1)+1):(J1*y.name.ind)
  beta.sel <- 2*y.name.ind-1
```

```
    b.row.sel <- rownames(b)==as.character(data.Y$id[n])
    as.matrix(data.Y[n, colnames(data.Y)%in%varnames.fixed])
      %*%alpha[alpha.sel] + beta[beta.sel] + beta[beta.sel+1]
      *data.Y$y.t[n] + b[b.row.sel,beta.sel]
      + b[b.row.sel,beta.sel+1]*data.Y$y.t[n]
  } )
  if (sum(abs(data.Y$y-mu)) > 1e-6) {
    for (idx1 in 1:(length(alpha))) {
      x.idx1 <- ifelse(idx1%%J1==0, J1, idx1%%J1)
      y.idx1 <- (idx1-x.idx1)/J1+1

      iter <- 0

      data.tmp <- data.Y[data.Y$y.name==y.idx1,]

      drvdrv <- sum(data.Y[data.Y$y.name==y.idx1,
        colnames(data.tmp)==varnames.fixed[x.idx1]]^2)/sigma.e2

      alpha.est <- alpha
      drv <- 1

      while (iter < 10 & sum(drv^2) > 1e-10) {
        mu <- sapply(1:nrow(data.tmp), function(n){
          y.name.ind <- data.tmp$y.name[n]
          alpha.sel <- (J1*(y.name.ind-1)+1):(J1*y.name.ind)
          beta.sel <- 2*y.name.ind-1
          b.row.sel <- rownames(b)==as.character(data.tmp$id[n])
          as.matrix(data.tmp[n,colnames(data.Y)%in%varnames.fixed])
            %*%alpha.est[alpha.sel] + beta[beta.sel] + beta[beta.sel+1]
            *data.tmp$y.t[n] + b[b.row.sel,beta.sel]
            + b[b.row.sel,beta.sel+1]*data.tmp$y.t[n]
        } )

        drv <- -sum(((data.tmp$y-mu)*as.matrix(data.tmp[,
          colnames(data.tmp)==varnames.fixed[x.idx1]])))/sigma.e2

        alpha.est.old <- alpha.est
        alpha.est[idx1] <- alpha.est.old[idx1] - drv/drvdrv

        iter <- iter+1
      }
      alpha <- alpha.est
```

```
  }
}

print(paste("alpha",alpha,",","iter",iter))


# update beta

iter <- 0

mu <- sapply(1:nrow(data.Y), function(n){
  y.name.ind <- data.Y$y.name[n]
  alpha.sel <- (J1*(y.name.ind-1)+1):(J1*y.name.ind)
  beta.sel <- 2*y.name.ind-1
  b.row.sel <- rownames(b)==as.character(data.Y$id[n])
  as.matrix(data.Y[n,colnames(data.Y)%in%varnames.fixed])
    %*%alpha[alpha.sel] + beta[beta.sel] + beta[beta.sel+1]
    *data.Y$y.t[n] + b[b.row.sel,beta.sel]
    + b[b.row.sel,beta.sel+1]*data.Y$y.t[n]
} )

drv <- -t(as.matrix(data.Y$y-mu))%*%Z/sigma.e2
H <- t(Z)%*%Z/sigma.e2

beta.est <- beta

while (iter < 10 & sum(abs(drv)) > 1e-6) {
  mu <- sapply(1:nrow(data.Y), function(n){
    y.name.ind <- data.Y$y.name[n]
    alpha.sel <- (J1*(y.name.ind-1)+1):(J1*y.name.ind)
    beta.sel <- 2*y.name.ind-1
    b.row.sel <- rownames(b)==as.character(data.Y$id[n])
    as.matrix(data.Y[n,colnames(data.Y)%in%varnames.fixed])
      %*%alpha[alpha.sel] + beta.est[beta.sel] + beta.est[beta.sel+1]
      *data.Y$y.t[n] + b[b.row.sel,beta.sel]
      + b[b.row.sel,beta.sel+1]*data.Y$y.t[n]
  } )
  drv <- -t(data.Y$y-mu)%*%Z/sigma.e2

  beta.est.old <- beta.est
  beta.est <- beta.est.old - solve(H)%*%t(drv)
```

```
    iter <- iter+1
  }

  print(paste("beta",beta,",","iter",iter))

# update sigma.e2

  iter <- 0

  drv1 <- sum(unlist(lapply(idRange, function(id) {
    data.Y.row.sel <- data.Y$id==id
    data <- data.Y[data.Y.row.sel,]
    N <- nrow(data)
    X <- as.matrix(data[,colnames(data)%in%varnames.fixed])
    data.T.row.sel <- data.T$id==id
    time <- data.T[data.T.row.sel,]$time
    delta <- data.T[data.T.row.sel,]$delta

    t <- data[data$y.name==1,]$y.t
    n <- length(t)
    Zi <- matrix(c(rep(1,n),t),n,2)
    for (j in 2:J) {
      t <- data[data$y.name==j,]$y.t
      n <- length(t)
      z <- matrix(c(rep(1,n),t),n,2)
      Zi <- superMatrix(list(Zi,z))
    }
    eta <- c(as.numeric(X[1,])%*%gamma0
      + gamma%*%b[rownames(b)==as.character(id),])
    H <- -t(Zi)%*%Zi/sigma.e2
      - h0*time*exp(eta)*gamma%*%t(gamma) - solve(Sigma)

    drv1i <- tr(solve(H)%*%t(Zi)%*%Zi)/(sigma.e2^2)/2
    drv1i
  })))

  mu <- sapply(1:nrow(data.Y), function(n){
    y.name.ind <- data.Y$y.name[n]
    alpha.sel <- (J1*(y.name.ind-1)+1):(J1*y.name.ind)
    beta.sel <- 2*y.name.ind-1
    b.row.sel <- rownames(b)==as.character(data.Y$id[n])
    as.numeric(data.Y[n,colnames(data.Y)%in%varnames.fixed])
```

```
     %*%alpha[alpha.sel] + beta[beta.sel]
     + beta[beta.sel+1]*data.Y$y.t[n] + b[b.row.sel,beta.sel]
     + b[b.row.sel,beta.sel+1]*data.Y$y.t[n]
  })
  drv2 <- nrow(data.Y)*1/sigma.e2/2
       - c(t(data.Y$y-mu)%*%(data.Y$y-mu)/(sigma.e2^2*2))
  drv <- drv1 + drv2

  if (sum(abs(drv)) > 1e-5) {
    if (f - negPL.fun(sigma.e2.est = sigma.e2-sign(drv)*1e-5)
       > (1e-5)^2/4)
      {

      iter <- 0
      step <- 0.01*0.5^k
      sigma.e2.update <- sigma.e2 - step*drv
      while (sigma.e2.update < 0) {
        step <- scale*step
        iter <- iter + 1
        sigma.e2.update <- sigma.e2 - step*drv
      }
      if (abs(step*drv) >1e-5) {
        f.new <- negPL.fun(sigma.e2.est=sigma.e2.update)
        while ( abs(step*drv) >1e-5 & f - f.new < step*drv^2/4) {
          step <- scale*step
          iter <- iter+1
          sigma.e2.update <- sigma.e2 - step*drv
          f.new <- negPL.fun(sigma.e2.est = sigma.e2.update)
        }
      }

      sigma.e2 <- ifelse(abs(step*drv)<1e-5, sigma.e2, sigma.e2.update)
      f <- ifelse(abs(step*drv)<1e-5, f, f.new)
    }
  }

  print(paste("sigma.e2",sigma.e2,",","iter",iter))

# update gamma0

  drv1 <- rep(0,length(gamma0))
  drv1.list <- lapply(idRange, function(id){
```

```
  data <- data.Y[data.Y$id==id,]
  N <- nrow(data)
  X <- as.matrix(data[,colnames(data)%in%varnames.fixed])
  time <- data.T[data.T$id==id,]$time
  delta <- data.T[data.T$id==id,]$delta

  t <- data[data$y.name==1,]$y.t
  n <- length(t)
  Zi <- matrix(c(rep(1,n),t),n,2)
  for (j in 2:J) {
    t <- data[data$y.name==j,]$y.t
    n <- length(t)
    z <- matrix(c(rep(1,n),t),n,2)
    Zi <- superMatrix(list(Zi,z))
  }
  eta <- c(as.numeric(X[1,])%*%gamma0
   + gamma%*%b[rownames(b)==as.character(id),])
  H <- -t(Zi)%*%Zi/sigma.e2
   - h0*time*exp(eta)*gamma%*%t(gamma) - solve(Sigma)
  drv1i <- -tr(solve(H)%*%gamma%*%t(gamma))*h0*time*exp(eta)*X[1,]/2
  drv1i
})
drv1 <- apply(matrix(unlist(drv1.list),
  nrow=length(gamma0), byrow=TRUE), 1, sum)

drv2 <- -t(data.T$delta)%*%
 as.matrix(data.X[,colnames(data.X)%in%varnames.fixed]) + t(h0
 *data.T$time*exp(as.matrix(data.X[,colnames(data.X)%in%varnames.fixed])
 %*%gamma0+b%*%gamma))%*%
 as.matrix(data.X[,colnames(data.X)%in%varnames.fixed])

drv <- drv1 + drv2


for (i in 1:length(gamma0)) {
  if (abs(drv[i]) > 1e-5) {
    iter <- 0
    if (f - negPL.fun(gamma0.est = replace(gamma0,
        i,gamma0[i]-sign(drv[i])*1e-5)) > (1e-5)^2/4)
     {

      step <- 0.01*0.5^k
```

```
      while ( abs(step*drv[i])>1e-5 & f - f.new  < step*drv[i]^2/4){
        step <- scale*step
        iter <- iter+1
        f.new <- negPL.fun(gamma0.est = replace(gamma0,
        i,gamma0[i]-step*drv[i]))
      }

      gamma0[i] <- ifelse(abs(step*drv[i])<1e-5,
        gamma0[i], gamma0[i] - step*drv[i])
      f <- ifelse(abs(step*drv[i])<1e-5, f, f.new)
    }
  }
  print(paste("i",i))
}

print(paste("gamma0",gamma0,",","iter",iter))


# update Sigma
system.time({
  iter <- 0
  for (i in 1:ncol(b)) {
    for (j in 1:i){

      Sigma.inv <- solve(Sigma)
      Sigma.drv <- matrix(numeric(ncol(b)*ncol(b)), ncol=ncol(b))
      Sigma.drv[i,j] <- Sigma.drv[j,i] <- 1

      drv <- sum(unlist(lapply(idRange, function(id){
        data.Y.row.sel <- data.Y$id==id
        data <- data.Y[data.Y.row.sel,]
        N <- nrow(data)
        X <- as.matrix(data[,colnames(data)%in%varnames.fixed])
        data.T.row.sel <- data.T$id==id
        time <- data.T[data.T.row.sel,]$time
        delta <- data.T[data.T.row.sel,]$delta

        t <- data[data$y.name==1,]$y.t
        n <- length(t)
        Zi <- matrix(c(rep(1,n),t),n,2)
```

```
    for (jj in 2:J) {
      t <- data[data$y.name==jj,]$y.t
      n <- length(t)
      z <- matrix(c(rep(1,n),t),n,2)
      Zi <- superMatrix(list(Zi,z))
    }
    eta <- c(as.numeric(X[1,])%*%gamma0
      + gamma%*%b[rownames(b)==as.character(id),])
    H <- -t(Zi)%*%Zi/sigma.e2
      - h0*time*exp(eta)*gamma%*%t(gamma) - Sigma.inv

    drv1i <- tr(Sigma.inv%*%Sigma.drv)/2 -
    t(b[rownames(b)==as.character(id),])%*%Sigma.inv%*%Sigma.drv
    %*%Sigma.inv%*%b[rownames(b)==as.character(id),]/2

    drv2i <- tr(solve(H)%*%Sigma.inv%*%Sigma.drv%*%Sigma.inv)/2
    drv1i + drv2i
  })))

drv <- drv
 + (i!=j)*lambda2*sign(Sigma[i,j])*length(idRange)*abs(Sigma[i,j])


 if (abs(drv) > 1e-5) {
  if (f - negPL.fun(Sigma.est = replace(Sigma,
     c((j-1)*ncol(b)+i,(i-1)*ncol(b)+j),Sigma[i,j]-sign(drv)*1e-5))
     > (1e-5)^2/4)
  {
     iter <- 0
     step <- 0.01*0.5^k
     if (abs(step*drv) > 1e-5 ) {

      Sigma.update <- replace(Sigma,
        c((j-1)*ncol(b)+i,(i-1)*ncol(b)+j),Sigma[i,j]-step*drv)
      while (det(Sigma.update) < 0) {
        step <- scale*step
        iter <- iter + 1
        Sigma.update <- replace(Sigma,
          c((j-1)*ncol(b)+i,(i-1)*ncol(b)+j),Sigma[i,j]-step*drv)
       }

      f.new <- negPL.fun(Sigma.est = Sigma.update)
```

```
              while ( abs(step*drv) > 1e-5 & f - f.new < step*drv^2/4) {
                step <- scale*step
                iter <- iter+1
                Sigma.update <- replace(Sigma,
                  c((j-1)*ncol(b)+i,(i-1)*ncol(b)+j),Sigma[i,j]-step*drv)
                f.new <- negPL.fun(Sigma.est = Sigma.update)
              }
             }

            Sigma[i,j] <- Sigma[j,i] <-
            ifelse(abs(step*drv) < 1e-5, Sigma[i,j], Sigma[i,j]-step*drv)

            f <- ifelse(abs(step*drv)<1e-5, f, f.new)

          }
        }
        print(paste("i",i,"j",j))
      }
    }
    print(Sigma)})


  negPL <- f

  rdiff <- c(abs(negPL - negPL.old)/(abs(negPL.old)+0.000001))
  k <- k+1

  print(paste("rdiff",rdiff,",","iter",k-1))

  print(list(alpha=alpha, beta=beta, gamma=gamma, gamma0=gamma0,
       Sigma=Sigma, sigma.e2=sigma.e2, h0=h0, negPL=negPL,
       rdiff=rdiff, iter=k))

}
list(alpha=alpha, beta=beta, gamma=gamma, gamma0=gamma0,
     Sigma=Sigma, sigma.e2=sigma.e2, h0=h0, negPL=negPL,
     rdiff=rdiff, iter=k)
}

lasso.JM(lambda=0.005, lambda2=0.1)
```

Bibliography

[1] Hirotugu Akaike. Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory*, pages 267–281. Akademinai Kiado, 1973.

[2] Howard D Bondell, Arun Krishna, and Sujit K Ghosh. Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, 66(4):1069–1077, 2010.

[3] Daniel J Brotman, Lori D Bash, Rehan Qayyum, Deidra Crews, Eric A Whitsel, Brad C Astor, and Josef Coresh. Heart rate variability predicts ESRD and CKD-related hospitalization. *Journal of the American Society of Nephrology*, 21(9):1560–1570, 2010.

[4] Elizabeth R Brown, Joseph G Ibrahim, and Victor DeGruttola. A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73, 2005.

[5] Peter Lukas Bühlmann, Sara A van de Geer, and Sara Van de Geer. *Statistics for high-dimensional data*. Springer, 2011.

[6] P. J. Catalano. Bivariate modeling of clustered continuous and ordered categorical outcomes. *Statistics in Medicine*, 16(8):883–900, 1997.

[7] Allan J Collins, Robert N Foley, Blanche Chavers, David Gilbertson, Charles Herzog, Areef Ishani, Kirsten Johansen, Bertram L Kasiske, Nancy Kutner, Jiannong Liu, et al. US renal data system 2013 annual data report. *American Journal of Kidney Diseases*, 63(1 Suppl):A7, 2014.

[8] Josef Coresh, Brad C Astor, Tom Greene, Garabed Eknoyan, and Andrew S Levey. Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third national health and nutrition examination survey. *American Journal of Kidney Diseases*, 41(1):1–12, 2003.

[9] Victor De Gruttola and Xin Ming Tu. Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, 50(4):1003–1014, 1994.

[10] Alexander R De Leon and Keumhee Carrière Chough. *Analysis of Mixed Data: Methods & Applications*. CRC Press, 2013.

[11] Penelope H Dennehy, Margaret M Cortese, Rodolfo E Bégué, Jenifer L Jaeger, Nancy E Roberts, Rongping Zhang, Philip Rhodes, John Gentsch, Richard Ward, David I Bernstein, et al. A case-control study to determine risk factors for hospitalization for rotavirus gastroenteritis in us children. *The Pediatric Infectious Disease Journal*, 25(12):1123–1131, 2006.

[12] David B Dunson and Amy H Herring. Bayesian latent variable models for mixed discrete outcomes. *Biostatistics*, 6(1):11–25, 2005.

[13] Mark D Eisner, Patricia P Katz, Edward H Yelin, Stephen C Shiboski, and Paul D Blanc. Risk factors for hospitalization among adults with asthma: the influence of sociodemographic factors and asthma severity. *Respiratory Research*, 2(1):53–60, 2000.

[14] Emre Erdem, C Kaya, A Karata^{o}, Melda Dilek, and Tekin Akpolat. Neutrophil to lymphocyte ratio in predicting short-term mortality in hemodialysis patients. *Journal of Experimental & Clinical Medicine*, 30:129–32, 2013.

[15] Jianqing Fan, Gang Li, and Runze Li. An overview on variable selection for survival analysis. *Contemporary Multivariate Analysis and Design of Experiments*, pages 315–336, 2005.

[16] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[17] Cheryl L Faucett and Duncan C Thomas. Simultaneously modeling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–1685, 1996.

[18] Garrett M Fitzmaurice and Nan M Laird. Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American statistical Association*, 90(431):845–852, 1995.

[19] Dean P Foster and Edward I George. The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22(4):1947–1975, 1994.

[20] Jerome Friedman, Trevor Hastie, Holger Höfling, Robert Tibshirani, et al. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332, 2007.

[21] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Satistical Software*, 33(1):1–22, 2010.

[22] Wenjiang J Fu. Penalized regressions: the bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7(3):397–416, 1998.

[23] Luke O Hansen, Robert S Young, Keiki Hinami, Alicia Leung, and Mark V Williams. Interventions to reduce 30-day rehospitalization: a systematic review. *Annals of Internal Medicine*, 155(8):520–528, 2011.

[24] Patricia L Harrison, Pamela A Hara, James E Pope, Michelle C Young, and Elizabeth Y Rula. The impact of postdischarge telephonic follow-up on hospital readmissions. *Population Health Management*, 14(1):27–32, 2011.

[25] Jiang He and Paul K Whelton. Elevated systolic blood pressure and risk of cardiovascular and renal disease: overview of evidence from observational epidemiologic studies and randomized controlled trials. *American Heart Journal*, 138(3):S211–S219, 1999.

[26] Zangdong He, Wanzhu Tu, Sijian Wang, Haoda Fu, and Zhangsheng Yu. Simultaneous variable selection for joint models of longitudinal and survival outcomes. *Biometrics*, 71(1):178–187, 2015.

[27] Robin Henderson, Peter Diggle, and Angela Dobson. Joint modeling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

[28] Chi-Yuan Hsu, Charles E McCulloch, Carlos Iribarren, Jeanne Darbinian, and Alan S Go. Body mass index and risk for end-stage renal disease. *Annals of Internal Medicine*, 144(1):21–28, 2006.

[29] C-Y Huang, J Qin, and M-C Wang. Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics*, 66(1):39–49, 2010.

[30] Xuelin Huang and Robert A Wolfe. A frailty model for informative censoring. *Biometrics*, 58(3):510–520, 2002.

[31] Joseph G Ibrahim, Ming-Hui Chen, and Debajyoti Sinha. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, 14(3):863–884, 2004.

[32] Joseph G Ibrahim, Haitao Chu, and Liddy M Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796–2801, 2010.

[33] Joseph G Ibrahim, Hongtu Zhu, Ramon I Garcia, and Ruixin Guo. Fixed and random effects selection in mixed effects models. *Biometrics*, 67(2):495–503, 2011.

[34] JK Inrig, EZ Oddone, V Hasselblad, Barbara Gillespie, UD Patel, D Reddan, R Toto, J Himmelfarb, JF Winchester, J Stivelman, et al. Association of intradialytic blood pressure changes with hospitalization and mortality rates in prevalent ESRD patients. *Kidney International*, 71(5):454–461, 2007.

[35] Stephen F Jencks, Mark V Williams, and Eric A Coleman. Rehospitalizations among patients in the medicare fee-for-service program. *New England Journal of Medicine*, 360(14):1418–1428, 2009.

[36] Michel P Kossovsky, Thomas V Perneger, François P Sarasin, Filippo Bolla, François Borst, and Jean-Michel Gaspoz. Comparison between planned and unplanned readmissions to a department of internal medicine. *Journal of Clinical Epidemiology*, 52(2):151–156, 1999.

[37] Tony Lancaster and Orna Intrator. Panel data with survival: hospitalization of hiv-positive patients. *Journal of the American Statistical Association*, 93(441):46–53, 1998.

[38] Ramon C Littell, Walter W Stroup, George A Milliken, Russell D Wolfinger, and Oliver Schabenberger. *SAS for mixed models*. SAS Institute, 2006.

[39] Lei Liu and Xuelin Huang. 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 (Applied Statistics)*, 58(1):65–81, 2009.

[40] Lei Liu, Robert A Wolfe, and Xuelin Huang. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3):747–756, 2004.

[41] Colin L Mallows. Some comments on C_p. *Technometrics*, 15(4):661–675, 1973.

[42] Charles McCulloch. Joint modelling of mixed outcome types using latent variables. *Statistical Methods in Medical Research*, 17(1):53–73, 2008.

[43] Scot E Moss, Ronald Klein, and Barbara EK Klein. Risk factors for hospitalization in people with diabetes. *Archives of Internal Medicine*, 159(17):2053–2057, 1999.

[44] Anne Pfuntner, Lauren M Wier, and Claudia Steiner. *Costs for hospital stays in the United States, 2011*. Agency for Health Care Policy and Research (US), 2013.

[45] Paul J Phelan, Patrick O'Kelly, Joseph J Walshe, and Peter J Conlon. The importance of serum albumin and phosphorous as predictors of mortality in esrd patients. *Renal Failure*, 30(4):423–429, 2008.

[46] William H Press. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press, 2007.

[47] James O Ramsay. Monotone regression splines in action. *Statistical Science*, 3(4):425–441, 1988.

[48] Dimitris Rizopoulos. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33, 2010.

[49] Mary Dupuis Sammel, Louise M Ryan, and Julie M Legler. Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):667–678, 1997.

[50] Rajiv Saran, Yi Li, Bruce Robinson, Kevin C Abbott, LY Agodoa, John Ayanian, Jennifer Bragg-Gresham, Rajesh Balkrishnan, JL Chen, Elizabeth Cope, et al. US renal data system 2015 annual data report: epidemiology of kidney disease in the united states. *American Journal of Kidney Diseases*, 67(3 suppl 1):S1–S434, 2016.

[51] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[52] Shirish Krishnaj Shevade and S Sathiya Keerthi. A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, 19(17):2246–2253, 2003.

[53] Anne Dichmann Sorknæs, Hanne Madsen, Jesper Hallas, Peder Jest, and Michael Hansen-Nord. Nurse tele-consultations with discharged copd patients reduce early readmissions–an interventional study. *The Clinical Respiratory Journal*, 5(1):26–34, 2011.

[54] Els Steeman, Philip Moons, Koen Milisen, Nele De Bal, Sabina De Geest, Cécile De Froidmont, Véronique Tellier, Christiane Gosset, and Ivo Abraham. Implementation of discharge management for geriatric patients at risk of readmission or institutionalization. *International Journal for Quality in Health Care*, 18(5):352–358, 2006.

[55] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[56] AA Tsiatis, Victor Degruttola, and MS Wulfsohn. 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(429):27–37, 1995.

[57] Anastasios A Tsiatis and Marie Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458, 2001.

[58] Anastasios A Tsiatis and Marie Davidian. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834, 2004.

[59] Len A Usvyat, Jeroen P Kooman, Frank M van der Sande, Yuedong Wang, Franklin W Maddux, Nathan W Levin, and Peter Kotanko. Dynamics of hospitalizations in hemodialysis patients: results from a large us provider. *Nephrology Dialysis Transplantation*, 29(2):442–448, 2014.

[60] Mei-Cheng Wang, Jing Qin, and Chin-Tsang Chiang. Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, 96(455):1057–1065, 2001.

[61] Yan Wang and Jeremy M G Taylor. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455):895–905, 2001.

[62] MS Wulfsohn and AA Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339, 1997.

[63] Teng Zhang and Hui Zou. Sparse precision matrix estimation via lasso penalized d-trace loss. *Biometrika*, 101:103–120, 2014.

[64] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Satistical Association*, 101(476):1418–1429, 2006.

[65] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.