Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jae Eui Soh Date

Nonparametric Regression for Assessing Time-Varying Effects in Survival Analysis

By

Jae Eui Soh
Doctor of Philosophy

Biostatistics and Bioinformatics

Yijian (Eugene) Huang, Ph.D.
Advisor

David H. Howard, Ph.D.
Committee Member

Robert H. Lyles, Ph.D.
Committee Member

Limin Peng, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Nonparametric Regression for Assessing Time-Varying Effects in Survival Analysis

By

Jae Eui Soh
M.A., Yonsei University, 2013
B.A., Yonsei University, 2011

Advisor: Yijian (Eugene) Huang, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics and Bioinformatics
2019

Abstract

Nonparametric Regression for Assessing Time-Varying Effects in Survival Analysis
By Jae Eui Soh

Most regression models in survival analysis tacitly assume constant effects of covariates on event times. However, this assumption may not always be realistic in practice. In a clinical study for AIDS patients, for example, a treatment might take time to reach its full efficacy rather than right after randomization; meanwhile, the treatment effect might also erode over time as drug resistance develops, e.g., Eshleman et al. (2001). In this dissertation, we present three projects to develop regression models that accommodate time-varying effects of covariates; two projects are for the analysis of recurrent events, and the third one is for the analysis of univariate survival data.

In the first project, we propose a varying-coefficient model for the mean frequency of recurrent events. We develop an estimation procedure that fully exploits observed data, and a resampling-based inference procedure. Consistency and weak convergence of the proposed estimator are established. Simulation studies demonstrate utility of the estimator with practical sample sizes. Two real data analyses are presented for illustration of the proposed method.

Most models for recurrent events consider the study-time scale, but gap times between recurrent events are of natural interest in many applications. The second project is concerned with the gap-time scale of recurrent events, and we propose a marginal varying-coefficient model for the cumulative hazard function of the gap time. Estimation and inference procedures are developed. We establish consistency and weak convergence of the proposed estimator, and Monte Carlo simulations demonstrate utility of the proposed estimator. An analysis of the bladder tumor trial data is presented for illustration.

In the third project, we propose a semiparametric survival regression model for the analysis of univariate survival data. With a mixture of time-varying and constant effects of covariates, the proposed model generalizes the proportional hazards model of Cox (1972), while being a sub-model of the temporal survival regression of Peng and Huang (2007). We develop an iterative estimation procedure and an inference procedure. Extensive simulations are conducted to assess finite-sample behaviors of the proposed estimator. The proposed method is illustrated by an analysis of the Veterans' Administration lung cancer trial data.

Nonparametric Regression for Assessing Time-Varying Effects in Survival Analysis

By

Jae Eui Soh
M.A., Yonsei University, 2013
B.A., Yonsei University, 2011

Advisor: Yijian (Eugene) Huang, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics and Bioinformatics
2019

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

1.1 Time-To-Event Data

Survival Time

Lifetime events such as death and disease have been of natural interest in human history. Endeavors to understand such random events and furthermore to establish risk factors have been primary pursuit in quantitative research. In a clinical trial, for example, to assess an experimental drug's preventive effect on death may be of primary interest. In health services research, public health practitioners would want to understand what socioeconomic risk factors affect patients' access to health care services.

In this monograph, we restrict our attention to random events with a well-defined follow-up time origin, e.g., treatment randomization in a clinical trial and initial disease diagnosis date for a patient. In this way, an event occurrence can be represented in terms of the time elapsed from the initiation of follow-up to the event of interest. Regardless of one's favor on an event of interest, the elapsed time is conventionally called *survival time* in survival analysis, thus the term will be used in the sequel without loss of generality. For example, the times to wedding, graduation, cancer,

or death would be understood and analyzed in the same framework using the term 'survival' time.

Censored Survival Time Data

Since 'waiting' is needed to observe an event occurrence, survival data often come in a mixture of complete and incomplete survival times. By the incomplete survival times, we mean the right-censored survival times, which are not actual event times but 'censored' times. Reasons of right censoring include the end of study and patient drop-out before an event occurs. Standard statistical methods such as a t-test or ordinary regressions cannot be naively applied to censored survival data. This is because even the mean of a random survival time could not be estimated unbiasedly in the presence of an incomplete observation. For the analysis of censored survival data, we need right mathematical concepts and statistical methods, as will be discussed from the next Section.

Recurrent Events and Time Scale

Some events, so-called recurrent events, occur repeatedly over time. They are often observed in follow-up studies where a subject may experience multiple occurrences of the same event. Examples of recurrent events in biomedical research include series of heart attacks in patients with cardiovascular disease and opportunistic infections in AIDS patients.

Times to event occurrences, i.e., recurrence times, may be represented and analyzed on different time scales. Two most widely used time scales in the recurrent event literature are the study-time scale and the gap-time scale. The study-time scale is in the time elapsed from a study origin to the event of interest. On the other hand, the gap-time scale is in the time elapsed from a preceding event to the next event. Either time-scale may be preferred depending on research interest. For example, the

gap-time scale may be more appealing choice if research interest is in prediction of the next event.

Recurrent events are of primary interest in this dissertation. Two of the three projects in this monograph are for the analysis of recurrent events, and they are concerned with the study-time scale and the gap-time scale, respectively, as described in Chapters 2 and 3. On the other hand, the third project presented in Chapter 4 deals with univariate survival data.

1.1.1 Survival Time Example: the VA Lung Cancer Trial Data

The Veterans' Administration lung cancer trial data set was originally discussed in Prentice (1973) and analyzed by many researchers including Kalbfleisch and Prentice (2002) and Peng and Huang (2007). A total of 137 male patients with advanced inoperable lung cancer was randomly assigned to either a standard or a test chemotherapy. The time from randomization to death was recorded for each patient. A total of 128 events of death was observed, and the other 9 patients' survival times were censored. Heterogeneity in the patients was measured by some covariates including age in year, histological type of tumor, and Karnofsky score at enrollment that is a measure of performance status.

Figure 1.1 shows 5 complete and 5 censored survival times on the study-time scale. Heterogeneity across the patients in the sample is not adjusted in this simple display.

This data set is analyzed in Chapter 4 for method illustration.

1.1.2 Recurrence Time Example: the Bladder Tumor Trial Data

Byar (1980) reported a randomized clinical trial which was to evaluate the effects of

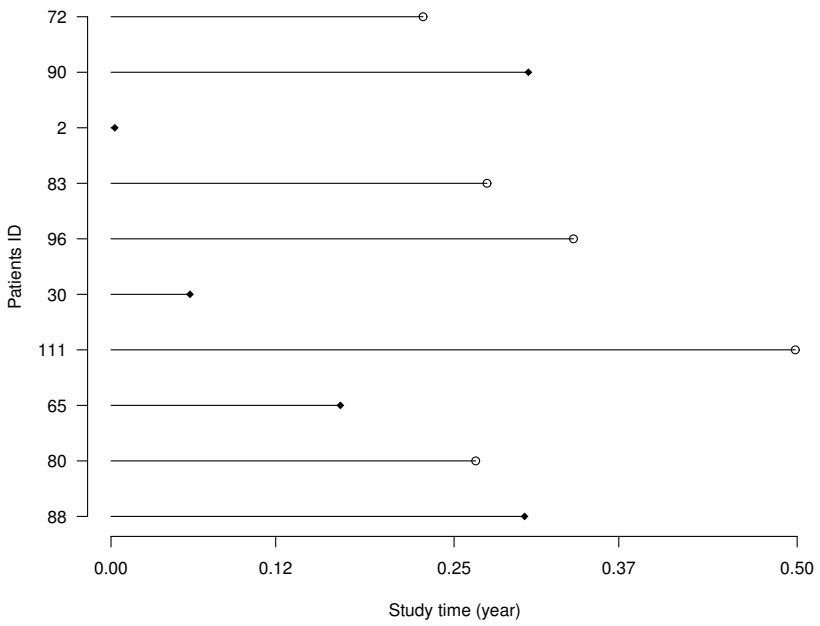

Figure 1.1: The Veterans Administration lung cancer trial data. Randomly selected 5 complete and 5 censored survival times on the study-time scale. Black dots and the empty circles indicate death and being censored, respectively.

new treatments on tumor recurrences. A total of 118 patients with superficial bladder tumors entered the study after removal of their initial tumors. The patients were randomly assigned to three treatment arms: pyridoxine, thiotepa, and placebo. During the follow-up, sixty-two patients experienced tumor recurrences, and the maximum number of recurrences was 9. Follow-up times varied from 1 to 64 months. Other information was available including the size of a largest initial tumor, the number of tumors at enrollment.

Figure 1.2 shows times to tumor recurrences for 10 randomly selected patients on the study-time scale, with censoring time information. Some important models for recurrent event analysis are discussed in Section 1.4. Nevertheless, depending on research interest a subset of the whole recurrent event data may be analyzed. For example, one may be interested in the 'first' tumor recurrence from randomization. In this case, classical survival data analysis methods may be applied to the subset data. Some classical models for univariate survival analysis are discussed in Section 1.3.

The bladder tumor trial data set is analyzed in Chapters 2 and 3 to illustrate the proposed methods for the analysis of recurrent events.

1.2 Nonparametric Estimators for the One-Sample Problem

Right mathematical concepts and methods are needed to deal with censored survival data objectively. In the literature, two quantities of the distribution of the random survival time have been of great interest. One is the survival function of the survival time, say, T,

$$S(t) \equiv Pr(T > t),$$

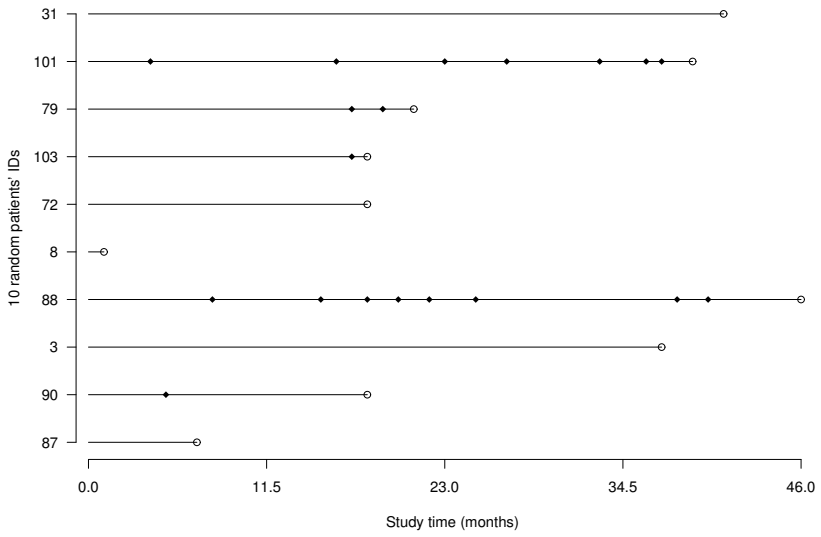

Figure 1.2: The bladder tumor trial data. Ten randomly selected patients' times to tumor recurrences and censoring times. The black dots and empty circles indicate the occurrences of tumor and censoring times

and the other one is the hazard function

$$\lambda(t) \equiv \frac{f(t)}{S(t-)} = \lim_{h \downarrow 0} \frac{1}{h} Pr(t \leq T < t + h | T \geq t),$$

where $f(t)$ is the density function of the survival time. Alternatively, the hazard function is represented in the form of the cumulative hazard function, namely

$$\Lambda(t) \equiv \int_0^t \lambda(s)\, ds.$$

Each of the quantities uniquely determines the distribution of the survival time, and either one can be represented by the other, e.g., $S(t) = \exp\{-\int_0^t \lambda(s)\, ds\}$.

1.2.1 Counting Processes and Nonparametric Estimators

It is natural to see the events we consider as stochastic processes since the events are intrinsically a random function of time. Moreover, the number of events between an

origin and a given time t is a natural measure of event times. Therefore, we denote the counting process for an underlying event process by $N^*(t)$, which is a random function of time measuring the number of events occurring in time interval $[0, t]$ for a subject. Also, we define the *at-risk* process $Y(t)$ as one if an event for a subject can be observed at time t or afterwards, and as zero otherwise. The underlying counting process $N^*(t)$ can be uniquely decomposed into model and error parts; that is,

$$N^*(t) = \Lambda(t) + M(t),$$

where the model part $\Lambda(t)$ is called the compensator and the error part $M(t)$ is called the martingale.

Theories for counting processes have been well-developed and widely used in the statistical literature. In particular, large-sample properties of many estimators have been well-studied based on martingale theory, e.g., consistency and asymptotic normality proofs of the standard Cox estimator by Andersen and Gill (1982). Under a less stringent condition with a mean-zero stochastic error process, but without a martingale error process, empirical process theory may be used to establish large-sample behaviors, e.g, van der Vaart and Wellner (1996).

In Chapters 2 and 3, we utilize empirical process theory to establish uniform consistency and weak convergence of the proposed estimators.

Kaplan-Meier and Nelson-Aalen Estimators

If survival time T is subject to the random censoring, say, at C, what are observed in univariate survival data are the observed survival time $X \equiv T \wedge C$ and the indicator $\Delta \equiv I(T \leq C)$, where $\wedge$ is the minimum operator and $I(\cdot)$ is the indicator function. Alternatively, the censored survival data can be represented using the counting process notation. Define the counting process for an observed event process as

$N(t) \equiv I(X \leq t, \Delta = 1)$, and let the *at-risk* process for a subject be $Y(t) = I(X \geq t)$. The censored survival data consist of $\{N_i(\cdot), Y_i(\cdot)\}_{i=1}^n$, which are n independent replicates of $\{N(\cdot), Y(\cdot)\}$.

The Kaplan-Meier estimator (a.k.a. the product-limit estimator) is a widely used nonparametric estimator for the survival function. Under the independence assumption between T and C, the Kaplan-Meier estimator estimates the survival function unbiasedly from the censored survival data. Let $0 < x_1 < x_2 < \cdots$ be the observed survival times in the sample of size n. Then, the Kaplan-Meier estimator can be defined in the counting process notation

$$\widehat{S}(t) \equiv \prod_{j:x_j \leq t} \{1 - \frac{\sum_{i=1}^n dN_i(x_j)}{\sum_{i=1}^n Y_i(x_j)}\}.$$

On the other hand, the Nelson-Aalen estimator can estimate the cumulative hazard function from the censored survival data. That is,

$$\widehat{\Lambda}(t) \equiv \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{k=1}^n Y_k(s)} = \sum_{j:x_j \leq t} \frac{\sum_{i=1}^n dN_i(x_j)}{\sum_{i=1}^n Y_i(x_j)}.$$

The two estimators are related in the same way that the survival function and the cumulative hazard function are connected. Let K time points partition the time interval $(0, t]$ such that $0 < x_1 < x_2 < \cdots < x_K = t$. It can be shown that

$$\widehat{S}(t) = \pi_{s \leq t}\{1 - d\widehat{\Lambda}(s)\} \equiv \lim_{M \to 0} \prod_{k=1}^K [1 - \{\widehat{\Lambda}(x_k) - \widehat{\Lambda}(x_{k-1})\}],$$

where π is the product-integral and $M \equiv \max_k |x_k - x_{k-1}|$ is the length of the longest subinterval. Uniform consistency and weak convergence of the nonparametric estimators can be shown based on the martingale theory, e.g., Andersen, Borgan, Gill, and Keiding (1993).

Figure 1.3 displays the Kaplan-Meier estimates and the Nelson-Aalen estimates

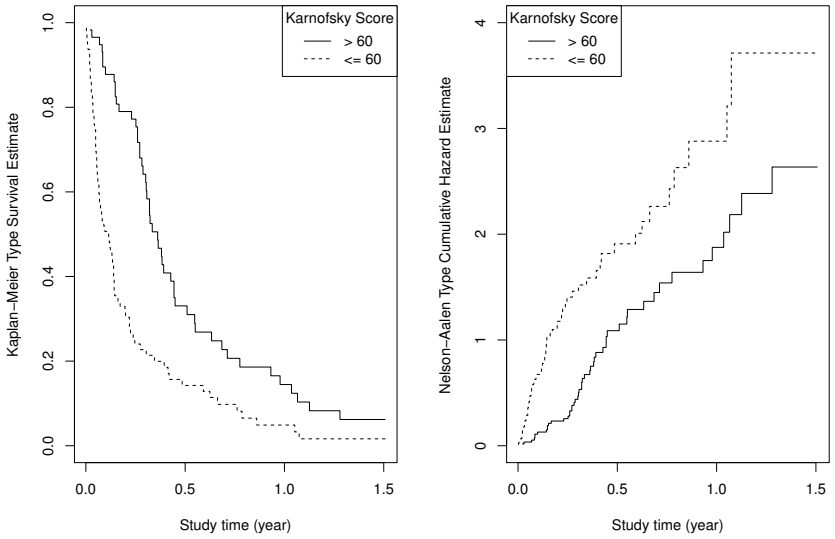

Figure 1.3: The Veterans Administration lung cancer trial data. (a) Kaplan-Meier estimates for the survival functions; (b) Nelson-Aalen estimates for the cumulative hazard functions for patients with baseline Karnofsky performance score ≤ 60 and > 60.

for two groups of patients with baseline Karnofsky score ≤ 60 and > 60 in the VA lung cancer trial data. Overall, the estimates indicate that patients who had higher Karnofsky score at enrollment lived longer in the data.

In the recurrent event data, what are observed are individual-level recurrent events and censoring times. Alternatively, the recurrent events can be represented using the counting process for an observed event process $N(\cdot) = N^*(\cdot \wedge C)$ and the *at-risk* process for a subject $Y(\cdot) = I(C \geq \cdot)$. Specifically, the recurrent event data of size n consist of $\{N_i(\cdot), Y_i(\cdot)\}_{i=1}^n$, which are n independent replicates of $\{N(\cdot), Y(\cdot)\}$. Let $\mu(t) \equiv E(N^*(t))$ be the mean frequency of recurrent events. Using the counting process notation, the Nelson-Aalen estimator for the cumulative hazard function $\Lambda(\cdot)$ can be extended directly to an estimator for the mean frequency function $\mu(\cdot)$. To be specific, the Nelson-Aalen-type nonparametric estimator for the mean frequency

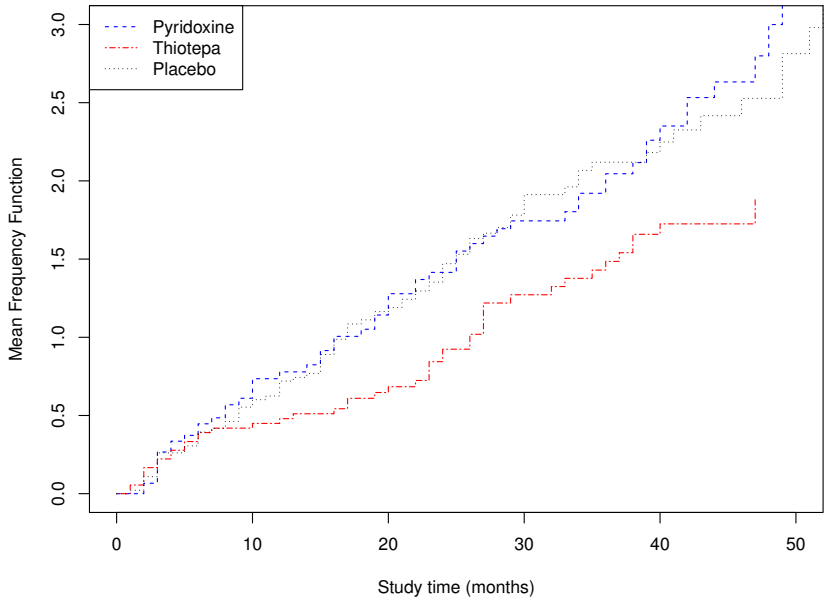

Figure 1.4: The bladder tumor trial data. The Nelson-Aalen type mean frequency function estimates for three treatment arms: pyridoxine; thiotepa; and placebo.

function is

$$\widehat{\mu}(t) \equiv \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{k=1}^{n} Y_k(s)} = \sum_{j:x_j \le t} \frac{\sum_{i=1}^{n} dN_i(x_j)}{\sum_{i=1}^{n} Y_i(x_j)},$$

where $0 < x_1 < x_2 < \cdots$ are unique event times from all individuals in the sample.

Figure 1.4 shows the Nelson-Aalen type mean frequency function estimates for three groups of patients by treatment. On the whole, the estimates indicate that patients in thiotepa treatment arm tended to have less frequent tumor occurrences over time compared to the other patients.

1.3 Existing Regression Models in Survival Analysis

1.3.1 Constant-Effects Regression Models for Univariate Survival Time

Since the advent of the proportional hazards model of Cox (1972), many regression models and methods for the censored survival time data have been developed as an extension or alternative of the proportional hazards model. Let $\boldsymbol{Z}$ be a p-dimensional vector of covariates. The Cox model for the hazard function of the survival time T given $\boldsymbol{Z}$ takes a form

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{\boldsymbol{b}_0^\top \boldsymbol{Z}\},$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and $\boldsymbol{b}_0$ is a p-dimensional regression coefficients. The proportional hazards model postulates constant effects of covariates on the survival time. As an alternative of the Cox model, the accelerated failure time model (a.k.a. the AFT model) has been considered directly referring to the logarithm of the survival time. Specifically, the accelerated failure time model takes a form

$$\log T = \boldsymbol{b}_0^\top \boldsymbol{Z} + \epsilon,$$

where ϵ is a residual from an unspecified distribution. This model postulates constant multiplicative effects of covariates on the survival-time scale change; see Buckley and James (1979) and Prentice (1978) among others. Although such 'accelerated time' modeling seems to provide a simple and direct physical interpretation on the survival time, a caution is needed when the estimated effects of covariates are interpreted as the mean effects for $\log T$. This is because the survival data may actually contain incomplete observations. Besides, a class of the semiparametric linear transformation

models has been considered by some authors; Dabrowska and Doksum (1988) and Chen et al. (2002) among others. A class of transformation models takes a form

$$g(T) = -\boldsymbol{b}_0^\top \boldsymbol{Z} + \epsilon,$$

where g is an unspecified monotone transformation function, and ϵ is a random variable from a known distribution. When ϵ has an extreme value distribution, this class of models reduces to the original Cox model. Also, the proportional odds model is a special case of this transformation model when ϵ follows the standard logistic regression.

1.3.2 Varying-Coefficient Models

Most regression models in survival analysis assume constant effects of covariates on the survival time. This simplistic assumption is often unrealistic in practice. In a clinical study for AIDS patients, for example, a treatment might take time to reach its full efficacy rather than right after randomization; meanwhile, the treatment effect might also erode over time as drug resistance develops, e.g., Eshleman et al. (2001) and Wu et al. (2005). In the VA lung cancer data, a changing ratio of two Nelson-Aalen estimates is observed in Figure 1.5. The patients with high Karnofsky performance score at enrollment tended to have higher survival rate at the beginning of the follow-up, but the association appeared to be weaken afterwards.

Such circumstances call for a more general model to accommodate time-varying evolving effects of covariates. There have been attempts to address the changing effects of covariates in a way of extending the Cox model. As a natural generalization of the original Cox model, a varying-coefficient Cox model with time-varying regression coefficients $\boldsymbol{b}_0(t)$, in place of constant $\boldsymbol{b}_0$, has been studied by many researchers; see Zucker and Karr (1990), Cai and Sun (2003), and Tian et al. (2005) among others.

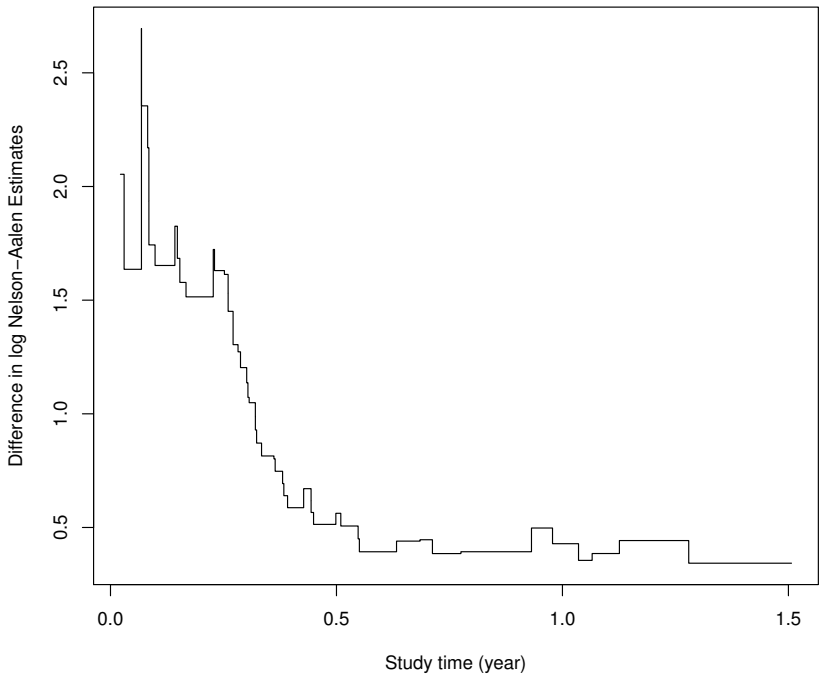

Figure 1.5: The Veterans Administration lung cancer trial data. Difference in log Nelson-Aalen estimates for survival time between patients with baseline Karnofsky performance score ≤ 60 and > 60.

The varying-coefficient Cox model takes a form

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{\boldsymbol{b}_0(t)^\top \boldsymbol{Z}\}.$$

Alternatively, Aalen (1989) proposed the additive hazards model, which specifies the 'additive' time-varying effects of covariates on the survival time. To be specific, the additive hazards model of Aalen (1989) takes a form

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) + \boldsymbol{b}_0(t)^\top \boldsymbol{Z}.$$

Unlike the varying-coefficient Cox model, the jth element of $\boldsymbol{b}_0(t)$ represents an excessive effect of the jth covariate as the increase in hazard at time t for a unit increase of the covariate. However, the additive hazards model has not been used as much as the Cox-type multiplicative effect models. Less popularity of this model might have been because the estimated conditional hazards can possibly deviate from the positive range, i.e., negative values of the hazards.

Both of the varying-coefficient Cox model and the additive hazards model specify the hazard of event over infinitesimal time intervals. Although such hazard function modeling enables a flexible mathematical specification, the effects of covariates may not be clinically meaningful in practice. Moreover, the methods for the varying-coefficient model and the additive hazards model are complicated as they require a kernel smoothing. Instead of the hazard function modeling, Peng and Huang (2007) considered the survival probability and proposed an alternative varying-coefficient model for the survival function. Specifically, the temporal survival regression model of Peng and Huang (2007) takes the following form

$$S(t|\boldsymbol{Z}) = \exp\big[-\exp\{\log\Lambda_0(t) + \boldsymbol{b}_0(t)^\top \boldsymbol{Z}\}\big], \qquad \textit{for all } t \geq 0,$$

where $\Lambda_0(\cdot)$ is an unspecified baseline cumulative hazard function. Basically, the temporal survival regression model postulates time-varying multiplicative effects of covariates on the baseline cumulative hazards $\Lambda_0(\cdot)$, and thereby on the unspecified baseline survival function $S_0(\cdot) \equiv \exp\{-\Lambda_0(\cdot)\}$. In the special case with constant regression coefficients, i.e., $\boldsymbol{b}_0(\cdot) = \boldsymbol{b}_0$, the temporal survival regression reduces to the original Cox model.

Alternatively, a quantile regression model has been studied by some authors; see Peng and Huang (2008), Huang (2010), Qian and Peng (2010) among others. Given $\boldsymbol{Z}$, conditional quantile function $Q(\tau|\boldsymbol{Z}) \equiv \sup\{t : 1 - S(t|\boldsymbol{Z}) \leq \tau\}$ for $\tau \in [0, 1)$ is specified as

$$Q(\tau|\boldsymbol{Z}) = Q_0(\tau) + \boldsymbol{b}_0(\tau)^\top \boldsymbol{Z},$$

where $Q_0(\cdot)$ is an unspecified baseline quantile function. Essentially, each of the baseline quantiles is modeled to be changing in its scale as covariates affect in the linear combination of their coefficients of corresponding cumulative probability τ. Since the quantile regression model is specified on the probability scale, clinical interpretation may not be as straightforward as other models specified on the study-time scale.

A concern with fully functional varying-coefficient models is efficiency loss due to the increasing model generality. When clinical or biological knowledge implies constant effects for some of the covariates, a mixture effect model with constant and time-varying effects would be useful. In Chapter 4, we propose a semiparametric mixture effect model for the survival function. The model is a sub-model of the temporal survival regression model, and therefore has direct interpretation of covariate effects on the survival function. Moreover, an iterative estimation procedure is developed, which does not involve smoothing.

1.4 Existing Models in Recurrent Events

To investigate association between covariates and recurrent events, earlier model developments focused on the intensity function modeling; see Prentice et al. (1981) and Andersen and Gill (1982) among others. In particular, the multiplicative intensity model of Andersen and Gill (1982) directly extends the idea of the Cox model for recurrent event data by adapting counting processes. To be specific, the multiplicative intensity model of Andersen and Gill (1982) takes a form

$$E(\, dN^*(t)|F_{t-}) = \lambda_0(t) \exp\{\boldsymbol{b}_0^\top \boldsymbol{Z}\}\, dt,$$

where F_{t-} is the history filtration right before time t, i.e., all the available information up to right before time t. The model presumes the independent increment structure of the non-homogeneous Poisson process, and postulates constant multiplicative effects of covariates on the next event. Without imposing the independent increment assumption on recurrent events, Prentice, Williams, and Peterson (1981) proposed the stratified proportional intensity model, which takes a form

$$E(\, dN^*(t)|F_{t-}) = \lambda_{0\nu}(t) \exp\{\boldsymbol{b}_{0\nu}^\top \boldsymbol{Z}\}\, dt,$$

where $\lambda_{0\nu}(t)$ is an unspecified stratum-specific baseline intensity function. Notation ν may be specified based on the event process history, e.g., $\nu \equiv N^*(t-)$ the number of previous events in time interval $[0, t)$.

1.4.1 Intra-Individual Correlation and Marginal Models

By the multiplicity nature of recurrent events, the intra-individual correlation between events is often exhibited in the data. The intra-individual correlation is due to not only observed covariates but also random effects. Since the intensity function

modeling typically requires that the baseline event process is either Markov or semi-Markov, it is tacitly assumed that the intra-individual correlation between events is entirely due to the observed covariates while the intensity function is formulated.

To relax the Markovian assumption, some marginal models have been developed on the study-time scale; see Pepe and Cai (1993), Lawless and Nadeau (1995), and Lin et al. (2000) among others. In marginal modeling, effects of covariates are formulated on a population-level quantity such as the rate or mean frequency function of recurrent events. Therefore, the individual-level dependence structure is left completely unspecified. In particular, the proportional means model of Lin, Wei, Yang, and Ying (2000) takes a form

$$E(N^*(t)|\boldsymbol{Z}) = \mu_0(t)\exp\{\boldsymbol{b}_0^\top \boldsymbol{Z}\},$$

where $\mu_0(t)$ is an unspecified baseline mean frequency function. This model marginally specifies the constant multiplicative effects of covariates on the expected frequency function, without conditional on the event process history. By averaging out the event history process, this marginal model generalizes the the multiplicative intensity model of Andersen and Gill (1982). From this perspective, marginal models can accommodate more general circumstances, such as frailty, facilitating robust method developments.

Some marginal models for recurrent events have also been developed on the gap-time scale between the successive events. Particularly, Huang and Chen (2003) proposed a marginal proportional hazards model for gap times between recurrent events. Under the renewal assumption on event processes (i.e., gap times for a subject are iid replicates of, say, T), the recurrent gap times are marginally modeled with observed covariates. To be specific, the marginal cumulative hazard function of the gap time

T given $\boldsymbol{Z}$ takes a form

$$\Lambda(t|\boldsymbol{Z}) = \Lambda_0(t)\exp\{\boldsymbol{b}_0^\top \boldsymbol{Z}\},$$

where $\Lambda_0(t)$ is an unspecified marginal baseline cumulative gap-time hazard function. Such gap-time modeling strategy is useful especially when the occurrence of next event is of research interest in recurrent event analysis. In the bladder tumor tiral data, for instance, researchers may be interested in the effect of a prophylatic treatment on the time to next tumor recurrence.

Besides, Lin, Wei, and Ying (1998) proposed the accelerated failure time model for recurrent event data. The model formulates the effects of covariates on the time scale change of the mean frequency function. Specifically, the accelerated failure time model of Lin et al. (1998) takes a form

$$E(N^*(t)|\boldsymbol{Z}) = \mu_0\{t\exp(\boldsymbol{b}_0^\top \boldsymbol{Z})\}.$$

As noted previously, the multiplicative intensity model of Andersen and Gill (1982) generalizes the proportional hazards model of Cox (1972). By adapting the counting process notation, this model of Lin et al. (1998) similarly generalizes the accelerated failure time model for survival data to accommodate recurrent events.

1.4.2 Varying-Coefficient Models

Most existing models for the analysis of recurrent events assume constant effects of covariates on recurrence times. As discussed in Section 1.3.2, the constant effect assumption may not always be realistic in practice. In the bladder tumor trial data, moreover, a changing ratio of nonparametric mean frequency function estimates is observed in Figure 1.6. The changing ratio indicates that the effect of the number of initial tumor changed over time on tumor recurrence times.

Several models and methods have been developed for recurrent event analysis to

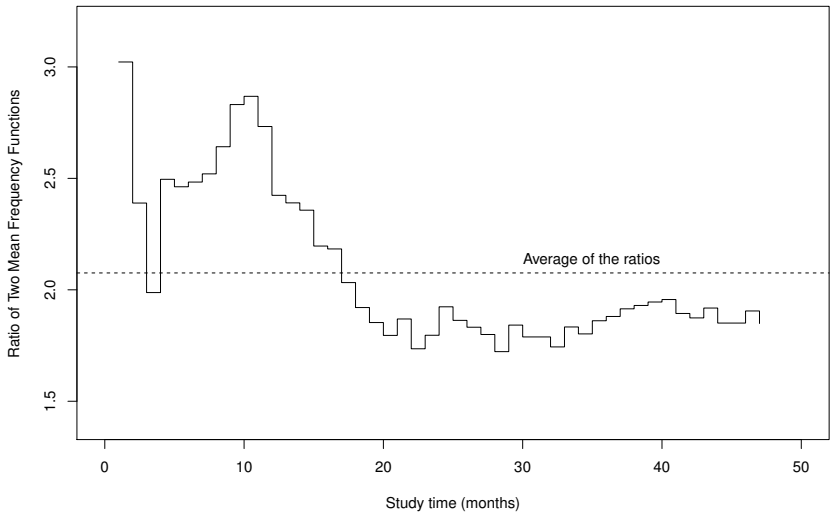

Figure 1.6: The bladder tumor trial data. The ratio of the Nelson-Aalen-type mean frequency function estimates between people with at most one initial tumor and people with 2 or more initial tumors.

accommodate time-varying effects of covariates. In particular, the temporal process regression of Fine, Yan, and Kosorok (2004) formulates temporal effects of covariates on a baseline mean frequency at a given time point. Chiang and Wang (2009) proposed a varying-coefficient model for the overall rate of recurrent events and developed a kernel smoothing-based estimation procedure. Alternatively, Huang and Peng (2009) proposed the accelerated recurrence time model in which time-varying effects of covariates are formulated through time-scale change of the mean frequency function. This model generalizes the accelerated failure time model of Lin et al. (1998). Along a similar line, Sun et al. (2016) proposed a generalization of the univariate censored quantile regression of Peng and Huang (2008) for recurrent events.

We propose two marginal varying-coefficient models on study-time and gap-time scales of recurrent events, respectively in Chapters 2 and 3. Both models are globally specified over time, and therefore accompanied smoothing-free estimation procedures fully utilize the event times in the sample. Such estimation is expected to be more efficient than local estimation, e.g., Fine et al. (2004), and implementation can be

straightforward than smoothing-based estimation, e.g., Chiang and Wang (2009).

1.5 Overview

In Chapter 2, we develop a dynamic regression model to target the mean frequency of recurrent events. By allowing regression coefficients to vary over time, it generalizes the proportional means model of Lin et al. (2000). Meanwhile, our proposed model can be viewed as a special case of the temporal process regression of Fine et al. (2004). We develop estimation and inference procedures and establish large sample properties of the proposed estimator. The performance of the proposed estimator is evaluated through Monte Carlo simulations. Two real data analyses are presented for illustration of the proposed method.

In Chapter 3, we propose a marginal varying-coefficient model for gap times between recurrent events. The proposed model accommodates evolving effects of covariates on gap times marginally at the population level. The proposed model directly generalizes the semiparametric model of Huang and Chen (2003) by replacing constant regression coefficients with time-varying ones. Estimation and inference procedures for the time-varying coefficients are developed, and we establish large sample properties of the proposed estimator. Finite-sample behaviors of the proposed estimator are evaluated through Monte Carlo simulations. An analysis of the bladder tumor trial data is presented to illustrate the proposed method.

In Chapter 4, we propose a semiparametric survival regression model with a mixture of time-varying and constant effects of covariates. The proposed model is a sub-model of the temporal survival regression of Peng and Huang (2007), and therefore provides a middle ground between the proportional hazards model of Cox (1972) and the temporal survival regression model. We develop a smoothing-free iterative estimation procedure and a nonparametric resampling-based inference procedure.

Finite-sample behaviors of the proposed estimator are investigated via Monte Carlo simulations. Also when the original Cox model holds, efficiency loss of the proposed estimators is investigated through simulation studies. An analysis of the well-known Veterans' Administration lung cancer data is presented for illustration.

We conclude with final summary and future work in Chapter 5. Detailed proofs of consistency and weak convergence results can be found in Appendices.

Chapter 2

Dynamic Regression with Recurrent Events

Most regression models with recurrent events may be categorized as either conditional or marginal modeling. In conditional regression modeling, the intensity function of a recurrent event process is specified, e.g., the models of Prentice et al. (1981) and Andersen and Gill (1982). Thus conditional models could be more useful in making an individual-level prediction. On the other hand, in marginal regression modeling, either the overall rate or the mean frequency of recurrent event processes is specified by covariates, e.g., the proportional means model by Lin et al. (2000).

Most of the existing models with recurrent events assume constant effects of covariates over time, but this simplistic assumption is often unrealistic in practice. To accommodate evolving effects of covariates, several regression models have been developed for the analysis of recurrent events. In particular, the temporal process regression by Fine et al. (2004) accommodates temporal effects of covariates on a baseline mean frequency at a given time point.

In this Chapter, we develop a dynamic regression model to target the mean frequency of recurrent events. By allowing regression coefficients to vary over time, it

generalizes the proportional means model of Lin et al. (2000). Meanwhile, our proposed model can be viewed as a special case of the temporal process regression of Fine et al. (2004). The model is described in Section 2.1. We propose estimation and inference procedures in Section 2.2, also present large sample properties of the proposed estimator. The performance of the proposed estimator is evaluated through Monte Carlo simulations in Section 2.3. In Section 2.4, two real data analyses are presented for illustration of the proposed method.

2.1 Model

Let $N^*(t)$ denote the number of recurrent events of interest for a subject in time interval $[0, t]$ and $\boldsymbol{Z}$ be a p-dimensional covariate vector. We develop a dynamic regression model with recurrent events

$$E\{N^*(t)|\boldsymbol{Z}\} = \exp\{\boldsymbol{\beta}_0(t)^\top \widetilde{\boldsymbol{Z}}\}, \qquad \textit{for all } t \geq 0, \tag{2.1}$$

where $\boldsymbol{\beta}_0(t) = \{\log \mu_0(t),\ \boldsymbol{b}_0(t)^\top\}^\top$ and $\widetilde{\boldsymbol{Z}} = (1,\ \boldsymbol{Z}^\top)^\top$. Here $\mu_0(t)$ denotes the baseline mean frequency of recurrent events, i.e., with all covariates being zero; and $\boldsymbol{b}_0(t)$ is the p-dimensional time-varying regression coefficient vector. Model (2.1) postulates time-varying multiplicative effects of covariates on the baseline mean frequency function. Note $\mu_0(0) = 0$ because $N^*(0) = 0$.

In many circumstances the underlying counting process $N^*(\cdot)$ for a subject is only observed up to a random follow-up time, say, C. Therefore what are observed are an observed event process $N(\cdot) = N^*(\cdot \wedge C)$ and an *at-risk* process $Y(\cdot) = I(C \geq \cdot)$, where $\wedge$ and $I(\cdot)$ are the minimization operator and the indicator function, respectively. We adopt the conditional independence assumption on the censoring mechanism

$$N^*(\cdot) \perp C \mid \boldsymbol{Z}.$$

Model (2.1) is a structure-imposed nonparametric model as the slope coefficients depend on time. In the special case of a k-sample problem, it is a saturated model, in which no structure is actually imposed. Note that our proposed model with time-varying coefficients generalizes the proportional means model of Lin et al. (2000). Meanwhile, our model turns out to be the special case of the temporal process regression of Fine et al. (2004) when log link and recurrent events are considered.

2.2 Estimation and Inference

2.2.1 Point Estimation

Suppose an observed data set $\{N_i(\cdot); Y_i(\cdot); C_i; \boldsymbol{Z}_i\}_{i=1}^n$ consists of n independent replicates of $\{N(\cdot); Y(\cdot); C; \boldsymbol{Z}\}$. Under model (2.1) with the conditional independence assumption on the censoring mechanism, it follows $E\{N(t)|\boldsymbol{Z}\} = E\big[\int_0^t Y(s)\, d\exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}\}|\boldsymbol{Z}\big]$. This motivates our proposed estimating integral equation

$$n^{-1}\sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i\big[N_i(t) - \int_0^t Y_i(s)\, d\exp\{\boldsymbol{\beta}(s)^\top \widetilde{\boldsymbol{Z}}_i\}\big] = 0, \qquad \textit{for all } t \geq 0. \tag{2.2}$$

This equation is similar to the one for the univariate survival model in Peng and Huang (2007). However, under the proposed model (2.1) the left-hand side of equation (2.2) does not have a martingale structure when $\boldsymbol{\beta}(\cdot)$ is $\boldsymbol{\beta}_0(\cdot)$, unlike the case for univariate survival data.

Equation (2.2) admits a càdlàg solution, $\widehat{\boldsymbol{\beta}}(t)$, which has jumps only at observed event times. Write $0 < x_1 < x_2 < \cdots < x_M$ as the observed event times of all subjects in the sample, and denote $\boldsymbol{Z}_{(1)}, \boldsymbol{Z}_{(2)}, \ldots, \boldsymbol{Z}_{(M)}$ as the associated covariate vectors. Note that one individual may have multiple contributions to $\{x_j, \boldsymbol{Z}_{(j)}\}_{j=1}^M$. The initial value, $\widehat{\boldsymbol{\beta}}(0)$, satisfies $\exp\{\widehat{\boldsymbol{\beta}}(t)^\top \widetilde{\boldsymbol{Z}}_i\} = 0$ for all i. Then, at time x_1, the

estimating integral equation (2.2) reduces to

$$\widetilde{\boldsymbol{Z}}_{(1)} - \sum_{i=1}^{n} Y_i(x_1)\widetilde{\boldsymbol{Z}}_i \exp\{\boldsymbol{\beta}(x_1)^\top \widetilde{\boldsymbol{Z}}_i\} = 0. \tag{2.3}$$

Sequentially at the observed event times $(x_j)_{j=2}^{M}$, the estimating integral equation (2.2) reduces to

$$\widetilde{\boldsymbol{Z}}_{(j)} - \sum_{i=1}^{n} Y_i(x_j)\widetilde{\boldsymbol{Z}}_i\big[\exp\{\boldsymbol{\beta}(x_j)^\top \widetilde{\boldsymbol{Z}}_i\} - \exp\{\widehat{\boldsymbol{\beta}}(x_{j-1})^\top \widetilde{\boldsymbol{Z}}_i\}\big] = 0, \qquad j = 2, \ldots, M. \tag{2.4}$$

Both estimating equations (2.3) and (2.4) have good computational properties since their left-hand sides are monotone functions.

If censoring is absent up to time t, the estimating integral equation (2.2) implies

$$n^{-1}\sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i\big[N_i(t) - \exp\{\boldsymbol{\beta}(t)^\top \widetilde{\boldsymbol{Z}}_i\}\big] = 0.$$

This estimating equation coincides with the one in Fine et al. (2004) when the natural logarithm is adopted for the link function in their model.Consequently an identical estimator is obtained. When censoring is present, however, equation (2.2) fully utilizes the observed data, whereas the estimating equation of Fine et al. (2004), $\sum_{i=1}^{n} Y_i(t)\widetilde{\boldsymbol{Z}}_i\big[N_i(t) - \exp\{\boldsymbol{\beta}(t)^\top \widetilde{\boldsymbol{Z}}_i\}\big] = 0$, uses the data only from uncensored observation as of the given time point of interest. Therefore, our estimator tends to be more efficient as shown later in our simulation studies.

2.2.2 Large Sample Properties

We establish the uniform consistency and the weak convergence of the proposed estimator $\widehat{\boldsymbol{\beta}}(\cdot)$. Write $\|\cdot\|$ and eigmin$(\cdot)$ as the Euclidean norm and the minimum eigenvalue of a positive semidefinite matrix, respectively. Write $\boldsymbol{\beta}_0(t)$ as the true value of $\boldsymbol{\beta}(t)$. We postulate the following regularity conditions:

C1. there exists a time point $\kappa > 0$ such that $Pr(C > \kappa) = 1$;

C2. $N(\tau)$ is bounded, where τ is a fixed time point satisfying $Pr(C > \tau) > 0$ for $\kappa < \tau < \infty$;

C3. $\|\boldsymbol{Z}\|$ is bounded;

C4. $\sup_{t\in[\kappa,\tau]} \|\boldsymbol{\beta}_0(t)\|$ is bounded;

C5. $\boldsymbol{\beta}_0(t)$ is continuously differentiable for $t \in (\kappa, \tau]$;

C6. $\inf_{t\in[0,\tau]}\text{eigmin}E\{Y(t)\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top\}$ is bounded away from zero.

The logarithm of a baseline mean frequency function is negative infinity, i.e., $\log \mu_0(0) = -\infty$, which creates a challenge in the asymptotic study. To avoid this issue, we focus on the properties of $\widehat{\boldsymbol{\beta}}(t)$ over $t \in [\kappa, \tau]$ for some small κ and impose condition C1. As such, $\widehat{\boldsymbol{\beta}}(\kappa)$ is the same as an estimator of Fine et al. (2004) and its properties follow the standard M-estimation theory. Other conditions, C2 to C6, are technical assumptions and fairly standard with varying-coefficient regression models.

Theorem 1. Under regularity conditions C1-C6, $\sup_{t\in[\kappa,\tau]} \|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\| \longrightarrow 0$, almost surely.

Theorem 2. Under regularity conditions C1-C6, $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\cdot) - \boldsymbol{\beta}_0(\cdot)\}$ on $(\kappa, \tau]$ weakly converges to a mean-zero Gaussian process.

The proofs are in the Appendix A.

2.2.3 Interval Estimation

Inference may be based on Theorem 2. In particular for a finite-dimensional quantity, e.g., $\boldsymbol{\beta}_0(t)$ for a given t, asymptotic variance can be estimated and subsequently confidence interval may be constructed. However, this approach is difficult if not impossible to construct confidence band for infinite-dimensional quantities such as

$\boldsymbol{\beta}_0(\cdot)$; see related discussion in Lin et al. (1994) and Huang (2014) in different contexts. We instead propose a nonparametric bootstrap inference procedure using multiplier bootstrap; see Rubin (1981), Kosorok (2008) and Huang (2014). To be specific, we propose a bootstrap estimating integral equation

$$n^{-1}\sum_{i=1}^{n}\xi_i\widetilde{\boldsymbol{Z}}_i\big[N_i(t)-\int_0^t Y_i(s)\,d\exp\{\boldsymbol{\beta}(s)^\top\widetilde{\boldsymbol{Z}}_i\}\big]=0,\qquad \textit{for all } t\geq 0, \tag{2.5}$$

where $(\xi_i)_{i=1}^n$ are independent and identically distributed, following the standard exponential distribution. Denote a solution to this equation (2.5) by $\boldsymbol{\beta}^*(\cdot)$. Then based on B bootstrap resamples, a set of bootstrap solutions $\{\boldsymbol{\beta}_b^*(\cdot)\}_{b=1}^B$ can be obtained. At any given time t, the variance of estimator $\widehat{\boldsymbol{\beta}}(t)$ can be estimated by the sample variance of $\{\boldsymbol{\beta}_b^*(t)\}_{b=1}^B$. Provided this variance estimate, a $100(1-\alpha)\%$ point-wise confidence interval for $\boldsymbol{\beta}_0(t)$ can be constructed by the normal approximation centered at $\widehat{\boldsymbol{\beta}}(t)$. Alternatively, a confidence interval can be simply constructed with the $(\alpha/2)$th and $(1-\alpha/2)$th quantiles of the empirical distribution of $\boldsymbol{\beta}^*(t)$. Given the solutions $\widehat{\boldsymbol{\beta}}(\cdot)$ and $\{\boldsymbol{\beta}_b^*(\cdot)\}_{b=1}^B$, we propose to construct a $100(1-\alpha)\%$ confidence band for $\{\boldsymbol{\beta}_0(t):\ t\in(l,u]\}$ with $\{\widehat{\boldsymbol{\beta}}(t)\pm\gamma_\alpha:\ t\in(l,u]\}$. Here γ_α is the $(1-\alpha/2)$th quantile of the empirical distribution of J^*, where $J_b^*=\sup_{t\in[l,u]}|\boldsymbol{\beta}_b^*(t)-\widehat{\boldsymbol{\beta}}(t)|$ for $b=1,\ldots,B$.

2.2.4 Average Effect and Test for Varying Effect

Our estimator $\widehat{\boldsymbol{\beta}}(\cdot)$ can provide a profile of time-varying effects of covariates over time. However, one may be interested in average effects of covariates over time interval $(l,u]$ for $0<l<u<\infty$. That is, $\overline{\boldsymbol{\beta}}_0(l,u)\equiv(u-l)^{-1}\int_l^u\boldsymbol{\beta}_0(t)\,dt$; see Peng and Huang (2008) for its usage in quantile regression. Therefore based on $\widehat{\boldsymbol{\beta}}(\cdot)$, we propose a natural estimator of the average effects $\overline{\widehat{\boldsymbol{\beta}}}(l,u)=(u-l)^{-1}\int_l^u\widehat{\boldsymbol{\beta}}(t)\,dt$.

Moreover, we consider the null hypothesis testing for constant vs. time-varying

effects

$$H_0 : \boldsymbol{c}^\top \boldsymbol{\beta}_0(t) \text{ is a constant for all } t \in (l, u],$$

where $\boldsymbol{c}$ is a known $(p+1)$-dimensional vector. In the special case of testing the varying effect of the ith covariate, the $(i+1)$th component of $\boldsymbol{c}$ is one and zeros elsewhere. To evaluate H_0, we propose a test statistic $\mathcal{T} = n^{1/2} \int_l^u \boldsymbol{c}^\top \{\widehat{\boldsymbol{\beta}}(t) - \overline{\widehat{\boldsymbol{\beta}}}(l,u)\} \Xi(t)\, dt$, where $\Xi(t)$ is a known nonnegative weight function. The identity function $\Xi(t) = t$ is considered to be a good candidate for the weight when a linear changing effect is reasonable. Note that the limiting distribution of $\mathcal{T}$ is a Gaussian because $\mathcal{T}$ is a linear function of $\widehat{\boldsymbol{\beta}}(\cdot)$. For variance estimation of the limiting distribution of $\mathcal{T}$, we consider $\mathcal{T}^* = n^{1/2} \int_l^u \boldsymbol{c}^\top \{\boldsymbol{\beta}^*(t) - \overline{\boldsymbol{\beta}^*}(l,u)\} \Xi(t)\, dt$ in a bootstrap sample as the counter part of $\mathcal{T}$, where $\overline{\boldsymbol{\beta}^*}(l,u) = (u-l)^{-1} \int_l^u \boldsymbol{\beta}^*(t)\, dt$. Since the limiting distribution of $\mathcal{T}$ under H_0 is equivalent to the limiting conditional distribution of $\mathcal{T}^*$ given the observed data, a p-value for a constant effect test can be computed based on the bootstrap replicates of $\mathcal{T}^*$.

2.3 Simulation Studies

We conducted Monte Carlo simulations to investigate finite-sample properties of the proposed estimator. Throughout the simulations, the underlying baseline mean frequency function was the identity function of time t, i.e., $\mu_0(t) = t$. To induce an intra-individual correlation, a subject-level random frailty η was generated from the unit-mean gamma distribution with $Var(\eta) = \sigma^2 = 0$, 0.5, and 1. Note that larger σ^2 leads to higher intra-individual correlation. The sample size n in each data set was 200, and the bootstrap size B was 500. Under each frailty scenario, 1,000 Monte Carlo data sets were simulated.

For comparison with the proposed method, we also considered the temporal process regression method of Fine et al. (2004) and the proportional means model of Lin

et al. (2000). For the proportional means model, we used the same estimates from the conventional estimation based on the whole observed data, regardless of the time interval under consideration for averaging.

2.3.1 Simulation 1: Single Covariate with Constant Effect

We first considered a single covariate with constant multiplicative effect on the baseline mean frequency. The covariate followed the uniform[0, 1] distribution, and its fixed regression coefficient $b_0(\cdot)$ was 0.5. Random follow-up time C was from the uniform[0, 3] distribution.

Table 2.1 reports a summary of the Monte Carlo simulations for both of the proposed method and the temporal process regression method: empirical biases, empirical standard deviations, average standard error estimates, and empirical coverage probabilities of Wald-type 95% confidence intervals at the predetermined time points. First of all, it is shown that all the estimates were close to the estimands over time. Moreover, the averages of the estimated standard errors agreed quite well with the empirical standard deviations. This indicates that the estimated standard errors based on the proposed resampling procedure approximate the true sampling variation well. When the variance of random frailty was increased, the empirical standard deviations increased and the empirical coverage probabilities decreased on the whole. This indicates that an extra variability was introduced in estimation by the increased dependence between events. On the whole, the Monte Carlo biases of estimator of Fine et al. (2004) were comparable to the ones based on the proposed method. However, relatively large empirical standard deviations were observed for the temporal process regression estimator, even more so at the end of the follow-up. This low efficiency of the temporal process regression estimator was attributed to the fact that in its estimation procedure, the data only from uncensored observation at a given time are used in cross-sectional fashion.

Table 2.1: Summary Results of Simulation 1 at the Prespecified Time Points

	$\log \mu_0(t) = \log t$								$b_0(t) = 0.5$							
	Proposed Method				Fine et al. (2004)				Proposed Method				Fine et al. (2004)			
t	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Zero intra-individual correlation																
0.50	0	192	194	94.7	0	200	202	94.9	-6	320	317	95.2	-6	337	331	95.0
1.00	-4	143	144	94.8	-6	160	160	94.4	1	233	235	95.4	4	264	261	94.7
1.50	-3	125	126	95.7	-4	152	150	94.3	2	206	205	94.6	0	251	245	93.8
2.00	-3	119	118	94.7	-9	162	158	93.7	1	196	193	94.0	9	267	258	93.0
2.25	-4	120	118	94.1	-10	182	171	92.7	1	199	192	94.5	7	297	280	92.6
2.50	-6	120	120	94.8	-8	225	197	89.9	7	200	197	94.6	7	366	322	89.6
2.75	-10	131	127	94.7	-37	325	267	87.7	9	220	208	93.7	21	539	435	86.3
Gamma frailty of variance 0.5																
0.50	0	222	219	94.1	-1	233	230	93.8	-14	367	360	93.6	-14	382	379	93.8
1.00	-3	183	179	93.5	-7	212	201	93.6	-11	311	296	93.1	-6	354	332	94.1
1.50	-4	168	166	94.3	-10	206	205	94.9	-6	287	277	94.3	2	350	342	95.0
2.00	-5	167	164	95.0	-18	236	233	94.3	-7	286	275	93.3	3	396	389	94.6
2.25	-8	170	166	94.8	-31	276	261	93.4	-1	291	278	94.1	23	456	437	93.9
2.50	-7	177	170	92.9	-49	346	310	92.9	-3	302	284	92.6	39	569	517	91.8
2.75	-6	187	178	93.9	-67	539	421	88.6	-5	318	296	92.4	19	881	703	88.3
Gamma frailty of variance 1																
0.50	-20	248	240	94.9	-19	263	253	94.3	11	409	399	94.6	6	436	419	94.3
1.00	-25	211	206	94.0	-25	239	233	93.3	20	352	345	94.2	17	397	390	93.6
1.50	-17	206	197	93.9	-23	260	244	92.9	9	344	332	93.5	9	433	411	92.6
2.00	-13	209	197	93.0	-32	299	282	93.2	2	347	334	93.6	11	500	476	93.9
2.25	-18	215	200	93.6	-41	353	317	92.1	9	356	339	93.2	14	585	534	91.6
2.50	-15	222	205	93.5	-52	442	376	91.0	0	369	345	92.6	5	748	633	91.2
2.75	-20	234	212	92.6	-125	733	525	88.2	6	391	358	93.1	47	1228	881	86.7

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald 95% confidence interval (×100). Based on 1,000 Monte Carlo replications.

Table 2.2 reports the three methods' estimation performance for the average effect of covariate. For each of the three intervals under consideration, the Monte Carlo empirical biases were close to zero relative to the corresponding empirical standard deviations over the different intra-individual correlation scenarios. In case of no dependence between events $\sigma^2 = 0$, the proportional means model of Lin et al. (2000) had smallest empirical standard deviations as expected. However, the Monte Carlo standard deviations increased by a great deal when an intra-individual correlation was introduced, i.e., with $\sigma^2 = 0.5$ and 1. Indeed, the finite-sample efficiency of estimator of Lin et al. (2000) became comparable to the ones based on the other two methods. On the other hand, for the time interval which includes the end of the follow-up, i.e., $(l, u] = (0, 3]$, the temporal process regression estimator of Fine et al. (2004) had larger Monte Carlo biases and big discrepancies between the empirical standard deviations and the average of the estimated standard errors. The temporal process regression is a robust method in general providing unbiased estimates in large sample. However, our simulation results with practical sample size indicate that its estimator tended to be highly variable when a time interval of interest included the end of the study time, in which less events were observed.

2.3.2 Simulation 2: Single Covariate with Time-Varying Effect

In the second simulation, we considered a single covariate with decreasing time-varying effect. The covariate followed the uniform[0, 1], and its time-varying coefficient $b_0(t)$ was $\exp\{-t/\exp(1)\}$. Random follow-up time C was generated from the uniform[0, 3].

Table 2.3 shows that the empirical biases of the considered estimators were close to zero relative to the corresponding Monte Carlo standard deviations. Notably all of the empirical standard deviations of the proposed estimator were smaller than those based

Table 2.2: Summary Results of Simulation 1 for the Average Effect of Covariate $\overline{\boldsymbol{\beta}}_0(l,u)$

	$b_0(t)=0.5$											
	Proposed Method				Fine et al. (2004)				Lin et al. (2000)			
	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Interval (0, 3]												
$\sigma^2=0$	4	194	190	94.6	13	596	212	83.1	3	176	176	94.8
$\sigma^2=0.5$	-7	284	263	94.3	-4	1151	310	71.8	-6	271	265	94.0
$\sigma^2=1$	9	337	317	93.9	137	1580	377	61.8	7	332	328	94.4
Interval (0, 2.5]												
$\sigma^2=0$	4	208	203	94.4	5	230	219	94.2				
$\sigma^2=0.5$	-5	288	271	93.7	1	322	301	93.6				
$\sigma^2=1$	10	344	322	92.8	11	391	357	92.0				
Interval (0.5, 2.5]												
$\sigma^2=0$	1	187	185	94.5	4	218	210	94.8				
$\sigma^2=0.5$	-8	273	262	93.6	1	318	306	94.2				
$\sigma^2=1$	9	326	318	94.2	12	394	371	92.1				

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald 95% confidence interval (×100). Based on 1,000 Monte Carlo replications. For the proportional means model of Lin et al. (2000), the same estimates from the conventional estimation were used based on the whole observed data regardless of the time interval under consideration for averaging; thus the results for intervals (0, 2.5] and (0.5, 2.5] are not presented as they are equivalent to the ones for interval (0, 3].

on the temporal process regression method. Table 2.4 summarizes the simulation results for the average of the time-varying coefficient $\overline{\boldsymbol{\beta}}_0(l,u)$. Regardless of degrees of intra-individual correlations, the Monte Carlo biases of the proposed average effect estimator $\overline{\widehat{\boldsymbol{\beta}}}(l,u)$ were close to zero relative to the empirical standard deviations. By contrast, the average estimator based on the temporal process regression had larger empirical biases and empirical standard deviations, especially when the time interval of interest contained the end-part of the study follow-up, i.e., interval $(l,u]=(0,3]$; the similar pattern was observed in Table 2.2.

2.3.3 Simulation 3: Two Covariates with Constant and Time-Varying Effects

We further considered two covariates, one with varying effect and the other with constant effect. Covariate $Z^{(1)}$ was randomly generated from the trimmed normal

Table 2.3: Summary Results of Simulation 2 at the Prespecified Time Points

	$\log \mu_0(t) = \log t$								$b_0(t) = \exp\{-t/\exp(1)\}$							
	Proposed Method				Fine et al. (2004)				Proposed Method				Fine et al. (2004)			
t	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Zero intra-individual correlation																
0.50	-16	190	187	94.2	-20	201	196	94.3	11	306	292	94.1	17	326	307	93.2
1.00	2	140	141	94.7	-3	160	157	94.1	-10	227	224	94.3	-4	259	249	94.6
1.50	2	125	124	95.1	2	154	149	93.9	-9	203	199	95.0	-13	253	241	94.0
2.00	0	123	118	93.6	-4	171	159	93.2	-7	198	190	94.3	-8	280	261	92.7
2.25	-3	121	117	93.9	-8	179	173	93.5	-5	196	189	93.9	-3	300	284	93.5
2.50	-4	125	119	94.2	-14	220	199	91.3	-4	203	194	93.9	3	362	330	92.2
2.75	-4	136	126	92.1	-45	327	266	88.3	-4	218	205	92.4	37	541	448	88.0
Gamma frailty of variance 0.5																
0.50	-5	213	211	94.6	-3	228	222	93.8	-10	354	340	93.4	-15	379	358	93.2
1.00	-3	178	174	94.5	-5	200	196	93.9	-12	297	286	93.5	-12	334	323	93.8
1.50	-4	166	163	94.2	-9	206	202	93.3	-10	280	271	93.7	-12	348	338	93.1
2.00	-1	166	161	94.3	-9	244	229	93.1	-14	279	269	93.7	-21	417	386	93.3
2.25	-1	164	162	94.3	-13	274	255	92.6	-15	276	271	93.1	-20	468	431	93.0
2.50	-2	167	166	94.6	-27	323	301	92.2	-13	282	276	94.3	-7	550	508	92.6
2.75	-2	178	172	94.2	-69	510	424	88.4	-15	298	286	94.2	3	859	715	90.3
Gamma frailty of variance 1																
0.50	-8	239	234	93.9	-16	250	247	94.5	-11	392	379	93.5	-1	409	400	94.4
1.00	-10	208	203	94.1	-14	235	229	93.8	-8	352	335	93.2	-3	399	380	91.9
1.50	-15	203	195	93.2	-24	261	243	92.5	3	344	326	92.6	12	442	408	92.3
2.00	-13	202	195	93.4	-35	304	280	92.6	1	344	326	92.4	18	519	473	92.9
2.25	-14	211	197	92.5	-41	363	315	90.8	2	359	330	91.3	10	620	533	91.1
2.50	-14	219	201	92.7	-53	435	371	90.0	4	369	336	91.6	8	747	630	90.0
2.75	-20	234	208	90.4	-112	688	518	88.1	12	393	347	92.0	29	1156	880	87.3

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald 95% confidence interval (×100). Based on 1,000 Monte Carlo replications.

Table 2.4: Summary Results of Simulation 2 for the Average Effect of Covariate $\overline{\boldsymbol{\beta}}_0(l,u)$

	$b_0(t) = \exp\{-t/\exp(1)\}$											
	Proposed Method				Fine et al. (2004)				Lin et al. (2000)			
	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Interval (0, 3]												
$\sigma^2 = 0$	2	199	186	93.0	18	606	209	81.7	-102	180	176	91.1
$\sigma^2 = 0.5$	-7	279	258	92.4	24	1163	316	68.5	-109	261	260	91.8
$\sigma^2 = 1$	1	339	313	92.3	-2	1834	385	59.8	-96	337	322	91.4
Interval (0, 2.5]												
$\sigma^2 = 0$	3	213	196	92.8	7	240	214	91.6	-150	180	176	85.5
$\sigma^2 = 0.5$	21	292	265	90.7	18	324	294	91.8	-130	279	261	90.7
$\sigma^2 = 1$	-1	342	317	92.5	8	399	354	90.4	-145	337	322	90.5
Interval (0.5, 2.5]												
$\sigma^2 = 0$	-7	189	182	94.7	-5	226	209	93.6	-85	180	176	92.3
$\sigma^2 = 0.5$	-14	265	257	93.8	-16	321	302	93.6	-93	261	260	92.7
$\sigma^2 = 1$	-3	333	313	93.0	6	413	368	90.6	-80	337	322	92.0

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald 95% confidence interval (×100). Based on 1,000 Monte Carlo replications. For the proportional means model of Lin et al. (2000), the same estimates from the conventional estimation were used based on the whole observed data regardless of the time interval under consideration for averaging.

distribution over range [0, 1] with mean 1/3 and variance 1, and had time-varying coefficient $b_0^{(1)} = \exp\{-t/\exp(1)\}$. Covariate $Z^{(2)}$ was from the uniform$[0,1]$ distribution, having the constant coefficient $b_0^{(2)}(t) = 0.5$. Random follow-up time C was generated conditionally on the second covariate such that $C|[Z^{(2)} \geq 0.5]$ from the uniform[0, 3] and $C|[Z^{(2)} < 0.5]$ from the uniform[2, 3].

Tables 2.5 and 2.6 report that the simulation results were similar to the ones discussed in the previous simulations, respectively for the covariate effects over time and the averaged effects of covariates.

2.4 Real Data Analyses

Our proposed method is illustrated by two real data analyses of the bladder tumor trial data (Byar, 1980) and the DISC study data (Tangpricha et al., 2017).

Table 2.5: Summary Results of Simulation 3 at the Prespecified Time Points

	$\log \mu_0(t) = \log t$								$b_0^{(1)}(t) = \exp\{-t/\exp(1)\}$								$b_0^{(2)}(t) = 0.5$							
	Proposed Method				Fine et al. (2004)				Proposed Method				Fine et al. (2004)				Proposed Method				Fine et al. (2004)			
t	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Zero intra-individual correlation																								
0.50	-6	219	206	93.3	-5	222	209	93.2	-2	264	254	94.3	-3	271	261	94.6	-5	263	247	92.9	-6	269	253	92.5
1.00	-7	161	152	94.3	-6	167	157	93.3	0	199	191	93.7	-1	212	201	93.3	0	198	186	94.1	-4	211	197	93.4
1.50	-4	134	129	93.2	-2	141	136	93.3	-2	172	165	93.8	-6	188	180	93.8	-2	171	162	92.4	-6	193	179	92.4
2.00	-1	119	116	94.0	2	129	126	93.6	-3	152	150	94.8	-9	176	171	94.2	-3	157	150	92.9	-6	194	179	91.4
2.25	-3	115	112	93.7	2	143	138	93.3	1	147	146	94.7	-6	198	186	92.4	-3	154	147	93.4	-9	215	195	91.6
2.50	-7	115	110	93.5	-5	173	160	92.5	7	150	146	94.2	4	240	217	91.1	-2	154	147	93.7	-14	252	228	92.1
2.75	-4	121	113	92.5	-1	245	211	89.0	4	158	150	93.8	-1	340	291	89.6	-3	165	153	93.5	-16	371	307	89.0
Gamma frailty of variance 0.5																								
0.50	-22	258	246	94.0	-21	264	250	93.9	-2	319	310	93.8	-4	333	318	93.1	23	315	303	94.0	21	322	311	94.3
1.00	-22	217	202	94.1	-21	228	210	93.5	9	282	260	93.2	6	302	277	94.1	17	265	256	93.9	16	289	273	93.5
1.50	-18	194	185	94.1	-15	202	197	94.9	7	256	243	93.7	5	277	268	92.9	13	249	239	93.5	3	284	270	94.5
2.00	-17	185	176	94.3	-14	204	194	93.8	4	247	234	94.0	2	285	271	94.1	13	244	233	93.5	1	307	287	93.8
2.25	-15	181	174	94.3	-16	227	217	93.6	3	246	233	93.4	3	331	304	92.3	9	242	232	94.3	-9	342	323	92.4
2.50	-15	182	173	93.9	-20	288	259	91.6	1	247	233	93.2	2	407	363	91.5	10	248	234	93.0	-12	426	385	92.1
2.75	-16	188	175	93.3	-36	422	355	89.6	6	253	236	93.0	13	586	495	89.1	9	257	239	92.9	-35	643	530	88.5
Gamma frailty of variance 1																								
0.50	-6	270	277	94.4	-7	279	282	95.1	-3	359	355	94.7	0	374	366	94.7	-18	351	346	94.2	-21	364	356	94.1
1.00	-6	242	240	94.7	-8	254	249	94.7	-2	324	315	95.0	0	351	336	94.5	-10	315	306	92.8	-13	339	328	93.4
1.50	-6	234	226	94.0	-11	252	241	93.6	2	313	301	95.2	9	353	333	94.3	-15	304	294	93.9	-19	357	334	93.6
2.00	-8	228	218	93.4	-13	255	241	92.1	8	311	295	94.9	19	368	341	93.2	-17	306	289	92.6	-37	398	359	90.9
2.25	-7	226	216	93.5	-17	289	271	92.7	7	309	294	94.6	8	416	385	93.8	-18	306	289	93.1	-30	449	403	91.9
2.50	-8	224	215	93.7	-21	357	322	91.4	7	310	294	94.4	5	507	459	91.4	-17	311	291	92.7	-34	549	480	90.6
2.75	-7	228	217	92.7	-16	516	438	89.9	5	318	297	93.7	-27	720	621	90.3	-18	320	296	92.3	-88	798	660	89.4

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald 95% confidence interval (×100). Based on 1,000 Monte Carlo replications.

Table 2.6: Summary Results of Simulation 3 for the Average Effects of Covariates: $\overline{\boldsymbol{\beta}}_0^{(1)}(l,u)$ and $\overline{\boldsymbol{\beta}}_0^{(2)}(l,u)$

	$b_0^{(1)}(t) = \exp\{-t/\exp(1)\}$												$b_0^{(2)}(t) = 0.5$											
	Proposed Method				Fine et al. (2004)				Lin et al. (2000)				Proposed Method				Fine et al. (2004)				Lin et al. (2000)			
	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Interval (0,3]																								
$\sigma^2 = 0$	0	171	158	91.6	12	361	174	82.5	-157	142	141	80.2	-1	173	155	92.0	5	437	169	82.6	-2	147	140	93.3
$\sigma^2 = 0.5$	11	253	237	93.4	30	845	255	73.3	-154	241	232	89.2	15	253	233	93.3	2	902	262	73.2	12	238	231	94.3
$\sigma^2 = 1$	5	311	294	94.4	46	1007	318	70.6	-154	303	296	89.8	-15	306	287	92.6	-90	1135	320	67.4	-16	301	291	93.6
Interval (0,2.5]																								
$\sigma^2 = 0$	-1	184	169	91.4	-5	202	185	91.8	-206	142	141	68.2	-1	186	165	92.1	-4	200	176	91.1				
$\sigma^2 = 0.5$	12	262	244	93.3	12	276	254	92.0	-202	241	232	84.5	17	261	240	93.1	7	289	252	93.1				
$\sigma^2 = 1$	5	317	300	94.5	9	339	313	93.5	-202	303	296	87.6	-15	312	293	92.7	-21	340	308	91.8				
Interval (0.5,2.5]																								
$\sigma^2 = 0$	-3	157	154	94.2	-7	170	164	93.3	-141	142	141	83.5	-3	161	151	92.8	-7	176	162	92.9				
$\sigma^2 = 0.5$	4	250	235	93.8	2	269	251	92.8	-137	241	232	90.1	14	242	231	93.3	4	269	251	93.8				
$\sigma^2 = 1$	1	305	294	94.8	6	337	314	94.2	-138	303	296	90.5	-15	298	287	93.2	-23	338	311	92.4				

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald 95% confidence interval (×100). Based on 1,000 Monte Carlo replications. For the proportional means model of Lin et al. (2000), the same estimates from the conventional estimation were used based on the whole observed data regardless of the time interval under consideration for averaging; thus the results for intervals (0, 2.5] and (0.5, 2.5] are not presented as they are equivalent to the ones for interval (0, 3].

2.4.1 Analysis of the Bladder Tumor Trial Data

Byar (1980) reported a randomized clinical trial assessing the effects of new treatments on tumor recurrence. A total of 118 patients with superficial bladder tumors entered the study after removal of their initial tumors. They were randomly assigned to three treatment arms (pyridoxine, thiotepa, and placebo). Follow-up time varied from 1 to 64 months. During the follow-up, sixty-two patients experienced tumor recurrences, and the maximum number of recurrences was 9. We applied our dynamic regression model with four baseline covariates: the size of a largest initial tumor, the number of tumors at enrollment, and two treatment indicators for pyridoxine and thiotepa.

Figure 2.1 shows the estimated effects of the covariates and corresponding Wald-type 95% point-wise bootstrap confidence intervals. The effects of initial number and size of tumors appeared to be changing over time. Specifically, the initial number of tumors was associated with lower recurrence rate around the middle of study follow-up but with higher recurrence rate toward the end. In contrast, the initial tumor size increased the mean frequency of tumor occurrence by about $\exp(0.1 * \text{size})$ around 5 months after the removal of initial tumors, and the size decreased the mean frequency by the same factor after 50-month of follow-up. The p-values for the proposed constant effect test were 0.105 and 0.030, respectively, for $(l, u] = (5, 53]$. On the other hand, pyridoxine and thiotepa had relatively constant effects, with p-values 0.960 and 0.691 respectively. Thiotepa appeared to reduce the tumor recurrence compared to the placebo, whereas pyridoxine did not. This observation is consistent with the results of other bladder tumor data analyses discussed in Therneau and Grambsch (2000).

2.4.2 Analysis of the DISC Trial Data

Cystic fibrosis is a chronic genetic disorder, which occurs at a rate of 3 cases per 10,000 Caucasian newborns each year in the U.S.; see Strausbaugh and Davis (2007).

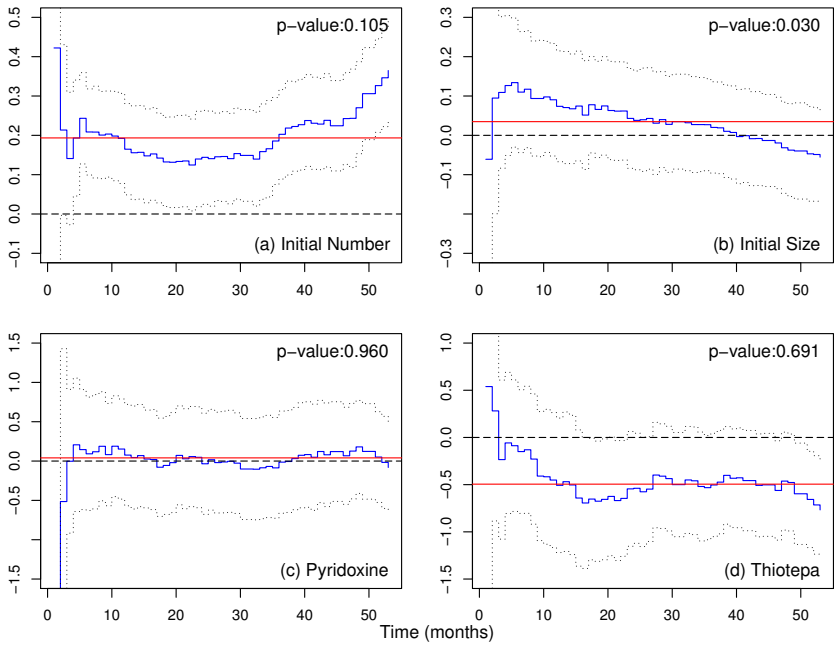

Figure 2.1: The bladder tumor trial data. Estimates for the effects of the considered covariates. The rugged solid lines and the dotted lines denote the point estimates and the Wald-type 95% point-wise bootstrap confidence intervals, respectively. The horizontal solid lines are the estimated average effects over (5, 53]. The p-values for the constant effect tests are calculated based on time interval (5, 53].

Also, it is reported that vitamin D deficiency in cystic fibrosis patients is highly prevalent, e.g., Wolfenden et al. (2008). In order to assess the effect of vitamin D on recurrence of pulmonary exacerbation, ninety-one adults with cystic fibrosis were randomly assigned to either high-dose vitamin D group or placebo group in the DISC trial of Tangpricha et al. (2017). During the follow-up, pulmonary exacerbations were observed as many as 9 times in each patient. In addition to vitamin D treatment, we considered additional five covariates in the model: gender, Caucasian, age, body mass index (BMI), and forced expiratory volume (FEV) at enrollment.

Figure 2.2 shows the estimated effects of covariates and corresponding Wald-type 95% point-wise bootstrap confidence intervals. The age effect appeared to vary over time on recurrence of pulmonary exacerbation although the test for the constant effect was not significant with p-value 0.163 for $(l, u] = (50, 500]$ (days). Younger people tended to have more frequent pulmonary exacerbation, perhaps even more so at the end of the study follow-up. On the other hand, the effects of the other covariates, including vitamin D treatment, did not appear to change over time in the sample. As one reviewer pointed out, the relatively constant effect estimates suggest that a time-independent coefficient model fits well. This shows another utility of the proposed methods, to provide justification for such a simpler model.

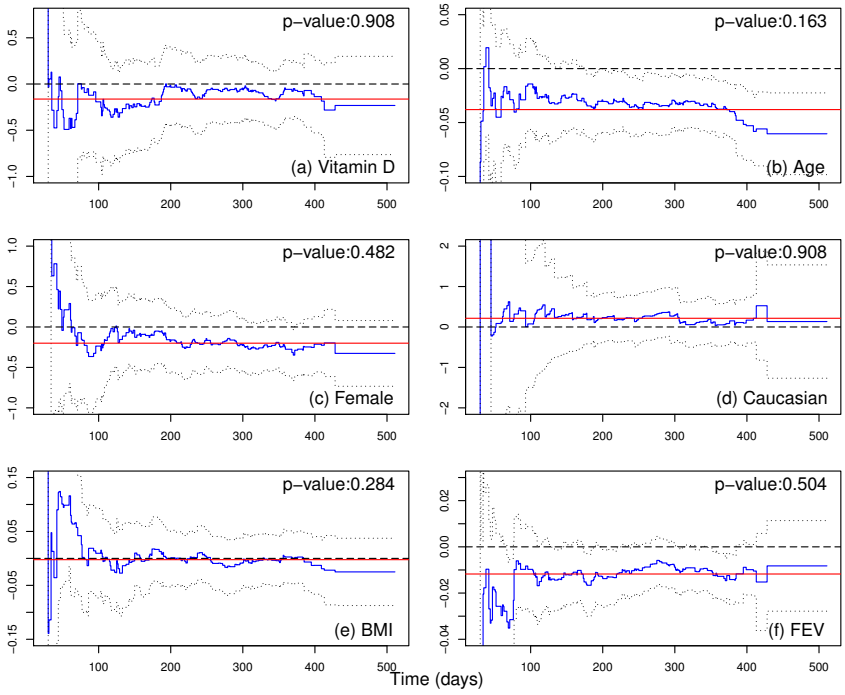

Figure 2.2: The DISC trial data. Estimates for the effects of covariates. The rugged solid lines and the dotted lines denote the point estimates and the Wald-type 95% point-wise bootstrap confidence intervals, respectively. The horizontal solid lines are the estimated average effects over (50, 500]. The p-values for the constant effect tests are calculated based on time interval (50, 500].

Chapter 3

A Varying-Coefficient Model for Gap Times Between Recurrent Events

To evaluate effects of risk factors on recurrent events, many regression models have been proposed in a way of extending the proportional hazards model of Cox (1972) for univariate survival data, e.g., the models of Prentice et al. (1981) and Andersen and Gill (1982). These extended Cox models for recurrent events specify the intensity function as conditional on the process history. Such so-called conditional modeling may target the study-time scale (i.e., the time scale from the study origin), or the gap-time scale (i.e., the time elapsed from the last event). By the multiplicity nature of recurrent events, however, the intra-individual correlation between events is often exhibited in the data. Since conditional modeling typically requires that the baseline recurrent event process is either Markov or semi-Markov, the intra-individual correlation is tacitly presumed to be entirely due to the observed covariates when the intensity function is formulated.

To relax the Markovian assumption, some marginal models have been developed

on the study-time scale, e.g., the models of Pepe and Cai (1993), Lawless and Nadeau (1995), and Lin et al. (2000) among others. In marginal models, effects of covariates are formulated on a population-level quantity such as the rate or mean frequency of recurrent events so that the individual-level dependence structure is left completely unspecified. Such partial specification of event processes facilitates robust method developments. For example, the proportional means model of Lin et al. (2000) specifies the mean frequency function conditional on observed covariates. In this way, arbitrary intra-individual dependence structure is allowed among recurrent events.

Some marginal models for recurrent events have also been developed on the gap-time scale. In particular, Huang and Chen (2003) proposed a marginal proportional hazards model for gap times between recurrent events. Under the renewal assumption on event processes (i.e., gap times for a subject are iid replicates of, say, T), the recurrent gap times are marginally modeled with observed covariates, say, $\boldsymbol{Z}$, a p-dimensional vector of covariates. To be specific, the marginal cumulative hazard function of T given $\boldsymbol{Z}$ takes a form

$$\Lambda(t|\boldsymbol{Z}) = \Lambda_0(t) \exp\{\boldsymbol{b}_0^\top \boldsymbol{Z}\},$$

where $\Lambda_0(t)$ is an unspecified baseline cumulative gap-time hazard and $\boldsymbol{b}_0$ is a p-dimensional constant vector of regression coefficients.

Most existing models for recurrent events assume constant effects of covariates over time. This simplistic assumption can be unrealistic in practice. To accommodate varying effects of covariates, in this Chapter, we propose a marginal varying-coefficient model for gap times between recurrent events. The proposed model accommodates evolving effects of covariates on gap times marginally at the population level. The model is described in Section 3.1. Estimation and inference procedures for the time-varying coefficients are developed in Section 3.2. Also, in Section 3.2, we establish

large sample properties of the proposed estimator. Finite-sample behaviors of the proposed estimator are evaluated through Monte Carlo simulations in Section 3.3. In Section 3.4, an analysis of the bladder tumor trial data is presented to illustrate the proposed method.

3.1 Model

Suppose that the study follow-up starts with an initial recurrent event. Assign indexes $j = 0, 1, 2, \ldots$ to the recurrent events for a subject. Denote the gap time between $(j-1)$st and jth events by $T_{(j)}$ for $j \geq 1$. Therefore the recurrent event process can be represented by a collection of the gap times, i.e., $\boldsymbol{T} \equiv \{T_{(j)} : j = 1, 2, \ldots\}$.

By the multiplicity nature of recurrent events, gap times between recurrent events for a subject often exhibit certain homogeneity. Thus we consider the events for a subject are from a renewal process. That way, gap times $\{T_{(j)}\}_j$ as within a subject are considered as iid replicates of, say, T. However, sample is a mixture of possibly heterogeneous renewal processes. Moreover, we consider the heterogeneity among renewal processes is owing to both the observed covariates and unobserved random effect, i.e., frailty. We model the gap times marginally with the observed covariates. Specifically, we propose a marginal varying-coefficient model for the cumulative hazard function of the gap time T, namely

$$\Lambda(t|\boldsymbol{Z}) = \exp\{\boldsymbol{\beta}_0(t)^\top \widetilde{\boldsymbol{Z}}\}, \qquad \textit{for all } t \geq 0, \tag{3.1}$$

where $\boldsymbol{\beta}_0(t) = [\log \Lambda_0(t),\ \boldsymbol{b}_0(t)^\top]^\top$ and $\widetilde{\boldsymbol{Z}} = [1,\ \boldsymbol{Z}^\top]^\top$. Here, $\boldsymbol{b}_0(t)$ is a p-dimensional time-varying regression coefficient vector for covariates $\boldsymbol{Z}$. Technically, model (3.1) marginally postulates time-varying multiplicative effects of covariates on the baseline cumulative gap-time hazards $\Lambda_0(\cdot)$. Note that our proposed model is a structure-imposed nonparametric regression model, and it directly generalizes the semipara-

metric model of Huang and Chen (2003) by replacing constant $\boldsymbol{b}_0$ with varying $\boldsymbol{b}_0(\cdot)$. In a special case of a homogeneous sample with no covariates, our model reduces to the recurrent survival function considered in Wang and Chang (1999).

We consider recurrent events for a subject are observable up to a random follow-up time C, and adopt the conditional independent assumption on the censoring mechanism,

$$\mathbf{T} \perp C \mid \boldsymbol{Z}.$$

3.2 Estimation and Inference

Due to the right censoring, for subject i, we observe the first, say, $M_i - 1$ complete gap times $\{T_{i(j)}\}_{j=1}^{M_i-1}$ and a censored gap, that is, $T^+_{i(M_i)} \equiv C_i - \sum_{j=1}^{M_i-1} T_{i(j)}$, where $\sum^0 = 0$. Thus the data set consists of $\{T_{i(j)} : j = 1, 2, \ldots, M_i - 1; T^+_{i(M_i)}; \boldsymbol{Z}_i\}$, $i = 1, \ldots, n$, which are n independent replicates of $\{T_{(j)} : j = 1, 2, \ldots, M - 1; T^+_{(M)}; \boldsymbol{Z}\}$.

In our estimation and inference procedures, a subset of the data is used. Essentially, the subset leaves out the censored gap times for subjects with at least one complete gap time, i.e., subjects with $M - 1 > 0$. The rationale for using this subset is based on the established connection between a subset of observed gap times and clustered survival times with informative cluster size; see Huang and Chen (2003, Section 2). To be specific, with each individual i, given C_i and M_i, the complete gap times $\{T_{i(j)}\}_{j=1}^{M_i-1}$ have an exchangeable distribution under the assumption of renewal process and conditional independence censorship. This connection facilitates the use of the subset data as the clustered survival data with informative cluster size, as described in the following Section. For notational convenience, we introduce $S_i \equiv \max(M_i - 1, 1)$, $\Delta_i \equiv I(M_i - 1 \geq 1)$, and

$$X_{i(j)} \equiv \begin{cases} T_{i(j)} & \text{for} \quad \Delta_i = 1 \\ T^+_{i(j)} & \text{for} \quad \Delta_i = 0 \end{cases}, \quad j = 1, \ldots, S_i,$$

where $I(\cdot)$ is the indicator function. Then the subset of the data consists of $\{X_{i(j)} : j = 1, 2, \ldots, S_i; \Delta_i; \boldsymbol{Z}_i\}$, $i = 1, \ldots, n$, which are n independent replicates of $\{X_{(j)} : j = 1, 2, \ldots, S; \Delta; \boldsymbol{Z}\}$.

3.2.1 Point Estimation

The subset can be alternatively represented in counting process notation. Let $N_{(j)}(t) \equiv I(X_{(j)} \leq t, \Delta = 1)$ be the counting process that indicates whether jth gap time for a subject is observed or not, between $(j-1)$st event and the time t on the gap-time scale. Also, define the *at-risk* process $Y_{(j)}(t) \equiv I(X_{(j)} \geq t)$, which indicates whether jth gap time for a subject can be observed in the gap time t or later, from $(j-1)$st event.

Under model (3.1) and the conditional independent censoring assumption, it follows $E(dN_{(1)}(t)|\boldsymbol{Z}) = Y_{(1)}(t)d\Lambda(t|\boldsymbol{Z}) = Y_{(1)}(t)\, d\exp\{\boldsymbol{\beta}_0(t)^\top \widetilde{\boldsymbol{Z}}\}$. Thus the method of Peng and Huang (2007), which was originally developed for univariate survival time data, can be applied to the first gap time. The estimating equation of Peng and Huang (2007) takes a form

$$n^{-1}\sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \big[\, dN_{i(1)}(t) - Y_{i(1)}(t)\, d\exp\{\boldsymbol{\beta}(t)^\top \widetilde{\boldsymbol{Z}}_i\}\big] = 0, \qquad \textit{for all } t \geq 0, \tag{3.2}$$

where $N_{i(1)}(t) = I(X_{i(1)} \leq t, \Delta_i = 1)$ and $Y_{i(1)}(t) = I(X_{i(1)} \geq t)$.

Due to the exchangeability of the observed complete gap times, the subset data can be treated as observed clustered survival data. This motivates our estimating integral equation

$$n^{-1}\sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \big[N_{i(j)}(t) - \int_0^t Y_{i(j)}(s)\, d\exp\{\boldsymbol{\beta}(s)^\top \widetilde{\boldsymbol{Z}}_i\}\big] = 0, \qquad \textit{for all } t \geq 0, \tag{3.3}$$

where $N_{i(j)}(t) = I(X_{i(j)} \leq t, \Delta_i = 1)$ and $Y_{i(j)}(t) = I(X_{i(j)} \geq t)$. Basically, the

observed first gap time in equation (3.2) is replaced with the average of all observed gap times from an individual.

Equation (3.3) admits a càdlàg solution, $\widehat{\boldsymbol{\beta}}(t)$, which has jumps only at observed gap times. Write $0 < x_1 < x_2 < \cdots < x_G$ as the unique observed gap times from all individuals in the sample. The initial value, $\widehat{\boldsymbol{\beta}}(0-)$, satisfies $\exp\{\widehat{\boldsymbol{\beta}}(0-)^\top \widetilde{\boldsymbol{Z}}_i\} = 0$ for all i. Then, at time x_1, the estimating integral equation (3.3) reduces to

$$n^{-1}\sum_{i=1}^{n} S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\Big(\, dN_{i(j)}(x_1) - Y_{i(j)}(x_1)\exp\{\boldsymbol{\beta}(x_1)^\top\widetilde{\boldsymbol{Z}}_i\}\Big) = 0. \tag{3.4}$$

Sequentially at the observed event times $x_g, g = 2, \ldots, G$, the estimating integral equation (3.3) reduces to

$$n^{-1}\sum_{i=1}^{n} S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\Big(\, dN_{i(j)}(x_g) - Y_{i(j)}(x_g)\big[\exp\{\boldsymbol{\beta}(x_g)^\top\widetilde{\boldsymbol{Z}}_i\} - \exp\{\widehat{\boldsymbol{\beta}}(x_{g-1})^\top\widetilde{\boldsymbol{Z}}_i\}\big]\Big) = 0. \tag{3.5}$$

In this way, our estimator $\widehat{\boldsymbol{\beta}}(\cdot)$ is obtained as the solutions to the equations (3.4) and (3.5), which have good computational properties since the left-hand sides of both equations are monotone functions of $\boldsymbol{\beta}$.

Since all the observed complete gap times are used in equation (3.3), some efficiency gain in estimation is expected compared to the estimation based on equation (3.2). This will be investigated in the Section 3.3 via Monte Carlo simulations.

3.2.2 Large Sample Properties

We establish uniform consistency and weak convergence of the proposed estimator $\widehat{\boldsymbol{\beta}}(\cdot)$. Write $\|\cdot\|$ and eigmin$(\cdot)$ as the Euclidean norm and the minimum eigenvalue of a positive semidefinite matrix, respectively. Write $\boldsymbol{\beta}_0(t)$ as the true value of $\boldsymbol{\beta}(t)$. We impose the following regularity conditions:

C1. $\sum_{j=1}^{\infty} N_{(j)}(\tau)$ is bounded, where τ is a fixed time point satisfying $Pr(C > \tau) > 0$;

C2. $\|\boldsymbol{Z}\|$ is bounded;

C3. $\sup_{t\in[0,\tau]} \|\boldsymbol{\beta}_0(t)\|$ is bounded;

C4. $\boldsymbol{\beta}_0(t)$ is continuously differentiable for $t \in (0, \tau]$;

C5. eigmin$E\{Y_{(1)}(\tau)\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top\}$ is bounded away from zero.

Condition C3 implies that $\boldsymbol{\beta}_0(0)$ is bounded and thereby the gap time T has a probability mass as time 0, i.e., $Pr(T = 0|\boldsymbol{Z}) > 0$. This facilitates a proof of the consistency of $\widehat{\boldsymbol{\beta}}(\cdot)$ at time 0 and furthermore its uniform consistency over time $[0, \tau]$. The other conditions are technical assumptions and fairly standard with varying-coefficient regression models.

Theorem 3. Under regularity conditions C1-C5, $\sup_{t\in[0,\tau]} \|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\| \longrightarrow 0$, almost surely.

Theorem 4. Under regularity conditions C1-C5, $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\cdot) - \boldsymbol{\beta}_0(\cdot)\}$ on $[0, \tau]$ weakly converges to a mean-zero Gaussian process.

Detailed proofs of Theorems 3 and 4 are in the Appendix B.

3.2.3 Interval Estimation

Inference about $\boldsymbol{\beta}_0(t)$ for a given time t may be based on a variance estimate as can be obtained through the asymptotic study. In constructing confidence band for infinite-dimensional quantities such as $\boldsymbol{\beta}_0(\cdot)$, however, such an asymptotics-based approach is difficult if not impossible; see related discussion in Lin et al. (1994) and Huang (2014) in different contexts. For this reason, we instead propose a nonparametric inference procedure using the multiplier bootstrap; see Rubin (1981), Kosorok (2008)

and Huang (2014). Specifically, we propose a bootstrap estimating integral equation

$$n^{-1}\sum_{i=1}^{n}\xi_i S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\big[N_{i(j)}(t)-\int_0^t Y_{i(j)}(s)\,d\exp\{\boldsymbol{\beta}(s)^\top\widetilde{\boldsymbol{Z}}_i\}\big]=0,\qquad \textit{for all } t\geq 0, \tag{3.6}$$

where $(\xi_i)_{i=1}^n$ are a random sample from a distribution with unit mean and unit variance, e.g., the standard exponential distribution. Denote the solution path to this equation (3.6) by $\boldsymbol{\beta}^*(\cdot)$. Then based on ,say, B bootstrap resamples, a set of bootstrap solutions $\{\boldsymbol{\beta}_b^*(\cdot)\}_{b=1}^B$ is obtained. The variance of $\widehat{\boldsymbol{\beta}}(t)$ can be estimated by the sample variance of $\{\boldsymbol{\beta}_b^*(t)\}_{b=1}^B$. Provided this, the Wald-type $100(1-\alpha)\%$ pointwise confidence interval for $\boldsymbol{\beta}_0(t)$ can be constructed by the normal approximation centered at $\widehat{\boldsymbol{\beta}}(t)$. Alternatively, a confidence interval may be constructed with the corresponding $(\alpha/2)$th and $(1-\alpha/2)$th quantiles of the empirical distribution of $\boldsymbol{\beta}^*(t)$. We also propose to construct a $100(1-\alpha)\%$ confidence band for $\{\boldsymbol{c}^\top\boldsymbol{\beta}_0(t):\ t\in(l,u]\}$ with $\{\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}(t)\pm\gamma_\alpha\ :\ \ t\in(l,u]\}$, where $\boldsymbol{c}$ is a given $(p+1)$-dimensional vector. Here γ_α is the $(1-\alpha/2)$th quantile of the empirical distribution of J^*, where $J_b^*=\sup_{t\in[l,u]}|\boldsymbol{c}^\top\{\boldsymbol{\beta}_b^*(t)-\widehat{\boldsymbol{\beta}}(t)\}|$ for $b=1,\ldots,B$.

3.2.4 Average Effect and Test for Varying Effect

Our proposed estimator $\widehat{\boldsymbol{\beta}}(\cdot)$ estimates evolving effects of covariates over time. Nevertheless one may be interested in an averaged effect of the covariates. We consider $\overline{\boldsymbol{\beta}}_0(l,u)\equiv(u-l)^{-1}\int_l^u\boldsymbol{\beta}_0(t)\,dt$ for $0<l<u<\infty$; see Peng and Huang (2008) for its usage in quantile regression. A natural estimator for $\overline{\boldsymbol{\beta}}_0(l,u)$ based on our estimator $\widehat{\boldsymbol{\beta}}(\cdot)$ is $\overline{\widehat{\boldsymbol{\beta}}}(l,u)=(u-l)^{-1}\int_l^u\widehat{\boldsymbol{\beta}}(t)\,dt$. Moreover, we consider a test for constant vs. time-varying effects

$$H_0:\boldsymbol{c}^\top\boldsymbol{\beta}_0(t)\ \textit{is a constant for all } t\in(l,u],$$

where $\boldsymbol{c}$ is a known $(p+1)$-dimensional vector. Note that in the special case of testing the varying effect of the kth covariate only, the $(k+1)$th component of $\boldsymbol{c}$ is set to be one and zeros elsewhere. To evaluate H_0, we propose a test statistic $\mathcal{T} = n^{1/2}\int_l^u \boldsymbol{c}^\top\{\widehat{\boldsymbol{\beta}}(t) - \overline{\widehat{\boldsymbol{\beta}}}(l,u)\}\Xi(t)\,dt$, where $\Xi(t)$ is a known nonnegative weight function. For example, the identity function $\Xi(t) = t$ is a good candidate for the weight when a linear changing effect is reasonable. Note that the limiting distribution of $\mathcal{T}$ is Gaussian because $\mathcal{T}$ is a linear function of $\widehat{\boldsymbol{\beta}}(\cdot)$. For variance estimation of the limiting distribution of $\mathcal{T}$, we consider $\mathcal{T}^* = n^{1/2}\int_l^u \boldsymbol{c}^\top\{\boldsymbol{\beta}^*(t) - \overline{\boldsymbol{\beta}^*}(l,u)\}\Xi(t)\,dt$ in a bootstrap sample to be the counter part of $\mathcal{T}$, where $\overline{\boldsymbol{\beta}^*}(l,u) = (u-l)^{-1}\int_l^u \boldsymbol{\beta}^*(t)\,dt$. Since the limiting distribution of $\mathcal{T}$ is equivalent to the limiting conditional distribution of $\mathcal{T}^*$ given the observed data, a p-value for a constant effect test can be calculated based on the bootstrap replicates of $\mathcal{T}^*$. Specifically, one can estimate the variance of the limiting distribution of $\mathcal{T}$ by an empirical variance of the bootstrap replicates of $\mathcal{T}^*$. Then under H_0, a p-value can be obtained based on Gaussian distribution with mean zero and the estimated variance.

3.3 Simulation Studies

We conducted Monte Carlo simulations to investigate finite-sample properties of the proposed estimator. We adapted the gap-time data generation strategy of Huang and Chen (2003), in which various degrees of intra-individual correlation can be accommodated in Monte Carlo sample. Monte Carlo gap times were generated from a marginal cumulative gap-time hazard function $\Lambda(\cdot)$ in the form of model (3.1), i.e., $\Lambda(t|\boldsymbol{Z}) = \exp\{\log t + \boldsymbol{b}_0(t)^\top\boldsymbol{Z}\}$. Specifically, Monte Carlo gap times were generated from $\Lambda^{-1}(U)$, where U was a random variable from the standard exponential distribution. In order to address the intra-individual correlation, random variable U was set to $-\log\{1-\Phi(A+B)\}$ for a mixture of heterogeneous renewal processes, where

individual-specific random variable A and episode-specific random variable B were independently generated from mean-zero normal distributions with variance ρ and $1-\rho$, respectively, for $\rho \in [0,1]$. Here $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Note that $\rho = 0$ leads to the same baseline renewal process across the subjects, whereas $\rho = 1$ generates gap times of the same length within a subject. That way, larger $\rho \in [0,1]$ accommodates higher intra-individual correlation in the sample.

The size of each Monte Carlo sample was $n = 200$, and the bootstrap size B was 500. A total of 1,000 Monte Carlo data sets was generated under each intra-individual correlation scenario, $\rho = 0.25$, 0.5, or 0.75. To compare estimation performance of the proposed method, we also considered the methods of Peng and Huang (2007) and Huang and Chen (2003). Note the method of Huang and Chen (2003) is valid only when all effects are constant. However, their method was applied in average effect estimation even when the effect under consideration was time-varying; and the conventional estimates were used based on the whole observed data over interval [0, 3] regardless of averaging time interval under consideration.

3.3.1 Simulation 1: Single Covariate with Constant Effect

Firstly, we considered a single covariate with constant effect on gap times. The covariate Z was from the uniform[-1, 1] distribution, and its fixed regression coefficient $b_0(\cdot)$ was one. Random follow-up time C was generated from the uniform[0, 3] distribution.

Table 3.1 reports summaries of the Monte Carlo estimates from our proposed method and the method of Peng and Huang (2007) in terms of the empirical bias, empirical standard deviation of the estimates, average standard error estimate, and empirical coverage probability of Wald-type 95% confidence intervals at prespecified time points. On the whole, the Monte Carlo biases of the estimates from both methods were close to zero. Moreover, the averages of the standard errors agreed well with the

empirical standard deviations of the estimates. In particular, relatively large empirical standard deviations were observed for the high intra-individual correlation scenario, i.e., $\rho = 0.75$. These observations were owing to increased unobserved heterogeneity across the subjects given the observed covariates. On the other hand, the proposed estimator appeared to have smaller empirical standard deviations, compared to the application of the method of Peng and Huang (2007) to the firstly observed complete gap times. This is because our estimation utilizes all the complete gap times, not only the firstly observed event times.

Table 3.2 shows summary results of the average effect estimates from three methods of Peng and Huang (2007), Huang and Chen (2003), and ours. For the methods of Peng and Huang (2007) and ours, we considered the average effect estimator proposed in Section 3.2.4, over averaging time interval (0.5, 2.0]. All the average effect estimates were close to the true value 1, with less than 2% Monte Carlo biases. Compared to our proposed average effect estimates, the estimates from the methods of Huang and Chen (2003) and Peng and Huang (2007), respectively, had smaller and bigger empirical standard deviations.

3.3.2 Simulation 2: Single Covariate with Time-Varying Effect

In the second simulation, we considered a single covariate with time-varying effect. The covariate Z followed the uniform[-1, 1] distribution, and its time-varying regression coefficient was a decreasing function of time; that was, $b_0(t) = (1 + 2t)^{-1}$. Random follow-up time C was generated from the uniform[0, 3].

Table 3.3 shows similar summary results to the ones in Table 3.1. Likewise, the Monte Carlo biases were close to zero over the prespecified time points. Furthermore, the averages of the standard errors agreed well with the empirical standard deviations, and the empirical coverage probabilities of the Wald-type 95% confidence intervals

Table 3.1: Summary Results of Simulation 1 Over Prespecified Time Points

	$\log\Lambda_0(t) = \log t$								$b_0(t) = 1.0$							
	Proposed Method				Peng and Huang (2007)				Proposed Method				Peng and Huang (2007)			
t	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Intra-individual correlation ($\rho = 0.25$)																
0.25	-20	133	133	95.1	-25	171	167	94.8	16	215	216	94.3	7	278	273	94.5
0.50	-11	106	108	95.7	-14	125	127	95.4	4	182	181	94.9	0	220	215	94.0
0.75	-7	100	101	95.8	-10	115	113	95.1	4	179	173	93.6	-4	203	197	94.5
1.00	-5	103	100	93.8	-8	114	109	93.5	2	183	175	94.3	-4	205	194	93.7
1.25	-2	104	103	95.2	-10	113	110	94.0	10	195	185	93.6	0	214	201	93.2
1.50	1	114	110	94.2	-8	124	115	94.0	17	215	201	93.5	1	231	216	93.5
1.75	-3	129	121	94.4	-15	140	126	94.4	17	240	224	93.4	-1	264	239	92.2
Intra-individual correlation ($\rho = 0.5$)																
0.25	-18	139	134	95.2	-34	173	168	94.7	10	230	219	94.7	7	286	275	94.9
0.50	-15	113	110	94.6	-25	131	128	93.5	15	187	185	95.1	12	219	216	94.7
0.75	-10	106	102	93.9	-17	120	114	93.6	3	171	176	95.5	0	197	197	95.0
1.00	-7	106	101	94.8	-13	116	109	93.6	11	183	179	94.6	5	201	194	94.3
1.25	-4	110	104	94.2	-11	117	110	93.6	10	196	188	93.7	-2	212	200	93.7
1.50	-6	116	111	94.4	-15	124	115	93.9	7	218	203	93.0	-9	236	213	93.7
1.75	-9	129	121	94.2	-21	130	124	94.5	8	252	224	92.4	-9	263	234	92.9
Intra-individual correlation ($\rho = 0.75$)																
0.25	-13	143	139	94.1	-24	174	167	94.4	8	241	230	94.4	12	291	272	93.6
0.50	-6	115	113	94.7	-13	131	127	94.8	6	199	192	94.4	4	226	215	94.5
0.75	-5	110	105	94.5	-12	119	114	94.6	6	193	182	93.0	1	213	197	93.4
1.00	-3	108	104	94.0	-9	113	109	94.2	3	196	183	92.9	1	209	194	93.0
1.25	1	110	106	94.9	-5	113	110	95.0	-2	202	192	93.0	-6	216	200	92.4
1.50	-1	120	113	93.4	-8	122	115	93.4	4	225	208	93.5	-7	235	214	92.8
1.75	-1	140	124	92.8	-11	140	126	93.3	4	264	232	92.1	-9	267	237	92.2

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald-type 95% point-wise confidence interval (×100). Based on 1,000 Monte Carlo replications.

Table 3.2: Summary Results of Simulation 1 for the Average Effect of Covariate $\overline{\boldsymbol{\beta}}_0(l, u)$

	$\overline{\boldsymbol{\beta}}_0(l, u) = 1.0$											
	Proposed Method				Peng and Huang (2007)				Huang and Chen (2003)			
	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Interval (0.5,2.0]												
$\rho = 0.25$	9	167	160	93.4	-3	178	170	93.4	3	156	151	94.1
$\rho = 0.5$	9	169	162	94.2	-1	176	169	94.8	10	158	154	94.2
$\rho = 0.75$	4	179	166	92.7	-3	184	169	92.7	15	158	157	95.0

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald-type 95% confidence interval (×100). Based on 1,000 Monte Carlo replications. For the method of Huang and Chen (2003), the same estimates from the conventional estimation were used based on the whole observed data over interval [0, 3].

Table 3.3: Summary Results of Simulation 2 Over Prespecified Time Points

	$\log \Lambda_0(t) = \log t$								$b_0(t) = (1+2t)^{-1}$							
	Proposed Method				Peng and Huang (2007)				Proposed Method				Peng and Huang (2007)			
t	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Intra-individual correlation ($\rho = 0.25$)																
0.25	-16	127	126	94.8	-21	156	162	95.8	7	219	211	93.4	4	276	269	94.5
0.50	-11	104	102	94.2	-13	123	122	95.1	9	180	175	94.5	6	213	207	95.4
0.75	-5	98	95	94.8	-8	111	108	94.4	9	172	164	93.3	8	194	186	93.6
1.00	-3	97	94	94.9	-7	105	103	94.3	6	167	163	94.9	2	187	178	94.1
1.25	-1	98	97	95.0	-7	102	102	95.4	5	175	167	93.7	1	189	178	93.3
1.50	1	106	101	94.2	-8	112	105	92.9	0	190	176	92.7	-4	200	183	93.1
1.75	1	117	108	92.9	-10	120	110	93.1	-3	207	187	93.1	-10	216	192	92.7
Intra-individual correlation ($\rho = 0.5$)																
0.25	-22	132	128	94.3	-35	163	163	96.1	5	223	216	93.4	8	281	271	93.7
0.50	-13	107	105	94.3	-19	124	122	94.7	-5	179	179	94.4	-5	209	208	94.7
0.75	-8	96	97	95.5	-13	108	108	94.3	-7	169	168	94.3	-7	191	186	94.6
1.00	-8	95	96	95.5	-13	103	103	95.6	-3	167	166	94.5	-6	182	178	94.5
1.25	-9	96	98	95.9	-15	102	102	95.2	1	177	170	93.7	-1	185	178	93.0
1.50	-7	102	102	94.8	-12	107	104	93.5	2	194	178	92.4	-1	196	182	92.2
1.75	-9	112	108	93.9	-16	114	109	93.7	-1	209	189	92.5	-8	213	192	91.0
Intra-individual correlation ($\rho = 0.75$)																
0.25	-12	137	134	94.5	-21	164	161	95.5	-5	219	227	95.7	-12	262	269	95.8
0.50	-10	112	108	94.2	-16	126	122	94.0	2	189	187	95.3	-2	208	208	95.7
0.75	-5	106	100	94.0	-11	112	108	94.4	-1	176	173	94.6	-4	190	187	93.9
1.00	0	101	98	94.3	-7	104	103	94.7	-1	175	171	94.5	-8	185	179	94.8
1.25	2	104	100	94.0	-4	106	102	94.5	2	186	174	93.7	-2	194	179	93.3
1.50	3	109	104	93.2	-4	109	104	94.0	0	194	181	94.2	-3	198	183	93.6
1.75	3	117	109	93.5	-4	118	109	93.4	-1	213	192	92.8	-3	216	193	91.7

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald-type 95% point-wise confidence interval (×100). Based on 1,000 Monte Carlo replications.

were close to the nominal level 95%. These indicate that our proposed nonparametric inference procedure worked well. On the other hand, the empirical standard deviations of the proposed estimator were smaller than the ones based on the estimation of Peng and Huang (2007), as discussed in the previous simulation.

Table 3.4 shows summary results of the averaged effect estimates for $\overline{\boldsymbol{\beta}}_0(l, u)$, cf., $\overline{\boldsymbol{\beta}}_0(0.5, 2.0) = 0.305$ and $\overline{\boldsymbol{\beta}}_0(0, 3) = 0.324$. On the whole, relatively small Monte Carlo biases were observed. Moreover, the closeness between the empirical standard deviations and the average of standard errors of our method indicates the proposed nonparametric inference procedure worked well.

Table 3.4: Summary Results of Simulation 2 for the Average Effect of Covariate $\overline{\boldsymbol{\beta}}_0(l,u)$

	$\overline{\boldsymbol{\beta}}_0(0.5,2.0)=0.305$											
	Proposed Method				Peng and Huang (2007)				Huang and Chen (2003)			
	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Interval (0.5,2.0]												
$\rho=0.25$	3	153	147	93.0	-1	161	152	93.5	36	150	145	94.2
$\rho=0.5$	-3	156	149	93.2	-6	159	152	93.2	28	148	147	94.0
$\rho=0.75$	-1	159	152	94.3	-5	160	153	93.5	34	151	149	93.9

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald-type 95% confidence interval (×100). Based on 1,000 Monte Carlo replications. For the method of Huang and Chen (2003), the same estimates from the conventional estimation were used based on the whole observed data over interval [0, 3], cf., $\overline{\boldsymbol{\beta}}_0(0,3)=0.324$.

3.3.3 Simulation 3: Two Covariates with Constant and Time-Varying Effects

In simulation 3, we considered two covariates, one with constant effect and the other with time-varying effect. Specifically, covariates $Z^{(1)}$ and $Z^{(2)}$ were independently generated from the uniform [-1, 1] distribution. $Z^{(1)}$ had a constant effect $b_0^{(1)}(\cdot)=1$ on the baseline hazard function, and $Z^{(2)}$ had a time-varying effect $b_0^{(2)}(t)=(1+2t)^{-1}$. Random follow-up time C was generated conditional on the second covariate such that C followed the uniform[0, 3] if $Z^{(2)}\geq 0$, otherwise from the uniform[1, 3].

Table 3.5 reports summary results of the Monte Carlo estimates, in which similar patterns were observed to the ones discussed in the previous simulations, cf., Tables 3.1 and 3.3. These observations demonstrate good finite-sample properties of the proposed estimator.

3.4 Analysis of the Bladder Tumor Trial Data

Byar (1980) reported a randomized clinical trial assessing experimental treatments' effects on tumor recurrence. In the trial, a total of 118 patients with superficial bladder tumors participated in the study after removal of their initial tumors. The patients were randomly assigned to three treatment arms: pyridoxine, thiotepa, and

Table 3.5: Summary Results of Simulation 3 Over Prespecified Time Points

	$\log \Lambda_0(t) = \log t$								$b_0^1(t) = 1.0$								$b_0^2(t) = (1+2t)^{-1}$							
	Proposed Method				Peng and Huang (2007)				Proposed Method				Peng and Huang (2007)				Proposed Method				Peng and Huang (2007)			
t	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95	B	SD	SE	CP95
Intra-individual correlation ($\rho = 0.25$)																								
0.25	-13	129	126	93.9	-32	176	168	94.7	-2	205	201	94.7	-4	267	265	94.9	-1	194	183	93.0	2	274	254	93.4
0.50	-12	104	103	94.7	-22	128	125	94.8	6	171	170	95.5	4	209	211	95.5	-2	165	158	94.3	-7	210	200	94.2
0.75	-7	93	95	95.6	-14	106	109	95.4	7	166	164	94.6	3	196	193	94.6	-2	162	153	93.8	-7	192	182	94.3
1.00	-7	94	93	94.6	-15	103	103	95.5	7	168	164	94.0	1	191	186	94.4	-4	166	156	92.7	-9	186	176	93.4
1.25	-6	98	95	94.0	-15	105	103	94.3	13	174	171	95.0	6	192	188	94.7	-5	175	162	93.1	-10	190	178	92.9
1.50	-8	104	101	93.8	-20	112	107	92.4	13	198	184	93.5	0	211	199	93.7	-7	190	173	92.7	-12	200	185	93.4
1.75	-4	121	111	93.2	-20	126	117	93.4	21	232	204	91.8	2	247	218	92.1	-3	202	187	93.0	-12	210	197	93.4
Intra-individual correlation ($\rho = 0.5$)																								
0.25	-12	132	128	94.3	-30	174	168	95.2	3	225	207	92.8	9	288	265	93.4	10	201	191	93.5	5	268	252	94.1
0.50	-5	105	105	94.7	-16	123	125	94.8	1	193	175	91.9	1	232	210	92.2	4	172	164	94.2	2	211	199	94.9
0.75	-5	99	97	94.3	-12	109	109	95.3	2	179	168	93.4	-4	203	192	93.8	2	165	159	93.5	-5	188	182	94.2
1.00	-6	94	96	95.5	-16	101	103	96.0	-3	179	169	93.3	-12	199	186	93.3	1	167	161	93.8	-9	184	176	93.1
1.25	-7	100	98	95.0	-16	108	103	93.4	-1	189	176	93.2	-9	203	188	92.7	5	172	166	94.6	-2	184	178	93.4
1.50	-7	109	104	94.2	-20	115	108	93.8	7	211	188	92.2	-7	223	199	92.3	2	192	176	92.4	-2	202	184	93.4
1.75	-6	124	114	93.6	-20	126	118	93.2	7	242	208	91.3	-7	253	218	91.4	8	205	189	93.3	2	212	196	93.0
Intra-individual correlation ($\rho = 0.75$)																								
0.25	-15	140	136	94.6	-35	170	170	96.1	16	215	219	94.3	19	268	266	94.4	5	220	207	93.5	12	264	253	93.7
0.50	-6	111	110	94.9	-21	129	125	95.1	1	191	184	94.0	0	219	211	94.4	3	181	174	93.2	3	212	199	93.3
0.75	-7	102	101	95.8	-21	114	110	94.6	2	183	175	94.5	-6	203	192	93.4	1	170	165	93.5	0	186	182	94.8
1.00	-8	103	98	93.9	-20	108	104	94.0	3	182	175	94.5	-7	195	186	94.2	1	174	165	93.6	-3	185	177	93.3
1.25	-5	106	100	92.8	-18	108	103	93.8	8	190	181	93.8	-2	200	189	93.3	8	179	171	93.9	0	185	178	93.9
1.50	-4	119	106	92.1	-16	121	108	91.3	14	205	194	93.2	4	213	200	93.4	5	193	180	93.5	-3	197	185	93.6
1.75	-6	131	116	93.2	-20	131	118	92.8	13	232	215	93.4	-2	235	219	93.0	4	211	194	92.7	-6	215	197	92.6

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); CP95: empirical coverage probability of the Wald-type 95% point-wise confidence interval (×100). Based on 1,000 Monte Carlo replications.

placebo. Follow-up times ranged between 1 and 64 months. Sixty-two patients experienced tumor recurrences during the follow-up, and the number of recurrences was as many as 9. In our analysis, we considered four covariates: the size of a largest initial tumor, the number of tumors at enrollment, and two treatment indicators for pyridoxine and thiotepa.

Figure 3.1 shows the estimated effects of covariates on the occurrence of next tumor, and the corresponding Wald-type 95% point-wise bootstrap confidence intervals. On the whole, thiotepa appeared to be most effective treatment, and the number of initial tumors was observed to be highly associated with the successive tumor recurrence; these observations were consistent with other analysis results discussed in Therneau and Grambsch (2000). Notice that thiotepa effect appeared to be changing over time. Specifically, the magnitude of its estimated preventive effect on next tumor recurrence was increasing. This was consistent with our constant effect test with p-value 0.031 over time interval $(l, u] = (1, 20]$. In contrast, effects of the number of initial tumors, the size of a largest tumor, and treatment pyridoxine appeared to be relatively constant, with large p-values 0.827, 0.753, and 0.473, respectively.

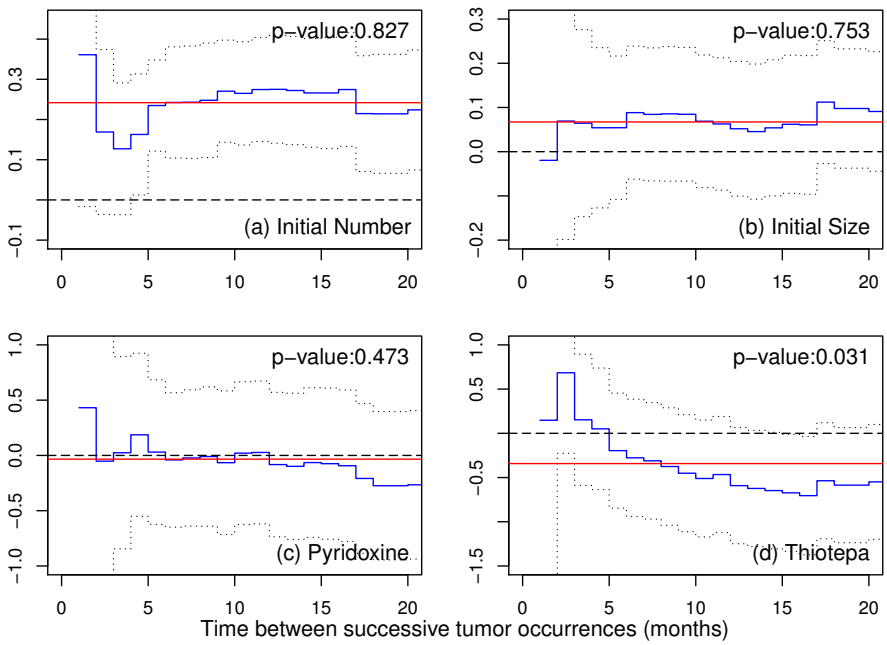

Figure 3.1: The bladder tumor trial data. Estimated effects of covariates on the baseline cumulative hazard function of gap times between successive tumor occurrences. The rugged solid lines and the dotted lines denote the point estimates and the Wald-type 95% point-wise bootstrap confidence intervals, respectively. The horizontal solid lines are the estimated average effects over (1, 20]. The p-values for the constant effect tests are calculated over time interval (1, 20].

Chapter 4

Semiparametric Survival Regression with a Mixture of Time-Varying and Constant Effects

In the analysis of survival time data, the proportional hazards model of Cox (1972) has been the most popular choice of regression models and a foundation of many other model developments. One key assumption of the original Cox model is that all covariates have constant multiplicative effects on survival times. Specifically, the Cox model for the hazard function of the survival time T takes a form

$$\lambda(t|\boldsymbol{W}) = \lambda_0(t)\exp(\boldsymbol{\alpha}_0^\top \boldsymbol{W}),$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and $\boldsymbol{W}$ is a vector of $(p+q)$-dimensional covariates. However, the proportionality assumption may not always be realistic in practice, but the effects may actually vary over time. To accommodate time-varying effects of covariates, a varying-coefficient Cox model with time-varying regression coefficients $\boldsymbol{\alpha}_0(t)$, in place of constant $\boldsymbol{\alpha}_0$, has been studied by many researchers; see Zucker and Karr (1990), Cai and Sun (2003), and Tian et al. (2005)

among others. On the other hand, Peng and Huang (2007) proposed an alternative varying-coefficient model for the survival function, rather than the hazard function, allowing all effects of covariates to vary over time. To be specific, the temporal survival regression model of Peng and Huang (2007) takes a form

$$S(t|\boldsymbol{W}) = \exp\big[-\exp\{\log\Lambda_0(t) + \boldsymbol{\alpha}_0(t)^\top \boldsymbol{W}\}\big], \qquad \textit{for all } t \geq 0, \tag{4.1}$$

where $\Lambda_0(\cdot)$ is an unspecified baseline cumulative hazard function. Essentially, model (4.1) postulates multiplicative time-varying effects of covariates on the baseline cumulative hazards $\Lambda_0(\cdot)$, and thereby on the unspecified baseline survival function $S_0(t) \equiv \exp\{-\Lambda_0(t)\}$. In a special case of constant regression coefficients, i.e., $\boldsymbol{\alpha}_0(\cdot) = \boldsymbol{\alpha}_0$, the temporal survival regression model reduces to the proportional hazards model.

Estimation efficiency is expected to decrease in such varying-coefficient models due to the increased model generality. Especially, when sample size is small and the number of covariates is relatively large, estimation of functional regression coefficients may not have good precision. Therefore, a model with a mixture of time-varying and time-constant effects can be a useful alternative in practice. This is the motivation of our semiparametric model development.

In Section 4.1, we propose a semiparametric survival regression model with a mixture of time-varying and constant effects of covariates. The proposed model is a sub-model of the temporal survival regression of Peng and Huang (2007), and it provides a middle ground between the proportional hazards model and the temporal survival regression model. We develop a smoothing-free iterative estimation procedure and a nonparametric resampling-based inference procedure in Section 4.2. In Section 4.3, finite-sample behaviors of the proposed estimator are investigated via Monte Carlo simulations. Also in Section 4.4, when the original Cox model holds, efficiency-loss of the estimators from the methods of Peng and Huang (2007) and

ours is investigated through simulation studies. In Section 4.5, an analysis of the well-known Veterans' Administration lung cancer data is presented for illustration.

4.1 Model

Split $\boldsymbol{W}$ into two covariate vectors, say, p- and q-dimensional $\boldsymbol{Z}$ and $\boldsymbol{V}$, according to whether their effects are time-varying or constant, and we propose a semiparametric survival regression model

$$S(t|\boldsymbol{Z},\boldsymbol{V}) = \exp\big[-\exp\{\boldsymbol{\beta}_0(t)^\top \widetilde{\boldsymbol{Z}} + \boldsymbol{\gamma}_0^\top \boldsymbol{V}\}\big], \qquad \textit{for all } t \geq 0, \tag{4.2}$$

where $\boldsymbol{\beta}_0(t) = [\log \Lambda_0(t),\ \boldsymbol{b}_0(t)^\top]^\top$ and $\widetilde{\boldsymbol{Z}} = [1,\ \boldsymbol{Z}^\top]^\top$. Here, $\boldsymbol{b}_0(t)$ is a p-dimensional vector of time-varying regression coefficients, and $\boldsymbol{\gamma}_0$ is a q-dimensional constant vector. In model (4.2), new terms $\boldsymbol{\beta}_0(\cdot)$ and $\widetilde{\boldsymbol{Z}}$ are introduced for notational convenience as both the unspecified baseline cumulative hazard function $\Lambda_0(\cdot)$ and the regression coefficients $\boldsymbol{b}_0(\cdot)$ are time-varying. With both types of coefficients $\boldsymbol{b}_0(\cdot)$ and $\boldsymbol{\gamma}_0$, model (4.2) postulates time-varying and time-constant multiplicative effects of covariates on the baseline cumulative hazard $\Lambda_0(\cdot)$, and thereby on the baseline survival function. In this way, the proposed model accommodates evolving effects of some covariates on survival times while formulating constant effects of the rest of the covariates. Note model (4.2) can be viewed as a sub-model of the temporal survival regression of Peng and Huang (2007) when some of the covariates are considered to have constant effects. Further, the proposed model reduces to the proportional hazards model when all covariates have constant effects. We consider the survival time is subject to the right censoring, say, at C; and adopt the conditional independent assumption on the censoring mechanism, namely

$$T \perp C \mid (\boldsymbol{Z},\boldsymbol{V}).$$

Adopting a similar strategy to address the both types of covariate effects, several models have been proposed for other quantities. Martinussen et al. (2002) proposed a semiparametric mixture effect model for the hazard function of the survival time. This model is a sub-model of the varying-coefficient Cox model. On the other hand, McKeague and Sasieni (1994) and Qian and Peng (2010) proposed semiparametric mixture effect models in additive hazards regression and censored quantile regression, respectively.

4.2 Estimation and Inference

Write the observed survival time as $X \equiv T \wedge C$ and define $\Delta \equiv I(T \leq C)$, where $\wedge$ is the minimization operator and $I(\cdot)$ is the indicator function. Then, the censored survival data consist of $\{X_i; \Delta_i; \boldsymbol{Z}_i; \boldsymbol{V}_i\}_{i=1}^n$, which are n independent replicates of $\{X; \Delta; \boldsymbol{Z}; \boldsymbol{V}\}$. Alternatively, the data can be represented using the counting process notation. Denote the counting process for an observed event process by $N(t) = I(X \leq t, \Delta = 1)$, and the *at-risk* process by $Y(t) = I(X \geq t)$. Then the observed data set consists of $\{N_i(\cdot); Y_i(\cdot); \boldsymbol{Z}_i; \boldsymbol{V}_i\}_{i=1}^n$, which are n independent replicates of $\{N(\cdot); Y(\cdot); \boldsymbol{Z}; \boldsymbol{V}\}$.

Under model (4.2) and the conditional independence assumption on the censoring mechanism, it follows $E\{dN(t)|\boldsymbol{Z}, \boldsymbol{V}\} = Y(t)\, d\Lambda(t|\boldsymbol{Z}, \boldsymbol{V}) = Y(t)\, d\exp\{\boldsymbol{\beta}_0(t)^\top \widetilde{\boldsymbol{Z}} + \boldsymbol{\gamma}_0^\top \boldsymbol{V}\}$. This equation motivates our proposed estimating integral equations, namely, for all $t \geq 0$,

$$\boldsymbol{S}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, t) \equiv n^{-1} \sum_{i=1}^n \widetilde{\boldsymbol{Z}}_i \big[N_i(t) - \int_0^t Y_i(s)\, d\exp\{\boldsymbol{\beta}(s)^\top \widetilde{\boldsymbol{Z}}_i + \boldsymbol{\gamma}^\top \boldsymbol{V}_i\} \big] = \boldsymbol{0} \qquad (4.3)$$

and

$$\boldsymbol{U}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) \equiv n^{-1} \sum_{i=1}^n \boldsymbol{V}_i \big[N_i(\tau) - \int_0^\tau Y_i(s)\, d\exp\{\boldsymbol{\beta}(s)^\top \widetilde{\boldsymbol{Z}}_i + \boldsymbol{\gamma}^\top \boldsymbol{V}_i\} \big] = \boldsymbol{0}, \qquad (4.4)$$

where τ is a fixed time point satisfying $Pr(C > \tau) > 0$. We propose an iterative recursive estimation procedure in the following Subsection 4.2.1.

When no covariates are considered for constant effects, estimating integral equation (4.3) reduces to the estimating equation of Peng and Huang (2007). This leads our varying-effect estimator to be identical with the estimator of Peng and Huang (2007). On the other hand, when no covariates are considered to have time-varying effect, our constant-effect estimator for $\boldsymbol{\gamma}_0$ and estimator for $\log \Lambda_0(\cdot)$ reduce the maximum partial-likelihood estimator of Cox (1975) and the logarithm of the Breslow estimator for $\Lambda_0(\cdot)$, respectively.

4.2.1 Estimation Procedure

Write $0 < x_1 < x_2 < \cdots < x_M$ as the observed event times in the sample, and denote the associated covariate vectors by $(\boldsymbol{Z}_{(1)}^\top, \boldsymbol{V}_{(1)}^\top)^\top, (\boldsymbol{Z}_{(2)}^\top, \boldsymbol{V}_{(2)}^\top)^\top, \ldots, (\boldsymbol{Z}_{(M)}^\top, \boldsymbol{V}_{(M)}^\top)^\top$. We develop a smoothing-free estimation procedure for $\boldsymbol{\beta}_0(\cdot)$ and $\boldsymbol{\gamma}_0$; that is,

Step 1. Set $k = 0$ and choose initial value $\widehat{\boldsymbol{\gamma}}^{(k)}$.

Step 2. Obtain estimates $\widehat{\boldsymbol{\beta}}^{(k)}(x_j, \widehat{\boldsymbol{\gamma}}^{(k)})$ at $x_j, j = 1, \ldots, M$ as the solution to equation

$$n^{-1} \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \big[\, dN_i(x_j) - Y_i(x_j) \, d \exp\{\boldsymbol{\beta}(x_j)^\top \widetilde{\boldsymbol{Z}}_i + \widehat{\boldsymbol{\gamma}}^{(k)\top} \boldsymbol{V}_i\} \big] = \mathbf{0}. \tag{4.5}$$

Step 3. Obtain estimate $\widehat{\boldsymbol{\gamma}}^{(k+1)}$ by solving the following equation

$$n^{-1} \sum_{i=1}^{n} \boldsymbol{V}_i \big[N_i(\tau) - \exp(\boldsymbol{\gamma}^\top \boldsymbol{V}_i) \int_0^\tau Y_i(s) \, d \exp\{\widehat{\boldsymbol{\beta}}^{(k)}(s, \widehat{\boldsymbol{\gamma}}^{(k)})^\top \widetilde{\boldsymbol{Z}}_i\} \big] = \mathbf{0}. \tag{4.6}$$

Step 4. Update k to $k + 1$

Step 5. Repeat Steps 2 and 4 until a predetermined convergence criterion.

Estimating equation (4.5) is reduced from the estimating integral equation (4.3) at the observed event times in the sample. Moreover, at time x_1, estimating equation (4.5) becomes

$$\widetilde{\boldsymbol{Z}}_{(1)} - \sum_{i=1}^{n} Y_i(x_1)\widetilde{\boldsymbol{Z}}_i \exp\{\widehat{\boldsymbol{\gamma}}^{(k)\top}\boldsymbol{V}_i\} \exp\{\boldsymbol{\beta}(x_1)^\top \widetilde{\boldsymbol{Z}}_i\} = \mathbf{0}. \tag{4.7}$$

Also, at event times $x_j, j = 2, \ldots, M$, the equation (4.5) sequentially becomes

$$\widetilde{\boldsymbol{Z}}_{(j)} - \sum_{i=1}^{n} Y_i(x_j)\widetilde{\boldsymbol{Z}}_i \exp\{\widehat{\boldsymbol{\gamma}}^{(k)\top}\boldsymbol{V}_i\} \big[\exp\{\boldsymbol{\beta}(x_j)^\top \widetilde{\boldsymbol{Z}}_i\} - \exp\{\widehat{\boldsymbol{\beta}}^{(k)}(x_{j-1}, \widehat{\boldsymbol{\gamma}}^{(k)})^\top \widetilde{\boldsymbol{Z}}_i\} \big] = \mathbf{0}. \tag{4.8}$$

Therefore $\widehat{\boldsymbol{\beta}}^{(k)}(x_j, \widehat{\boldsymbol{\gamma}}^{(k)})$ is obtained as the solution to the estimating equations (4.7) and (4.8) at x_j, $j = 1, \ldots, M$. Note in equation (4.8), the estimate $\widehat{\boldsymbol{\beta}}^{(k)}(x_{j-1}, \widehat{\boldsymbol{\gamma}}^{(k)})$ from the previous estimation is used in estimation of $\boldsymbol{\beta}^{(k)}(x_j, \widehat{\boldsymbol{\gamma}}^{(k)})$. On the other hand, given the estimates $\widehat{\boldsymbol{\beta}}^{(k)}(x_j, \widehat{\boldsymbol{\gamma}}^{(k)})$ at $x_j, j = 1, \ldots, M$, estimate $\widehat{\boldsymbol{\gamma}}^{(k+1)}$ is obtained as a root of the equation (4.6).

Once a predetermined convergence criteria is met, we estimate $\boldsymbol{\gamma}_0$ and $\boldsymbol{\beta}_0(\cdot)$ by our final estimates $\widehat{\boldsymbol{\gamma}}^{(k+1)}$ and $\widehat{\boldsymbol{\beta}}^{(k+1)}(\cdot, \widehat{\boldsymbol{\gamma}}^{(k+1)})$, respectively; and they are denoted by $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\beta}}(\cdot)$. The proposed estimator $\widehat{\boldsymbol{\beta}}(\cdot)$ will be a càdlàg function that may jump only at the observed event times. Initial value $\widehat{\boldsymbol{\beta}}(0)$ for time interval $[0, x_1)$ is set to satisfy $\exp\{\widehat{\boldsymbol{\beta}}(0-)^\top \widetilde{\boldsymbol{Z}}_i + \widehat{\boldsymbol{\gamma}}^\top \widetilde{\boldsymbol{V}}_i\} = 0$ for all i, i.e., $\widehat{S}(0 - |\boldsymbol{Z}, \boldsymbol{V}) = 1$ to meet the condition $S(0 - |\boldsymbol{Z}, \boldsymbol{V}) = 0$. In addition, both estimating equations (4.5) and (4.6) have good computational properties since their left-hand sides are monotone functions of $\boldsymbol{\beta}(\cdot)$ and $\boldsymbol{\gamma}$.

4.2.2 Interval Estimation

We propose a practical nonparametric resampling-based inference procedure by adapting the Bayesian bootstrap of Rubin (1981); see more in Kosorok (2008) and Huang (2014). To be specific, we consider the following bootstrap estimating integral equa-

tions, for all $t \geq 0$,

$$\boldsymbol{S}_n^*(\boldsymbol{\beta}, \boldsymbol{\gamma}, t) \equiv n^{-1} \sum_{i=1}^{n} \xi_i \widetilde{\boldsymbol{Z}}_i \big[N_i(t) - \int_0^t Y_i(s)\, d \exp\{\boldsymbol{\beta}(s)^\top \widetilde{\boldsymbol{Z}}_i + \boldsymbol{\gamma}^\top \boldsymbol{V}_i\} \big] = \mathbf{0} \qquad (4.9)$$

and

$$\boldsymbol{U}_n^*(\boldsymbol{\beta}, \boldsymbol{\gamma}) \equiv n^{-1} \sum_{i=1}^{n} \xi_i \boldsymbol{V}_i \big[N_i(\tau) - \int_0^\tau Y_i(s)\, d \exp\{\boldsymbol{\beta}(s)^\top \widetilde{\boldsymbol{Z}}_i + \boldsymbol{\gamma}^\top \boldsymbol{V}_i\} \big] = \mathbf{0}. \qquad (4.10)$$

where $(\xi_i)_{i=1}^n$ are a non-negative random sample from a distribution with unit-mean and unit-variance, e.g., the standard exponential distribution. Denote by $\boldsymbol{\beta}^*(\cdot)$ and $\boldsymbol{\gamma}^*$, convergent solutions to equations (4.9) and (4.10) from an iterative root finding algorithm. Based on, say, B bootstrap resamples, a set of bootstrap solutions $[\{\boldsymbol{\beta}_b^*(\cdot)^\top, \boldsymbol{\gamma}_b^{*\top}\}^\top]_{b=1}^B$ is obtained. The variance of estimator $\{\widehat{\boldsymbol{\beta}}(t)^\top, \widehat{\boldsymbol{\gamma}}^\top\}^\top$ for $t > 0$ can be estimated by the sample variance of $[\{\boldsymbol{\beta}_b^*(t)^\top, \boldsymbol{\gamma}_b^{*\top}\}^\top]_{b=1}^B$. Provided a variance estimate, a Wald-type $100(1-\alpha)\%$ point-wise confidence interval for $\{\boldsymbol{\beta}_0(t)^\top, \boldsymbol{\gamma}_0^\top\}^\top$ can be constructed based on the normal approximation centered at $\{\widehat{\boldsymbol{\beta}}(t)^\top, \widehat{\boldsymbol{\gamma}}^\top\}^\top$. Alternatively, a confidence interval can be simply constructed with the corresponding $(\alpha/2)$th and $(1-\alpha/2)$th quantiles of the empirical distribution of $\{\boldsymbol{\beta}^*(t)^\top, \boldsymbol{\gamma}^{*\top}\}^\top$. Moreover, a $100(1-\alpha)\%$ confidence band for $\big[\boldsymbol{c}^\top\{\boldsymbol{\beta}_0(t)^\top, \boldsymbol{\gamma}_0^\top\}^\top : \ t \in (l, u]\big]$ can be constructed with $\big[\boldsymbol{c}^\top\{\widehat{\boldsymbol{\beta}}(t)^\top, \widehat{\boldsymbol{\gamma}}^\top\}^\top \pm \eta_\alpha : \ t \in (l, u]\big]$, where $\boldsymbol{c}$ is a $(p+q+1)$-dimensional known constant vector. Here η_α is the $(1-\alpha/2)$th quantile of the empirical distribution of J^*, where $J_b^* = \sup_{t \in [l,u]} |\boldsymbol{c}^\top[\{\boldsymbol{\beta}_b^*(t)^\top, \boldsymbol{\gamma}_b^{*\top}\}^\top - \{\widehat{\boldsymbol{\beta}}(t)^\top, \widehat{\boldsymbol{\gamma}}^\top\}^\top]|$ for $b = 1, \ldots, B$.

4.3 Monte Carlo Simulations under the Mixture Effect Model

We conducted Monte Carlo simulations to investigate finite-sample behaviors of the proposed estimator. Two covariates with time-varying and constant effects were con-

sidered. Specifically, event times were generated from the following survival function in the form of model (4.2)

$$S(t|Z,V) = \exp\big[-t\exp\{(1+2t)^{-1}Z - V\}\big],$$

where covariate Z had time-varying coefficient $b_0(t) = (1+2t)^{-1}$; and covariate V had constant coefficient $\gamma_0 = -1$. Covariates Z and V were independently generated from the uniform[-1, 1] distribution. The underlying baseline cumulative hazard function was the identity function of time t, i.e., the baseline hazard $\lambda_0(\cdot) = 1$. Random follow-up time C was from the uniform $[c_0, 3]$ distribution, where $c_0 \in (0, 1)$ for two censoring scenarios. Considered sample sizes were n =50, 100, 200, 400, and 800, and bootstrap size B was 100. For each combination of sample sizes and censoring scenarios, 1,000 Monte Carlo samples were generated. For estimation performance comparison with our method proposal, we also considered the temporal survival regression method of Peng and Huang (2007).

Table 4.1 reports summary results of the Monte Carlo estimates at the predetermined time points, which are summarized in terms of empirical bias, empirical standard deviation, average standard error estimate, and empirical coverage probability of Wald-type 95% point-wise confidence intervals. On the whole, empirical biases were close to zero over time. Moreover, the averages of the estimated standard errors from our method agreed well with the corresponding empirical standard deviations. This indicates that our proposed bootstrap method approximated the sampling distribution of the proposed estimator well in our simulations. When sample size was relatively large, e.g., $n \geq 200$, the coverage probabilities were close to the nominal level of 95% though in small sample, e.g., $n \leq 100$, the coverage probabilities tended to be smaller than 95%, regardless of method and type of covariate effect. The empirical standard deviations for our proposed estimator, especially in estimation of

constant effect γ_0, appeared to be smaller than the ones for the temporal survival regression estimator. This indicates that our proposed estimation procedure attained more efficient estimation, especially for constant-effect estimation in our simulations.

4.4 Efficiency-Loss Study when the Cox Model Holds

Loss in estimation efficiency is expected with the increasing model generality. Therefore, we conducted simulation studies to assess the extent of the efficiency loss of the proposed estimator when the Cox model holds. Three baseline hazard functions, which are constant, increasing, and decreasing, were considered to address diverse scenarios in the Cox model. Table 4.2 and Figure 4.1 display the considered hazard functions. For each baseline hazards scenario, Monte Carlo samples were generated from the Cox model in the form of (4.2); that is,

$$S(t|V^{(1)}, V^{(2)}) = \exp\big[-\Lambda_0(t)\exp\{\gamma_0^{(1)}V^{(1)} + \gamma_0^{(2)}V^{(2)}\}\big],$$

where covariates $V^{(1)}$ and $V^{(2)}$ were independently generated from the uniform [-1, 1] distribution; and true effects $\gamma_0^{(1)}$ and $\gamma_0^{(2)}$ were 1 and -1, respectively. Random censoring time C followed the uniform $[c_0, 3]$ distribution for $c_0 \in (0, 1, 2)$, and sample sizes were n = 50, 100, 200, 400, and 800. For each combination of the simulation set-ups, 1,000 Monte Carlo samples were generated.

To compare efficiency loss of the proposed estimator, the estimator of Peng and Huang (2007) was also considered. For the functional estimators of Peng and Huang (2007) and ours, an averaged effect estimator was adapted in constant effect estimation, e.g., $\widehat{\boldsymbol{\gamma}}(l,u) \equiv (u-l)^{-1}\int_l^u \widehat{\boldsymbol{\beta}}(t)\,dt$ for $0 < l < u < \infty$. For the constant effect estimators of Cox (1972) and ours, on the other hand, the estimates from the original

Table 4.1: Summary Results for Monte Carlo Simulation under the Mixture Effect Model

		$b_0(t) = (1+2t)^{-1}$								$\gamma_0 = -1$							
		Peng and Huang (2007)				Proposed Method				Peng and Huang (2007)				Proposed Method			
n	t	B	SD	SEE	CP95	B	SD	SEE	CP95	B	SD	SEE	CP95	B	SD	SEE	CP95
50	0.25	2	594	512	92.2	20	598	518	91.7	16	616	533	93.2				
50	0.50	-28	466	414	92.8	-12	456	408	93.0	6	482	435	91.4				
50	0.75	-12	456	392	91.3	7	440	380	91.5	5	471	416	92.2				
50	1.00	2	477	397	91.2	15	443	378	91.5	20	509	426	89.8	1	366	333	92.7
50	1.25	9	527	419	90.4	25	469	390	90.7	9	605	467	89.3				
50	1.50	-39	737	463	88.6	-9	616	420	89.7	3	799	539	88.0				
50	1.75	-22	950	505	84.9	-2	775	466	87.5	29	1058	609	84.4				
100	0.25	-20	378	354	92.9	-10	377	358	93.5	4	400	369	93.0				
100	0.50	-4	305	287	93.5	8	305	287	93.7	14	316	299	94.1				
100	0.75	-10	284	267	93.0	3	280	265	93.4	3	305	281	92.5				
100	1.00	-4	275	265	94.7	6	269	260	95.0	-1	317	283	91.9	-13	239	229	93.7
100	1.25	0	287	272	93.2	5	275	265	93.1	-1	332	296	91.9				
100	1.50	5	334	288	91.3	8	311	275	91.6	-7	390	327	91.5				
100	1.75	1	393	315	90.5	4	348	296	91.4	-4	508	383	90.8				
200	0.25	-5	261	250	94.8	2	262	252	94.5	13	265	261	94.4				
200	0.50	9	211	202	94.0	12	210	202	94.0	-7	223	213	94.2				
200	0.75	6	196	189	93.3	9	196	188	93.0	-8	204	199	94.9				
200	1.00	5	190	186	94.4	9	190	184	94.4	-10	215	198	93.1	-11	169	161	94.7
200	1.25	3	192	189	94.4	6	193	187	93.8	-5	214	206	93.8				
200	1.50	-1	209	197	93.8	2	206	193	93.5	-2	231	221	93.2				
200	1.75	5	234	211	93.2	8	227	204	92.9	-4	267	245	93.1				
400	0.25	1	196	180	92.9	4	197	181	92.8	-6	189	188	95.0				
400	0.50	-4	156	144	92.8	-1	156	144	92.8	0	153	151	94.4				
400	0.75	-5	140	134	93.6	-4	140	134	93.4	-2	144	141	94.1				
400	1.00	1	138	132	93.8	3	136	131	93.6	-1	144	140	93.3	-6	116	114	94.9
400	1.25	3	142	134	94.1	5	141	133	94.1	0	151	144	93.9				
400	1.50	5	148	140	93.9	7	145	139	94.2	0	164	153	94.0				
400	1.75	5	153	149	94.8	7	151	147	94.7	-3	183	168	93.2				
800	0.25	-2	133	127	93.5	-1	133	127	93.3	11	134	132	94.4				
800	0.50	-2	104	102	94.4	-2	104	102	94.2	5	108	107	94.4				
800	0.75	-2	97	95	93.7	-1	97	94	93.8	3	102	100	93.9				
800	1.00	-2	95	93	94.6	-2	95	93	94.5	5	98	99	95.1	4	80	80	94.4
800	1.25	-1	100	95	93.6	0	99	95	93.3	5	105	102	94.0				
800	1.50	0	101	100	94.7	1	100	99	95.0	5	112	108	93.7				
800	1.75	3	113	106	93.5	3	111	105	93.2	6	123	118	93.7				

NOTE: B: empirical bias (×1000); SD: empirical standard deviation of estimates (×1000); SEE: average of standard error estimates (×1000); CP95: empirical coverage probability of the Wald-type 95% point-wise confidence interval (×100). Based on 1,000 Monte Carlo replications.

Table 4.2: The Three Risk Scenarios under the Cox Model in the Form of Model (4.2)

Risk Scenario	$\lambda_0(t)$	$S(t\|\boldsymbol{V})$
1	1	$\exp\left[-t\exp\{V^{(1)}-V^{(2)}\}\right]$
2	$\frac{2}{3}t$	$\exp\left[-\frac{1}{3}t^2\exp\{V^{(1)}-V^{(2)}\}\right]$
3	$-\frac{2}{3}t+2$	$\exp\left[-\left(-\frac{1}{3}t^2+2t\right)\exp\{V^{(1)}-V^{(2)}\}\right]$

See Figure 4.1 for graphical display of the baseline hazards

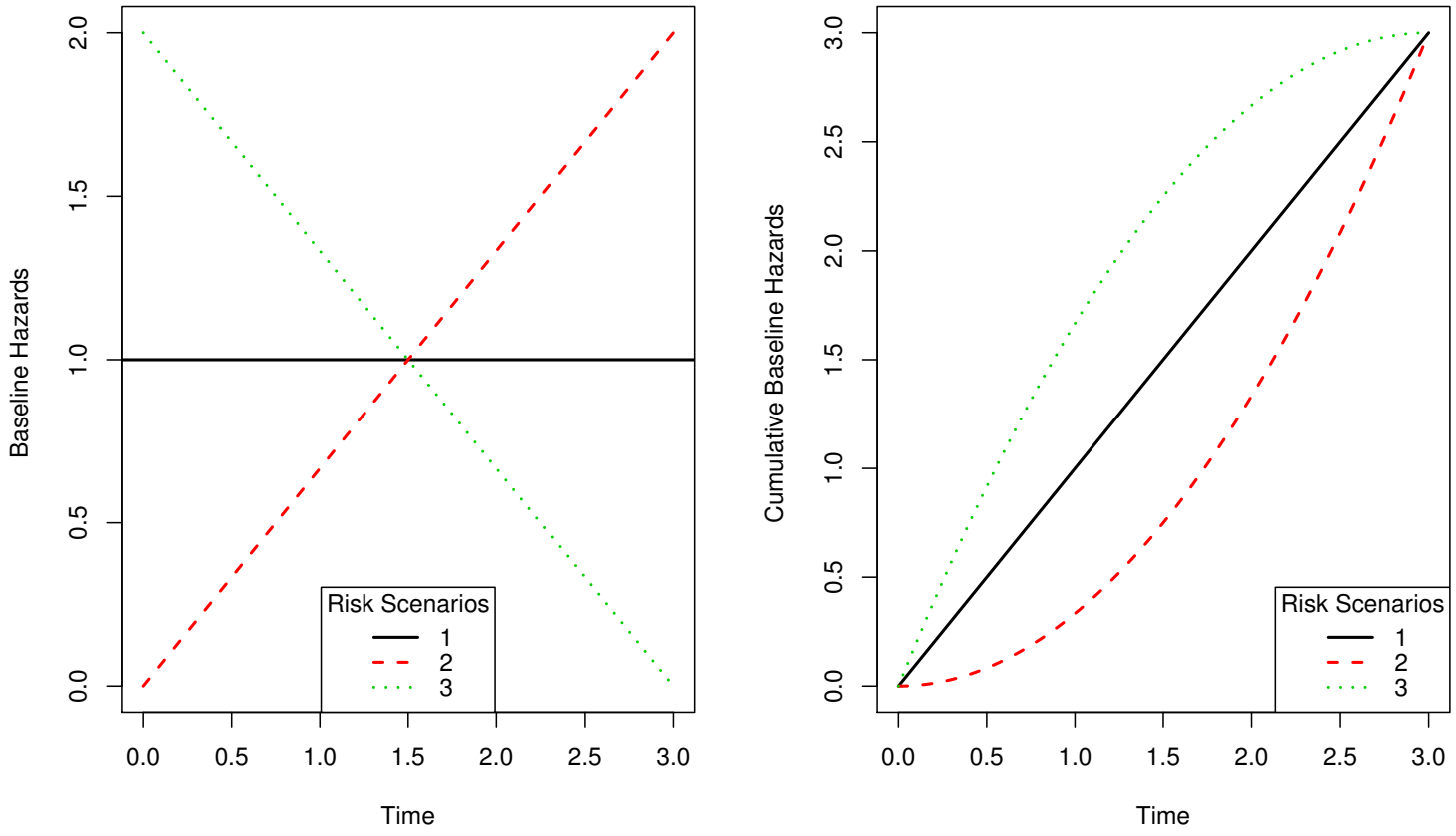

Figure 4.1: Monte Carlo baseline risk scenarios 1, 2, and 3. (a) three baseline hazard functions; and (b) the corresponding baseline cumulative hazard functions

estimation were used based on the whole observed data regardless of the averaging time interval under consideration.

Table 4.3 shows summary results of the Monte Carlo estimates from the risk scenario 1, i.e., $\lambda_0(t) = 1$. The results are reported in terms of empirical bias, empirical standard deviation, and empirical relative efficiency of the Cox estimator relative to the estimator under consideration, e.g., the empirical variance of our averaged effect estimates divided by the empirical variance of the Cox estimates. On the whole, empirical biases for all estimators were close to zero. Specifically, the Monte Carlo biases were at most 5.5% for sample size $n = 50$ regardless of method, and were under 1% for $n = 800$.

In estimation of $\gamma_0^{(2)}$, on the other hand, the empirical relative efficiency of the Cox estimator to our constant effect estimator was close to 1. This indicates that our proposed constant effect estimator had competitive efficiency to the Cox estimator in our simulations. In contrast, our averaged effect estimator for $\gamma_0^{(1)}$ appeared to be not as efficient as the Cox estimator. To be specific, the relative efficiency of the Cox estimator to our proposed averaged effect estimator ranged from 1.02 to 1.75. However, our averaged effect estimator for $\gamma_0^{(1)}$ had better empirical efficiency, compared to the averaged effect estimator of Peng and Huang (2007) to which the relative efficiency of the Cox estimator ranged from 1.06 to 2.12.

The summary results in Table 4.3 were generally similar with the results from the other simulations with different baseline hazard scenarios. For example, Tables 4.4 and 4.5 show summary results from the risk scenarios 2 and 3, with the increasing and decreasing baseline hazards. The empirical biases were close to zero, and the empirical standard deviations of our constant effect estimator for $\gamma^{(2)}$ were comparable to the ones of the Cox estimator. In estimation of $\gamma^{(1)}$, our averaged effect estimator had consistently better efficiency over the averaged effect estimator of Peng and Huang (2007) even though ours was not as much efficient as the Cox estimator.

Table 4.3: Summary Results for Risk Scenario 1 under the Cox Model

	$\gamma_0^{(1)} = 1$								$\gamma_0^{(2)} = -1$							
	Cox (1972)		Peng and Huang*			Proposed Method*			Cox (1972)		Peng and Huang*			Proposed Method		
Sample Size	B	SD	B	SD	relEff	B	SD	relEff	B	SD	B	SD	relEff	B	SD	relEff
Censoring (2, 3]																
50	48	343	2	358	1.09	5	361	1.11	-36	325	7	371	1.30	-16	322	0.98
100	31	223	12	232	1.08	16	229	1.05	-26	220	-4	236	1.15	-15	221	1.01
200	6	151	-8	156	1.07	-5	153	1.03	-9	146	3	151	1.07	-4	146	1.00
400	8	106	0	110	1.08	2	107	1.02	0	102	9	107	1.10	5	103	1.02
800	3	72	-1	74	1.06	-1	73	1.03	-2	73	2	74	1.03	1	73	1.00
Censoring (1, 3]																
50	54	330	3	402	1.48	18	381	1.33	-43	334	-3	424	1.61	-23	335	1.01
100	17	222	-13	254	1.31	-5	249	1.26	-12	222	20	246	1.23	1	221	0.99
200	13	152	-1	165	1.18	1	163	1.15	-10	155	4	172	1.23	-3	156	1.01
400	4	108	-4	113	1.09	-4	111	1.06	-2	106	5	112	1.12	2	107	1.02
800	2	75	-2	81	1.17	-2	79	1.11	-3	76	2	80	1.11	1	76	1.00
Censoring (0, 3]																
50	45	370	8	539	2.12	16	489	1.75	-55	383	-22	564	2.17	-32	384	1.01
100	16	246	6	316	1.65	9	309	1.58	-38	243	-20	324	1.78	-27	242	0.99
200	16	169	0	200	1.40	6	187	1.22	-16	178	7	209	1.38	-7	177	0.99
400	13	113	4	135	1.43	7	127	1.26	-15	113	-5	130	1.32	-10	114	1.02
800	5	84	-1	93	1.23	0	91	1.17	-5	85	0	94	1.22	-1	85	1.00

NOTE: B: empirical bias (×1000); SD: empirical standard deviation of estimates (×1000); relEff: estimated relative efficiency of Cox estimator to the estimator under consideration. Based on 1,000 Monte Carlo replications.

* Based on the averaged estimates over $(0.2, 2.8]$.

Table 4.4: Summary Results for Risk Scenario 2 under the Cox Model

	$\gamma_0^{(1)} = 1$								$\gamma_0^{(2)} = -1$							
	Cox (1972)		Peng and Huang*			Proposed Method*			Cox (1972)		Peng and Huang*			Proposed Method		
Sample Size	B	SD	B	SD	relEff	B	SD	relEff	B	SD	B	SD	relEff	B	SD	relEff
Censoring (2, 3]																
50	46	351	0	379	1.17	21	387	1.22	-37	333	9	370	1.23	-16	332	0.99
100	34	228	11	250	1.20	25	253	1.23	-27	222	6	246	1.23	-16	223	1.01
200	7	155	-15	168	1.17	-4	170	1.20	-11	149	8	161	1.17	-5	149	1.00
400	8	109	-7	123	1.27	-2	123	1.27	1	103	15	112	1.18	6	104	1.02
800	3	73	-10	80	1.20	-8	79	1.17	-2	74	9	82	1.23	2	73	0.97
Censoring (1, 3]																
50	59	360	34	444	1.52	51	423	1.38	-44	361	-15	431	1.43	-22	361	1.00
100	13	238	-10	283	1.41	6	277	1.35	-13	245	19	275	1.26	4	242	0.98
200	12	164	-1	183	1.25	7	181	1.22	-8	168	13	186	1.23	1	169	1.01
400	3	115	-13	123	1.14	-8	123	1.14	-2	115	13	127	1.22	3	115	1.00
800	2	81	-11	86	1.13	-10	86	1.13	-3	82	8	90	1.20	1	82	1.00
Censoring (0, 3]																
50	61	459	31	582	1.61	43	547	1.42	-79	471	-62	585	1.54	-52	469	0.99
100	30	290	3	344	1.41	10	333	1.32	-46	288	-12	345	1.44	-28	285	0.98
200	15	200	-16	214	1.14	-3	210	1.10	-23	202	0	219	1.18	-12	201	0.99
400	13	134	-6	144	1.15	1	142	1.12	-15	137	0	145	1.12	-8	138	1.01
800	7	96	-4	101	1.11	-3	98	1.04	-4	100	10	104	1.08	1	100	1.00

NOTE: B: empirical bias ($\times 1000$); SD: empirical standard deviation of estimates ($\times 1000$); relEff: estimated relative efficiency of Cox estimator to the estimator under consideration. Based on 1,000 Monte Carlo replications.

* Based on the averaged estimates over $(0.2, 2.8]$.

Table 4.5: Summary Results for Risk Scenario 3 under the Cox Model

	$\gamma_0^{(1)} = 1$								$\gamma_0^{(2)} = -1$							
	Cox (1972)		Peng and Huang*			Proposed Method*			Cox (1972)		Peng and Huang*			Proposed Method		
Sample Size	B	SD	B	SD	relEff	B	SD	relEff	B	SD	B	SD	relEff	B	SD	relEff
Censoring (2, 3]																
50	46	337	3	406	1.45	5	373	1.23	-33	319	10	436	1.87	-16	317	0.99
100	32	220	9	253	1.32	12	246	1.25	-28	217	-2	253	1.36	-17	218	1.01
200	7	149	-8	165	1.23	-6	161	1.17	-10	144	3	166	1.33	-6	144	1.00
400	9	104	2	118	1.29	3	113	1.18	1	100	9	117	1.37	4	101	1.02
800	3	71	1	81	1.30	1	78	1.21	-2	73	2	81	1.23	1	72	0.97
Censoring (1, 3]																
50	49	325	1	449	1.91	17	406	1.56	-43	320	-6	456	2.03	-23	325	1.03
100	16	214	-16	256	1.43	-10	249	1.35	-13	213	13	262	1.51	-1	213	1.00
200	12	146	-4	171	1.37	-1	167	1.31	-10	148	4	177	1.43	-3	150	1.03
400	4	103	-3	115	1.25	-2	112	1.18	-1	102	7	116	1.29	3	103	1.02
800	3	72	1	84	1.36	0	81	1.27	-3	74	-1	84	1.29	0	74	1.00
Censoring (0, 3]																
50	43	345	26	566	2.69	12	507	2.16	-50	355	-10	580	2.67	-29	354	0.99
100	13	230	2	326	2.01	0	302	1.72	-35	228	-24	329	2.08	-25	228	1.00
200	13	161	-11	206	1.64	-3	189	1.38	-13	166	15	206	1.54	-6	167	1.01
400	12	106	4	131	1.53	6	127	1.44	-14	106	-6	135	1.62	-9	107	1.02
800	4	79	-1	97	1.51	-1	94	1.42	-5	80	1	96	1.44	-1	80	1.00

NOTE: B: empirical bias (×1000); SD: empirical standard deviation of estimates (×1000); relEff: estimated relative efficiency of Cox estimator to the estimator under consideration. Based on 1,000 Monte Carlo replications.

* Based on the averaged estimates over $(0.2, 2.8]$.

Overall, all the empirical biases were small, and the efficiency trends were consistent over the simulation scenarios. In small sample, however, some variations were observed in summary results over the censoring distributions and baseline hazards. They indicate that survival times in small sample were sensitively affected by survival-time and censoring distributions.

4.5 Analysis of the VA Lung Cancer Trial Data

To illustrate the proposed method, we analyzed the Veterans' Administration lung cancer data, which were previously analyzed and discussed by some authors including Prentice (1973), Peng and Huang (2007), and Kalbfleisch and Prentice (2002). In the clinical trial, a total of 137 male patients with advanced inoperable lung cancer was randomly assigned to a standard or a test chemotherapy. Time to death from randomization was recorded as a primary endpoint for each patient, and there were 9 cases of incomplete follow-up.

Karnofsky performance score, which is a measure of performance status, and types of histological tumor (squamous, small, adenoma, or large cell) were covariates in our survival regression modeling. Specifically, Karnofsky performance score at enrollment was considered to have time-varying effect on survival times, but types of histological tumor were assumed to have constant effects. The covariate choice whether for constant or varying effect was based on the earlier analysis by Peng and Huang (2007), in which Karnofsky score appeared to be the only covariate to have varying effect. To be specific, we consider a mixture effect model in the form of model (4.2); that is, for all $t \geq 0$,

$$S(t|Z, \boldsymbol{V}) = \exp\big[- \Lambda_0(t) \exp\{b_0(t)\text{K.S.} + \gamma_0^{(1)} I(\text{squam.}) + \gamma_0^{(2)} I(\text{small}) + \gamma_0^{(3)} I(\text{adeno.})\}\big].$$

Figure 4.2 shows the estimated effects of covariates from the methods of Cox

(1972), Peng and Huang (2007), and ours. The estimates from all three methods were generally in the same direction on average. In particular, Karnofsky performance score appeared to decrease the death rate on the whole, e.g., the negative value from the Cox method. Nevertheless, its effect may be truly changing over time as indicated by the estimates from the methods of Peng and Huang (2007) and ours. To be specific, the size of estimated effect was relatively large until 3-month of follow-up, but the size diminished afterwards. This implies that the proportionality assumption in the Cox model may be too simple to adequately describe changing effects of covariates. In contrast, the effects of types of histological tumor appeared to be relatively constant. On the whole, people with small cell or adenoma tumor history tended to experience the event of death frequently, compared to people with large or squamous tumor history.

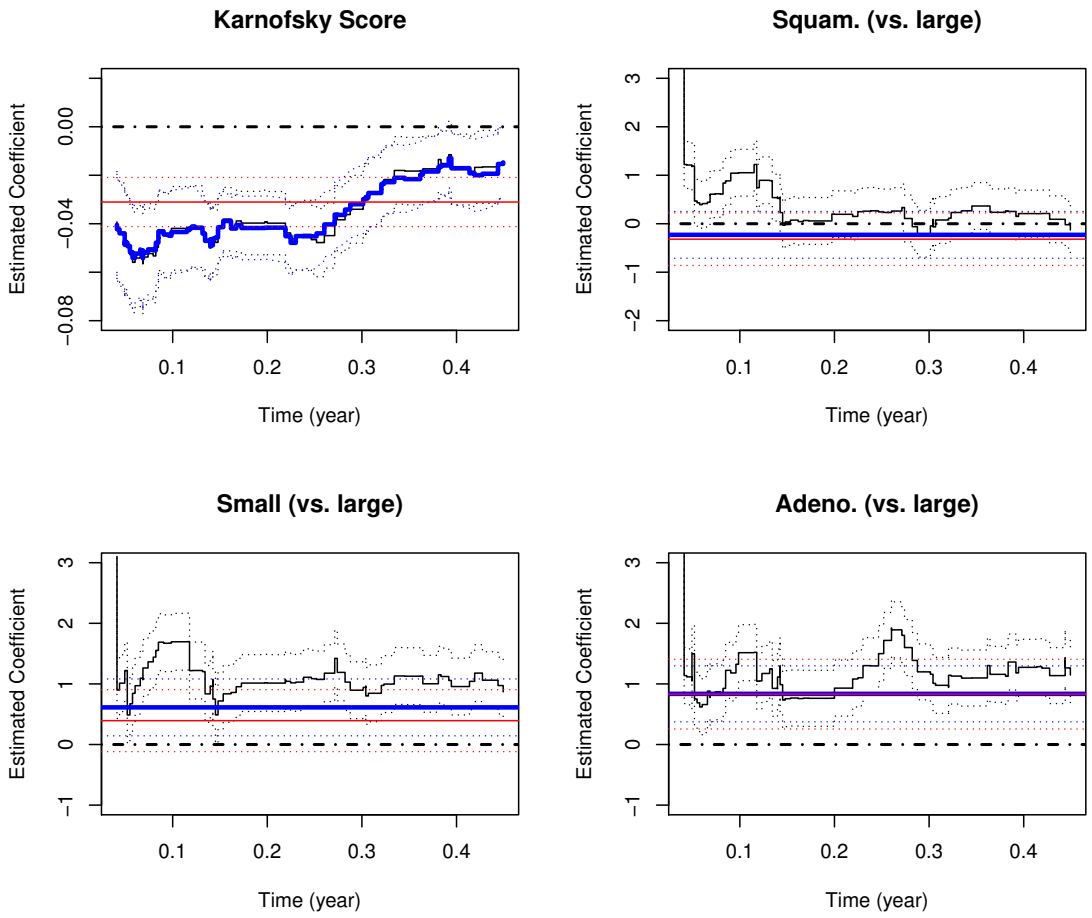

Figure 4.2: The Veterans' Administration lung cancer trial data. Solid lines are estimated effects of covariates from the three methods (Red: the Cox estimate; Black: Peng-Huang estimate; Thick Blue: proposed mixture effect estimate). The corresponding dotted lines denote the Wald-type 95% point-wise confidence intervals. Dash-Dotted lines are the reference line at 0, indicating no proportional effect.

Chapter 5

Summary and Future Work

5.1 Summary

We have carried out three projects to develop regression models that accommodate time-varying effects of covariates. The proposed models offer flexible modeling options in recurrent event and survival analysis.

We first consider a dynamic regression model for the mean frequency of recurrent events. With the time-varying regression coefficients, the model generalizes the proportional means model of Lin et al. (2000). Smoothing-free estimation and nonparametric inference procedures are developed. Large-sample properties of the proposed estimator are established based on the empirical process theory. Simulation studies demonstrate utility of the proposed method with practical sample sizes. Our analyses of the bladder tumor trial data and the DISC trial data illustrate the proposed method well.

Next, we develop a varying-coefficient model for the cumulative gap-time hazard function on the gap-time scale of recurrent events. The model generalizes the marginal proportional gap-time hazards model of Huang and Chen (2003). Large-sample and finite-sample behaviors of the proposed estimator are investigated through asymptotic

studies and Monte Carlo simulations. An analysis of the bladder tumor trial data is presented to illustrate the proposed method.

For the analysis of univariate survival data, we study a semiparametric survival regression model with a mixture of time-varying and constant effects of covariates. The model is a sub-model of the temporal process regression of Peng and Huang (2007), generalizing the proportional hazards model of Cox (1972). An iterative estimation algorithm and a resampling-based inference procedure are developed for time-varying and time-constant coefficients. Extensive simulations are carried out to study the efficiency loss of the proposed estimator. Our analysis of the VA lung cancer trial data is presented for illustration.

5.2 Future Work

Each of the proposed models in this dissertation may be generalized to a class of dynamic transformation models. Generalizing the model (2.1), specifically, a class of transformation models for the mean frequency of recurrent events can take a form

$$E\{N^*(t)|\boldsymbol{Z}\} = g\{\boldsymbol{\beta}_0(t)^\top \widetilde{\boldsymbol{Z}}\},$$

where g is a known function. Both of the estimation procedure and the resampling approach in Chapter 2 can be easily adapted. This class includes the additive regression model of Aalen et al. (2004) for recurrent event data as a special case when g is the identity function, and their estimator coincides with ours. Adopting a similar strategy, the dynamic cumulative gap-time hazards model (3.1) and the mixture effect survival regression model (4.2) can be extended to $\Lambda(t|\boldsymbol{Z}) = g\{\boldsymbol{\beta}_0(t)^\top \widetilde{\boldsymbol{Z}}\}$ and $S(t|\boldsymbol{Z}, \boldsymbol{V}) = g\{\boldsymbol{\beta}_0(t)^\top \widetilde{\boldsymbol{Z}} + \boldsymbol{\gamma}_0^\top \boldsymbol{V}\}$, respectively. The original estimation and inference procedures in Chapters 3 and 4 would be easily adapted.

We have considered time-independent covariates for our model and method devel-

opments. The mean frequency regression model (2.1) and the mixture effect survival regression model (4.2) can be extended to accommodate 'external' time-dependent covariates; see Kalbfleisch and Prentice (2002, pg. 196) for discussion about external time-dependent covariates. Then technically, the original estimation and inference methods can be easily adapted to the extended models. However, the interpretation would be difficult for the extended models with time-dependent covariates as both the covariates and their effects are time-varying.

We speculate that our proposed estimators for time-varying and constant effects in Chapter 4 are consistent and asymptotically normal, which seems to be supported by the conducted simulation studies. Large-sample behaviors of our estimator are left for our future work.

A mean frequency function of recurrent events and a cumulative gap-time hazard function do not decrease over time. On the other hand, a survival function does not increase. One concern is that lack of monotonicity respecting may arise in the estimated conditional functions. Recently Huang (2017) proposed the adaptive interpolation method for restoration of monotonicity respecting. That method should be applicable to each of our estimators as well.

By introducing weight processes in each of the proposed estimating integral equations (2.2), (3.3), (4.3), and (4.4), a class of estimating integral equations can be developed. The weight processes may depend on covariates as well as $\boldsymbol{\beta}(\cdot)$. To pursue more efficient estimation within this class is a future research topic.

As described in Sections 2.2, 3.2, and 4.2, our estimation algorithms solve the proposed estimating integral equations at each and every observed event time. A faster estimation may be achieved by solving estimating integral equations approximately over a grid of time points. Such algorithm can be a viable alternative especially when a data set has a large number of distinct event times.

Appendix A

Proofs of Consistency and Weak Convergence in Chapter 2

We first characterize the jump size of the proposed estimator $\widehat{\boldsymbol{\beta}}(\cdot)$ in Lemma 1, which paves the way for establishing the large sample properties of the estimator $\widehat{\boldsymbol{\beta}}(\cdot)$.

Lemma 1. Suppose $\widehat{\boldsymbol{\beta}}(t) \in \boldsymbol{B}$ for all $t \in [\kappa, \tau]$, where $\boldsymbol{B}$ is a compact parameter space in $\mathcal{R}^{p+1}$. Under conditions C1-C3, and C6, $\triangle\widehat{\boldsymbol{\beta}}(t) \equiv \widehat{\boldsymbol{\beta}}(t) - \widehat{\boldsymbol{\beta}}(t-)$ is $O(n^{-1})$, almost surely, uniformly over $t \in (\kappa, \tau]$

Proof. From the estimating integral equation (2.2), it follows that, for $t > 0$

$$0 = \frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \bigg(dN_i(t) - Y_i(t)\big[\exp\{\widehat{\boldsymbol{\beta}}(t)^\top \widetilde{\boldsymbol{Z}}_i\} - \exp\{\widehat{\boldsymbol{\beta}}(t-)^\top \widetilde{\boldsymbol{Z}}_i\}\big] \bigg).$$

There may be change in $\widehat{\boldsymbol{\beta}}(\cdot)$ at the observed event times $(x_j)_{j=1}^{M}$, whereas no jump occurs in $\widehat{\boldsymbol{\beta}}(\cdot)$ over time between any two adjacent event times. At the jth event

time, it follows that

$$\begin{aligned}\frac{1}{n}\widetilde{\boldsymbol{Z}}_{(j)} &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\widetilde{\boldsymbol{Z}}_i\big[\exp\{\widehat{\boldsymbol{\beta}}(t)^\top\widetilde{\boldsymbol{Z}}_i\} - \exp\{\widehat{\boldsymbol{\beta}}(t-)^\top\widetilde{\boldsymbol{Z}}_i\}\big] \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top\exp\{\boldsymbol{\beta}_i^{\circledast}(t)^\top\widetilde{\boldsymbol{Z}}_i\}\{\widehat{\boldsymbol{\beta}}(t)-\widehat{\boldsymbol{\beta}}(t-)\},\end{aligned} \tag{A.1}$$

where $\boldsymbol{\beta}_i^{\circledast}(t)^\top\widetilde{\boldsymbol{Z}}_i$ is on the line segment between $\widehat{\boldsymbol{\beta}}(t)^\top\widetilde{\boldsymbol{Z}}_i$ and $\widehat{\boldsymbol{\beta}}(t-)^\top\widetilde{\boldsymbol{Z}}_i$. Note $\boldsymbol{\beta}_i^{\circledast}(t)$ satisfies $\exp\{\boldsymbol{\beta}_i^{\circledast}(t)^\top\widetilde{\boldsymbol{Z}}_i\} = \displaystyle\int_0^1 \exp\left(\big[\widehat{\boldsymbol{\beta}}(t-)+r\{\widehat{\boldsymbol{\beta}}(t)-\widehat{\boldsymbol{\beta}}(t-)\}\big]^\top\widetilde{\boldsymbol{Z}}_i\right)dr$. Under C3, $\frac{1}{n}\widetilde{\boldsymbol{Z}}_{(j)}$ on the left-hand side of equation (A.1) is $O(n^{-1})$. The limit of the factor $\frac{1}{n}\sum_{i=1}^n Y_i(t)\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top\exp\{\boldsymbol{\beta}_i^{\circledast}(t)^\top\widetilde{\boldsymbol{Z}}_i\}$, on the right-hand side of (A.1) has a smallest eigenvalue bounded away from zero under C6 as every $\exp\{\boldsymbol{\beta}_i^{\circledast}(t)^\top\widetilde{\boldsymbol{Z}}_i\}$ is bounded away from zero. Consequently, at any $t \in [\kappa,\tau]$, the jump size $\triangle\widehat{\boldsymbol{\beta}}(t) = O(n^{-1})$ almost surely. □

A.1. Proof of Theorem 1: the uniform consistency of $\widehat{\boldsymbol{\beta}}(\cdot)$

Define classes $\boldsymbol{F}$ and $\boldsymbol{G}$ as $\boldsymbol{F} \equiv \{\widetilde{\boldsymbol{Z}}N(t) : t \in [0,\tau]\}$ and $\boldsymbol{G} \equiv \{Y(t)\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top\exp(\boldsymbol{\beta}^\top\widetilde{\boldsymbol{Z}}) : t \in [0,\tau], \boldsymbol{\beta} \in \boldsymbol{B}\}$, where $\mathbf{B}$ is a compact subset of $\mathcal{R}^{p+1}$. It is known that both $\{I(T \le t) : t \in \mathcal{R}\}$ and $\{I(C \ge t) : t \in \mathcal{R}\}$ are Donsker (Kosorok, 2008, Lemma 9.10). By permanence properties of the Donsker class, $\boldsymbol{F}$ and $\boldsymbol{G}$ are Donsker. Since Donsker implies Glivenko-Cantelli, it follows that

$$\sup_{t\in[0,\tau]}\|n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\big[N_i(t) - \int_0^t Y_i(s)\widetilde{\boldsymbol{Z}}_i^\top\exp\{\boldsymbol{\beta}_0(s)^\top\widetilde{\boldsymbol{Z}}_i\}\,d\boldsymbol{\beta}_0(s)\big]\| \longrightarrow 0, \text{ almost surely.} \tag{A.2}$$

Let $\boldsymbol{\psi}(t) \equiv n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\big[N_i(t) - \int_0^t Y_i(s)\,d\exp\{\boldsymbol{\beta}_0(s)^\top\widetilde{\boldsymbol{Z}}_i\}\big]$ for $t \in [0,\tau]$. With the solutions to the estimating equations (2.3) and (2.4), it follows that

$$\begin{aligned}
\boldsymbol{\psi}(t) - \boldsymbol{\psi}(\kappa) &= n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\int_\kappa^t Y_i(s)\,d\big[\exp\{\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\boldsymbol{Z}}_i\} - \exp\{\boldsymbol{\beta}_0(s)^\top\widetilde{\boldsymbol{Z}}_i\}\big] \\
&= n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top\int_\kappa^t Y_i(s)\,d\big[\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}\{\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s)\}\big] \\
&= \int_\kappa^t n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top Y_i(s)\{\widehat{\boldsymbol{\beta}}(s-) - \boldsymbol{\beta}_0(s-)\}\,d\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\} \\
&\quad + \int_\kappa^t n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top Y_i(s)\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}\,d\{\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s)\}, \qquad \text{(A.3)}
\end{aligned}$$

where $\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i$ is on the line segment between $\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\boldsymbol{Z}}_i$ and $\boldsymbol{\beta}_0(s)^\top\widetilde{\boldsymbol{Z}}_i$. Note that the subscript i in $\boldsymbol{\beta}_i^{\circledcirc}(s)$ indicates its dependency on $\widetilde{\boldsymbol{Z}}_i$.

Let $\widehat{\boldsymbol{A}}(t) \equiv \int_\kappa^t n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top Y_i(s)\,d\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}$ and $\widehat{\boldsymbol{B}}(s) \equiv n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top Y_i(s)\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}$. From equation (A.3), we have

$$d\boldsymbol{\psi}(t) = \widehat{\boldsymbol{A}}(dt)\{\widehat{\boldsymbol{\beta}}(t-) - \boldsymbol{\beta}_0(t-)\} + \widehat{\boldsymbol{B}}(t)d\{\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\}. \qquad \text{(A.4)}$$

By conditions C3, C4, and C6, $\widehat{\boldsymbol{B}}(t)^{-1}$ exists almost surely for sufficiently large n. By multiplying the inverse of $\widehat{\boldsymbol{B}}(t)$ on each side of the equation (A.4) followed by integration over $(\kappa, t]$, we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t) &= \{\widehat{\boldsymbol{\beta}}(\kappa) - \boldsymbol{\beta}_0(\kappa)\} \\
&\quad + \int_\kappa^t \widehat{\boldsymbol{B}}(s)^{-1}d\boldsymbol{\psi}(s) - \int_\kappa^t \widehat{\boldsymbol{B}}(s)^{-1}\widehat{\boldsymbol{A}}(ds)\{\widehat{\boldsymbol{\beta}}(s-) - \boldsymbol{\beta}_0(s-)\} \qquad \text{(A.5)}
\end{aligned}$$

Since this is a the Volterra integral equation (e.g., Andersen et al., 1993, Theorem

II.6.3), unique solution $\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)$ is obtained for $t \in [\kappa, \tau]$. That is,

$$\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t) = [\prod\nolimits_{[\kappa,t]} \{\boldsymbol{I} - \widehat{\boldsymbol{A}}(du)\widehat{\boldsymbol{B}}(u)^{-1}\}]^{\top} \{\widehat{\boldsymbol{\beta}}(\kappa) - \boldsymbol{\beta}_0(\kappa)\}$$

$$+ \int_{\kappa}^{t} [\prod\nolimits_{(s,t]} \{\boldsymbol{I} - \widehat{\boldsymbol{A}}(du)\widehat{\boldsymbol{B}}(u)^{-1}\}]^{\top} \widehat{\boldsymbol{B}}(s)^{-1} \boldsymbol{\psi}(ds), \qquad \text{(A.6)}$$

where $\prod$ denotes the product-integral.

Suppose that almost surely $\widehat{\boldsymbol{\beta}}(t)$ is in a compact subset of $\mathcal{R}^{p+1}$ for $t \in [\kappa, \tau]$ for sufficiently large n. By Lemma 1 $\triangle\widehat{\boldsymbol{\beta}}(t)$ and $\triangle\widehat{\boldsymbol{A}}(t)$ on $[\kappa, \tau]$ are $O(n^{-1})$, almost surely. As $\widehat{\boldsymbol{A}}(t)$ is differentiable on intervals between the adjacent event times, the derivative of $\widehat{\boldsymbol{A}}(t)$ is bounded. Therefore, $\widehat{\boldsymbol{A}}(\cdot)$ and $\int_{\kappa}^{t} \widehat{\boldsymbol{A}}(du)\widehat{\boldsymbol{B}}(u)^{-1}$ are of bounded variation on $[\kappa, \tau]$. Note the product integral $\prod\nolimits_{(s,t]} \{\boldsymbol{I} - \widehat{\boldsymbol{A}}(du)\widehat{\boldsymbol{B}}(u)^{-1}\}$ exists for $\kappa \leq s \leq t \leq \tau$ (Gill and Johansen, 1990, Theorem 1). Given the consistent estimator $\widehat{\boldsymbol{\beta}}(\kappa)$, the first term on the right-hand side of equation (A.6) is $o(1)$ almost surely; also, the second term on the right-hand side of equation (A.6) is $o(1)$ almost surely by the integration by parts. Thus, we have the result that $\sup_{t\in[\kappa,\tau]} \|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\| \longrightarrow 0$, almost surely.

Now it remains to show that almost surely $\sup_{t\in[\kappa,\tau]} \|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\| \longrightarrow 0$ if it is not true that $\widehat{\boldsymbol{\beta}}(t)$ is in a compact subset of $\mathcal{R}^{p+1}$ for $t \in [\kappa, \tau]$. There exists $t^* \in (\kappa, \tau]$ at which $\widehat{\boldsymbol{\beta}}(\cdot)$ has a jump and goes off the boundary of the compact set for the first time, since $\widehat{\boldsymbol{\beta}}(\kappa) - \boldsymbol{\beta}_0(\kappa) \to 0$ almost surely. As $\widehat{\boldsymbol{\beta}}(t)$ is in the compact set for all $t \in [\kappa, t^*]$, $\|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\|$ can be made arbitrarily small for $t \in [\kappa, t^*]$ with sufficiently large sample size. It implies that the jump size of $\widehat{\boldsymbol{\beta}}(\cdot)$ at t^* is bounded away from 0. This would result in a contradiction in light of the equation (A.1) in Lemma 1.

A.2. Proof of Theorem 2: the weak convergence of $\widehat{\boldsymbol{\beta}}(\cdot)$

Recall $\boldsymbol{\psi}(\cdot)$ defined in the proof of Theorem 1. For each $t \in (\kappa, \tau]$, almost surely,

$$\begin{aligned}
\boldsymbol{\psi}(t) - \boldsymbol{\psi}(\kappa) &= n^{-1} \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \int_{\kappa}^{t} Y_i(s)\, d\big[\exp\{\widehat{\boldsymbol{\beta}}(s)^\top \widetilde{\boldsymbol{Z}}_i\} - \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\}\big] \\
&= n^{-1} \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \int_{\kappa}^{t} Y_i(s)\, d\big[\exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\} \widetilde{\boldsymbol{Z}}_i^\top \{\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s)\}\big] + \widetilde{\boldsymbol{C}}(t) \\
&= \int_{\kappa}^{t} n^{-1} \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top Y_i(s) \{\widehat{\boldsymbol{\beta}}(s-) - \boldsymbol{\beta}_0(s-)\}\, d \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\} \\
&\quad + \int_{\kappa}^{t} n^{-1} \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top Y_i(s) \{\exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\}\, d\{\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s)\} + \widetilde{\boldsymbol{C}}(t),
\end{aligned}$$

for some $\widetilde{\boldsymbol{C}}(t) = o\{\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\}$. Define $\widetilde{\boldsymbol{A}}(t) \equiv \displaystyle\int_{\kappa}^{t} n^{-1} \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top Y_i(s)\, d \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\}$ and $\widetilde{\boldsymbol{B}}(s) \equiv n^{-1} \displaystyle\sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top Y_i(s) \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\}$. It follows that $d\boldsymbol{\psi}(t) = \widetilde{\boldsymbol{A}}(dt)\{\widehat{\boldsymbol{\beta}}(t-) - \boldsymbol{\beta}_0(t-)\} + \widetilde{\boldsymbol{B}}(t) d\{\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\} + d\widetilde{\boldsymbol{C}}(t)$. Under the regularity conditions, $\widetilde{\boldsymbol{B}}(t)^{-1}$ exists almost surely for sufficiently large n. By virtue of the Volterra integral equation,

$$
\begin{aligned}
n^{1/2}\{\widehat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t)\} &= n^{1/2}[\boldsymbol{\Pi}_{[\kappa,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^{\top}\{\widehat{\boldsymbol{\beta}}(\kappa)-\boldsymbol{\beta}_0(\kappa)\} \\
&\quad + n^{1/2}\int_{\kappa}^{t}[\boldsymbol{\Pi}_{(s,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^{\top}\widetilde{\boldsymbol{B}}(s)^{-1}d\{\psi(s)-\widetilde{\boldsymbol{C}}(s)\} \\
&= n^{1/2}[\boldsymbol{\Pi}_{[\kappa,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^{\top}[-\{\widetilde{\boldsymbol{B}}(\kappa)+o_p(1)\}^{-1}\boldsymbol{\psi}(\kappa)] \\
&\quad + n^{1/2}\int_{\kappa}^{t}[\boldsymbol{\Pi}_{(s,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^{\top}\widetilde{\boldsymbol{B}}(s)^{-1}d\{\boldsymbol{\psi}(s)-\widetilde{\boldsymbol{C}}(s)\} \\
&= -n^{1/2}[\boldsymbol{\Pi}_{[\kappa,t]}\{\boldsymbol{I}-\boldsymbol{A}_0(du)\boldsymbol{B}_0(u)^{-1}\}]^{\top}\boldsymbol{B}_0(\kappa)^{-1}\boldsymbol{\psi}(\kappa) \\
&\quad + n^{1/2}\int_{\kappa}^{t}[\boldsymbol{\Pi}_{(s,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^{\top}\boldsymbol{B}_0(s)^{-1}\boldsymbol{\psi}(ds) \\
&\quad + o_p[n^{1/2}\{\widehat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t)\}],
\end{aligned}
$$

where $\boldsymbol{A}_0(t) \equiv E\big[\int_{\kappa}^{t}\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^{\top}Y(s)\,d\exp\{\boldsymbol{\beta}_0(s)^{\top}\widetilde{\boldsymbol{Z}}\}\big]$ and $\boldsymbol{B}_0(s) \equiv E\big[\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^{\top}Y(s)\exp\{\boldsymbol{\beta}_0(s)^{\top}\widetilde{\boldsymbol{Z}}\}\big]$. Then, $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\cdot)-\boldsymbol{\beta}_0(\cdot)\}$ weakly converges to a Gaussian process as it is a linear mapping of $\boldsymbol{\psi}(\cdot)$, which is a Gaussian process itself.

Appendix B

Proofs of Consistency and Weak Convergence in Chapter 3

We first characterize the jump size of the proposed estimator in Lemma 2.

Lemma 2. Suppose $\widehat{\boldsymbol{\beta}}(t) \in \boldsymbol{B}$ for all $t \in [0, \tau]$, where $\boldsymbol{B}$ is a compact parameter space in $\mathcal{R}^{p+1}$. Under conditions C1-C2, and C5, $\triangle\widehat{\boldsymbol{\beta}}(t) \equiv \widehat{\boldsymbol{\beta}}(t) - \widehat{\boldsymbol{\beta}}(t-)$ is $O(n^{-1})$, almost surely, uniformly over $t \in (0, \tau]$

Proof of Lemma 2.

From the estimating integral equation (3.3), it follows that, for $t > 0$,

$$\frac{1}{n}\sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \Big[dN_{i(j)}(t) - Y_{i(j)}(t) \big[\exp\{\widehat{\boldsymbol{\beta}}(t)^\top \widetilde{\boldsymbol{Z}}_i\} - \exp\{\widehat{\boldsymbol{\beta}}(t-)^\top \widetilde{\boldsymbol{Z}}_i\} \big] \Big] = 0.$$

Step function $\widehat{\boldsymbol{\beta}}(\cdot)$ may jump only at the unique observed gap times $0 < x_1 < x_2 < \cdots < x_G$ from the all individuals in the sample. At the gth gap time, it follows that

$$\begin{aligned} \frac{1}{n} S_{(g)}^{-1} \widetilde{\boldsymbol{Z}}_{(g)} &= \frac{1}{n}\sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i Y_{i(j)}(t) \big[\exp\{\widehat{\boldsymbol{\beta}}(t)^\top \widetilde{\boldsymbol{Z}}_i\} - \exp\{\widehat{\boldsymbol{\beta}}(t-)^\top \widetilde{\boldsymbol{Z}}_i\} \big] \qquad \text{(B.1)} \\ &= \frac{1}{n}\sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} Y_{i(j)}(t) \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top \exp\{\boldsymbol{\beta}_i^{\circledast}(t)^\top \widetilde{\boldsymbol{Z}}_i\} \{\widehat{\boldsymbol{\beta}}(t) - \widehat{\boldsymbol{\beta}}(t-)\}, \end{aligned}$$

where $\widetilde{\boldsymbol{Z}}_{(g)}$ and $S_{(g)}$ are the associated covariate vector and the number of observed gap with the subject that contributed the gth gap-time in the sample. Here, $\boldsymbol{\beta}_i^{\circledast}(t)^\top \widetilde{\boldsymbol{Z}}_i$ is on the line segment between $\widehat{\boldsymbol{\beta}}(t)^\top \widetilde{\boldsymbol{Z}}_i$ and $\widehat{\boldsymbol{\beta}}(t-)^\top \widetilde{\boldsymbol{Z}}_i$. Note $\boldsymbol{\beta}_i^{\circledast}(t)$ satisfies $\exp\{\boldsymbol{\beta}_i^{\circledast}(t)^\top \widetilde{\boldsymbol{Z}}_i\} = \int_0^1 \exp\{[\widehat{\boldsymbol{\beta}}(t-) + r(\widehat{\boldsymbol{\beta}}(t) - \widehat{\boldsymbol{\beta}}(t-)]^\top \widetilde{\boldsymbol{Z}}_i\}\, dr$. Under C1 and C2, $\frac{1}{n} S_{(g)}^{-1} \widetilde{\boldsymbol{Z}}_{(g)}$ on the left-hand side of equation (B.1) is $O(n^{-1})$. Under C5, the limit of the factor,

$\frac{1}{n}\sum_{i=1}^n S_i^{-1} \sum_{j=1}^{S_i} Y_{i(j)}(t) \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top \exp\{\boldsymbol{\beta}_i^{\circledast}(t)^\top \widetilde{\boldsymbol{Z}}_i\}$, on the right-hand side of equation (B.1) has a smallest eigenvalue bounded away from zero because $\exp\{\boldsymbol{\beta}^{\circledast}(t)^\top \widetilde{\boldsymbol{Z}}\}$ is bounded away from zero. Consequently, at any $t \in [0, \tau]$, the jump size $\triangle \widehat{\boldsymbol{\beta}}(t)$ is asymptotically finite in the sense that $\triangle \widehat{\boldsymbol{\beta}}(t) = O(n^{-1})$. □

Proof of Theorem 3: the uniform consistency of $\widehat{\boldsymbol{\beta}}(\cdot)$.

Define classes $\boldsymbol{F}$ and $\boldsymbol{G}$ as $\boldsymbol{F} \equiv \{\widetilde{\boldsymbol{Z}} N(t) : t \in [0, \tau]\}$ and $\boldsymbol{G} \equiv \{Y(t) \widetilde{\boldsymbol{Z}} \widetilde{\boldsymbol{Z}}^\top \exp(\boldsymbol{\beta}^\top \widetilde{\boldsymbol{Z}}) : t \in [0, \tau], \boldsymbol{\beta} \in \boldsymbol{B}\}$, where $\boldsymbol{B}$ is a compact subset of $\mathcal{R}^{p+1}$. It is known that both $\{I(X \le t, \Delta = 1) : t \in \mathcal{R}\}$ and $\{I(X \ge t) : t \in \mathcal{R}\}$ are Donsker (Kosorok 2008, Lemma 9.10). By permanence properties of the Donsker class, $\boldsymbol{F}$ and $\boldsymbol{G}$ are Donsker.

Since Donsker implies Glivenko-Cantelli, it follows that, almost surely,

$$\sup_{t \in [0,\tau]} \| n^{-1} \sum_{i=1}^n S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \left[N_{i(j)}(t) - \int_0^t Y_{i(j)}(s) \widetilde{\boldsymbol{Z}}_i^\top \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\}\, d\boldsymbol{\beta}_0(s) \right] \| \longrightarrow 0. \tag{B.2}$$

Let $\boldsymbol{\psi}(t) \equiv n^{-1} \sum_{i=1}^n S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \left[N_{i(j)}(t) - \int_0^t Y_{i(j)}(s)\, d \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\} \right]$ for $t \in [0, \tau]$. Note $\boldsymbol{\psi}(t) = o(1)$, almost surely for $t \in (0, \tau]$. With the solutions to the estimating

equations (3.4) and (3.5), it follows that

$$\begin{aligned}
\boldsymbol{\psi}(t) &= n^{-1}\sum_{i=1}^{n} S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\int_0^t Y_{i(j)}(s)\,d\big[\exp\{\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\boldsymbol{Z}}_i\}-\exp\{\boldsymbol{\beta}_0(s)^\top\widetilde{\boldsymbol{Z}}_i\}\big]\\
&= n^{-1}\sum_{i=1}^{n} S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top\int_0^t Y_{i(j)}(s)\,d\big[\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}\{\widehat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s)\}\big]\\
&= \int_0^t n^{-1}\sum_{i=1}^{n} S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top Y_{i(j)}(s)\{\widehat{\boldsymbol{\beta}}(s-)-\boldsymbol{\beta}_0(s-)\}\,d\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}\\
&\quad+\int_0^t n^{-1}\sum_{i=1}^{n} S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top Y_{i(j)}(s)\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}\,d\{\widehat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s)\},
\end{aligned}\tag{B.3}$$

where $\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i$ is on the line segment between $\widehat{\boldsymbol{\beta}}(s)^\top\widetilde{\boldsymbol{Z}}_i$ and $\boldsymbol{\beta}_0(s)^\top\widetilde{\boldsymbol{Z}}_i$. Note that the subscript i in $\boldsymbol{\beta}_i^{\circledcirc}(s)$ indicates its dependency on $\widetilde{\boldsymbol{Z}}_i$.

Let

$$\widehat{\boldsymbol{A}}(t)\equiv\int_0^t n^{-1}\sum_{i=1}^{n} S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top Y_{i(j)}(s)\,d\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}$$

and

$$\widehat{\boldsymbol{B}}(s)\equiv n^{-1}\sum_{i=1}^{n} S_i^{-1}\sum_{j=1}^{S_i}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^\top Y_{i(j)}(s)\exp\{\boldsymbol{\beta}_i^{\circledcirc}(s)^\top\widetilde{\boldsymbol{Z}}_i\}.$$

It follows that

$$d\boldsymbol{\psi}(t)=\widehat{\boldsymbol{A}}(dt)\{\widehat{\boldsymbol{\beta}}(t-)-\boldsymbol{\beta}_0(t-)\}+\widehat{\boldsymbol{B}}(t)d\{\widehat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t)\}.\tag{B.4}$$

Under conditions C1 – C3, and C5, the limit of $\widehat{\boldsymbol{B}}(t)^{-1}$ exists. By multiplying the inverse of $\widehat{\boldsymbol{B}}(t)$ on each side of the equation (B.4) followed by integration over $(0,t]$, we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t) &= \{\widehat{\boldsymbol{\beta}}(0)-\boldsymbol{\beta}_0(0)\}\\
&\quad+\int_0^t\widehat{\boldsymbol{B}}(s)^{-1}d\psi(s)-\int_0^t\widehat{\boldsymbol{B}}(s)^{-1}\widehat{\boldsymbol{A}}(ds)\{\widehat{\boldsymbol{\beta}}(s-)-\boldsymbol{\beta}_0(s-)\}
\end{aligned}\tag{B.5}$$

The equation (B.5) is in the form of the Volterra integral equation (e.g., Andersen et al. 1993, Theorem II.6.3). Therefore, we obtain the unique solution $\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)$ for $t \in [0, \tau]$; that is,

$$\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t) = [\prod\nolimits_{[0,t]} \{\boldsymbol{I} - \widehat{\boldsymbol{A}}(du)\widehat{\boldsymbol{B}}(u)^{-1}\}]^\top \{\widehat{\boldsymbol{\beta}}(0) - \boldsymbol{\beta}_0(0)\}$$

$$+ \int_0^t [\prod\nolimits_{(s,t]} \{\boldsymbol{I} - \widehat{\boldsymbol{A}}(du)\widehat{\boldsymbol{B}}(u)^{-1}\}]^\top \widehat{\boldsymbol{B}}(s)^{-1} \boldsymbol{\psi}(ds), \qquad \text{(B.6)}$$

where $\prod$ denotes the product-integral.

Suppose that almost surely $\widehat{\boldsymbol{\beta}}(t)$ is in a compact subset of $\mathcal{R}^{p+1}$ for $t \in [0, \tau]$ for sufficiently large n. It follows from Lemma 2 that $\triangle\widehat{\boldsymbol{\beta}}(t)$ and $\triangle\widehat{\boldsymbol{A}}(t)$ on $[0, \tau]$ are $O(n^{-1})$, almost surely. As $\widehat{\boldsymbol{A}}(t)$ is differentiable over intervals between the adjacent event times, the derivative of $\widehat{\boldsymbol{A}}(t)$ is bounded. Therefore, $\widehat{\boldsymbol{A}}(\cdot)$ and $\int_0^t \widehat{\boldsymbol{A}}(du)\widehat{\boldsymbol{B}}(u)^{-1}$ are of bounded variation on $[0, \tau]$. Note the product integral $\prod_{(s,t]}(I - \{\widehat{\boldsymbol{A}}(du)\widehat{\boldsymbol{B}}(u)^{-1}\})$ exists for $0 \leq s \leq t \leq \tau$ (Gill and Johansen 1990, Theorem 1). Under C3, the solution vector to our proposed estimating integral equation (3.3) at time 0 is a consistent binary regression estimator for $\boldsymbol{\beta}_0(0)$. Given this consistent estimator $\widehat{\boldsymbol{\beta}}(0)$, the first term on the right-hand side of equation (B.6) is $o(1)$ almost surely; also, the second term on the right is $o(1)$ almost surely by the integration by parts. Thus, we have the result that $\sup_{t\in[0,\tau]} \|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\| \to 0$, almost surely.

If $\widehat{\boldsymbol{\beta}}(t)$ for $t > 0$ is not in a compact neighborhood of the parameter space, there must have been a time point between time 0 and t at which the estimator $\widehat{\boldsymbol{\beta}}(\cdot)$ deviates from the compact space. Let $t^* \in (0, \tau]$ be the time where $\widehat{\boldsymbol{\beta}}(\cdot)$ goes off the boundary of the compact set for the first time. As $\widehat{\boldsymbol{\beta}}(t)$ is in the compact set for all $t \in [0, t^*)$, $\|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\|$ can be made arbitrarily small for $t \in [0, t^*)$ with sufficiently large sample size. This implies that the jump size of $\widehat{\boldsymbol{\beta}}(\cdot)$ at t^* is bounded away from 0. It can be shown that this would result in a contradiction in light of the equation (B.1)

in Lemma 2. Therefore, $\widehat{\boldsymbol{\beta}}(t)$ for $t \in (0, \tau]$ residing out of the compact parameter space would not exist in large sample given the regular conditions. This completes the proof.

□

Proof of Theorem 4: the weak convergence of $\widehat{\boldsymbol{\beta}}(\cdot)$.

Recall $\boldsymbol{\psi}(\cdot)$ defined in the proof of Theorem 3. For each $t \in (0, \tau]$, almost surely,

$$\begin{aligned}
\boldsymbol{\psi}(t) &= n^{-1} \sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \int_0^t Y_{i(j)}(s)\, d\big[\exp\{\widehat{\boldsymbol{\beta}}(s)^\top \widetilde{\boldsymbol{Z}}_i\} - \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\} \big] \\
&= n^{-1} \sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \int_0^t Y_{i(j)}(s)\, d\big[\exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\} \widetilde{\boldsymbol{Z}}_i^\top \{\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s)\} \big] + \widetilde{\boldsymbol{C}}(t) \\
&= \int_0^t n^{-1} \sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top Y_{i(j)}(s) \{\widehat{\boldsymbol{\beta}}(s-) - \boldsymbol{\beta}_0(s-)\}\, d \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\} \\
&\quad + \int_0^t n^{-1} \sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top Y_{i(j)}(s) \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\}\, d\{\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s)\} + \widetilde{\boldsymbol{C}}(t),
\end{aligned}$$

for some $\widetilde{\boldsymbol{C}}(t) = o\{\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\}$ due to the consistency of $\widehat{\boldsymbol{\beta}}(t)$. Define $\widetilde{\boldsymbol{A}}(t) \equiv \int_0^t n^{-1} \sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top Y_{i(j)}(s)\, d \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\}$ and $\widetilde{\boldsymbol{B}}(s) \equiv n^{-1} \sum_{i=1}^{n} S_i^{-1} \sum_{j=1}^{S_i} \widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_i^\top Y_{i(j)}(s) \exp\{\boldsymbol{\beta}_0(s)^\top \widetilde{\boldsymbol{Z}}_i\}$. It follows that $d\boldsymbol{\psi}(t) = \widetilde{\boldsymbol{A}}(dt)\{\widehat{\boldsymbol{\beta}}(t-) - \boldsymbol{\beta}_0(t-)\} + \widetilde{\boldsymbol{B}}(t) d\{\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\} + d\widetilde{\boldsymbol{C}}(t)$. Under the regularity conditions, $\widetilde{\boldsymbol{B}}(t)^{-1}$ exists almost surely for sufficiently large n. By virtue of the Volterra integral equation,

$$
\begin{aligned}
n^{1/2}\{\widehat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t)\} &= n^{1/2}[\boldsymbol{\Pi}_{[0,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^\top\{\widehat{\boldsymbol{\beta}}(0)-\boldsymbol{\beta}_0(0)\} \\
&\quad + n^{1/2}\int_0^t [\boldsymbol{\Pi}_{(s,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^\top\widetilde{\boldsymbol{B}}(s)^{-1}d\{\psi(s)-\widetilde{\boldsymbol{C}}(s)\} \\
&= n^{1/2}[\boldsymbol{\Pi}_{[0,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^\top[-\{\widetilde{\boldsymbol{B}}(0)+o_p(1)\}^{-1}\boldsymbol{\psi}(0)] \\
&\quad + n^{1/2}\int_0^t [\boldsymbol{\Pi}_{(s,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^\top\widetilde{\boldsymbol{B}}(s)^{-1}d\{\boldsymbol{\psi}(s)-\widetilde{\boldsymbol{C}}(s)\} \\
&= -n^{1/2}[\boldsymbol{\Pi}_{[0,t]}\{\boldsymbol{I}-\boldsymbol{A}_0(du)\boldsymbol{B}_0(u)^{-1}\}]^\top\boldsymbol{B}_0(0)^{-1}\boldsymbol{\psi}(0) \\
&\quad + n^{1/2}\int_0^t [\boldsymbol{\Pi}_{(s,t]}\{\boldsymbol{I}-\widetilde{\boldsymbol{A}}(du)\widetilde{\boldsymbol{B}}(u)^{-1}\}]^\top\boldsymbol{B}_0(s)^{-1}\boldsymbol{\psi}(ds) \\
&\quad + o_p[n^{1/2}\{\widehat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t)\}],
\end{aligned}
$$

where $\boldsymbol{A}_0(t) \equiv E\big[\int_0^t \widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top Y(s)\, d\exp\{\boldsymbol{\beta}_0(s)^\top\widetilde{\boldsymbol{Z}}\}\big]$ and $\boldsymbol{B}_0(s) \equiv E\big[\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top Y(s)\exp\{\boldsymbol{\beta}_0(s)^\top\widetilde{\boldsymbol{Z}}\}\big]$. Then, $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\cdot)-\boldsymbol{\beta}_0(\cdot)\}$ weakly converges to a Gaussian process as it is a linear mapping of $n^{1/2}\boldsymbol{\psi}(\cdot)$, which is a Gaussian process itself.

□

Bibliography

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine* **8,** 907–925.

Aalen, O. O., Fosen, J., Weedon-Fekjær, H., Borgan, Ø., and Husebye, E. (2004). Dynamic analysis of multivariate failure time data. *Biometrics* **60,** 764–773.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10,** 1100–1120.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66,** 429–436.

Byar, D. P. (1980). *The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumours: Comparisons of Placebo, Pyridoxine and Topical Thiotepa*. Springer US, Boston, MA.

Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in cox's regression models. *Scandinavian Journal of Statistics* **30,** 93–111.

Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89,** 659–668.

Chiang, C.-T. and Wang, M.-C. (2009). Varying-coefficient model for the occurrence rate function of recurrent events. *Annals of the Institute of Statistical Mathematics* **61,** 197–213.

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

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association* **83,** 744–749.

Eshleman, S. H., Mracna, M., Guay, L., Deseyve, M., Cunningham, S., Mirochnick, M., et al. (2001). Selection and fading of resistance mutations in women and infants receiving nevirapine to prevent hiv-1 vertical transmission (hivnet 012). *AIDS* **15,** 1951–7.

Fine, J. P., Yan, J., and Kosorok, M. R. (2004). Temporal process regression. *Biometrika* **91,** 683–703.

Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics* **18,** 1501–1555.

Huang, Y. (2010). Quantile calculus and censored regression. *Ann. Statist.* **38,** 1607–1637.

Huang, Y. (2014). Bootstrap for the case-cohort design. *Biometrika* **101,** 465–476.

Huang, Y. (2017). Restoration of monotonicity respecting in dynamic regression. *Journal of the American Statistical Association* **112,** 613–622.

Huang, Y. and Chen, Y. Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Analysis* **9,** 293–303.

Huang, Y. and Peng, L. (2009). Accelerated recurrence time models. *Scandinavian journal of statistics* **36,** 636–648.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Inc.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag New York.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37,** 158–168.

Lin, D., Fleming, T., and Wei, L. (1994). Confidence bands for survival curves under the proportional: Hazards model. *Biometrika* **81,** 73–81.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 711–730.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85,** 605–618.

Martinussen, T., Scheike, T. H., and Skovgaard, I. M. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics* **29,** 57–74.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81,** 501–514.

Peng, L. and Huang, Y. (2007). Survival analysis with temporal covariate effects. *Biometrika* **94,** 719–733.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association* **103,** 637–649.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* **88,** 811–820.

Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika* **60,** 279–288.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65,** 167–179.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68,** 373–379.

Qian, J. and Peng, L. (2010). Censored quantile regression with partially functional effects. *Biometrika* **97,** 839–850.

Rubin, D. B. (1981). The bayesian bootstrap. *The Annals of Statistics* **9,** 130–134.

Strausbaugh, S. D. and Davis, P. B. (2007). Cystic fibrosis: A review of epidemiology and pathobiology. *Clinics in Chest Medicine* **28,** 279 – 288. Cystic Fibrosis.

Sun, X., Peng, L., Huang, Y., and Lai, H. J. (2016). Generalizing quantile regression for counting processes with applications to recurrent events. *Journal of the American Statistical Association* **111,** 145–156.

Tangpricha, V., Smith, E. M., Binongo, J., Judd, S. E., Ziegler, T. R., Walker, S., et al. (2017). The vitamin d for enhancing the immune system in cystic fibrosis (disc) trial: Rationale and design of a multi-center, double-blind, placebo-controlled trial of high dose bolus administration of vitamin d3 during acute pulmonary exacerbation of cystic fibrosis. *Contemporary Clinical Trials Communications* **6,** 39 – 45.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag New York.

Tian, L., Zucker, D., and Wei, L. J. (2005). On the cox model with time-varying regression coefficients. *Journal of the American Statistical Association* **100,** 172–183.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer New York.

Wang, M.-C. and Chang, S.-H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association* **94,** 146–153.

Wolfenden, L. L., Judd, S. E., Shah, R., Sanyal, R., Ziegler, T. R., and Tangpricha, V. (2008). Vitamin d and bone health in adults with cystic fibrosis. *Clinical Endocrinology* **69,** 374–381.

Wu, H., Huang, Y., P Acosta, E., L Rosenkranz, S., R Kuritzkes, D., J Eron, J., et al. (2005). Modeling long-term hiv dynamics and antiretroviral response: Effects of drug potency, pharmacokinetics, adherence, and drug resistance. *Journal of Acquired Immune Deficiency Syndromes* **39,** 272–83.

Zucker, D. M. and Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *Ann. Statist.* **18,** 329–353.