Semiparametric Functional Temporal Process Regression of Prevalence Outcomes

by

Tianyu Zhan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2017

Doctoral Committee:

Professor Douglas E. Schaubel, Chair
Research Assistant Professor Zhi He
Assistant Professor Christopher J. Sonnenday
Professor Alexander Tsodikov

Tianyu Zhan

tianyuzh@umich.edu

ORCID iD: 0000-0002-8572-4539

Acknowledgments

I would like to express my special appreciation and thanks to my thesis advisor Professor Douglas E. Schaubel, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a statistician. Your advice on both research as well as on my career have been priceless. I would also like to thank my committee members, professor Zhi (Kevin) He, professor Christopher J. Sonnenday and professor Alexander Tsodikov for serving as my committee members even at hardship. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you. My sincere thanks also goes to Professor Hui Jiang, for offering me the summer internship and research assistant opportunities in his group and leading me working on diverse exciting projects.

I would also like to thank all of my friends who supported me and incented me to strive towards my goal. To the very best of times we had in the last four years at University of Michigan.

A special thanks to my families. Words cannot express how grateful I am to my mother Jie Wang and my father Kai Zhan for all the support and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

We develop novel semiparametric methods for flexibly modeling a prevalence indicator process observed over follow-up time. The proposed methods are motivated by various settings arising in end-stage renal disease (ESRD) and end-stage liver disease (ESLD).

In Chapter 1, we consider the response Survival-Out-of-Hospital, defined as a temporal process taking the value 1 when the subject is currently alive and not hospitalized, and 0 otherwise. The semiparametric model we consider assumes multiplicative covariate effects and leaves unspecified the baseline probability of being alive-and-out-of-hospital. Asymptotic properties are derived, and simulation studies are performed. The proposed methods are applied to the Dialysis Outcomes and Practice Patterns Study (DOPPS), a prospective international ESRD study.

In Chapter 2, we extend the methods from Chapter 1 to accommodate dependent censoring. We derive a modification of Inverse Probability of Censoring Weighting which offers improved stability and reduced computational burden relative to the traditional IPCW. We show that the regression estimator is asymptotically normal, and that the baseline probability function estimator converges to a Gaussian process. The methods are used to model the probability of being alive and active on the liver transplant waiting list, which can be dependently censored by liver transplantation.

In Chapter 3, we jointly model prevalence conditional on survival, and the death hazard. A frailty is shared by the prevalence and hazard models, with the frailty effect scaled in the latter. We propose an iterative procedure for estimating the regression parameters by updating frailties from their implied estimating equations. The scale parameter and random effect variance are estimated by numerical integration. The algorithm is asymptotically equivalent to an EM algorithm which treats the frailty terms as missing data. We apply the methods to DOPPS data.

CHAPTER 1

Semiparametric Temporal Process Regression of Survival-Out-of-Hospital

1.1 Introduction

Often in clinical or epidemiological studies, both a recurrent event process and terminal event are of interest. This is particularly true given the recent proliferation of administrative databases available for secondary analysis. Correspondingly, a considerable number of methods have been developed for the recurrent/terminal event data structure. One option is to model the marginal mean/rate [Cook and Lawless, 1997, Ghosh and Lin, 2002, Schaubel et al., 2006, Cook et al., 2009], essentially averaging over the survival experience. Another class of approaches involves jointly modeling survival and the conditional recurrent event rate given survival [Huang and Wang, 2004, Liu et al., 2004, Ye et al., 2007, Zeng and Cai, 2010, Kalbfleisch et al., 2013].

In this report, we study a useful alternative framework for the analysis of recurrent/terminal event data. In particular, a frequently arising example of this data structure involves hospitalization representing the recurrent event, with death serving as the terminal event. Although hospital admission can be regarded as a point process, the length of stay for a hospitalization may be several days and, therefore, should not be ignored in the analysis. This concept is recognized in the works of Hu et al. [2011] and Zhu et al. [2014], for example. Survival-out-of-hospital may be viewed as a refinement of survival time in the study of chronic illness such as end-stage renal disease (ESRD), with the refinement being the incorporation of each patient's hospital admission and length of stay

information. An appealing characteristic of survival-out-of-hospital is that it incorporates morbidity data in an objective and easily understood manner. Data of this structure are increasingly available, given the relatively recent proliferation of publicly available health administrative data sets [Holland and Lam, 2000, Sands et al., 2006, Carson et al., 2009].

Temporal process regression would appear to be a natural conceptualization of survival-out-of-hospital. Several process regression methods have been proposed in the last decade. For example, Fine et al. [2004] developed functional generalized linear models, for which covariates effects are completely unspecified and are estimated nonparametrically over time. Such an approach has also been generalized to multivariate survival settings to model both mean and association structures [Yan and Fine, 2005]. To increase precision, the partly functional temporal process model has been developed, with covariate effects being nonparametric for some covariate elements and parametric for others [Yan and Huang, 2009, Estes et al., 2015]. These functional generalized linear models generally focus on time-varying covaiate effects. In addition, martingale-based estimating equations have been proposed for directly modeling survival function by solving a sequence of monotone equations [Peng and Huang, 2007]. Approaches listed in this paragraph are part of the inspiration for the methods we propose in this report. However, as will become more clear later in our report, none of these approaches are applicable to our setting given the specifics of analytic objectives, along with the assumed data structure and model of interest.

Various other existing methods are pertinent to the data structure of our interest, but not applicable to our research question. For instance, methods proposed in Andersen et al. [2003] and Scheike and Zhang [2007] involved direct modeling of a state transition probability or a state occupation probability in a multi-state model. In these approaches, each state is assumed to be visited not more than once. As such, none of these approaches are directly applicable to our particular research question and data structure, since patients can move in and out of hospital many times prior to death. In addition, pseudo-observation approaches [Andersen et al., 2003, Grand and Putter, 2016], despite their utility, generally only allow for censoring that does not dependent on the covariate vector, an assumption that may be violated in observational studies. As described in

Section 2, the methods we propose allow censoring to depend on the covariate vector.

In this report, we propose a semiparametric temporal process regression method, where covariates have multiplicative effects on a completely unspecified baseline probability function. This model can be thought of a process version of the generalized linear model with the process indexed by time. In terms of estimation, the regression parameter is estimated by the solution to an estimating equation which is free of the baseline probability function. We propose a nonparametric estimator for the baseline probability process, a closed form for which can be computed after estimating the regression parameter. The estimating functions do not require inverse weighting in the setting where censoring times are known (e.g., if all censoring was administrative, occurring at the same calendar date). To accommodate the more commonly occurring scenario where censoring is random, we employ multiple imputation [Little and Rubin, 2002] to recover censoring times unobserved due to death [Schaubel and Zhang, 2010].

Our method has several distinguishing features. First, covariate effects are on the relative risk scale, as opposed to odds ratio. The baseline probability process is specified as a nonparametric function of follow-up time, to increase flexibility and robustness. Through a development that parallels the derivation of the Cox partial likelihood score function [Cox, 1972a], we derive an estimating function for the regression parameter that is free of the baseline probability function. This reduces the complexity of computation and permits the use of standard statistical software. Moreover, the baseline probability function can be subsequently estimated at any specific time point before the maximum observation time. We also propose estimating the integral of the baseline probability, which can then be used to predict expected survival time out-of-hospital over a finite time interval.

In this article, we are interested in the joint event of being out-of-hospital and being alive. This response variable takes quality of life information into consideration, while leaves the dependent structure between temporal indicator and terminal event completely unspecified. More model details and challenges are illustrated in the following section. Our method is different from the current practice in clinical trials which use the time to the first recurrent event or terminal event

[Lewis, 1999, Pfeffer et al., 2003], and is also distinct from a weighted composite endpoint of all recurrent and terminal events [Neaton et al., 2005, Mao and Lin, 2016].

The remainder of the article is organized as follows. In Section 2, we introduce notations, formulate model assumptions, and propose estimating procedures for parameters and baseline probability function. In Section 3, we show that the regression parameter estimator converges to a Normal distribution, and estimator of the baseline probability function converges to a Gaussian process. Simulation studies are performed to evaluate our method under various scenarios in Section 4. We also illustrate our method through an analysis of survival-out-of-hospital using data from the Dialysis Outcomes and Practice Patterns Study (DOPPS), a long-running international prospective study of ESRD patients. Finally, concluding remarks are provided at Section 6.

1.2 Model and Methods

Suppose there are a total of n independent subjects. Let D_i denote the death (terminal event) time of subject i ($i = 1, 2, ..., n$). We let C_i be the censoring time, and let $\boldsymbol{Z}_i(t)$ be a p-dimensional covariate vector which may contain time-varying elements (assumed to be external; Kalbfleisch and Prentice [2002], p.g. 196). Here, we consider follow-up time $t \in [0, \tau]$, where τ is a pre-specified constant satisfying $\Pr(C_i \geq \tau) > 0$ for $i = 1, 2, ..., n$. In practice, τ could be chosen as the maximum of observation time, $C_i \wedge D_i$, where $a \wedge b = \min(a, b)$.

Let $H_i(t) = 1$ if subject i is in the hospital at time t, and 0 if out of hospital. The probability of interest is the probability that a subject i is alive and out-of-hospital at time t, i.e.,

$$\pi_i(t) = P\{H_i(t) = 0, D_i > t | \boldsymbol{Z}_i(t)\}. \tag{1.1}$$

We assume that the covariates have multiplicative effects on an unspecified baseline probability function $\pi_0(t)$, such that

$$\pi_i(t) = \pi_0(t)\exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)\}, \tag{1.2}$$

where $\boldsymbol{\beta}_0$ is a p-dimensional vector. This is similar to the Cox proportional hazard assumption [Cox, 1972a], but $\pi_i(t)$ is interpreted as the probability of being out of hospital and alive for subject i at time t. Note the distinction between (1.2) and an intensity [Andersen and Gill, 1982] or marginal rate function [Lin et al., 2000].

For brevity of notation, we define $A_i(t) = I\{D_i > t\}$ as the alive indicator, and $A_i^0(t) = I\{H_i(t) = 0, D_i > t\}$ as the survival-out-of-hospital indicator. Similar we could define $A_i^1(t) = I\{H_i(t) = 1, D_i > t\}$ as survival-hospitalized indicator. In this report we are interested in $A_i^0(t)$, while people could also focus on $A_i^1(t)$. It is important to note that information is still available on subject i before the censoring time, even after the terminal event has occurred. Note that $A_i^0(t) = 0$ for $t \in [D_i, C_i]$ if $D_i < C_i$.

1.2.1 Known Censoring

To begin, suppose that censoring time C_i is always known. This would be the case in a closely monitored prospective study from which patients could not be randomly lost to follow-up. In such cases, C_i is known even if $C_i > D_i$. This set-up does not match most observational studies, but it is a useful starting point. We assume that censoring time is independent of our target event $A_i^0(t)$, conditional on $\boldsymbol{Z}_i(t)$; more explicitly, we can express the assumption as follows,

$$E\{A_i^0(t)|\boldsymbol{Z}_i(t), C_i \geq t\} = E\{A_i^0(t)|\boldsymbol{Z}_i(t)\}. \tag{1.3}$$

Consider the following two estimating functions,

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i(t)[A_i^0(t) - \pi_i(t)]I(C_i \geq t)dt \tag{1.4}$$

$$\sum_{i=1}^{n} [A_i^0(t) - \pi_i(t)]I(C_i \geq t). \tag{1.5}$$

These estimating functions have expectation zero under model (1.2) and assuming conditionally independent censoring (1.3). Note that equation (1.5) could be evaluated at any specific time point

$t \in [0, \tau]$, and does not have to be an observed event time. Solving equation (1.5) for $\pi_0(t)$ treating $\boldsymbol{\beta}_0$ as known, then substituting the resulting $\widehat{\pi}_0(t)$ back into equation (1.4) yields the following zero-mean estimating equations for $\boldsymbol{\beta}_0$,

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\beta})\} A_i^0(t) I(C_i \geq t) dt = \boldsymbol{0}, \tag{1.6}$$

where $\bar{\boldsymbol{Z}}(t;\boldsymbol{\beta}) = \boldsymbol{S}^{(1)}(t;\boldsymbol{\beta})/S^{(0)}(t;\boldsymbol{\beta})$, $\boldsymbol{S}^{(k)}(t;\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} \boldsymbol{Z}_i(t)^{\otimes k} I(C_i \geq t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}$ for $k = 0, 1, 2$, where $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. We also define the following notations. Let $\boldsymbol{s}^{(k)}(t;\boldsymbol{\beta}) = E[\boldsymbol{Z}_1(t)^{\otimes k} I(C_1 \geq t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_1(t)\}]$ and $\bar{\boldsymbol{z}}(t;\boldsymbol{\beta}) = \boldsymbol{s}^{(1)}(t;\boldsymbol{\beta})/s^{(0)}(t;\boldsymbol{\beta})$ be the limiting value of $\boldsymbol{S}^{(k)}(t;\boldsymbol{\beta})$ and $\bar{\boldsymbol{Z}}(t;\boldsymbol{\beta})$, respectively. Moreover, define
$\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{S}^{(2)}(t;\boldsymbol{\beta})/S^{(0)}(t;\boldsymbol{\beta}) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\beta})^{\otimes 2}\} A_i^0(t) I(C_i \geq t) dt$ and its limiting value $\boldsymbol{\Omega}(\boldsymbol{\beta}) = E[\int_0^{\tau} \{\boldsymbol{s}^{(2)}(t;\boldsymbol{\beta})/s^{(0)}(t;\boldsymbol{\beta}) - \bar{\boldsymbol{z}}(t;\boldsymbol{\beta})^{\otimes 2}\} A_1^0(t) I(C_1 \geq t) dt]$. Observed data for subject i can be summarized by $F_i = \{H_i(t), D_i \wedge C_i, I(D_i \geq C_i), \boldsymbol{Z}_i(t), t \in [0, X_i]\}$ for $i = 1, 2, ..., n$, where $X_i = D_i \wedge C_i$.

Equation (1.6) is reminiscent of the Cox regression score function, a property which can be exploited computationally (i.e., to ease programming effort). For instance, if the time scale is days (like our real-data application in Section 5) or some other discrete measure, then standard proportional hazards software (e.g., coxph(·) in R, proc phreg in SAS) can be used after augmenting the data set. Specifically, the augmented data would contain one record for each time unit t a subject is uncensored, with the event indicator for time unit t would be $A_i^0(t)$. Such data would left truncated such that, within subject, the $(j+1)$th record begins (i.e., its left subinterval boundary) where the jth records ended (i.e., its right interval boundary). Alternatively, if the time scale were truly continuous, then the augmented data would contain, for each subject, one record for every time at which $\boldsymbol{Z}_i(t)$, $\bar{\boldsymbol{Z}}(t;\boldsymbol{\beta})$, or $A_i^0(t)$ changes. Additionally, for the jth record of subject i spanning, say, $(t_{j-1}, t_j]$, a weight of $(t_j - t_{j-1})$ would be used, along with offset $\log(t_j - t_{j-1})$. Note that the Breslow approximation (for tie-handling) would be used, as is the default in SAS's phreg. Naturally, an alternative is to write explicit code to solve (1.6) using Newton-Raphson.

Denoting the solution to (1.6) by $\widehat{\boldsymbol{\beta}}$, one can estimate $\pi_0(t)$ through the closed form,

$$\widehat{\pi}_0(t) = \frac{\sum_{i=1}^{n} A_i^0(t) I(C_i \geq t)}{\sum_{i=1}^{n} I(C_i \geq t) \exp[\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)]}. \tag{1.7}$$

Note that $\widehat{\pi}_0(t)$ is closely related to the Breslow estimator of baseline cumulative hazard function [Breslow, 1972]. However, $\pi_0(t)$ is bounded above by 1, since it corresponds to a probability. With respect to leveraging standard software, $\widehat{\pi}_0(t)$ can be computed directly through successive differencing in the previously described scenarios wherein the time scale is discrete (with one record per uncensored time unit per subject). If time is continuous, then $\widehat{\pi}_0(t)$ would need to be computed explicitly, with the value changing each time, for any subject i in the data set, either of $A_i^0(t)$, or $I(C_i \geq t)$ or $\boldsymbol{Z}_i(t)$ changes its value.

Note that, due to the use of the log link, $\widehat{\pi}_0(t) > 1$ is possible. This could occur, for example, in settings wherein hospitalization and mortality rates are low. One solution is to cap $\widehat{\pi}_0(t)$ by 1.

We also define the integral of the baseline survival-out-of-hospital probability $\Pi_0(L)$ as

$$\Pi_0(L) = \int_0^L \pi_0(t) dt \tag{1.8}$$

which could be interpreted as the expected length of being alive and out-of-hospital up to time L for subjects with baseline covariate. We could estimate $\Pi_0(L)$ by $\widehat{\Pi}_0(L) = \int_0^L \widehat{\pi}_0(t) dt$. Note that $\widehat{\pi}_0(t)$ in (1.7) would jump up when subjects are discharged from the hospital, and jump down if subjects are admitted to the hospital or dead. This property would facilitate computation in real data application, because $\widehat{\Pi}_0(L)$ would be a sum of rectangular area with height $\widehat{\pi}_0(t)$ and length in recorded unit, for example day.

1.2.2 Random Censoring

Now, consider the more typical scenario in which the censoring time C_i is not fixed at $t = 0$, with its randomness implying that C_i is not known in cases where D_i is observed for subject i. In this

set-up, one cannot carry out estimation through (1.6) and (1.7), defined in the preceding subsection, due to C_i being missing for subjects observed to die. A simple solution is to set censoring time as the maximum follow-up time across all subjects, which may lead to substantial bias in estimating $\boldsymbol{\beta}_0$ and $\pi_0(t)$. Similarly, setting censoring time as D_i for subjects with $C_i > D_i$ would generally lead to invalid inference.

One could use weighting techniques to recover missing censoring time, essentially weights subjects with $C_i > D_i$ by a conditional survival probability [Ghosh and Lin, 2002, Mao and Lin, 2016]. However, a weighted version of estimating equation (1.6) is tedious to carry out, since the time line is continuous. In this report we consider an alternative imputation approach which is easy to implement by standard software.

Our solution is to multiply impute C_i when $D_i < C_i$ and, for this purpose, we assume the following proportional hazard model on the censoring time,

$$\lambda_i^C(t) = \lambda_0^C(t)\exp\{\boldsymbol{\theta}_0^T \boldsymbol{Z}_i(t)\}, \tag{1.9}$$

where $\lambda_0^C(t)$ is an unspecified baseline hazard function for C_i. In light of the model for C_i, we now define the corresponding counting process $N_i^C(t) = I(C_i \leq t \wedge D_i)$, and its increment $dN_i^C(t) = N_i^C(t^- + dt) - N_i^C(t^-)$. Similarly, the counting process for death time is defined by $N_i^D(t) = I(D_i \leq t \wedge C_i)$. Let $Y_i(t) = I(X_i \geq t)$ be the at risk process. Standard partial likelihood [Cox, 1975] techniques can be fitted to the observed censoring time data $\{X_i, I(C_i \leq D_i), \boldsymbol{Z}_i(t);\ \ t \in [0,\tau]\}_{i=1}^n$ to compute $\widehat{\boldsymbol{\theta}}$, which can be shown to be a strongly consistent estimator of $\boldsymbol{\theta}_0$ [Andersen and Gill, 1982]. The baseline cumulative hazard function for $\Lambda_0^C(t)$ is estimated through the method of Breslow [Breslow, 1972].

We will create M imputed datasets, where normally $M = 10$ would suffice. The proposed methods are also valid for $M = 1$, although this would be less efficient. Consider the mth imputed dataset, for subjects with $C_i \leq D_i$, we set imputed censoring time as the known censoring time. For subjects $C_i > D_i$, which means that the censoring time is not observed, we impute $\widehat{C}_i^{\langle m \rangle}$ from

the estimated conditional survival function,

$$\widehat{G}(t;\widehat{\boldsymbol{\theta}}) = I(t \geq D_i)\text{exp}[-\widehat{\Lambda}_i^C(t;\widehat{\boldsymbol{\theta}}) + \widehat{\Lambda}_i^C(D_i;\widehat{\boldsymbol{\theta}})] \tag{1.10}$$

More explicitly, we could express the imputed censoring time $C_i^{\langle m\rangle}$ as

$$C_i^{\langle m\rangle} = [1 - N_i^D(X_i)]C_i + N_i^D(X_i)\widehat{C}_i^{\langle m\rangle}.$$

We make the following notations. First consider the uncensored indicator $I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta})$, where we include $\boldsymbol{\theta}$ in the parenthesis to emphasize that $C_i^{\langle m\rangle}$ depends on imputation parameter $\boldsymbol{\theta}$. In finite samples, we impute $\widehat{C}_i^{\langle m\rangle}$ from $\widehat{G}(t;\widehat{\boldsymbol{\theta}})$ for subjects with $C_i > D_i$ to obtain $I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})$. Define $G(t;\boldsymbol{\theta}_0) = I(t \geq D_i)\text{exp}[-\Lambda_i^C(t;\boldsymbol{\theta}_0) + \Lambda_i^C(D_i;\boldsymbol{\theta}_0)]$. Therefore, $I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)$ refers to a hypothetical scenario where we impute $\widehat{C}_i^{\langle m\rangle}$ from the true underlying $G(t;\boldsymbol{\theta}_0)$ when $C_i > D_i$. When (1.9) is correctly specified, $I(C_i \geq t)$ and $I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)$ follow exact the same distribution. The notation $I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)$ is useful for the establishment of asymptotic properties. Also note that $A_i^0(t)I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}}) = A_i^0(t)I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)$ because $A_i^0(t) = 0$ for $t \in [D_i,\tau]$. Let $\boldsymbol{S}^{(k)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i(t)^{\otimes k}I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta})\text{exp}\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}$, and $\boldsymbol{s}^{(k)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = E[\boldsymbol{Z}_1(t)^{\otimes k}I(C_1^{\langle 1\rangle} \geq t;\boldsymbol{\theta})\text{exp}\{\boldsymbol{\beta}^T\boldsymbol{Z}_1(t)\}]$ be the limiting value of $\boldsymbol{S}^{(k)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$ for $k = 0,1,2$. Also let $\bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = \boldsymbol{S}^{(1)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})/S^{(0)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$, and denote its limiting value as $\bar{\boldsymbol{z}}^{\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = \boldsymbol{s}^{(1)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})/s^{(0)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$. Moreover, define
$\widehat{\boldsymbol{\Omega}}^{\langle m\rangle}(\boldsymbol{\beta},\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n\int_0^\tau\{\boldsymbol{S}^{(2)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})/S^{(0)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) - \bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})^{\otimes 2}\}A_i^0(t)I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta})dt$ and its limiting value
$\boldsymbol{\Omega}^{\langle 1\rangle}(\boldsymbol{\beta},\boldsymbol{\theta}) = E[\int_0^\tau\{\boldsymbol{s}^{(2)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})/s^{(0)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) - \bar{\boldsymbol{z}}^{\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})^{\otimes 2}\}A_1^0(t)I(C_1^{\langle 1\rangle} \geq t;\boldsymbol{\theta})dt]$.

For the mth imputed data set, the estimation of $\boldsymbol{\beta}_0$ and $\pi_0(t)$ is as defined in Section 1.2.1, but with C_i replaced by $C_i^{\langle m\rangle}$. More explicitly, $\widehat{\boldsymbol{\beta}}^{\langle m\rangle}$ is computed by solving the following estimating equation,

$$\boldsymbol{U}^{\langle m\rangle}(\boldsymbol{\beta},\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n\int_0^\tau\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta},\widehat{\boldsymbol{\theta}})\}A_i^0(t)I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})dt = 0 \tag{1.11}$$

$\widehat{\pi}_0^{\langle m\rangle}(t)$ is given by

$$\widehat{\pi}_0^{\langle m\rangle}(t) = \frac{\sum_{i=1}^n A_i^0(t) I(C_i^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\theta}})}{\sum_{i=1}^n \{I(C_i^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\theta}}) \text{exp}[\boldsymbol{Z}_i^T(t) \widehat{\boldsymbol{\beta}}^{\langle m\rangle}]\}} \tag{1.12}$$

Having computed $\widehat{\boldsymbol{\beta}}^{\langle m\rangle}$ and $\widehat{\pi}_0^{\langle m\rangle}(t)$ $(m = 1, 2, ..., M)$, $\boldsymbol{\beta}_0$ and $\pi_0(t)$ will be estimated by a pooled estimator $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{\pi}_0^M(t)$,

$$\widehat{\boldsymbol{\beta}}^M = M^{-1} \sum_{m=1}^M \widehat{\boldsymbol{\beta}}^{\langle m\rangle} \tag{1.13}$$

$$\widehat{\pi}_0^M(t) = \frac{\sum_{i=1}^n \sum_{m=1}^M A_i^0(t) I(C_i^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\theta}})}{\sum_{i=1}^n \sum_{m=1}^M I(C_i^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\theta}}) \text{exp}[\boldsymbol{Z}_i^T(t) \widehat{\boldsymbol{\beta}}^M]} \tag{1.14}$$

The estimator $\widehat{\pi}_0^M(t)$ given in (1.14) is easier for us to develop the asymptotic properties, because both the numerator and denominator can be expressed as empirical processes indexed by time t. The imputed $\widehat{\Pi}_0^M(L)$ could be defined similarly to $\widehat{\Pi}_0(L)$, but with $\widehat{\pi}_0(t)$ replaced by $\widehat{\pi}_0^M(t)$.

1.3 Asymptotic Properties

We assume the following set of regularity conditions:

(a) $\{H_i(t), X_i, I(C_i \leq D_i), \boldsymbol{Z}_i(t)\}$ for $t \in [0, X_i]$, $i = 1, 2, ..., n$ are independent and identically distributed.

(b) $Pr(C_i \geq \tau) > 0$ for $i = 1, 2, ..., n$, where τ is a pre-specified constant.

(c) $|\boldsymbol{Z}_{ij}(0)| + \int_0^\tau |d\boldsymbol{Z}_{ij}(t)| < c_{\boldsymbol{Z}} < \infty$ almost surely for $i = 1, 2, ..., n$, $j = 1, 2, ..., p$, i.e., $\boldsymbol{Z}_i(t)$ has bounded total variations.

(d) $\boldsymbol{\Omega}(\boldsymbol{\beta}_0) = E[\int_0^\tau \{\boldsymbol{s}^{(2)}(t; \boldsymbol{\beta}_0)/s^{(0)}(t; \boldsymbol{\beta}_0) - \bar{\boldsymbol{z}}(t; \boldsymbol{\beta}_0)^{\otimes 2}\} A_1^0(t) I(C_1 \geq t) dt]$ is positive definite.

(e) For $\boldsymbol{\beta} \in \mathscr{B}_\delta$, where $\mathscr{B}_\delta$ is a small neighborhood around $\boldsymbol{\beta}_0$, $s^{(0)}(t; \boldsymbol{\beta})$, and $\boldsymbol{s}^{(1)}(t; \boldsymbol{\beta})$ are bounded away from zero.

(f) For $\boldsymbol{\beta} \in \mathscr{B}_\delta$, $k = 0, 1, 2$, $\boldsymbol{s}^{(k)}(t; \boldsymbol{\beta})$ are continuous uniformly on $t \in [0, \tau]$, and are bounded on $[0, \tau] \times \mathscr{B}_\delta$.

We summarize the essential asymptotic properties of $\widehat{\boldsymbol{\beta}}$, $\widehat{\pi}_0(t)$ and imputed versions $\widehat{\boldsymbol{\beta}}^M$, $\widehat{\pi}_0^M(t)$ in the following theorems. Proofs are sketched in the Supplemental Materials.

1.3.1 Known Censoring

We begin by describing the asymptotic properties of estimators applicable to the known-censoring set-up from Section 1.2.1.

Theorem 1. *Under assumptions (1.2) and (1.3) and the afore-listed regularity conditions, $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$, and $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a mean-zero Normal random variable with a variance-covariance matrix*

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}_0) = \boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1} E[\boldsymbol{u}_1(\boldsymbol{\beta}_0)\boldsymbol{u}_1(\boldsymbol{\beta}_0)^T]\boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1}, \tag{1.15}$$

where $\boldsymbol{u}_i(\boldsymbol{\beta}) = \int_0^\tau \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{z}}(t;\boldsymbol{\beta})\} dM_i(t;\boldsymbol{\beta})$, and $dM_i(t;\boldsymbol{\beta}) = A_i^0(t) I(C_i \geq t) dt - exp[\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)] I(C_i \geq t) dt$. A consistent estimator of $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$ is given by $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}) = n^{-1} \sum_{i=1}^n [\widehat{\boldsymbol{f}}_i^{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}) \widehat{\boldsymbol{f}}_i^{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}})^T]$, where $\widehat{\boldsymbol{f}}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1} \int_0^\tau \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\beta})\} dM_i(t;\boldsymbol{\beta})$.

We now describe the asymptotic behavior of $\widehat{\pi}_0(t)$ from (1.7).

Theorem 2. *Under assumptions (1.2) and (1.3), $n^{1/2}(\widehat{\pi}_0 - \pi_0)$ converges weakly to a mean-zero Gaussian process with a variance and covariance matrix between $n^{1/2}[\widehat{\pi}_0(s) - \pi_0(s)]$ and $n^{1/2}[\widehat{\pi}_0(t) - \pi_0(t)]$ given by $\sigma(s,t) = E[\xi_1(s)\xi_1(t)]$, where*

$$\xi_i(t) = \frac{f_i^{\pi_1}(t;\boldsymbol{\beta}_0) - f_i^{\pi_2}(t;\boldsymbol{\beta}_0)}{s^{(0)}(t;\boldsymbol{\beta}_0)}$$

and $f_i^{\pi_1}(t;\boldsymbol{\beta}) = A_i^0(t) I(C_i \geq t) - exp[\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)] I(C_i \geq t) \pi_0(t)$, $f_i^{\pi_2}(t;\boldsymbol{\beta}) = \boldsymbol{s}^{(1)}(t;\boldsymbol{\beta})^T \pi_0(t) \boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta})$, $\boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \boldsymbol{\Omega}(\boldsymbol{\beta})^{-1} \boldsymbol{u}_i(\boldsymbol{\beta})$. A consistent estimator of $\sigma(s,t)$ is given by its empirical counterparts $\widehat{\sigma}(s,t) = n^{-1} \sum_{i=1}^n \widehat{\xi}_i(s) \widehat{\xi}_i(t)$, where

$$\widehat{\xi}_i(t) = \frac{f_i^{\pi_1}(t;\widehat{\boldsymbol{\beta}}) - \widehat{f}_i^{\pi_2}(t;\widehat{\boldsymbol{\beta}})}{S^{(0)}(t;\widehat{\boldsymbol{\beta}})}$$

and $\widehat{f}_i^{\pi_2}(t;\boldsymbol{\beta}) = \boldsymbol{S}^{(1)}(t;\boldsymbol{\beta})^T\widehat{\pi}_0(t)\widehat{\boldsymbol{f}}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta})$.

1.3.2 Random Censoring

Now we describe the limiting behavior of estimator for the random-censoring set-up from Section 1.2.2.

Theorem 3. *If censoring mechanism (1.9), assumptions (1.2) and (1.3) are correctly specified, then under the previously listed regularity conditions,* $\widehat{\boldsymbol{\beta}}^M$ *is a consistent estimator of* $\boldsymbol{\beta}_0$*, and* $n^{1/2}(\widehat{\boldsymbol{\beta}}^M - \boldsymbol{\beta}_0)$ *converges in distribution to a mean-zero Normal random variable with a variance-covariance matrix*

$$\boldsymbol{\Sigma}_M(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) = [\boldsymbol{\Omega}(\boldsymbol{\beta}_0)]^{-1}E\{[M^{-1}\sum_{m=1}^{M}\boldsymbol{u}_1^{\langle m\rangle}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]^{\otimes 2}\}[\boldsymbol{\Omega}(\boldsymbol{\beta}_0)]^{-1}$$

where $\boldsymbol{u}_i^{\langle m\rangle}(\boldsymbol{\beta},\boldsymbol{\theta}) = \int_0^\tau\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{z}}^{\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})\}dM_i^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$,
and $dM_i^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = A_i^0(t)I(C_i^{\langle m\rangle}\geq t)dt - exp[\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)]I(C_i^{\langle m\rangle}\geq t;\boldsymbol{\theta})dt$.

A consistent estimator of $\boldsymbol{\Sigma}_M(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ *is given by*

$$\widehat{\boldsymbol{\Sigma}}_M(\widehat{\boldsymbol{\beta}}^M,\widehat{\boldsymbol{\theta}}) = n^{-1}\sum_{i=1}^{n}[\widehat{\boldsymbol{f}}_i^{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}^M,\widehat{\boldsymbol{\theta}},M)]^{\otimes 2}$$

where $\widehat{\boldsymbol{f}}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\theta},M) = [M^{-1}\sum_{k=1}^{M}\widehat{\boldsymbol{\Omega}}^{\langle k\rangle}(\boldsymbol{\beta},\boldsymbol{\theta})]^{-1}M^{-1}\sum_{m=1}^{M}\int_0^\tau\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})\}dM_i^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$.

The asymptotic behavior of imputed version $\widehat{\pi}_0^M(t)$ from (1.14) is summarized in the following theorem.

Theorem 4. *If censoring mechanism (1.9), assumptions (1.2) and (1.3) are correctly specified, then* $n^{1/2}(\widehat{\pi}_0^M - \pi_0)$ *converges weakly to a mean-zero Gaussian process with a variance and covariance matrix between* $n^{1/2}[\widehat{\pi}_0^M(s) - \pi_0(s)]$ *and* $n^{1/2}[\widehat{\pi}_0^M(t) - \pi_0(t)]$ *given by* $\sigma_M(s,t) = E[\xi_{1M}(s)\xi_{1M}(t)]$*, where*

$$\xi_{iM}(t) = \frac{f_i^{\pi_1}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,M) - f_i^{\pi_2}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,M)}{s^{(0)\langle 1\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)}$$

and $f_i^{\pi_1}(t;\boldsymbol{\beta},\boldsymbol{\theta},M) = M^{-1}\sum_{m=1}^{M} I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta})A_i^0(t) - M^{-1}\sum_{m=1}^{M} I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta})exp[\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)]\pi_0(t)$, $f_i^{\pi_2}(t;\boldsymbol{\beta},\boldsymbol{\theta},M) = \boldsymbol{s}^{(1)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})^T\pi_0(t)\boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\theta},M)$, $\boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\theta},M) = [\boldsymbol{\Omega}(\boldsymbol{\beta})]^{-1}M^{-1}\sum_{m=1}^{M}\boldsymbol{u}_i^{\langle m\rangle}(\boldsymbol{\beta},\boldsymbol{\theta})$. *A consistent estimator of* $\sigma_M(s,t)$ *is given by its empirical counterparts* $\widehat{\sigma}_M(s,t) = n^{-1}\sum_{i=1}^{n}\widehat{\xi}_{iM}(s)\widehat{\xi}_{iM}(t)$, *where*

$$\widehat{\xi}_{iM}(t) = \frac{f_i^{\pi_1}(t;\widehat{\boldsymbol{\beta}}^M,\widehat{\boldsymbol{\theta}},M) - \widehat{f}_i^{\pi_2}(t;\widehat{\boldsymbol{\beta}}^M,\widehat{\boldsymbol{\theta}},M)}{M^{-1}\sum_{m=1}^{M} S^{(0)\langle m\rangle}(t;\widehat{\boldsymbol{\beta}}^M,\widehat{\boldsymbol{\theta}})}$$

where $\widehat{f}_i^{\pi_2}(t;\boldsymbol{\beta},\boldsymbol{\theta},M) = [M^{-1}\sum_{m=1}^{M}\boldsymbol{S}^{(1)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})^T]\widehat{\pi}_0^M(t)\widehat{\boldsymbol{f}}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\theta},M)$.

1.4 Simulation Studies

We evaluate the finite-sample performance of our method through simulation studies. For each setting, $n = 500$ subjects are generated. Elements of the covariate vector, $\boldsymbol{Z} = (Z_1, Z_2)^T$, follow either a Bernoulli(0.5) or a Uniform$(0,1)$ distribution. The target model is survival-out-of-hospital probability, $\pi(t) = \pi_0(t)\text{exp}\{\boldsymbol{\beta}_0^T\boldsymbol{Z}\}$, where $\pi_0(t) = 0.3 - 0.0025t$, for $t = 1, 2, ..., 100$. This model of the joint outcome can be generated through the hazard function for death time D, $\lambda^D(t)$, and $P\{H(t) = 0|D > t, \boldsymbol{Z}\}$. We set $\lambda^D(t) = \lambda^D = \lambda_0^D\text{exp}(\boldsymbol{\alpha}_0^T\boldsymbol{Z})$, where $\lambda_0^D = 0.012$ or 0.006 and $\boldsymbol{\alpha}_0 = (0.405, -0.405)^T$. In this case, the out-of hospital event given the subject is alive is sampled from the conditional probability $P\{H(t) = 0|D > t, \boldsymbol{Z}\} = \pi(t)\text{exp}(\lambda^D t)$. Censoring time C is generated from hazard function $\lambda^C(t) = \lambda^C = \lambda_0^C\text{exp}(\boldsymbol{\gamma}_0^T\boldsymbol{Z})$, where $\lambda_0^C = 0.008$, $\boldsymbol{\gamma}_0 = (-0.693, 0.693)^T$, such that C and the target event $I\{H(t) = 0, D > t\}$ are independent given covariates $\boldsymbol{Z}$.

In Table (1.1), we consider the scenario where censoring times are known for all subjects. For Table (1.2) and (1.3), censoring time C is unknown in cases where death time D is observed, i.e. $C > D$, and imputation method with $M = 1$ and 10 are evaluated. The baseline terminal event hazard function λ_0^D are considered at 0.012 and 0.006, which result in about 49% and 68% subjects are censored, respectively. Three different magnitudes of $\boldsymbol{\beta}_0$ are considered: 0.693, 0.405 and 0.

In each setting, the biases of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Pi}_0(50)$ are very small, indicating that our estimators are consistent. Moreover, empirical standard deviations (ESDs) are generally close to the average

Table 1.1: Simulations Results for Known Censoring Time based on $n = 500$ and 500 Replicates

		$\widehat{\boldsymbol{\beta}}$				
λ_0^D	$\boldsymbol{Z}$	True	BIAS	ASE	ESD	ECP
0.012	Binary	(0.693, -0.693)	(-0.005, 0.000)	(0.069, 0.069)	(0.069, 0.067)	(0.964, 0.956)
		(0.405, -0.405)	(0.001, 0.002)	(0.066, 0.066)	(0.065, 0.067)	(0.954, 0.956)
		(0.000, 0.000)	(-0.001, 0.000)	(0.062, 0.062)	(0.064, 0.062)	(0.944, 0.956)
	Uniform	(0.693, -0.693)	(-0.002, -0.010)	(0.114, 0.114)	(0.120, 0.115)	(0.942, 0.942)
		(0.405, -0.405)	(-0.014, 0.005)	(0.110, 0.110)	(0.104, 0.107)	(0.966, 0.944)
		(0.000, 0.000)	(0.003, 0.009)	(0.105, 0.105)	(0.100, 0.108)	(0.950, 0.944)
0.006	Binary	(0.693, -0.693)	(0.001, 0.000)	(0.048, 0.048)	(0.049, 0.050)	(0.944, 0.930)
		(0.405, -0.405)	(-0.001, 0.002)	(0.046, 0.046)	(0.046, 0.048)	(0.952, 0.932)
		(0.000, 0.000)	(0.001, 0.002)	(0.044, 0.044)	(0.045, 0.042)	(0.946, 0.968)
	Uniform	(0.693, -0.693)	(-0.005, 0.001)	(0.080, 0.080)	(0.084, 0.084)	(0.936, 0.938)
		(0.405, -0.405)	(-0.003, -0.003)	(0.077, 0.077)	(0.081, 0.080)	(0.946, 0.948)
		(0.000, 0.000)	(0.000, 0.001)	(0.075, 0.075)	(0.076, 0.078)	(0.956, 0.932)

		$\widehat{\Pi}_0(50)$				
λ_0^D	$\boldsymbol{Z}$	True	BIAS	ASE	ESD	ECP
0.012	Binary	11.812	0.041	0.594	0.561	0.962
			0.015	0.586	0.597	0.946
			0.010	0.584	0.587	0.932
	Uniform		0.075	0.957	1.021	0.942
			0.054	0.940	0.875	0.958
			-0.054	0.916	0.879	0.952
0.006	Binary		-0.023	0.435	0.436	0.958
			0.002	0.426	0.428	0.952
			-0.008	0.426	0.401	0.952
	Uniform		0.026	0.683	0.732	0.926
			0.045	0.670	0.697	0.950
			0.007	0.661	0.674	0.924

asymptotic standard errors (ASEs), showing that our proposed variance estimator appear to be applicable to finite samples. The empirical coverage probabilities (ECPs) are also around 0.95, implying the accuracy of large-sample confidence intervals.

1.5 Real Data Analysis

We applied the proposed methods to data from the Dialysis Outcomes and Practice Patterns Study (DOPPS) Phase 5. The DOPPS is a prospective, observational study designed to elucidate as-

Table 1.2: Simulations Results for Unknown Proportional Censoring with $M = 1$ Time based on $n = 500$ and 500 Replicates

		$\widehat{\boldsymbol{\beta}}^1$				
λ_0^D	$\boldsymbol{Z}$	True	BIAS	ASE	ESD	ECP
0.012	Binary	(0.693, -0.693)	(-0.005, 0.000)	(0.069, 0.069)	(0.070, 0.074)	(0.938, 0.932)
		(0.405, -0.405)	(-0.001, 0.001)	(0.066, 0.066)	(0.070, 0.065)	(0.944, 0.954)
		(0.000, 0.000)	(0.000, 0.003)	(0.062, 0.062)	(0.061, 0.061)	(0.956, 0.956)
	Uniform	(0.693, -0.693)	(0.001, -0.008)	(0.114, 0.114)	(0.114, 0.116)	(0.958, 0.942)
		(0.405, -0.405)	(-0.001, 0.002)	(0.110, 0.110)	(0.114, 0.100)	(0.946, 0.960)
		(0.000, 0.000)	(0.003, 0.001)	(0.106, 0.105)	(0.104, 0.111)	(0.956, 0.928)
0.006	Binary	(0.693, -0.693)	(0.001, 0.002)	(0.048, 0.048)	(0.047, 0.049)	(0.950, 0.942)
		(0.405, -0.405)	(-0.003, 0.002)	(0.046, 0.046)	(0.046, 0.045)	(0.942, 0.960)
		(0.000, 0.000)	(0.002, -0.002)	(0.044, 0.044)	(0.045, 0.046)	(0.950, 0.940)
	Uniform	(0.693, -0.693)	(0.001, 0.003)	(0.080, 0.080)	(0.078, 0.081)	(0.948, 0.948)
		(0.405, -0.405)	(-0.006, -0.006)	(0.078, 0.078)	(0.075, 0.076)	(0.958, 0.948)
		(0.000, 0.000)	(-0.001, -0.007)	(0.075, 0.075)	(0.074, 0.076)	(0.952, 0.946)

		$\widehat{\Pi}_0^1(50)$				
λ_0^D	$\boldsymbol{Z}$	True	BIAS	ASE	ESD	ECP
0.012	Binary	11.812	0.036	0.593	0.616	0.926
			0.010	0.585	0.597	0.936
			-0.005	0.585	0.585	0.946
	Uniform		0.041	0.958	1.001	0.934
			0.028	0.938	0.936	0.948
			-0.008	0.917	0.982	0.936
0.006	Binary		-0.021	0.432	0.418	0.950
			0.000	0.428	0.429	0.956
			-0.009	0.425	0.444	0.936
	Uniform		-0.007	0.679	0.671	0.960
			0.078	0.678	0.654	0.956
			0.060	0.662	0.658	0.950

Table 1.3: Simulations Results for Unknown Proportional Censoring with $M = 10$ Time based on $n = 500$ and 500 Replicates

		$\widehat{\boldsymbol{\beta}}^{10}$				
λ_0^D	$\boldsymbol{Z}$	True	BIAS	ASE	ESD	ECP
0.012	Binary	(0.693, -0.693)	(0.003, 0.000)	(0.068, 0.068)	(0.067, 0.068)	(0.952, 0.954)
		(0.405, -0.405)	(0.004, -0.001)	(0.065, 0.064)	(0.062, 0.064)	(0.952, 0.952)
		(0.000, 0.000)	(0.004, -0.001)	(0.060, 0.060)	(0.059, 0.062)	(0.948, 0.954)
	Uniform	(0.693, -0.693)	(-0.002, 0.008)	(0.111, 0.111)	(0.114, 0.107)	(0.936, 0.954)
		(0.405, -0.405)	(-0.001, 0.007)	(0.107, 0.107)	(0.112, 0.103)	(0.934, 0.960)
		(0.000, 0.000)	(-0.003, 0.007)	(0.103, 0.102)	(0.107, 0.098)	(0.936, 0.968)
0.006	Binary	(0.693, -0.693)	(-0.002, 0.003)	(0.047, 0.047)	(0.047, 0.046)	(0.948, 0.950)
		(0.405, -0.405)	(0.000, 0.001)	(0.045, 0.045)	(0.046, 0.043)	(0.954, 0.952)
		(0.000, 0.000)	(0.000, 0.001)	(0.043, 0.043)	(0.043, 0.042)	(0.952, 0.960)
	Uniform	(0.693, -0.693)	(-0.001, 0.000)	(0.078, 0.078)	(0.079, 0.080)	(0.950, 0.934)
		(0.405, -0.405)	(-0.001, 0.000)	(0.075, 0.075)	(0.076, 0.077)	(0.954, 0.936)
		(0.000, 0.000)	(-0.001, 0.001)	(0.073, 0.073)	(0.073, 0.076)	(0.958, 0.944)

		$\widehat{\Pi}_0^{10}(50)$				
λ_0^D	$\boldsymbol{Z}$	True	BIAS	ASE	ESD	ECP
0.012	Binary	11.812	-0.019	0.575	0.583	0.932
			-0.017	0.569	0.574	0.940
			-0.015	0.566	0.571	0.946
	Uniform		-0.042	0.923	0.900	0.952
			-0.039	0.906	0.880	0.952
			-0.031	0.889	0.870	0.944
0.006	Binary		-0.013	0.421	0.411	0.954
			-0.002	0.417	0.408	0.954
			-0.001	0.415	0.407	0.954
	Uniform		0.015	0.662	0.669	0.952
			0.015	0.654	0.660	0.952
			0.013	0.645	0.661	0.946

pects of hemodialysis practice that are associated with the best outcomes for hemodialysis patients [Young et al., 2000]. In particular, Phase 5 data were collected between 2012 and 2015. Our research interests include identifying demographic and clinical variables that are associated with survival-out-of-hospital probability, and characterizing the underlying survival-out-of-hospital process.

The study population for DOPPS is of prevalent patients. In the interests of having t reflect time-since-dialysis-initiation instead of time-since-DOPPS-entry, we restricted our study sample to include the $n = 5,298$ patients who entered DOPPS within 3 months of initiating dialysis. Patients included in our analysis were from 470 hemodialysis units across 11 different countries, with the counties including: Belgium, Canada, China, Gulf Coast Consortium, Germany, Italy, Japan, Spain, Sweden, the United Kingdom and the U.S.. Covariates include age, race, gender, height, time on dialysis at study entry, as well as the following list of comorbid conditions: coronary artery disease (CAD), cancer, cardiovascular disease (CVD), stroke, congestive heart failure (CHF), diabetes, hypertension, chronic obstructive pulmonary disease (COPD), psychiatric disorder and peripheral vascular disease (PVD).

Since hospitalization and death times are recorded in days, t represents day (i.e, day post DOPPS entry) in our analysis. The mean number of hospital admissions was 0.595 per patient, while the median length of stay per visit was 5 days. Observed follow-up time had a median of 326 days. Censoring time cannot realistically be considered to be fixed, since patients are frequently lost to follow-up for reasons other than death or closure of the DOPPS database. To recover missing censoring time, we created $M = 10$ imputed data sets as described in Section 1.2.2.

At some early time points (i.e., $t < 30$ days), $\widehat{\pi}_0(t) > 1$ due to hospitalization and mortality rates being low. We therefore set such probabilities to 1. As shown in Table (1.4), coronary artery disease ($p = 0.042$) and cancer ($p = 0.043$) have significant negative effects on survival-out-of-hospital. Patients from Italy have significantly lower survival-out-of-hospital probability ($p = 0.046$) than patients from the U.S. (reference). For continuous variables, age had a significant negative effect ($p = 0.006$), while height had significant positive effect ($p = 0.031$). The estimated

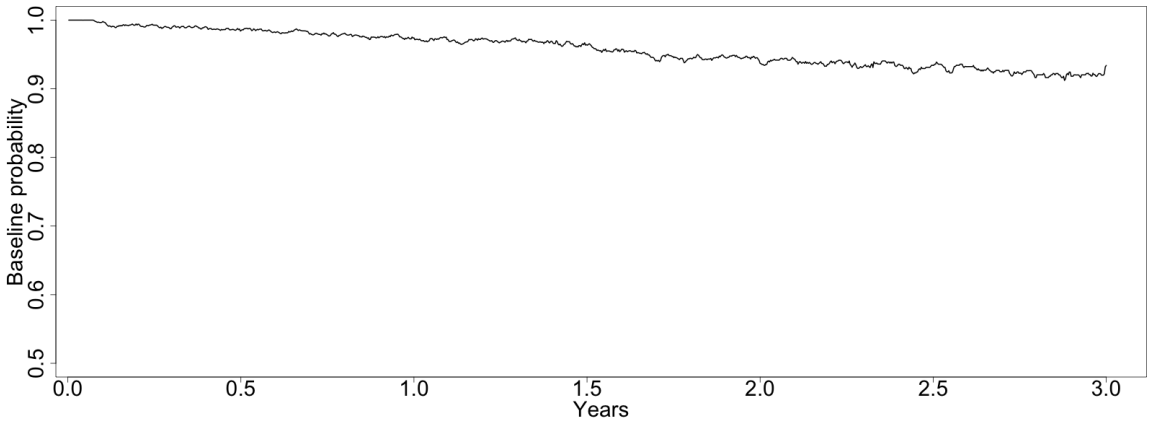

Figure 1.1: Fitted Baseline Survival-Out-of-Hospital Probability

baseline survival-out-of-hospital probability $\widehat{\pi}_0^{10}(t)$ is shown in Figure (1.1). The curve generally decreases as follow-up time increases, although the decrease is not monotone.

1.6 Concluding Remarks

In this report, we propose semiparametric methods for analyzing the probability of survival-out-of-hospital, a novel end-point pertinent to a frequently arising instance of the recurrent/terminal event data structure. Estimation proceeds through estimating equations which are analogous to those employed in Cox regression. Multiple imputation is implemented to accommodate missing censoring times that are unobserved due to the subject dying. Asymptotic properties of estimators are derived, and simulations studies show the proposed methods have satisfactory finite sample performance.

We applied the methods to data from Phase 5 of the Dialysis Outcomes and Practice Patterns Study in order to identify significant predictors of survival-out-of-hospital. Coronary artery disease, cancer and psychiatric disorder are found to be comorbidity factors with significant negative effects on survival-out-of-hospital events for a typical DOPPS patient in the U.S.. Patients at Italy have significantly lower survival-out-of-hospital probability than patients in the U.S.. Moreover, increasing age had a significantly negative effect, while height had significantly positive effect.

A key advantage of the proposed methods is that the baseline probability process does not

need to be specified. This is important, since this temporal process will often represent a nuisance parameter. Modeling the baseline parametrically could be tedious to carry out accurately, and could lead to bias in the regression parameter (of chief interest) if an incorrect parametric form is assumed. Note that, although the proposed baseline estimator shares a structure similar to that of the Breslow (1972) estimator, there are some important differences. First, our baseline estimator is not monotone, consistent with the temporal process it is targeting. Second, the estimator can be computed for any value of t, since the baseline probability reflects prevalence, as opposed to an intensity or occurrence rate.

An alternative estimator of baseline probability $\pi_0(t)$ is given by the empirical estimator $\pi_0^*(t) = \sum_{i=1}^n I\{H_i(t) = 0, D_i > t, C_i \geq t\} I(\boldsymbol{Z}_i = 0) / \sum_{i=1}^n I(C_i \geq t, \boldsymbol{Z}_i = 0)$ if censoring time C_i is known or fixed. However, $\pi_0^*(t)$ would generally be less efficient. Moreover, the estimator would be impossible to compute when few or none of the subjects have covariates at the baseline level, which would generally be the case in the presence of continuous predictors.

The multiple imputation method we use for unobserved censoring times represents so-called improper imputation, in the sense that the imputation model parameters are fixed at their estimated values. For an alternative strategy, proper imputation could also be considered, where the imputation model parameters are drawn from their estimated predictive posterior distribution [Little and Rubin, 2002]. In such cases, more randomness are introduced to the model since imputation parameters are drawn instead of being fixed. Both imputation methods would lead to consistent estimators of covariate effects but, potentially, with slightly different efficiency [Schaubel and Cai, 2006]. The proposed methods are valid under single imputation, $M = 1$, which would not be the case under proper imputation. Note that the deterministic variance structures of $\widehat{\boldsymbol{\beta}}^1$ and $\widehat{\pi}_0^1$ for single imputation are generally larger than that for multiple imputations $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{\pi}_0^M$, i.e., $\Sigma_1'(\boldsymbol{\beta}_0) > \Sigma_M'(\boldsymbol{\beta}_0)$ and $\sigma_1'(t,t) > \sigma_M'(t,t)$, for $M > 1$, which can be proved through the Cauchy-Schwartz inequality. An intuitive argument is that $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{\pi}_0^M$ are the means of M imputed estimators, and are generally more efficient than single imputed estimators.

We proposed the Cox (1972) model for imputing censoring times, due to its flexibility and dom-

inance in the analysis of censored epidemiologic data. Parameters estimated through the proposed methods are consistent only provided the censoring model is correctly specified. Assumptions on the censoring model could be assessed through standard techniques, such as Schoenfeld and Martingale residuals (Kalbfleisch and Prentice [2002]).

A potential problem of our method is that $\widehat{\pi}_0(t)$ could be larger than one, for example when hospitalization and mortality rates are low as we encounter in the real data analysis. In such cases, it may be useful to model $P\{H_i(t) = 1 \text{ or } D_i \leq t | \boldsymbol{Z}_i(t)\} = 1 - \pi_i(t)$. Future research could involve the use of a different link function, such that $\pi_i(t) = \pi_0(t) g^{-1}\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)\}$, where $g(\cdot)$ is a continuous function. Note that our method is a special case when $g^{-1}(\cdot)$ is the exponential function. In particular, choosing $g^{-1}(\cdot)$ as a function mapping from real line to $[1, \infty)$, for example $\exp(\cdot) + 1$, results in $\widehat{\pi}_0(t)$ having the desirable property of being bounded by 0 and 1.

Table 1.4: Analysis of DOPPS Data: Covariate Effects on the Survival-Out-of-Hospital (based on $M = 10$ Imputations)

Covariate	$\widehat{\beta}$	$\widehat{SE}(\widehat{\beta})$	p	exp($\widehat{\beta}$)
Age (per 5 years)	-0.003	0.001	0.006*	0.997
Time on dialysis (years)	-0.023	0.037	0.549	0.977
Height (per 10 cm)	0.009	0.004	0.031*	1.009
Female	0.009	0.008	0.339	1.009
CAD	-0.027	0.012	0.042*	0.973
Cancer	-0.034	0.015	0.043*	0.966
CVD	-0.030	0.016	0.117	0.971
Stroke	-0.012	0.018	0.534	0.989
CHF	-0.005	0.009	0.547	0.995
Diabetes	0.007	0.006	0.273	1.007
Hypertension	0.001	0.006	0.625	1.001
COPD	-0.038	0.021	0.089	0.962
Psychiatric Disorder	-0.030	0.017	0.108	0.970
PVD	-0.008	0.011	0.422	0.992
Belgium	-0.012	0.023	0.609	0.988
Canada	-0.010	0.016	0.520	0.990
China	-0.007	0.017	0.678	0.993
Gulf	-0.016	0.016	0.388	0.984
Germany	-0.021	0.016	0.279	0.980
Italy	-0.046	0.022	0.046*	0.955
Japan	-0.000	0.012	0.807	1.000
Spain	0.003	0.022	0.684	1.003
Sweden	-0.008	0.020	0.714	0.992
UK	0.004	0.014	0.498	1.004

a: Coronary artery disease
b: Cardiovascular disease
c: Congestive heart failure
d: Chronic obstructive pulmonary disease
e: Peripheral vascular disease

CHAPTER 2

Semiparametric Regression Methods for Temporal Processes subject to Dependent Censoring

2.1 Introduction

In biomedical applications, the response of interest can often be cast as an binary indicator process indexed by time. We consider the setting wherein the indicator at time t takes the value 1 (denoting 'success' in some form) when the patient is alive and in a particular state, and 0 otherwise. Examples include the following: (i) In a study of leukaemia patients, the response could be coded as 1 if the patient is alive and in remission t days following diagnosis, and 0 otherwise. (ii) In a study of morbidity among end-stage renal disease patients, the response at time t equals 1 if the patient is alive-and-not-hospitalized at time t, and 0 otherwise. (iii) As another example from the organ failure setting, an end-stage liver disease (ESLD) patient could be coded as having response 1 if active on the liver transplant waiting list (at time t days after initial wait list registration), and 0 otherwise. When covariate effect are of chief interest, temporal process regression is a natural way to cast the afore-described data structure. Although several methods amenable to this data structure have been developed in the last 10-15 years (arguably beginning with the work of Fine et al. [2004]), relatively few modeling choices are available relative to the frequency with which this data structure arises in practice. In this report, we develop semiparametric process regression methods which can be used to model settings such as (i), (ii) and (iii) above, in a flexible manner and making fewer assumptions regarding the censoring process.

Formalizing the above-described data structure, suppose that (for a hypothetical subject) D represents time of death, and $\mathcal{E}(t)$ is an indicator taking the value 1 if the subject is in the state of interest at follow-up time t (and 0 if not). We can define $A(t) = \mathcal{E}(t)I(D > t)$, such that the survival time in the state of interest, D^A, can be written $D^A = \int_{[0,\infty)} A(t)dt = \int_{[0,D]} \mathcal{E}(t)dt$. This can be considered a special case of quality adjusted survival time [Gelber et al., 1989, Glasziou et al., 1990], where 'quality' is defined as yes versus no (1 vs. 0). Nonparametric methods [Zhao and Tsiatis, 1997, Huang and Louis, 1999, Murray and Cole, 2000] and regression methods [Cole et al., 1993, Laan and Hubbard, 1999, Andrei and Murray, 2007, Zhao and Wang, 2008] have been proposed to estimate this quality of life measure. In medical cost analysis, if $A(t)$ denotes the cost at time t, then D^A corresponds to the total cost for a patient from the start of treatment to death without censoring. Lin et al. [1997], Bang and Tsiatis [2000], Willan et al. [2002] focus on the estimation of mean cost, $E(D^A)$. In this report, we do not wish to add to the rich body of techniques for analyzing more general versions of D^A. Instead, we develop methods for the related but distinct goal of analyzing the process, $E[A(t)]$.

The methods we propose are motivated by the end-stage liver disease (ESLD) setting. The preferred therapy for ESLD is deceased-donor liver transplantation. However, due to a shortage of donor livers, medically suitable patients are placed on a waiting list. A wait-listed patient is eligible to receive a transplant only when 'active'; patients may be deactivated for several reasons, most of which are related to a decline in health status which renders the patient at least temporarily unsuitable for transplantation. Hence, keeping the patient active on the wait list represents a successful outcome, in the sense that the patient not only continues to survive but also remains eligible for the preferred treatment. A patient's active process may be censored by liver transplantation, with such censoring representing dependent censoring due to mutual correlation between $A(t)$ and liver transplantation. Note that, due to the nature of the liver allocation system in the U.S., a patient's rank on the wait list is determined by their Model for End-stage Liver Disease (MELD) score. In particular, the wait list is sequenced in decreasing order of (current) MELD score. Since higher MELD scores correspond to higher pre-transplant mortality, a model of pre-transplant outcomes

based on baseline (time 0) patient characteristics will generally be subject to dependent censoring.

The response we consider could be framed as a temporal process, $A(t)$, where t is continuous. In contrast to a counting process, $A(t)$ need not be a non-decreasing function. In the context of our afore-described motivating example, we let $A(t)$ be the indicator of being both alive and active on the transplant waiting list at time t. In several existing methods developed for temporal process regression, the expectation of $A(t)$ is linked to linear components through a continuous link function; for example, Fine et al. [2004], Yan and Fine [2005], Yan and Huang [2009]. This could be viewed as a generalized linear model indexed by time. The regression coefficients $\boldsymbol{\beta}(t)$ could be solved at observed jump points. In our work, we consider a semiparametric model for $E[A(t)]$, where covariates have multiplicative effects on a completely unspecified probability function indexed by time.

In this manuscript, we develop semiparametric regression methods for a temporal process subject to dependent censoring. Two types of censoring are considered. Specifically, we let C_1 denote censoring which is independent conditional on external covariates. Dependent censoring, denoted by C_2, is correlated with the process of interest even given covariates introduced in the process regression model. To avoid bias due to dependent censoring, we derive a variant of Inverse Probability of Censoring Weighting (IPCW; Robins and Rotnitzky [1992]) based on a semiparmametric additive hazard model [Lin and Ying, 1994]. We also derive a stabilized version of the proposed inverse weights [Hernán et al., 2000, Robins and Finkelstein, 2000, Zhang and Schaubel, 2011] to simplify calculations and, hence, considerably reduce computing time in large data sets. Analogous to a weighted partial likelihood score equation [Cox, 1972b, Sasieni, 1993], the regression estimator could be estimated by the solution to an estimating equation free of the baseline probability.

Our methods have several novel features. First, the baseline probability function is represented in the model nonparametrically. This is a potentially big advantage, since covariate effects typically take center stage in process regression (and other regression settings), with little interest in modeling the baseline probability. We essentially profile out the baseline probability function, which

results in major computation reduction relative to a fully parametric probability model. Second, the response indicator we consider is the joint event of survival and state occupation. The limited number of process regression methods which considered a terminating event typically modeled the state indicator conditional on survival. Notwithstanding the utility of such approaches, it is useful to develop methods for the joint outcome of survival and state occupation (a response for which few methods have been developed). Third, existing process regression methods typically assume independent censoring, while the proposed methods allow for both independent and dependent censoring. Fourth, in contrast with the vast majority of methods which accommodate dependent censoring, we construct the inverse weight under an additive hazards model.

The remainder of the article is organized as follows. We set up notation and describe our proposed methods in the next section. In Section 3, we derive the asymptotic properties of the regression parameter estimator and baseline probability function estimator, with proofs provided in the Supplemental Materials document. Simulation studies are performed to evaluate our method in finite samples in Section 4 under various scenarios. In Section 5, we apply the proposed methods to national ESLD data. Finally, concluding remarks are given in Section 6.

2.2 Model and Methods

We begin by formalizing the data structure described in Section 1. We then describe the proposed inference methods.

2.2.1 Notation and Assumed Models

Suppose that there are n independent subjects ($i = 1, 2, \ldots, n$). Let D_i be the death (terminal event) time of subject i, and let $\mathcal{E}_i(t)$ be a indicator function taking value 1 when subject i is occupying the state of interest at follow-up time t. The outcome of interest is the joint event, being alive and occupying the state of interest, which we denote by $A_i(t) = \mathcal{E}_i(t)I(D_i > t)$. In the end-stage liver disease example, D_i represents death (in the absence of liver transplantation), while

$\mathcal{E}_i(t) = 1$ if subject i is active on the liver transplant waiting list as of t days following initial wait list registration, and 0 otherwise. We let $\boldsymbol{Z}_i(t)$ be a p-dimensional covariate vector, with any time-dependent elements being external [Kalbfleisch and Prentice, 2002]. The probability of interest is the probability that a subject i is alive and occupying the state of interest at time t,

$$\pi_i(t) = P[A_i(t) = 1 | \boldsymbol{Z}_i(t)]. \tag{2.1}$$

We assume that the covariate $\boldsymbol{Z}_i(t)$ has a multiplicative effect on a completely unspecified baseline probability function, $\pi_0(t)$, such that

$$\pi_i(t) = \pi_0(t)\text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z_i}(t)], \tag{2.2}$$

where $\boldsymbol{\beta}_0$ is the p-dimensional parameter vector of chief interest. Model (2.2) is reminiscent of the Cox proportional hazards model. However, there are some important differences, including the fact that $\pi_i(t)$ is interpreted as a marginal probability, rather than a conditional probability rate, and that $\pi_i(t)$ need not be monotone.

Two types of censoring are considered. Let C_{1i} be the administrative censoring, which is assumed to be independent of $A_i(t)$ given $\boldsymbol{Z}_i(t)$; i.e.,

$$E[A_i(t) | \boldsymbol{Z}_i(t), C_{1i}, C_{1i} \geq t] = E[A_i(t) | \boldsymbol{Z}_i(t)]. \tag{2.3}$$

This is also known as covariate-dependent censoring, in the sense that C_{1i} is allowed to depend on the covariate employed in the model of interest. We let C_{2i} represent dependent censoring time; that is, C_{2i} is not assumed to be conditionally independent $A_i(t)$ given $\boldsymbol{Z}_i(t)$. For example, in the context of our motivating example, a patient's pre-transplant $A_i(t)$ process is censored if and when the patient receives a liver transplant; i.e., the liver transplant hazard and mortality hazard may be correlated, even conditional on $\boldsymbol{Z}_i(t)$. We let $C_i = C_{1i} \wedge C_{2i}$ represent the censoring time, where $a \wedge b = \min(a, b)$. Here we consider follow-up time $t \in [0, \tau]$, where τ is a pre-specified constant

satisfying $Pr(C_i \geq \tau) > 0$ for $i = 1, 2, ..., n$. In practice, τ could be chosen as the maximum observed censoring time. To further characterize C_{2i}, we let $\boldsymbol{X}_i^{\dagger}(t)$ represent the time-dependent covariate at time t. Note that $\boldsymbol{X}_i^{\dagger}(t)$ would typically contain the elements of $\boldsymbol{Z}_i(t)$, as well as additional factors (the most important being internal time-varying covariates assumed to predict both D_i and C_{2i}). We denote the covariate history as of time t by $\widetilde{\boldsymbol{X}}_i(t) = \{\boldsymbol{X}_i^{\dagger}(s), s \in [0, t)\}$. Finally, we let $\lambda_i^{C_2}(t)$ be the cause specific hazard function of C_{2i} which is defined as

$$\lambda_i^{C_2}(t) = \lim_{\delta \to 0} \frac{1}{\delta} Pr[t \leq C_{2i} < t + \delta | C_{2i} \geq t, D_i \geq t, \widetilde{\boldsymbol{X}}_i(t)]. \tag{2.4}$$

We assume that, conditional on $\widetilde{\boldsymbol{X}}_i(t)$, the cause-specific hazard of C_{2i} at time t does not further depend on the possibly unobserved, D_i or $\mathcal{E}_i(u), s \in (t, \tau]$, i.e.,

$$\lambda_i^{C}\{t|\widetilde{\boldsymbol{X}}_i(t)\} = \lambda_i^{C}\{t|\widetilde{\boldsymbol{X}}_i(t), C_{1i}, C_{1i} \geq t, D_i, D_i \geq t, \mathcal{E}_i(u), u \in (t, \tau]\}. \tag{2.5}$$

This represents the critical ‘no unmeasured confounders‘ for censoring assumption [Robins, 1993, Robins and Finkelstein, 2000] in our context. The following semiparametric additive hazards model [Lin and Ying, 1994] is assumed for dependent censoring C_{2i},

$$\lambda_i^{C_2}(t; \boldsymbol{\theta}_0) = \lambda_0^{C_2}(t) + \boldsymbol{\theta}_0^T \boldsymbol{X}_i(t), \tag{2.6}$$

where $\lambda_0^{C_2}(t)$ is the baseline hazard function for C_{2i} and the covariate $\boldsymbol{X}_i(t)$ is chosen (e.g., through model selection techniques) to satisfy $\lambda_i^{C_2}[t|\boldsymbol{X}_i(t)] = \lambda_i^{C_2}[t|\widetilde{\boldsymbol{X}}_i(t)]$. Note that $\boldsymbol{X}_i(t)$ need not be based on the covariate status at time t and could, in fact, contain elements representing the covariate history. Finally, we define $\Lambda_i^{C_2}(t) = \int_0^t \lambda_i^{C_2}(s)ds$ as the cumulative hazard function corresponding to C_{2i}.

The additive hazards model stated in (2.6) facilitates the calculation of the weight function, since the baseline cumulative hazard function can be canceled out after a particular stabilizing factor introduced. Detail is provided later in our report on this matter.

In the next subsection, we describe the proposed methods for the scenario where independent censoring time C_1 is known. We then subsequently describe the proposed techniques for the more frequently occurring set-up when C_1 is not known.

2.2.2 Known Censoring Time

We first consider the case where the independent censoring time, C_{1i}, is known for all subjects. In such cases, C_{1i} is known even if D_i occurs first. This set-up would apply, for example, in a clinical trial with staggered entry but no drop-out or random loss to follow-up. This does not match most observational data, but it is a useful starting point in terms of outlining the proposed estimation techniques. Consider the following two estimating equations,

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i(t)[A_i(t) - \pi_i(t)]I(C_i \geq t)dt \tag{2.7}$$

$$\sum_{i=1}^{n} [A_i(t) - \pi_i(t)]I(C_i \geq t). \tag{2.8}$$

These two estimating equations do not have expectation zero under model assumption (2.2) and conditionally independent censoring assumption (2.3) of C_{1i}, since dependent censoring C_{2i} is potentially correlated with $A_i(t)$, even conditional on $\boldsymbol{Z}_i(t)$. To handle this issue, we utilize Inverse Probability of Censoring Weighting (IPCW) [Robins and Rotnitzky, 1992] to accommodate dependent censoring. Define

$$W_i^A(t; \boldsymbol{\theta}_0) = I(C_{2i} \geq t)\text{exp}\{\Lambda_i^{C_2}[t \wedge D_i; \boldsymbol{\theta}_0]\} \tag{2.9}$$

as, heuristically, the inverse probability of being uncensored by C_{2i} as of time t. Note that, in the data structure we consider, dependent censoring cannot occur after death. This make sense intuitively, from the perspective that C_{2i} is driven by internal factors (including a subject's survival). For instance, in our motivating example, a patient cannot receive a liver transplant after dying.

Now, consider the revised estimating equations,

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i(t)[A_i(t) - \pi_i(t)]I(C_{1i} \geq t)W_i^A(t;\boldsymbol{\theta}_0)dt \tag{2.10}$$

$$\sum_{i=1}^{n} [A_i(t) - \pi_i(t)]I(C_{1i} \geq t)W_i^A(t;\boldsymbol{\theta}_0). \tag{2.11}$$

Under the no-unmeasured-confounders assumption given in (2.5), these two weighted estimating equations have expectation zero. Basically, the proof follows from the fact that $W_i^A(t;\boldsymbol{\theta}_0)$ can be written as one minus a Martingale component of C_{2i}, which is independent of $[A_i(t) - \pi_i(t)]I(C_{1i} \geq t)$ conditional on $\boldsymbol{X}_i(t)$ [Robins and Finkelstein, 2000]. Detailed proof are provided in the Section 4.6 of Supplemental Materials.

In contrast to the majority of the existing literature, we derive a stabilizer that is merely a function of t, which is valid since $E\{g(t)[A_i(t) - \pi_i(t)]I(C_{1i} \geq t)W_i^A(t;\boldsymbol{\theta}_0)|\boldsymbol{Z}_i(t)\} = 0$ will hold. We denote $g(t) = \exp[-\int_0^{t\wedge D_i} d\Lambda_0^{C_2}(s;\boldsymbol{\theta}_0)]$ for this purpose, such that

$$W_i^B(t;\boldsymbol{\theta}_0) = I(C_{2i} \geq t)\exp\left\{\Lambda_i^{C_2}[t \wedge D_i;\boldsymbol{\theta}_0] - \int_0^{t\wedge D_i} d\Lambda_0^{C_2}(s;\boldsymbol{\theta}_0)\right\}. \tag{2.12}$$

Under the assumed additive hazard model from (2.6), $W_i^B(t;\boldsymbol{\theta}_0) = \exp[\int_0^{t\wedge D_i} \boldsymbol{\theta}_0^T \boldsymbol{X}_i(s)ds]$ since $\Lambda_0^{C_2}(t)$ cancels out. This nice property enables us to get unbiased estimating equations without estimating the baseline cumulative hazard $\Lambda_0^{C_2}(t)$ and, hence, should increase computational efficiency.

Solving (2.11) for $\pi_0(t)$ by treating $\boldsymbol{\beta}$ as known, then substituting the estimated $\pi_0(t)$ into (2.10) gives us the following estimating equation,

$$\sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}, W(\boldsymbol{\theta}_0)]\}I(C_{1i} \geq t)W_i(t;\boldsymbol{\theta}_0)dt = 0, \tag{2.13}$$

where $\bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}, W(\boldsymbol{\theta})] = \boldsymbol{Z}^{(1)}[t;\boldsymbol{\beta}, W(\boldsymbol{\theta})]/Z^{(0)}[t;\boldsymbol{\beta}, W(\boldsymbol{\theta})]$, $\boldsymbol{Z}^{(k)}[t;\boldsymbol{\beta}, W(\boldsymbol{\theta})] = n^{-1}\sum_{i=1}^{n} \boldsymbol{Z}_i(t)^{\otimes k} I(C_{1i} \geq t)W_i(t;\boldsymbol{\theta})\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}$, for $k = 0, 1, 2$, where $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$, and with

$W_i(t;\boldsymbol{\theta})$ set to either $W_i^A(t;\boldsymbol{\theta})$ or $W_i^B(t;\boldsymbol{\theta})$, depending on which is preferred for the application at hand. Having estimated $\boldsymbol{\beta}_0$ through the root of equation (2.13), denoted by $\widehat{\boldsymbol{\beta}}$, we could estimate $\pi_0(t)$ by solving (2.11),

$$\hat{\pi}_0(t) = \frac{\sum_{i=1}^n A_i(t) I(C_{1i} \geq t) W_i(t;\boldsymbol{\theta}_0)}{\sum_{i=1}^n I(C_{1i} \geq t) W_i(t;\boldsymbol{\theta}_0) \exp[\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)]}. \tag{2.14}$$

Furthermore, one could also define the integral of baseline probability function up to time L,

$$P_0(L) = \int_0^L \pi_0(t) dt, \tag{2.15}$$

which could be interpreted as the restricted mean time survived in the state of interest for subject with covariate equal to the reference level for all elements. The quantity $P_0(L)$ is estimated by $\widehat{P}_0(L) = \int_0^L \widehat{\pi}_0(t) dt$.

Based on model (2.6), one could estimate $\boldsymbol{\theta}_0$ and $\Lambda_i^{C_2}(t;\boldsymbol{\theta}_0)$ by $\widehat{\boldsymbol{\theta}}$ and $\widehat{\Lambda}_i^{C_2}(t;\widehat{\boldsymbol{\theta}})$. From the works of Lin and Ying [1994], $\widehat{\boldsymbol{\theta}}$ and $\widehat{\Lambda}_i^{C_2}(t;\widehat{\boldsymbol{\theta}})$ are given by

$$\widehat{\boldsymbol{\theta}} = \frac{\sum_{i=1}^n \int_0^\infty Y_i(t)\{\boldsymbol{X}_i(t) - \bar{\boldsymbol{X}}(t)\} dN_i^{C_2}(t)}{\sum_{i=1}^n \int_0^\infty Y_i(t)\{\boldsymbol{X}_i(t) - \bar{\boldsymbol{X}}(t)\}^{\otimes 2} dt} \tag{2.16}$$

$$d\widehat{\Lambda}_0^{C_2}(t;\widehat{\boldsymbol{\theta}}) = \frac{\sum_{i=1}^n \{dN_i^{C_2}(t) - Y_i(t)\widehat{\boldsymbol{\theta}}^T \boldsymbol{X}_i(t) dt\}}{\sum_{i=1}^n Y_i(t)}, \tag{2.17}$$

where $N_i^{C_2}(t) = I(C_{2i} \leq t \wedge X_i)$, $dN_i^{C_2}(t) = N_i^{C_2}(t^- + dt) - N_i^{C_2}(t^-)$, $Y_i(t) = I(X_i \geq t)$ and $X_i = D_i \wedge C_i$. Let $\boldsymbol{X}^{(k)}(t) = n^{-1} \sum_{i=1}^n \boldsymbol{X}_i(t)^{\otimes k} Y_i(t)$, $\bar{\boldsymbol{X}}(t) = \boldsymbol{X}^{(1)}(t)/X^{(0)}(t)$. After that, the estimated weights $\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})$ or $\widehat{W}_i^B(t;\widehat{\boldsymbol{\theta}})$ can be calculated.

2.2.3 Random Censoring Time

Next we consider a more realistic scenario where independent censoring time is random, with the randomness implying that C_{1i} is unknown when D_i occurs first. Setting the missing censoring time to either D_i or the maximum follow up time, τ, would introduce bias, since the indicator

$I(C_{1i} > D_i)$ is correlated with the target process. As for C_{2i}, if D_i happens first, then C_{2i} could be treated as infinity or considered to be subject to a dependent censoring hazard of 0 for $t > D_i$. The reason is that C_{2i} relies on time varying covariate vector $\boldsymbol{X}_i$ containing internal covariates, which would shut down if death occurs. In this case, the hazard for C_{2i} is zero after D_i. Moreover, the inverse weighting function $W_i(t;\boldsymbol{\theta}_0)$ remains constant after D_i if it is observed earlier than C_i.

Our solution is to impute missing C_{1i} from its assumed model [Rubin, 2004]. Specifically, we assume it follows the proportional hazards model

$$\lambda_i^{C_1}(t;\boldsymbol{\gamma}_0) = \lambda_0^{C_1}(t)\exp[\boldsymbol{\gamma}_0^T \boldsymbol{Z}_i(t)]. \tag{2.18}$$

For subjects with $C_i \leq D_i$, we set the imputed censoring time as the known censoring time. In the mth imputed dataset, for subjects with $C_i > D_i$, we impute $\widehat{C}_{1i}^{\langle m\rangle}$ from the estimated conditional survival function,

$$\widehat{G}(t;\widehat{\boldsymbol{\gamma}}) = I(t \geq D_i)\exp[-\widehat{\Lambda}_i^{C_1}(t;\widehat{\boldsymbol{\gamma}}) + \widehat{\Lambda}_i^{C_1}(D_i;\widehat{\boldsymbol{\gamma}})].$$

Standard partial likelihood [Cox, 1975] techniques can be fitted to the observed censoring time data $\{X_i, I(C_{1i} \leq D_i \wedge C_{2i}), \boldsymbol{Z}_i(t); \quad t \in [0,\tau]\}_{i=1}^n$ to compute $\widehat{\boldsymbol{\gamma}}$. The baseline cumulative hazard function for $\Lambda_0^{C_1}(t)$ is estimated through the method of Breslow [1972]. Then, we set $C_{1i}^{\langle m\rangle} = I(C_i \leq D_i)C_i + I(C_i > D_i)\widehat{C}_{1i}^{\langle m\rangle}$. In total we will create M imputation datasets. Within each imputed dataset m, we substitute $C_{1i}^{\langle m\rangle}$ for C_{1i} and set C_{2i} as τ if $D_i < C_{1i}$. Estimators arising from the mth imputed data set are denoted by $\widehat{\boldsymbol{\beta}}^{\langle m\rangle}$ and $\widehat{\pi}_0^{\langle m\rangle}(t)$. We then estimate $\boldsymbol{\beta}_0$ and $\pi_0(t)$ by averaging the M imputation-specific estimators,

$$\widehat{\boldsymbol{\beta}}^M = M^{-1}\sum_{m=1}^M \widehat{\boldsymbol{\beta}}^{\langle m\rangle} \tag{2.19}$$

$$\widehat{\pi}_0^M(t) = \frac{\sum_{m=1}^M \sum_{i=1}^n A_i(t) I(C_{1i}^{\langle m\rangle} \geq t) W_i(t;\boldsymbol{\theta}_0)}{\sum_{m=1}^M \sum_{i=1}^n I(C_{1i}^{\langle m\rangle} \geq t) W_i(t;\boldsymbol{\theta}_0)\exp[\boldsymbol{Z}_i(t)^T \widehat{\boldsymbol{\beta}}^M]}. \tag{2.20}$$

Note that the multiple imputation method we employ does not sample the parameters assumed to underly the C_1 distribution but, instead, imputes $C_{1i}^{\langle m\rangle}$ values from the same estimated survival

curve. This procedure has been referred to as Improper Imputation [Wang and Robins, 1998]. As a consequence of this choice, the well-established variance formula for multiple imputation [Rubin, 2004] does not apply, necessitating an explicit derivation of variances estimators corresponding to (2.19) and (2.20). These issues are dealt with in the next section, along with our treatment of the large-sample properties of the proposed estimators.

Some commentary regarding our combination of inverse weighting and imputation is useful at this juncture. The sources of censoring, C_1 and C_2, have very different implications in terms of their impact on parameter estimation. Due to the marginal nature of $\pi_i(t)$, subjects contribute relevant follow-up until time C_{1i}, which may occur after D_i. A similar issue arises in Gray [1988] and Fine and Gray [1999] in the context of inference targeting the subdistribution hazard function in the competing risks setting. Both Gray [1988] and Fine and Gray [1999] used a weight function, rather then imputation. It is important to note that the weight was not an inverse weight; if anything, it could be described as an 'inverse-inverse' weight, since it corresponds to a conditional survival probability (as opposed to the inverse thereof). Essentially, the risk sets contributions are weighted with respect the conditional probability of remaining uncensored at time t (consider $t > D_i$), given that $C_{i1} > D_i$. From this perspective, imputing C_{1i} does in fact appear consistent in spirit with the weights use in subdistribution modeling, which can be cast heuristically as mean imputation at the risk set level. In fact, Ruan and Gray [2008] later proposed subdistribution hazard methods which involved imputing censoring times. An analogous imputation scheme was later employed by Schaubel and Zhang [2010].

Note that C_{1i} marks the end of relevant follow-up and, hence, is a variate we want to observe. Provided we either observe C_{1i} or can validly impute its value, the $[0, C_{1i}]$ experience could be analyzed without inverse weighting if not for dependent censoring, C_{2i}. In line with the setting where IPCW is typically employed, we inverse weight the uncensored experience to reflect data that would have been observed if C_{1i} were the only source of censoring. The events D_i and C_{2i} serve as competing risks in the sense of Prentice et al. [1978]. From this angle, dependent censoring does not occur after D_i, hence the weight function not incrementing for $t > D_i$, per (2.9).

2.3 Asymptotic Properties

We begin by considering the case where independent censoring time C_1 is known. The large sample properties are provided for the weight, $W_i^B(t;\boldsymbol{\theta}_0) = \exp[\int_0^{t\wedge D_i} \boldsymbol{\theta}_0^T \boldsymbol{X}_i(s)ds]$, with results for $W_i^A(t;\boldsymbol{\theta}_0)$ given in the Supplemental Materials. In addition to the regularity conditions for proportional hazards model of C_{1i} and additive hazards model of C_{2i} (each provided in the Section 3 of Supplemental Materials), we also assume the following set of regularity conditions:

(a) $\{A_i(t), X_i, \boldsymbol{Z}_i(t), \boldsymbol{X}_i(t), I(C_{1i} \leq D_i \wedge C_{2i}), I(C_{2i} \leq D_i \wedge C_{1i}); t \in [0, X_i]\}$ are independent and identically distributed for $i = 1, 2, \ldots, n$.

(b) $Pr(C_i \geq \tau) > 0$ for $i = 1, 2, ..., n$, where τ is a pre-specified constant.

(c) $|\boldsymbol{Z}_{ij}(0)| + \int_0^\tau |d\boldsymbol{Z}_{ij}(t)| < c_{\boldsymbol{Z}} < \infty$ almost surely for $i = 1, 2, ..., n$, $j = 1, 2, ..., p$; i.e., $\boldsymbol{Z}_i(t)$ has bounded total variation.

(d) $\boldsymbol{\Omega}(\boldsymbol{\beta}_0, W) = E[\int_0^\tau \{\boldsymbol{z}^{(2)}(t;\boldsymbol{\beta}_0, W)/z^{(0)}(t;\boldsymbol{\beta}_0, W) - \bar{\boldsymbol{z}}(t;\boldsymbol{\beta}_0, W)^{\otimes 2}\} I(C_1 \geq t) A_1(t)dt]$ is positive definite.

(e) For $\boldsymbol{\beta} \in \mathscr{B}_\delta$, where $\mathscr{B}_\delta$ is a small neighborhood around $\boldsymbol{\beta}_0$, $z^{(0)}(t;\boldsymbol{\beta}, W)$, and $\boldsymbol{z}^{(1)}(t;\boldsymbol{\beta}, W)$ are bounded away from zero.

(f) For $\boldsymbol{\beta} \in \mathscr{B}_\delta$, $k = 0, 1, 2$, $\boldsymbol{z}^{(k)}(t;\boldsymbol{\beta}, W)$ are continuous uniformly on $t \in [0, \tau]$, and are bounded on $[0, \tau] \times \mathscr{B}_\delta$ for $W^A(\boldsymbol{\theta}_0)$ and $W^B(\boldsymbol{\theta}_0)$.

We summarize the essential large sample properties of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\pi}_0(t)$ in the following theorems. Proofs are sketched in the Supplemental Materials.

Theorem 5. *Under assumptions (2.2), (2.3), and (2.5), $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$, and $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a mean-zero normal random variable with a variance-*

covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) = E[\boldsymbol{f}_1^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)^{\otimes 2}]$*, where*

$$\begin{aligned}
\boldsymbol{f}_i^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}, \boldsymbol{\beta}) &= \boldsymbol{\Omega}[\boldsymbol{\beta}, W^B(\boldsymbol{\theta})]^{-1}[\boldsymbol{\Phi}_{1i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}, \boldsymbol{\beta}) + \boldsymbol{\Phi}_{2i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}, \boldsymbol{\beta})] \\
\boldsymbol{\Phi}_{1i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}, \boldsymbol{\beta}) &= \int_0^\tau \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{z}}[t; \boldsymbol{\beta}, W^B(\boldsymbol{\theta})]\} W_i^B(t; \boldsymbol{\theta}) dM_i(t; \boldsymbol{\beta}) \\
\boldsymbol{\Phi}_{2i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}, \boldsymbol{\beta}) &= \boldsymbol{H}^B[\boldsymbol{\beta}, \boldsymbol{\theta}, W^B(\boldsymbol{\theta})][\boldsymbol{\Omega}^{C_2}]^{-1} \boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}).
\end{aligned}$$

Here $[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0)$ is the influence function for additive hazard model of C_{2i}, and is defined in the Supplemental Materials along with $dM_i(t; \boldsymbol{\beta})$ and $\boldsymbol{H}^B[\boldsymbol{\beta}, \boldsymbol{\theta}, W]$. The variance estimator $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}}) = n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{f}}_i^{\boldsymbol{\beta}_B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}})^{\otimes 2}$ could be obtained by substituting limiting values in $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$ by their corresponding empirical counterparts. However, the computation of $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}})$ are tedious due to the complexity of $\widehat{\boldsymbol{\Phi}}_{2i}^{\boldsymbol{\beta}_B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}})$. An attractive alternative is to treat estimated weights function $\widehat{W}_i^B(t; \boldsymbol{\theta})$ as fixed, and estimate $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$ by $n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\Phi}}_{1i}^{\boldsymbol{\beta}_B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\beta}})^{\otimes 2}$, where $\widehat{\boldsymbol{\Phi}}_{1i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \int_0^\tau \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t; \boldsymbol{\beta}, \widehat{W}^B(\boldsymbol{\theta})]\}\widehat{W}_i^B(t; \boldsymbol{\theta}) d\widehat{M}_i(t; \boldsymbol{\beta})$. In this case, the proposed variance estimator will be conservative in estimating the true variance [Hernán et al., 2000, Pan and Schaubel, 2008, Zhang and Schaubel, 2011].

We now describe the asymptotic properties of $\widehat{\pi}_0(t)$.

Theorem 6. *Under assumptions (2.2), (2.3), and (2.5),* $n^{1/2}(\widehat{\pi}_0 - \pi_0)$ *converges weakly to a mean-zero Gaussian process with a variance and covariance matrix between* $n^{1/2}[\widehat{\pi}_0(s) - \pi_0(s)]$ *and*

$n^{1/2}[\widehat{\pi}_0(t) - \pi_0(t)]$ *given by* $\sigma(s,t,\boldsymbol{\theta}_0,\boldsymbol{\beta}_0) = E[\xi_1(s,\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)\xi_1(t,\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)]$*, where*

$$\xi_i(t,\boldsymbol{\theta},\boldsymbol{\beta}) = \frac{\widetilde{f}_i^{\pi_1,B}(t,\boldsymbol{\theta}) - [E\widetilde{f}_1^{\pi_1,B}(t,\boldsymbol{\theta})]\widetilde{f}_i^{\pi_2,B}(t,\boldsymbol{\theta},\boldsymbol{\beta})}{E[\widetilde{f}_1^{\pi_2,B}(t,\boldsymbol{\theta},\boldsymbol{\beta})]}$$

$$\widetilde{f}_i^{\pi_1,B}(t,\boldsymbol{\theta}) = \widetilde{f}_i^{\pi_{11},B}(t,\boldsymbol{\theta}) + \widetilde{f}_i^{\pi_{12},B}(t,\boldsymbol{\theta})$$

$$\widetilde{f}_i^{\pi_{11},B}(t,\boldsymbol{\theta}) = I(C_{1i} \geq t)A_i(t)W_i^B(t;\boldsymbol{\theta})$$

$$\widetilde{f}_i^{\pi_{12},B}(t,\boldsymbol{\theta}) = E[I(C_{11} \geq t)A_1(t)W_1^B(t;\boldsymbol{\theta})\int_0^t \boldsymbol{X}_1^T(s)Y_1(s)ds][\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta})$$

$$\widetilde{f}_i^{\pi_2,B}(t,\boldsymbol{\theta},\boldsymbol{\beta}) = \widetilde{f}_i^{\pi_{21},B}(t,\boldsymbol{\theta},\boldsymbol{\beta}) + \widetilde{f}_i^{\pi_{22},B}(t,\boldsymbol{\theta},\boldsymbol{\beta}) + \widetilde{f}_i^{\pi_{23},B}(t,\boldsymbol{\theta},\boldsymbol{\beta})$$

$$\widetilde{f}_i^{\pi_{21},B}(t,\boldsymbol{\theta},\boldsymbol{\beta}) = I(C_{1i} \geq t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}W_i^B(t;\boldsymbol{\theta})$$

$$\widetilde{f}_i^{\pi_{22},B}(t,\boldsymbol{\theta},\boldsymbol{\beta}) = E[I(C_{11} \geq t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_1(t)}W_1^B(t;\boldsymbol{\theta})\int_0^t \boldsymbol{X}_1^T(s)Y_1(s)ds][\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta})$$

$$\widetilde{f}_i^{\pi_{23},B}(t,\boldsymbol{\theta},\boldsymbol{\beta}) = \boldsymbol{z}^{(1)}[t;\boldsymbol{\beta},W^B(\boldsymbol{\theta})]^T\boldsymbol{f}_i^{\boldsymbol{\beta}_B}(\boldsymbol{\theta},\boldsymbol{\beta}).$$

Similar to the calculation of $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\beta}})$, we also treat the estimated weights function as fixed. Therefore, the variance estimator $\widehat{\sigma}(s,t,\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\beta}})$ is calculated by estimating $\widetilde{f}_i^{\pi_1,B}(t,\boldsymbol{\theta}_0)$ by $\widehat{f}_i^{\pi_{11},B}(t,\widehat{\boldsymbol{\theta}})$, and $\widetilde{f}_i^{\pi_2,B}(t,\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)$ by $\widehat{f}_i^{\pi_{21},B}(t,\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\beta}})$, where $\widehat{f}_i^{\pi_{11},B}(t,\boldsymbol{\theta}) = I(C_{1i} \geq t)A_i(t)\widehat{W}_i^B(t;\boldsymbol{\theta})$, $\widehat{f}_i^{\pi_{21},B}(t,\boldsymbol{\theta},\boldsymbol{\beta}) = I(C_{1i} \geq t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}\widehat{W}_i^B(t;\boldsymbol{\theta})$.

Next we consider the scenario where C_{1i} is not known for all i, with M imputed censoring times used for subjects where C_{1i} is not observed.

Theorem 7. *Under assumptions (2.2), (2.3), (2.5) and (2.18)* $\widehat{\boldsymbol{\beta}}^M$ *is a consistent estimator of* $\boldsymbol{\beta}_0$*, and* $n^{1/2}(\widehat{\boldsymbol{\beta}}^M - \boldsymbol{\beta}_0)$ *converges in distribution to a mean-zero normal random variable with a variance-covariance matrix* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,M) = E[\boldsymbol{f}_1^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,M)^{\otimes 2}]$*, where*

$$\boldsymbol{f}_i^{\boldsymbol{\beta}_B}(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\gamma},M) = \boldsymbol{\Omega}[\boldsymbol{\beta},W^B(\boldsymbol{\theta})]^{-1}[\boldsymbol{\Phi}_{1i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\gamma},M) + \boldsymbol{\Phi}_{2i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta},\boldsymbol{\beta})]$$

$$\boldsymbol{\Phi}_{1i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\gamma},M) = \int_0^\tau \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{z}}[t;\boldsymbol{\beta},W^B(\boldsymbol{\theta})]\}W_i^B(t;\boldsymbol{\theta})\frac{1}{M}\sum_{m=1}^M dM_i^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\gamma})$$

$$dM_i^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\gamma}) = \left[A_i(t) - \pi_0(t)exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}\right]I(C_{1i}^{\langle m\rangle} \geq t;\boldsymbol{\gamma})I(C_{2i} \geq t)dt.$$

Note that $\mathbf{\Phi}_{2i}^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}, \boldsymbol{\beta})$ is the same as defined in Theorem 1.

For $\pi_0^M(t)$, we have the following result.

Theorem 8. *Under assumptions (2.2), (2.3), (2.5) and (2.18),* $n^{1/2}(\widehat{\pi}_0^M - \pi_0)$ *converges weakly to a mean-zero Gaussian process with a variance and covariance matrix between* $n^{1/2}[\widehat{\pi}_0^M(s) - \pi_0(s)]$ *and* $n^{1/2}[\widehat{\pi}_0^M(t) - \pi_0(t)]$ *given by* $\sigma(s, t, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, M) = E[\xi_1(s, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, M)\xi_1(t, \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, M)]$, *where*

$$\xi_i(t, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M) = \frac{\widetilde{f}_i^{\pi_1, B}(t, \boldsymbol{\theta}) - [E\widetilde{f}_1^{\pi_1, B}(t, \boldsymbol{\theta})]\widetilde{f}_i^{\pi_2, B}(t, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M)}{E[\widetilde{f}_1^{\pi_2, B}(t, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M)]}$$

$$\widetilde{f}_i^{\pi_2, B}(t, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M) = \widetilde{f}_i^{\pi_{21}, B}(t, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M) + \widetilde{f}_i^{\pi_{22}, B}(t, \boldsymbol{\theta}, \boldsymbol{\beta}) + \widetilde{f}_i^{\pi_{23}, B}(t, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M)$$

$$\widetilde{f}_i^{\pi_{21}, B}(t, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M) = \frac{1}{M}\sum_{m=1}^{M} I(C_{1i}^{\langle m \rangle} \geq t; \boldsymbol{\gamma}) I(C_{2i} \geq t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)} W_i^B(t; \boldsymbol{\theta})$$

$$\widetilde{f}_i^{\pi_{23}, B}(t, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M) = \boldsymbol{z}^{(1)}[t; \boldsymbol{\beta}, W^B(\boldsymbol{\theta})]^T \boldsymbol{f}_i^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, M).$$

Similar to the known censoring case, the variance could be estimated by its empirical counterpart and by treating weight function as known.

2.4 Simulation Studies

We report on simulations to evaluate performance of our methods. The results of three weights are evaluated: $W_i(t) = 1$, which does not correctly accommodate the censoring mechanisms and is included for comparison purposes only; $W_i^A(t; \boldsymbol{\theta}_0)$ defined in (2.9); and the stabilized weights, $W_i^B(t; \boldsymbol{\theta}_0)$, defined in (2.12).

For each simulation setting, two scenarios ($n = 300$, $n = 500$) are generated. The covariate $\boldsymbol{Z} = (Z_1, Z_2)'$ has elements which are Bernoulli(0.5). The terminal event, D, is generated by the hazard function $\lambda_0^D \exp\{\boldsymbol{\alpha}_0^T \boldsymbol{Z}\}$, where $\lambda_0^D = 0.02$ and $\boldsymbol{\alpha}_0 = (-0.609, 0.609)^T$. The target model represents the probability of being alive and active, $\pi(t) = \pi_0(t)\exp\{\boldsymbol{\beta}_0^T \boldsymbol{Z}\}$, where $\pi_0(t) = 0.3 - 0.0025t$, for $t = 1, 2, ..., 100$. The event of being active on waiting list given the subject is

alive is sampled from the conditional probability, $P\{\mathcal{E}(t) = 1|D > t, \boldsymbol{Z}\} = \pi(t)\text{exp}[\lambda^D \times t]$.

We generated two censoring times, C_1 and C_2, for each individual. The independent censoring time, C_1, is generated from the hazard function $\lambda_0^C \text{exp}\{\boldsymbol{\gamma}_0^T \boldsymbol{Z}\}$, where $\lambda_0^C = 0.015$ and $\boldsymbol{\gamma}_0 = (0.609, -0.609)^T$. For the dependent censoring time, C_2, we first generate X_t, where $X_t = \min\{D, -40 \times \log[Z_1\epsilon_1 + (1 - Z_2)(1 - \epsilon_1)] + 5\epsilon_2\}$, $\epsilon_1 = \int_1^{100} A(t)dt/100$, $\epsilon_2 \sim$ Uniform(0,1). Let $X(t) = I(X_t \geq t)$, with $X(t)$ being dependent on $A(t)$ even conditional on $\boldsymbol{Z}$ due to its mutual association with ϵ_1. Next, we generate time-dependent censoring time C_2 from the hazard function $\lambda_0^C + \phi_1 Z_1 + \phi_2 Z_2 + \phi_3 X(t)$. Note that ϵ_1 and ϵ_2 are mutually independent.

In Table 2.1, we consider the setting wherein C_1 is known. We set $\lambda_0^C = 0.006$, $\phi_1 = \phi_2 = -0.002$, and $\phi_3 = 0.035$ for heavy censoring C_2, which results in about 53% subjects being censored by C_2 and 37% subjects censored by C_1. Moreover, two magnitudes of $\boldsymbol{\beta}_0$ are considered: 0.916 and 0.405. Next we consider a light censoring case for C_2, where $\lambda_0^C = 0.003$, $\phi_1 = \phi_2 = -0.001$, and $\phi_3 = 0.015$. This set-up results in 33% subjects are censoring be C_2 and 48% subjects are censoring be C_1. In Table 2.2 we treat censoring time as random, and get imputed estimators from $M = 1$ imputation dataset.

In each setting, the biases of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Pi}_0(50)$ are very small for both $\widehat{W}_i^A(t; \widehat{\boldsymbol{\theta}})$ and $\widehat{W}_i^B(t; \widehat{\boldsymbol{\theta}})$, indicating that our estimators are consistent. Moreover, empirical standard deviations (ESDs) are generally close to, but larger than the average asymptotic standard errors (ASEs), because we treat the estimated weights function as fixed. The empirical coverage probabilities (ECPs) are also around 0.95, implying the accuracy of large-sample confidence intervals. Due the substantial biases of not adjusting time dependent confounder, the unweighted method exhibits bias and has inaccurate estimated variance and poor coverage probabilities.

2.5 Real Data Analysis

We applied the proposed methods to model the wait list active/inactive process using data obtained from the Scientific Registry of Transplant Recipients (SRTR). In the end-stage liver disease

(ESLD) setting, the number of available deceased-donor livers is always less than the number of patients in need of liver transplantation. Once an ESLD patient is wait listed, the patient's status can oscillate between active and inactive based on their medical condition. A wait listed patient can receive deceased-donor organ offers only when active. To date, there has been little study of the probability of remaining active on the wait list, for ESLD or any type of end-stage organ failure.

The sample size of this study is $n = 53,991$. Patients are subject to independent right censoring due to administrative censoring and living donor transplantation. Note that living-donor transplants are not allocated using the MELD system and, as such, are not determined by internal time-varying covariates. In contrast, deceased-donor transplantation is viewed as dependent censoring, since the transplant hazard and pre-transplant mortality hazard are correlated with time-varying MELD score. Therefore, deceased-donor transplantation is viewed as dependent censoring, while other causes of censoring are aggregated into C_1. Based on this classification, there were 13,180 subjects observed to die, 17,982 patients who were independently censored, and 22,829 subjects who were dependently censored. The median observation time is 201 days, while the median time spent in the active state is 160 days.

Baseline covariates include blood type, gender, race, BMI status, hospitalization, age, region, and values at wait listing for MELD score, serum albumin and serum sodium. Comorbid conditions are also included, for example hepatitis C, noncholestatic, cholestatic, acute hepatic necrosis, metastatic disease, malignant neoplasm. The covariate information at time zero are used to characterize the process of being alive and active on waiting list, and to implement the imputation of C_1. As for time dependent covariates, we include more predictors to the baseline covariates set: MELD all score, albumin levels, sodium, ascites, encephalitis, and dialysis status, and exclude baseline MELD, baseline albumin and baseline sodium. Moreover, continuous variables are centered at their mean values.

The stabilized weights, $W^B(t;\boldsymbol{\theta})$, were used to remove bias due to dependent censoring. To further mitigate the impact of outliers, the weights were capped by 150. For subjects with observed death, we imputed C_1 as describe in Section 2.2. Due to the size of the data set, we used $M = 1$.

Covariate effects along with p values are listed in Table 2.3. Among the diagnoses leading ESLD, acute hepatic necrosis was associated with a significant 12% increase in probability of being alive and active on the waiting list. Patients with ESLD resulting from malignant neoplasm experienced an 11% reduction in survival/active probability ($p = 0.01$). Obesity (i.e., BMI $\geq$30) was associated with a significant 5% reduction. Being hospitalized in the ICU at the time of wait listing was associated with an 18% increase ($p = 0.001$) in the alive/active probability, relative to not being hospitalized. Each 5-year increase in age at wait listing was associated with a significant 2% decrease in alive/active probability. Relative to the United Network for Organ Sharing (UNOS) Region with the greatest number of wait listed patients (Region 5), Region 1, 10 and 11 had significant reductions in alive/active probability, at 17%, 6%, and 16%, respectively. Region 7 had a 6% increase ($p = 0.01$). The probability of being alive and active on the wait list decreased with by 1% for each (integer) increase in MELD score, and increased by 9% per unit increase in serum albumin. We provide some interpretation of these results in Section 2.6.

The estimated baseline probability function is plotted in Figure 2.1. We estimated the integral of baseline probability of being alive and active over $[0, 5]$ at 3.92 years; this indicates that a 'baseline' patient (i.e., patient with all covariates equal to the reference) would be expected to be alive and active on the wait list for $\approx$4 of the first 5 years after wait listing, in the absence of liver transplantation.

2.6 Concluding Remarks

In this article, we propose semiparametric temporal process regression methods. Relative to existing process regression methods, our methods are distinguished by several features. In particular, the baseline probability (as a function of time) is unspecified and is essentially profiled out in the estimation of the regression coefficient (presumably of chief interest in most studies). The method accommodates dependent right censoring, and does so through a computationally attractive additive hazards model. In our context, the additive hazard model facilities the calculation of weight

function, since the baseline cumulative hazard function cancels out. Moreover, the proposed methods accommodate independent right censoring through imputation rather than a weight function.

We applied our proposed methods to the analysis of factors affecting the joint probability of being alive and active on the liver transplant waiting list. There are many reports in the literature regarding the factors affecting pre-liver transplant survival. The analysis we present in Section 2.5 is among the first to incorporate active/inactive status into the end-point. Survival in the absence of liver transplantation is, in its own right, an important process in the context of end-stage liver disease. Part of its importance derives from the fact that the preferred treatment for ESLD, liver transplantation, does not occur without the patient surviving long enough on the wait list to receive offers. However, survival on the wait list is not enough, since only a patient can only receive offers while active.

Our analysis revealed several factors as being significantly associated with the probability of being alive/active. Model of End-stage Liver Disease score is the basis of deceased-donor liver allocation among chronic ESLD patients, so its effect is not surprising. Serum albumin is a marker of nutritional status, and have been shown in several previous reports to be predictive of pre-transplant survival. The appearance that being hospitalized in the intensive care unit is possible due to selection bias. It is possible that sicker patients in the ICU die before having the opportunity to be wait listed. Or, the wait listing of such patients may sometimes deliberately not be pursued, in order to avoid futile transplantation. We do not have data to further evaluate either hypothesis. The differences in active/survival probability by UNOS Region may be due to differences in pre-transplant survival and/or discrepancies in deactivation protocols. Such differences are worthy of future investigation.

Our methods make the distinction between independent censoring, C_1, and dependent censoring, C_2. This is necessary since the variates play very different roles in our framework. C_1 represent the end of a subject's potential follow-up. This is the case in several existing methods, including the popular subdistribution hazard modeling of Fine and Gray (1999). The methods of Fine and Gray (1999) do not impute unobserved C_1 but, instead, apply a weight function which

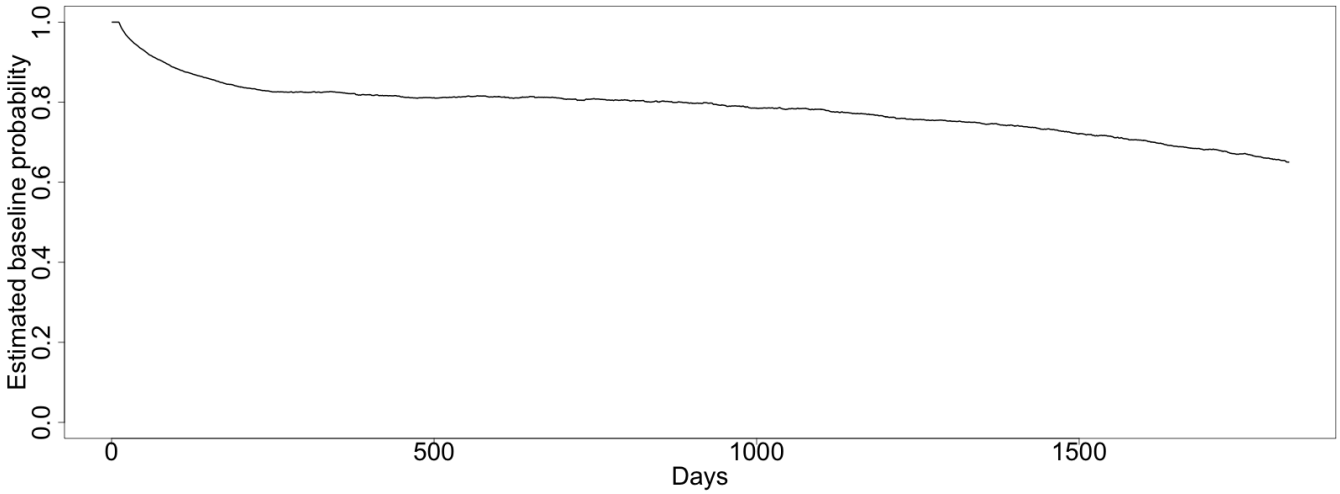

Figure 2.1: Estimated Baseline Probability of being Active and Alive

represent the conditional probability of being uncensored (given that the subject was uncensored at the time of death). This weight function is essentially playing the same role as our imputation of C_1. In contrast, C_2 is a nuisance process, with its corresponding inverse weight seeking to recover the data that would be observed if the process underling C_2 were absent.

In our framework, we assume additive hazards model [Lin and Ying, 1994] on dependent censoring time, which is a practical alternative to the proportional hazards model. These two models will normally provide adequate fit to data if appropriate time-dependent covariates are included.

Table 2.1: Simulations Results for Known Censoring Time based on 500 Replicates

				$\widehat{\boldsymbol{\beta}}$			
Censoring	n	$\boldsymbol{\beta}_0$	Weights	Bias	ASE	ESD	CP
Heavy	500	(0.916, -0.916)	1	(0.183, -0.125)	(0.095, 0.096)	(0.098, 0.095)	(0.520, 0.764)
			W^A	(0.000, 0.008)	(0.134, 0.134)	(0.149, 0.146)	(0.908, 0.928)
			W^B	(0.025, -0.010)	(0.119, 0.119)	(0.127, 0.128)	(0.932, 0.932)
		(0.405, -0.405)	1	(0.191, -0.132)	(0.095, 0.095)	(0.099, 0.090)	(0.480, 0.738)
			W^A	(0.031, 0.000)	(0.141, 0.140)	(0.157, 0.158)	(0.900, 0.926)
			W^B	(0.037, -0.022)	(0.123, 0.124)	(0.130, 0.122)	(0.918, 0.944)
	1000	(0.916, -0.916)	1	(0.187, -0.129)	(0.068, 0.068)	(0.066, 0.067)	(0.204, 0.536)
			W^A	(0.001, 0.016)	(0.100, 0.100)	(0.104, 0.110)	(0.936, 0.926)
			W^B	(0.024, -0.014)	(0.087, 0.087)	(0.088, 0.096)	(0.936, 0.934)
		(0.405, -0.405)	1	(0.187, -0.123)	(0.068, 0.068)	(0.068, 0.071)	(0.218, 0.544)
			W^A	(0.025, -0.007)	(0.101, 0.102)	(0.111, 0.112)	(0.918, 0.906)
			W^B	(0.036, -0.020)	(0.091, 0.091)	(0.094, 0.091)	(0.912, 0.940)
Light	500	(0.916, -0.916)	1	(0.103, -0.059)	(0.089, 0.089)	(0.092, 0.090)	(0.776, 0.896)
			W^A	(-0.003, 0.006)	(0.104, 0.104)	(0.103, 0.103)	(0.948, 0.952)
			W^B	(0.011, 0.001)	(0.098, 0.098)	(0.097, 0.097)	(0.960, 0.962)
		(0.405, -0.405)	1	(0.115, -0.075)	(0.090, 0.090)	(0.090, 0.099)	(0.750, 0.838)
			W^A	(0.006, -0.009)	(0.107, 0.108)	(0.114, 0.105)	(0.928, 0.960)
			W^B	(0.017, -0.017)	(0.100, 0.100)	(0.098, 0.103)	(0.952, 0.936)
	1000	(0.916, -0.916)	1	(0.100, -0.058)	(0.063, 0.063)	(0.063, 0.064)	(0.654, 0.844)
			W^A	(-0.005, 0.017)	(0.074, 0.074)	(0.071, 0.072)	(0.966, 0.952)
			W^B	(0.008, -0.003)	(0.069, 0.069)	(0.067, 0.069)	(0.946, 0.946)
		(0.405, -0.405)	1	(0.107, -0.068)	(0.064, 0.064)	(0.064, 0.066)	(0.624, 0.816)
			W^A	(0.011, -0.006)	(0.076, 0.076)	(0.078, 0.078)	(0.920, 0.950)
			W^B	(0.019, -0.015)	(0.072, 0.072)	(0.071, 0.074)	(0.932, 0.928)

					$\widehat{\Pi}_0(50)$			
Censoring	$\Pi_0(50)$	n	$\boldsymbol{\beta}_0$	Weights	Bias	ASE	ESD	CP
Heavy	11.812	500	(0.916, -0.916)	1	-2.182	0.849	0.843	0.280
				W^A	0.258	1.321	1.368	0.920
				W^B	-0.364	1.170	1.219	0.896
			(0.405, -0.405)	1	-2.247	0.854	0.831	0.258
				W^A	0.096	1.316	1.349	0.944
				W^B	-0.400	1.169	1.197	0.904
		1000	(0.916, -0.916)	1	-2.202	0.607	0.597	0.070
				W^A	0.185	0.946	0.895	0.954
				W^B	-0.347	0.844	0.838	0.900
			(0.405, -0.405)	1	-2.248	0.609	0.621	0.058
				W^A	0.154	0.945	0.929	0.954
				W^B	-0.369	0.842	0.831	0.924
Light	11.812	500	(0.916, -0.916)	1	-1.107	0.858	0.864	0.718
				W^A	0.215	1.023	0.992	0.946
				W^B	-0.140	0.956	0.972	0.934
			(0.405, -0.405)	1	-1.113	0.859	0.863	0.726
				W^A	0.275	1.020	1.026	0.938
				W^B	0.011	0.967	0.931	0.962
		1000	(0.916, -0.916)	1	-1.125	0.605	0.629	0.550
				W^A	0.172	0.714	0.676	0.960
				W^B	-0.073	0.680	0.682	0.950
			(0.405, -0.405)	1	-1.094	0.611	0.630	0.554
				W^A	0.191	0.715	0.727	0.930
				W^B	-0.059	0.681	0.649	0.950

Table 2.2: Simulations Results for Random Censoring Time based on $M = 1$, and 500 Replicates

				$\widehat{\boldsymbol{\beta}}^1$			
Censoring	n	$\boldsymbol{\beta}_0$	Weights	Bias	ASE	ESD	CP
Heavy	500	(0.916, -0.916)	1	(0.190, -0.141)	(0.096, 0.096)	(0.092, 0.093)	(0.472, 0.718)
			W^A	(0.011, -0.004)	(0.137, 0.135)	(0.136, 0.146)	(0.968, 0.936)
			W^B	(0.028, -0.023)	(0.122, 0.121)	(0.123, 0.123)	(0.938, 0.952)
		(0.405, -0.405)	1	(0.194, -0.136)	(0.096, 0.096)	(0.093, 0.098)	(0.484, 0.678)
			W^A	(0.036, -0.019)	(0.142, 0.142)	(0.151, 0.155)	(0.914, 0.930)
			W^B	(0.042, -0.040)	(0.124, 0.124)	(0.129, 0.127)	(0.912, 0.934)
	1000	(0.916, -0.916)	1	(0.191, -0.132)	(0.069, 0.069)	(0.069, 0.067)	(0.186, 0.518)
			W^A	(0.000, 0.005)	(0.102, 0.101)	(0.095, 0.107)	(0.964, 0.944)
			W^B	(0.020, -0.014)	(0.088, 0.088)	(0.085, 0.088)	(0.940, 0.942)
		(0.405, -0.405)	1	(0.189, -0.137)	(0.069, 0.068)	(0.068, 0.067)	(0.194, 0.494)
			W^A	(0.024, -0.019)	(0.104, 0.105)	(0.104, 0.108)	(0.924, 0.928)
			W^B	(0.037, -0.034)	(0.092, 0.092)	(0.092, 0.094)	(0.902, 0.904)
Light	500	(0.916, -0.916)	1	(0.098, -0.055)	(0.090, 0.089)	(0.088, 0.084)	(0.818, 0.926)
			W^A	(-0.007, 0.008)	(0.103, 0.104)	(0.100, 0.100)	(0.960, 0.958)
			W^B	(0.000, 0.000)	(0.098, 0.098)	(0.097, 0.094)	(0.956, 0.954)
		(0.405, -0.405)	1	(0.114, -0.065)	(0.091, 0.090)	(0.090, 0.088)	(0.764, 0.896)
			W^A	(0.010, -0.008)	(0.107, 0.107)	(0.103, 0.111)	(0.962, 0.936)
			W^B	(0.031, -0.032)	(0.101, 0.101)	(0.095, 0.097)	(0.950, 0.934)
	1000	(0.916, -0.916)	1	(0.100, -0.060)	(0.064, 0.063)	(0.063, 0.063)	(0.676, 0.860)
			W^A	(-0.008, 0.014)	(0.074, 0.074)	(0.071, 0.073)	(0.952, 0.948)
			W^B	(0.002, -0.001)	(0.070, 0.070)	(0.070, 0.066)	(0.944, 0.956)
		(0.405, -0.405)	1	(0.110, -0.074)	(0.065, 0.064)	(0.064, 0.062)	(0.608, 0.814)
			W^A	(0.015, -0.009)	(0.077, 0.077)	(0.075, 0.071)	(0.934, 0.956)
			W^B	(0.023, -0.021)	(0.072, 0.073)	(0.074, 0.073)	(0.930, 0.932)

					$\widehat{\Pi}_0^1(50)$			
Censoring	$\Pi_0(50)$	n	$\boldsymbol{\beta}_0$	Weights	Bias	ASE	ESD	CP
Heavy	11.812	500	(0.916, -0.916)	1	-2.304	0.847	0.820	0.244
				W^A	0.082	1.351	1.199	0.980
				W^B	-0.483	1.179	1.100	0.918
			(0.405, -0.405)	1	-2.401	0.846	0.843	0.204
				W^A	-0.012	1.335	1.260	0.948
				W^B	-0.478	1.178	1.100	0.922
		1000	(0.916, -0.916)	1	-2.362	0.602	0.605	0.038
				W^A	0.020	0.953	0.855	0.958
				W^B	-0.458	0.837	0.811	0.904
			(0.405, -0.405)	1	-2.368	0.604	0.601	0.038
				W^A	0.010	0.942	0.923	0.948
				W^B	-0.476	0.839	0.837	0.912
Light	11.812	500	(0.916, -0.916)	1	-1.182	0.851	0.848	0.706
				W^A	0.182	1.012	0.954	0.952
				W^B	-0.150	0.955	0.994	0.936
			(0.405, -0.405)	1	-1.234	0.849	0.826	0.666
				W^A	0.147	1.008	1.002	0.948
				W^B	-0.219	0.959	0.904	0.932
		1000	(0.916, -0.916)	1	-1.202	0.603	0.585	0.502
				W^A	0.087	0.715	0.664	0.962
				W^B	-0.156	0.679	0.662	0.942
			(0.405, -0.405)	1	-1.176	0.608	0.576	0.506
				W^A	0.064	0.715	0.668	0.960
				W^B	-0.168	0.678	0.682	0.952

Table 2.3: Analysis of Liver Transplant Data: Covariate Effects on being Active on the Waiting List and being Alive (based on $M = 1$ Imputation)

Covariate	Value	$\widehat{\boldsymbol{\beta}}$	$\widehat{\mathrm{SE}}(\widehat{\boldsymbol{\beta}})$	p	$\exp(\widehat{\boldsymbol{\beta}})$
Blood type (v.s. O)	A	0.0113	0.0140	0.421	1.01
	AB	-0.0057	0.0390	0.8843	0.99
	B	0.0160	0.0202	0.4265	1.02
Gender	Female	0.0234	0.0131	0.0748	1.02
Race (v.s. White)	Black	0.0036	0.0257	0.8898	1.00
	Hispanic	0.0266	0.0182	0.1441	1.03
	Asian	0.0284	0.0291	0.3287	1.023
	Other	-0.0396	0.0698	0.5707	0.96
Diagnosis	HCV	-0.0245	0.0215	0.2544	0.98
	Noncholestatic	0.0388	0.0204	0.0568	1.04
	Cholestatic	0.0450	0.0256	0.0792	1.05
	Acute hepatic necrosis	0.1116	0.0429	0.0094*	1.12
	Metastatic disease	0.0246	0.0500	0.6229	1.03
	Malignant neoplasm	-0.1130	0.0464	0.0149*	0.89
BMI (v.s. $(20, 25)$)	$[0, 20]$	0.0178	0.0277	0.5217	1.02
	$[25, 30)$	-0.0011	0.0159	0.9436	1.000
	$[30, \infty)$	-0.0513	0.0169	0.0024*	0.95
Hospitalization status	ICU	0.1652	0.0500	0.001*	1.18
(v.s. not hospitalized)	Not ICU	-0.0279	0.0323	0.3868	0.97
Age	per 5 years	-0.0205	0.0030	$<$.0001*	0.98
Region (v.s. 5)	1	-0.1924	0.0453	$<$.0001*	0.83
	2	0.0204	0.0213	0.3394	1.02
	3	-0.0259	0.0276	0.3493	0.97
	4	-0.0423	0.0222	0.0572	0.96
	6	-0.0646	0.0479	0.1772	0.94
	7	0.0611	0.0248	0.0135*	1.06
	8	-0.0511	0.0304	0.0932	0.95
	9	-0.0130	0.0232	0.5771	0.99
	10	-0.0587	0.0285	0.0397*	0.94
	11	-0.1753	0.0328	$<$.0001*	0.84
MELD	per unit score	-0.0078	0.0016	$<$.0001*	0.99
Albumin	per mmol/L	0.0850	0.0107	$<$.0001*	1.09
Sodium	per g/dL	0.0026	0.0018	0.1481	1.00

CHAPTER 3

Joint Modeling of State Prevalence and Mortality

3.1 Introduction

In the data structure we consider, subjects (while alive) can switch back and forth between two states, such that the subject is in one (and only one) of the two states until the terminating event occurs. The term 'prevalent' will refer to the subject being in the state of interest. The follow-up of each subject can be independently right censored.

In this chapter, we jointly model state prevalence (given survival) and death hazard by a shared frailty term. In Chapters 1 and 2, the survival outcome was built into the definition of the state analogous to 'prevalent'. The separation of the state prevalence and survival components in Chapter 3 implies that the residual correlation between the two processes must be taken into account.

The model of the prevalence outcome has a marginal interpretation, from the perspective that it is conditional on survival, baseline covariates, external time-varying covariates, and frailty terms, but not on the previous history. The linear predictor is connected to the probability of prevalence outcomes through a known link function (for example, the logit function in this manuscript). This model formulation is closely related to generalized linear mixed models (GLMMs) [McCulloch and Neuhaus, 2001], but we also consider a shared frailty term in the death model. Instead of constraining the same frailty term in both the prevalence and the death model, we incorporate an unknown scale parameter to the death model, which distinguishes the frailty effect on the death and prevalence process. Some previous methods have relaxed this assumption even further by

estimating the correlation between the random effects of the two processes [Rizopoulos et al., 2008].

Note that our proposed model accommodates a prevalent process indexed by a continuous timeline. For example, in end-stage renal disease (ESRD) setting, a subject may stay in the hospital for several days, and length of stay should not be ignored in the analysis. This unique data structure prevents direct application of many joint modeling approaches for longitudinal and survival data [Wulfsohn and Tsiatis, 1997, Tsiatis and Davidian, 2004, Yu et al., 2004, Chi and Ibrahim, 2006], including several models for binary outcomes [Pulkstenis et al., 1998, Albert, 2000, Larsen, 2004]. Turning our attention now to another branch of methodology, many existing methods considered a point process, implicitly assuming that the events have no duration [Liu et al., 2004, Ye et al., 2007]. The afore-mentioned methods consider a type of data structure where outcomes are observed at discrete time points. Moreover, we intend to build regression methods on the probability of binary response, in order to have a prevalence interpretation, in contrast to models for transition probabilities [Albert, 2000, Shirley et al., 2010], many of which are built on a Markov type assumption.

The method we propose in this chapter could be viewed as an extension to temporal process regression methods proposed by Fine et al. [2004]. Their paper modeled $P[Y(t)|S(t) = 1]$, where $Y(t)$ is observed on a continuous timeline and $S(t) = 1$ could be survival. The extension we make is simultaneously modeling the death hazard by incorporating a shared frailty term and an unknown scale parameter. Considering $S(t)$ as survival indicator is attractive, because in most cases where death is a terminal event, it will prevent the future occurrence of outcome $Y(t)$. Similar conditional modeling technique is employed in many recurrent/terminal event literature [Huang and Wang, 2004, Liu et al., 2004, Ye et al., 2007, Zeng and Cai, 2010, Kalbfleisch et al., 2013]. The big distinction is that, in our case, the "event" has a duration associated with it.

In models with unknown frailty terms, the primary difficulty involves the evaluation of marginal likelihood, in sense that random effect need to be integrated out. In general, there are three options in current literature. The first one is to obtain a tractable integration of the likelihood function by

using specific link-random-effect combinations, e.g., complementary log-log link models with conjugate gamma random effects [Pulkstenis et al., 1998, Rondeau et al., 2007, Ye et al., 2007]. Another option is using numerical approaches to approximate the marginal likelihood function. Examples include marginal quasi-likelihood (MQL) [Goldstein, 1991] and penalized quasi-likelihood (PQL) [Breslow and Clayton, 1993], which approximate generalized linear mixed models as linear mixed models so that iterative generalized least squares (IGLS) [Goldstein, 1986] technique can be applied. Gaussian quadrature can also be used to approximate the marginal likelihood by a sum of integrand times suitably corresponding weights [Liu and Pierce, 1994, Li et al., 2017]. In the presence of normal random effects, the marginal likelihood could also be approximated with a mixture of binomial distributions [Thomas et al., 1998]. Finally some previous works here employed Monte Carlo integration, for example importance sampling [Yu et al., 2004], and Metropolis sampling [Liu et al., 2004, Chi and Ibrahim, 2006].

In this chapter, we consider an alternative approach to estimate covariate effects combining both iterative updates and numerical approximation of the likelihood involving frailties. If frailties and the scale parameter are known, then treating them as offsets, we could apply partial likelihood [Cox, 1975] to estimate the regression coefficients, and apply generalized estimating equations (GEEs) to estimate covariate effects for the prevalence model [Liang and Zeger, 1986]. On the other hand, if those coefficients are known, we could estimate each frailty term from their corresponding estimating equation. The scale parameter and the variance of frailty terms could be further estimated by numerical approximation approach (for example, Gaussian quadrature [Liu and Pierce, 1994]). Following these ideas, we propose to iteratively update the frailty terms and regression coefficients until convergence. It can be shown that our algorithm is equivalent to the EM algorithm [Dempster et al., 1977], and, therefore, will maximize the marginal likelihood function by treating the frailty terms as missing data. More details are demonstrated in the following section. We also propose an efficient method to approximate the asymptotic variance of covariate effects which combines a Metropolis-Hastings algorithm and permutation methods.

The remainder of the chapter is organized as follows. We set up notation and describe our

proposed methods in the next section. Simulation studies are performed to evaluate our method in finite samples in Section 3 under various scenarios. In Section 4, we apply the proposed methods to national ESRD data. Finally, concluding remarks are given in Section 5.

3.2 Proposed Methods

Suppose that there are n subjects. Let D_i be the death (i.e., terminal event) time, and let C_i be the censoring time for subject i ($i = 1, 2, ..., n$). We let $\boldsymbol{Z}_i(t)$ be a p-dimensional external covariate vector for subject i at time t [Kalbfleisch and Prentice, 2002], and let $\boldsymbol{Z}_{1i}(t) = [1, \boldsymbol{Z}_i^T(t)]^T$. Follow-up time $t \in [0, \tau]$ is considered, where τ is a pre-specified constant satisfying $Pr(C_i \geq \tau) > 0$ for $i = 1, 2, ..., n$. In practice, one could choose τ as the maximum of observation time X_i, where $X_i = C_i \wedge D_i$, with $a \wedge b = \min(a, b)$.

Let $P_i(t)$ be an indicator function representing 'prevalent' at time t; i.e., $P_i(t) = 1$ if patient i is in the state of interest and 0 otherwise. For example, being hospitalized, or being active on the kidney transplant waiting list. We assume that the expectation of $P_i(t)$ given $\boldsymbol{Z}_{1i}(t)$, ν_i and subject i is alive has the following form,

$$E[P_i(t)|\boldsymbol{Z}_{1i}(t), \nu_i, D_i \geq t] = g^{-1}\left\{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{1i}(t) + \nu_i\right\}, \tag{3.1}$$

where ν_i is a random effect or frailty term to absorb the residual correlations between $P_i(t)$ and D_i within subject i. We assume a proportional hazard model with frailty terms ν_i on D_i,

$$\lambda_i^D[t|\boldsymbol{Z}_i(t), \nu_i] = \lambda_0^D(t)\exp\left[\boldsymbol{\alpha}_0^T \boldsymbol{Z}_i(t) + \gamma\nu_i\right], \tag{3.2}$$

where γ is an unknown scale parameter, and $\lambda_0^D(t)$ is an unspecified baseline hazard function for the death time.

The frailty term ν_i is assumed to follow an independent normal distribution with variance σ_ν^2. For identifiability, we set the normal distribution mean zero. Following the general set up in gen-

eralized linear models (GLMs), $g(\cdot)$ is a link function between the expectation of the outcome and the linear component. In this manuscript, we consider a logit link function, $g(\mu) = \log[\mu/(1-\mu)]$, to guarantee that fitted values of the conditional expectation are bounded by 0 and 1. Other link functions could also be considered; for example, the probit function. Models like (3.1) have been referred to as partial marginal in recurrent event analysis [Ye et al., 2007] and other areas of survival analysis (e.g., [Gong and Schaubel, 2013, 2017]), since the expectation is conditioning on less information than a intensity function [Andersen and Gill, 1982], but more information than a marginal mean model [Lin et al., 2000]. A practical convention is that the terminal event preclude subsequent prevalence, such that $P_i(t) = 0$ for $t \in [D_i, \tau]$.

If the frailty term ν_i and the scale parameter γ were known, then we could apply partial likelihood [Cox, 1975] for model (3.2) in order to estimate $\boldsymbol{\alpha}_0$, and generalized estimating equations (GEE; Liang and Zeger [1986]) to estimate $\boldsymbol{\beta}_0$. Challenges include ν_i not being known, and the lack of an explicitly specified error distribution.

3.2.1 Iterative Estimating Procedures

Before illustrating our estimating procedures, we first set up the requisite notation. Define the at-risk indicator $Y_i(t) = I(X_i \geq t)$, and an observed-death indicator $\Delta_i = I(D_i \leq C_i)$. Observed data for subject i are denoted by $O_i = \{P_i(t), X_i, \Delta_i, \boldsymbol{Z}_i(t), t \in [0, X_i]\}$.

We make the following assumptions for the joint model:

1. Observed O_1, ..., O_n are independent and identically distributed.

2. Two conditional independence assumptions for censoring:

   (a) $\lambda_i^D[t|\boldsymbol{Z}_i(t), \nu_i] = \lambda_i^D[t|\boldsymbol{Z}_i(t), \nu_i, C_i \geq t]$

   (b) $E[P_i(t)|D_i \geq t, \boldsymbol{Z}_i(t), \nu_i] = E[P_i(t)|X_i \geq t, \boldsymbol{Z}_i(t), \nu_i]$

Under models (3.1) and (3.2), and modeling assumptions (1) and (2) above, the likelihood function for subject i is given by

$$
\begin{aligned}
&L[\boldsymbol{\alpha},\boldsymbol{\beta}, \boldsymbol{\Lambda}_0^D, \sigma_\nu^2, \gamma; O_i, \nu_i] \\
&\quad = L_1\left[\boldsymbol{\beta}; (P_i(t), t \in (0, X_i])|\nu_i, \boldsymbol{Z}_{1i}(t), D_i = s, s \geq X_i\right] L_2\left[\boldsymbol{\alpha}, \boldsymbol{\Lambda}_0^D; (X_i, \Delta_i)|\nu_i, \gamma, \boldsymbol{Z}_i(X_i)\right] L_3\left[\sigma_\nu^2; \nu_i\right] \\
&\quad \propto \prod_{t\in(0,X_i]} \exp\left[\boldsymbol{\beta}^T \boldsymbol{Z}_{1i}(t) + \nu_i\right]^{P_i(t)} \left\{1 + \exp\left[\boldsymbol{\beta}^T \boldsymbol{Z}_{1i}(t) + \nu_i\right]\right\}^{-1} \\
&\quad \times \left\{\exp\left[\boldsymbol{\alpha}^T \boldsymbol{Z}_i(X_i) + \gamma\nu_i\right] d\Lambda_0^D(X_i)\right\}^{\Delta_i} \times \exp\left\{-\int_0^\tau Y_i(u)\exp\left[\boldsymbol{\alpha}^T \boldsymbol{Z}_i(u) + \gamma\nu_i\right] d\Lambda_0^D(u)\right\} \\
&\quad \times (\sigma_\nu^2)^{-1/2}\exp\left[-\frac{\nu_i^2}{2\sigma_\nu^2}\right]. \qquad (3.3)
\end{aligned}
$$

This likelihood function can be considered as hierarchical or h-likelihood [Lee and Nelder, 1996, Ha et al., 2001], and the log likelihood is therefore given by

$$
\begin{aligned}
&l[\boldsymbol{\alpha},\boldsymbol{\beta}, \boldsymbol{\Lambda}_0^D, \sigma_\nu^2, \gamma; O_i, \nu_i] \\
&\quad \propto \int_0^{X_i} \left(P_i(t)\left[\boldsymbol{\beta}^T \boldsymbol{Z}_{1i}(t) + \nu_i\right] - \log\left\{1 + \exp\left[\boldsymbol{\beta}^T \boldsymbol{Z}_{1i}(t) + \nu_i\right]\right\}\right) dt \\
&\quad + \Delta_i\left[\boldsymbol{\alpha}^T \boldsymbol{Z}_i(X_i) + \gamma\nu_i + \log d\Lambda_0^D(X_i)\right] - \int_0^\tau Y_i(u)\exp\left[\boldsymbol{\alpha}^T \boldsymbol{Z}_i(u) + \gamma\nu_i\right] d\Lambda_0^D(u) \\
&\quad - \frac{1}{2}\log(\sigma_\nu^2) - \frac{\nu_i^2}{2\sigma_\nu^2}. \qquad (3.4)
\end{aligned}
$$

Differentiating (3.4) with respect to ν_i yields the following estimating equation,

$$
\begin{aligned}
U_{1i}(\nu_i) = &\int_0^{X_i} \left(P_i(t) - \text{expit}\left[\boldsymbol{\beta}^T \boldsymbol{Z}_{1i}(t) + \nu_i\right]\right) dt \\
&+ \Delta_i\gamma - \int_0^\infty \gamma Y_i(u)\exp\left[\boldsymbol{\alpha}^T \boldsymbol{Z}_i(u) + \gamma\nu_i\right] d\Lambda_0^D(u) - \frac{\nu_i}{\sigma_\nu^2}, \qquad (3.5)
\end{aligned}
$$

where the expit function is defined by $\text{expit}(\mu) = \exp(\mu)/[1 + \exp(\mu)]$. We further denote $\boldsymbol{U}_1(\boldsymbol{\nu}) = [U_{11}(\nu_1), U_{12}(\nu_2), ..., U_{1n}(\nu_n)]^T$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, ..., \nu_n)^T$ which are both $n \times 1$ vectors. Given $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $d\Lambda_0(t), t \in [0, X_i]$, σ_ν^2 and γ, we could solve $\boldsymbol{U}_1(\boldsymbol{\nu})$ through a Newton-Raphson procedure to estimate $\boldsymbol{\nu}$, with starting values set to the estimates obtained based on using only the prevalence data; this would amount to a GLMM [McCulloch and Neuhaus, 2001].

Now consider a hypothetical scenario where ν and γ are known. In order to estimate $\boldsymbol{\alpha}_0$ from model (3.2), we could apply partial likelihood [Cox, 1975]. Specifically, we could solve the following estimating equation for $\boldsymbol{\alpha}$,

$$\boldsymbol{U}_2(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \boldsymbol{Z}_i(t) - \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\alpha}, t; \nu, \gamma)}{S^{(0)}(\boldsymbol{\alpha}, t; \nu, \gamma)} \right\} dN_i^D(t), \tag{3.6}$$

where $N_i^D(t) = I(D_i \leq C_i \wedge t)$, $dN_i^D(t) = N_i^D(t^- + dt) - N_i^D(t^-)$, $\boldsymbol{S}^{(k)}(\boldsymbol{\alpha}, t; \nu, \gamma) = n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i(t)^{\otimes k} Y_i(t) \exp[\boldsymbol{\alpha}^T \boldsymbol{Z}_i(t) + \gamma \nu_i]$ for $k = 0, 1, 2$, where $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. Let t_{D1}, t_{D2}, ..., t_{Dm} be the m distinct ordered failure times. We need to evaluate $d\Lambda_0^D(t)$ at those m distinct time points. We further define $d\boldsymbol{\Lambda}_0^D = [d\Lambda_0^D(t_{D1}), d\Lambda_0^D(t_{D2}), ..., d\Lambda_0^D(t_{Dm})]^T$. A Breslow method [Breslow, 1972] analog would be used to estimate $d\boldsymbol{\Lambda}_0^D$ by solving the following estimating equation for $d\boldsymbol{\Lambda}_0^D$, where the jth element is given by ($j = 1, 2, ..., m$),

$$U_{3j}[d\Lambda_0^D(t_{Dj})] = \sum_{i=1}^{n} \left(\frac{dN_i^D(t_{Dj})}{d\Lambda_0^D(t_{Dj})} - Y_i(t_{Dj}) \exp\left[\boldsymbol{\alpha}^T \boldsymbol{Z}_i(t_{Dj}) + \gamma \nu_i\right] \right). \tag{3.7}$$

Next we consider the estimation of $\boldsymbol{\beta}_0$ under the hypothetical scenario when $\boldsymbol{\nu}$ and γ are known. Having established $l[\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Lambda}_0^D, \sigma_\nu^2, \gamma; O_i, \nu_i]$ in (3.4), we could differentiate the sum of log likelihood function with respect to $\boldsymbol{\beta}$ to get the estimating equation,

$$\boldsymbol{U}_4(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{X_i} \boldsymbol{Z}_{1i}(t) \left(P_i(t) - \text{expit}\left[\boldsymbol{\beta}^T \boldsymbol{Z}_{1i}(t) + \nu_i\right] \right) dt, \tag{3.8}$$

which is analog to the GEE version of logistic regression [McCulloch and Neuhaus, 2001], under a working independence correlation structure.

Finally, we estimate σ_ν^2 and γ using MLE techniques. Specifically, Gauss-Hermite quadrature [Liu and Pierce, 1994] is used to approximate the integral of likelihood function (3.3) with respect to $\boldsymbol{\nu}$. We denote the score function as $\boldsymbol{U}_5(\sigma_\nu^2, \gamma)$. The basic idea of Gaussian quadrature is to approximate an integral by a group of standard normal distributions with quadrature weight and location of Gauss-Hermite quadrature. One could estimate σ_ν^2 from only data pertinent to model (3.1) to allow usage of standard software; for example, the package *lme4* [Bates et al., 2015] in R.

However, this would lose some efficiency and accuracy in the estimation.

In order to estimate all unknown parameters $\boldsymbol{\eta} = (\boldsymbol{\nu}^T, \boldsymbol{\alpha}^T, d\boldsymbol{\Lambda}_0^{DT}, \boldsymbol{\beta}^T, \sigma_\nu^2, \gamma)^T$, we need to solve $\boldsymbol{U} = [\boldsymbol{U}_1^T(\boldsymbol{\nu}), \boldsymbol{U}_2^T(\boldsymbol{\alpha}), \boldsymbol{U}_3^T(d\boldsymbol{\Lambda}_0^D), \boldsymbol{U}_4^T(\boldsymbol{\beta}), \boldsymbol{U}_5^T(\sigma_\nu^2, \gamma)]^T = \mathbf{0}$. However the dimension of $\boldsymbol{U}$ is $(n + 2p + m + 3) \times 1$, which increases as sample size n increases. Therefore, generalized estimating equations (GEEs) [Liang and Zeger, 1986] could not be directly applied. We propose an iterative estimating approach to recursively update each parameter in $\boldsymbol{\eta}$ by treating the other parameters known. The proposed procedure is summarized as follows;

1. Set initial values of $d\boldsymbol{\Lambda}_0^{D(0)}$, $\boldsymbol{\alpha}^{(0)}$, $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\nu}^{(0)}$, $\sigma_\nu^{2(0)}$ and $\gamma^{(0)}$. One could first fit GLMM to the prevalence process to set $\sigma_\nu^{2(0)}$ and $\boldsymbol{\nu}^{(0)}$, and let $\gamma^{(0)} = 1$. Then get $\boldsymbol{\alpha}^{(0)}$, $d\boldsymbol{\Lambda}_0^{D(0)}$ and $\boldsymbol{\beta}^{(0)}$ as the solutions to $\boldsymbol{U}_2(\boldsymbol{\alpha})$, $\boldsymbol{U}_3(d\boldsymbol{\Lambda}_0^D)$, and $\boldsymbol{U}_4(\boldsymbol{\beta})$ with $\boldsymbol{\nu}^{(0)}$, $\sigma_\nu^{2(0)}$ and $\gamma^{(0)}$.

2. Replace $[\boldsymbol{\alpha}^T, d\boldsymbol{\Lambda}_0^{DT}, \sigma_\nu^2, \gamma]^T$ by $[\boldsymbol{\alpha}^{(0)T}, d\boldsymbol{\Lambda}_0^{D(0)T}, \sigma_\nu^{2(0)}, \gamma^{(0)}]^T$ in $\boldsymbol{U}_1(\boldsymbol{\nu})$ to get updated estimate $\boldsymbol{\nu}^{(1)}$.

3. Treat $\boldsymbol{\nu}^{(1)}$ and $\gamma^{(0)}$ as known in $\boldsymbol{U}_2(\boldsymbol{\alpha})$, $\boldsymbol{U}_3(d\boldsymbol{\Lambda}_0^D)$, and $\boldsymbol{U}_4(\boldsymbol{\beta})$ to get estimates $d\boldsymbol{\Lambda}_0^{D(1)}$, $\boldsymbol{\alpha}^{(1)}$, $\boldsymbol{\beta}^{(1)}$.

4. Get estimates $\sigma_\nu^{2(1)}$, $\gamma^{(1)}$ by solving $\boldsymbol{U}_5(\sigma_\nu^2, \gamma) = 0$.

5. Replace $d\boldsymbol{\Lambda}_0^{D(0)}$, $\boldsymbol{\alpha}^{(0)}$, $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\nu}^{(0)}$, $\sigma_\nu^{2(0)}$ and $\gamma^{(0)}$ by $d\boldsymbol{\Lambda}_0^{D(1)}$, $\boldsymbol{\alpha}^{(1)}$, $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\nu}^{(1)}$, $\sigma_\nu^{2(1)}$ and $\gamma^{(1)}$, respectively. Repeat steps 2-4 until $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, σ_ν^2 and γ converge to get $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma_\nu^2}$ and $\widehat{\gamma}$. Denote $d\boldsymbol{\Lambda}_0^{D(1)}$ and $\boldsymbol{\nu}^{(1)}$ in the last iteration as $d\widehat{\boldsymbol{\Lambda}}_0^D$ and $\widehat{\boldsymbol{\nu}}$.

3.2.2 Robust Variance Estimators

If $\boldsymbol{\nu}$ is known, then $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ are $\sqrt{n}$ consistent estimators and the asymptotic variances could be estimated through a robust (sandwich) estimator [Lin et al., 2000, Liang and Zeger, 1986]. However, substituting $\widehat{\boldsymbol{\nu}}$ in the sandwich estimators would severely underestimate the true variance, since it does not account for the randomness in $\widehat{\boldsymbol{\nu}}$. Another challenge, as mentioned in the previous section, is that the dimension of $\boldsymbol{U}$ increases as sample size increases. A bootstrap method [Efron

and Tibshirani, 1994] could be applied, but requires intensive computation. Here we propose a more efficient method to approximate the asymptotic variance of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$.

For each frailty ν_i, the first step is to use the Metropolis-Hastings algorithm to generate realizations from $f[\nu_i|O_i]$, which is proportional to $f[O_i|\nu_i]f(\nu_i)$ as defined in (3.3). Similar idea is also used to approximate the information matrix with random effects [Liu et al., 2004]. The remaining parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $d\boldsymbol{\Lambda}_0^D$, σ_ν^2 and γ are fixed at their estimated values $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, $d\widehat{\boldsymbol{\Lambda}}_0^D$, $\widehat{\sigma}_\nu^2$ and $\widehat{\gamma}$ respectively. A total of 2000 samples are generated and the first 1000 samples serving only as a burn-in period. We then calculate the empirical standard error from the 1000 samples after the burn-in, and denote by $\widehat{\sigma}^2(\widehat{\nu}_i)$. The second step is to approximate the limiting distribution by permutation methods[Lin et al., 1994]. To be specific, we generate 100 permutation samples. Within jth permutation sample, we sample G_i^j from a standard normal distribution with mean zero and variance one. Then replace ν_i by $G_i^j \times \widehat{\sigma}(\widehat{\nu}_i)$ in the estimating equations (3.6) and (3.8), and estimate the asymptotic variance of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ by their corresponding sandwich estimators. Finally, the pooled variance estimators between 100 permutation samples are reported. In the next section, we evaluate our robust variance estimator through simulations studies.

3.3 Simulation Studies

For each simulation setting, $n = 500$ subjects are generated. Covariates Z_1 and Z_2 follow Bernoulli(0.5) distributions. Let $\boldsymbol{Z} = (Z_1, Z_2)^T$ and $\boldsymbol{Z}_1 = (1, Z_1, Z_2)^T$. We first generate random effects ν from a mean-zero normal distribution with variance σ_ν^2, where $\sigma_\nu^2 = 0.3$ or 0.5. The terminal event D is generated by transforming $\epsilon_1 \sim$ Uniform (0, 1) using hazard function $\lambda^D(t) = \lambda_0^D \exp[\boldsymbol{\alpha}_0^T \boldsymbol{Z} + \gamma\nu]$, where $\boldsymbol{\alpha}_0 = (0.693, -0.693)^T$. In the first censoring setup, the censoring time C is generated from an uniform distribution with minimum 90 and maximum 100, and $\lambda_0^D = 0.01$. This would result in approximately 31% subjects being censored. In the second censoring scenario, C is generated from a Uniform distribution with minimum 100 and maximum 150, and $\lambda_0^D = 0.015$, which lead to approximately 20% subjects being censored.

The prevalence outcome $P(t)$ is sampled from $E[P(t) = 1|\boldsymbol{Z}_1, \nu, X > t] = \text{expit}(\boldsymbol{\beta}_0^T \boldsymbol{Z}_1 + \nu)$, where t is discrete, i.e., $t = 1, 2, ..., T$, where $T = 100$ in the first censoring scenario and $T = 150$ in the second one. Two magnitudes of $\boldsymbol{\beta}_0$ (including intercept) are considered: $(1, 0.6, -0.6)^T$ and $(1, 1, -1)^T$. The variance of the linear predictor $\boldsymbol{Z}_1^T \boldsymbol{\beta}_0$ would be 0.18 and 0.5 correspondingly. Three levels of γ are considered: -0.5, 0, and 0.5.

Gauss-Hermite quadrature with 15 nodes is used to approximate the integral of likelihood function. Simulation results for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$ with 1000 replicates are summarized in Table 3.1 and 3.2. In all simulation scenarios, the bias is relatively small, and asymptotic standard errors (ASEs) are generally close to the corresponding empirical standard derivations (ESDs). The empirical coverage probabilities (CPs) are quite close to the nominal 0.95. In Table 3.3 we present results for $\widehat{\sigma}_\nu^2$ and $\widehat{\gamma}$. When γ is nonzero, the ASEs of $\widehat{\gamma}$ appear to slightly underestimate the corresponding ESDs.

3.4 Real Data Analysis

We applied the proposed methods to data from the Dialysis Outcomes and Practice Patterns Study (DOPPS) Phase 5. The DOPPS is a prospective, observational study designed to elucidate aspects of hemodialysis practice that are associated with the best outcomes for hemodialysis patients [Young et al., 2000]. In particular, Phase 5 data were collected between 2012 and 2015. Our research interests include identifying demographic and clinical variables that are associated with out-of-hospital process and death process, considering their inner correlations.

The study population for DOPPS generally consists of prevalent patients. In the interests of having t reflect time-since-dialysis-initiation instead of merely time-since-DOPPS-entry, we restricted our study sample to include the $n = 5,298$ patients who entered DOPPS within 3 months of initiating dialysis. Patients included in our analysis were from 470 hemodialysis units across 11 different countries, with the counties including: Belgium, Canada, China, Gulf Coast Consortium, Germany, Italy, Japan, Spain, Sweden, the United Kingdom and the U.S.. Covariates include

Table 3.1: Simulations Results for $\widehat{\boldsymbol{\beta}}$ based on $n = 500$ and 1000 Replicates

% C	σ_ν^2	γ	True	BIAS	ASE	ESD	CP
$\approx$ 30%	0.3	-0.5	(1, 0.6, -0.6)	(-0.022, -0.012, 0.017)	(0.055, 0.065, 0.065)	(0.051, 0.058, 0.059)	(0.953, 0.966, 0.966)
		0	(1, 0.6, -0.6)	(-0.018, -0.016, 0.014)	(0.056, 0.065, 0.065)	(0.051, 0.060, 0.057)	(0.949, 0.970, 0.974)
		0.5	(1, 0.6, -0.6)	(-0.017, -0.015, 0.014)	(0.055, 0.065, 0.065)	(0.051, 0.059, 0.058)	(0.949, 0.967, 0.966)
	0.5	-0.5	(1, 0.6, -0.6)	(-0.024, -0.022, 0.020)	(0.065, 0.077, 0.076)	(0.059, 0.073, 0.071)	(0.946, 0.949, 0.957)
		0	(1, 0.6, -0.6)	(-0.022, -0.018, 0.018)	(0.067, 0.078, 0.078)	(0.063, 0.074, 0.072)	(0.952, 0.961, 0.963)
		0.5	(1, 0.6, -0.6)	(-0.027, -0.018, 0.018)	(0.066, 0.077, 0.077)	(0.060, 0.073, 0.069)	(0.945, 0.954, 0.967)
	0.3	-0.5	(1, 1, -1)	(-0.023, -0.019, 0.025)	(0.055, 0.066, 0.066)	(0.051, 0.059, 0.059)	(0.949, 0.958, 0.960)
		0	(1, 1, -1)	(-0.019, -0.024, 0.022)	(0.056, 0.067, 0.067)	(0.052, 0.061, 0.057)	(0.947, 0.970, 0.970)
		0.5	(1, 1, -1)	(-0.018, -0.022, 0.021)	(0.056, 0.066, 0.066)	(0.052, 0.059, 0.060)	(0.947, 0.957, 0.956)
	0.5	-0.5	(1, 1, -1)	(-0.026, -0.031, 0.030)	(0.065, 0.078, 0.078)	(0.060, 0.074, 0.073)	(0.944, 0.946, 0.957)
		0	(1, 1, -1)	(-0.024, -0.028, 0.028)	(0.067, 0.080, 0.080)	(0.063, 0.075, 0.072)	(0.948, 0.956, 0.956)
		0.5	(1, 1, -1)	(-0.028, -0.026, 0.027)	(0.066, 0.078, 0.078)	(0.060, 0.073, 0.070)	(0.952, 0.948, 0.960)
$\approx$ 20%	0.3	-0.5	(1, 0.6, -0.6)	(-0.020, -0.017, 0.016)	(0.059, 0.070, 0.070)	(0.052, 0.062, 0.061)	(0.962, 0.965, 0.971)
		0	(1, 0.6, -0.6)	(-0.020, -0.016, 0.013)	(0.060, 0.071, 0.071)	(0.052, 0.060, 0.061)	(0.972, 0.972, 0.976)
		0.5	(1, 0.6, -0.6)	(-0.022, -0.019, 0.017)	(0.060, 0.070, 0.071)	(0.051, 0.062, 0.060)	(0.955, 0.964, 0.977)
	0.5	-0.5	(1, 0.6, -0.6)	(-0.025, -0.019, 0.023)	(0.069, 0.082, 0.082)	(0.062, 0.074, 0.074)	(0.947, 0.962, 0.966)
		0	(1, 0.6, -0.6)	(-0.030, -0.023, 0.028)	(0.071, 0.084, 0.084)	(0.061, 0.076, 0.072)	(0.960, 0.963, 0.966)
		0.5	(1, 0.6, -0.6)	(-0.031, -0.019, 0.025)	(0.071, 0.083, 0.083)	(0.063, 0.074, 0.073)	(0.948, 0.966, 0.958)
	0.3	-0.5	(1, 1, -1)	(-0.022, -0.024, 0.023)	(0.059, 0.071, 0.071)	(0.053, 0.063, 0.061)	(0.958, 0.968, 0.967)
		0	(1, 1, -1)	(-0.021, -0.024, 0.021)	(0.061, 0.073, 0.073)	(0.052, 0.060, 0.062)	(0.972, 0.969, 0.968)
		0.5	(1, 1, -1)	(-0.023, -0.026, 0.024)	(0.061, 0.072, 0.072)	(0.051, 0.063, 0.061)	(0.960, 0.957, 0.969)
	0.5	-0.5	(1, 1, -1)	(-0.027, -0.029, 0.034)	(0.070, 0.084, 0.084)	(0.062, 0.076, 0.074)	(0.954, 0.952, 0.965)
		0	(1, 1, -1)	(-0.032, -0.033, 0.038)	(0.072, 0.086, 0.086)	(0.062, 0.077, 0.073)	(0.956, 0.962, 0.959)
		0.5	(1, 1, -1)	(-0.033, -0.030, 0.036)	(0.071, 0.085, 0.085)	(0.063, 0.076, 0.075)	(0.948, 0.962, 0.950)

Table 3.2: Simulations Results for $\widehat{\boldsymbol{\alpha}}$ based on $n = 500$ and 1000 Replicates

% C	$\boldsymbol{\beta}$	σ_ν^2	γ	True	BIAS	ASE	ESD	CP
$\approx$ 30%	(1, 0.6, -0.6)	0.3	-0.5	(0.693, -0.693)	(0.012, -0.010)	(0.117, 0.117)	(0.118, 0.123)	(0.951, 0.945)
			0	(0.693, -0.693)	(-0.001, 0.006)	(0.117, 0.117)	(0.114, 0.115)	(0.957, 0.952)
			0.5	(0.693, -0.693)	(0.000, 0.007)	(0.117, 0.117)	(0.125, 0.118)	(0.935, 0.945)
		0.5	-0.5	(0.693, -0.693)	(0.011, -0.004)	(0.117, 0.117)	(0.125, 0.117)	(0.936, 0.944)
			0	(0.693, -0.693)	(0.000, 0.001)	(0.117, 0.117)	(0.117, 0.119)	(0.947, 0.942)
			0.5	(0.693, -0.693)	(-0.007, 0.002)	(0.117, 0.117)	(0.124, 0.119)	(0.931, 0.945)
	(1, 1, -1)	0.3	-0.5	(0.693, -0.693)	(0.014, -0.011)	(0.117, 0.117)	(0.119, 0.123)	(0.952, 0.939)
			0	(0.693, -0.693)	(-0.001, 0.005)	(0.117, 0.117)	(0.114, 0.115)	(0.955, 0.952)
			0.5	(0.693, -0.693)	(0.001, 0.006)	(0.117, 0.117)	(0.125, 0.119)	(0.935, 0.940)
		0.5	-0.5	(0.693, -0.693)	(0.014, -0.006)	(0.117, 0.117)	(0.125, 0.117)	(0.935, 0.944)
			0	(0.693, -0.693)	(0.001, 0.001)	(0.117, 0.117)	(0.117, 0.119)	(0.947, 0.940)
			0.5	(0.693, -0.693)	(-0.005, 0.001)	(0.117, 0.117)	(0.124, 0.119)	(0.932, 0.946)
$\approx$ 20%	(1, 0.6, -0.6)	0.3	-0.5	(0.693, -0.693)	(0.011, -0.007)	(0.102, 0.102)	(0.105, 0.103)	(0.933, 0.946)
			0	(0.693, -0.693)	(-0.001, 0.009)	(0.102, 0.102)	(0.099, 0.101)	(0.952, 0.954)
			0.5	(0.693, -0.693)	(-0.006, -0.002)	(0.102, 0.102)	(0.104, 0.103)	(0.941, 0.944)
		0.5	-0.5	(0.693, -0.693)	(-0.000, -0.003)	(0.102, 0.102)	(0.108, 0.108)	(0.938, 0.942)
			0	(0.693, -0.693)	(-0.005, 0.002)	(0.102, 0.102)	(0.103, 0.101)	(0.948, 0.949)
			0.5	(0.693, -0.693)	(-0.006, 0.010)	(0.102, 0.102)	(0.114, 0.105)	(0.924, 0.936)
	(1, 1, -1)	0.3	-0.5	(0.693, -0.693)	(0.012, -0.008)	(0.102, 0.102)	(0.106, 0.103)	(0.933, 0.944)
			0	(0.693, -0.693)	(-0.001, 0.008)	(0.102, 0.102)	(0.099, 0.101)	(0.953, 0.948)
			0.5	(0.693, -0.693)	(-0.005, -0.002)	(0.102, 0.102)	(0.104, 0.103)	(0.944, 0.944)
		0.5	-0.5	(0.693, -0.693)	(0.002, -0.006)	(0.102, 0.102)	(0.109, 0.108)	(0.935, 0.942)
			0	(0.693, -0.693)	(-0.004, 0.002)	(0.102, 0.102)	(0.103, 0.101)	(0.948, 0.948)
			0.5	(0.693, -0.693)	(-0.006, 0.010)	(0.102, 0.102)	(0.114, 0.105)	(0.924, 0.937)

Table 3.3: Simulations Results for $\widehat{\sigma}_{\nu}^{2}$ and $\widehat{\gamma}$ based on $n = 500$ and 1000 Replicates

% C	$\boldsymbol{\beta}$	σ_{ν}^{2}	BIAS	ASE	ESD	CP	γ	BIAS	ASE	ESD	CP
$\approx 30\%$	(1, 0.6, -0.6)	0.3	-0.000	0.028	0.028	0.937	-0.5	-0.021	0.131	0.133	0.939
		0.3	-0.001	0.027	0.027	0.939	0	-0.028	0.130	0.134	0.935
		0.3	-0.002	0.028	0.027	0.951	0.5	-0.031	0.132	0.138	0.930
		0.5	-0.002	0.043	0.042	0.948	-0.5	-0.016	0.097	0.104	0.932
		0.5	-0.002	0.042	0.042	0.951	0	-0.016	0.095	0.100	0.939
		0.5	-0.009	0.042	0.042	0.935	0.5	-0.024	0.099	0.100	0.930
	(1, 1, -1)	0.3	-0.000	0.028	0.029	0.944	-0.5	-0.025	0.133	0.134	0.946
		0.3	-0.002	0.028	0.028	0.938	0	-0.033	0.132	0.137	0.932
		0.3	-0.002	0.028	0.027	0.953	0.5	-0.033	0.134	0.140	0.929
		0.5	-0.003	0.043	0.043	0.949	-0.5	-0.020	0.098	0.104	0.929
		0.5	-0.003	0.043	0.042	0.944	0	-0.018	0.096	0.102	0.935
		0.5	-0.010	0.042	0.042	0.936	0.5	-0.025	0.100	0.100	0.939
$\approx 20\%$	(1, 0.6, -0.6)	0.3	0.001	0.029	0.029	0.950	-0.5	-0.013	0.112	0.123	0.911
		0.3	-0.001	0.029	0.029	0.944	0	-0.024	0.110	0.112	0.938
		0.3	-0.003	0.029	0.031	0.922	0.5	-0.032	0.114	0.117	0.934
		0.5	-0.003	0.044	0.044	0.953	-0.5	-0.010	0.086	0.086	0.936
		0.5	-0.006	0.044	0.043	0.944	0	-0.012	0.083	0.084	0.941
		0.5	-0.012	0.044	0.043	0.940	0.5	-0.023	0.087	0.089	0.928
	(1, 1, -1)	0.3	0.001	0.029	0.030	0.944	-0.5	-0.018	0.113	0.124	0.925
		0.3	-0.001	0.029	0.029	0.951	0	-0.026	0.111	0.114	0.926
		0.3	-0.004	0.029	0.031	0.925	0.5	-0.034	0.115	0.119	0.924
		0.5	-0.004	0.045	0.044	0.944	-0.5	-0.012	0.086	0.087	0.938
		0.5	-0.007	0.044	0.043	0.932	0	-0.014	0.083	0.084	0.940
		0.5	-0.014	0.044	0.043	0.928	0.5	-0.026	0.087	0.089	0.921

age, race, gender, height, time on dialysis at study entry, as well as the following list of comorbid conditions: coronary artery disease (CAD), cancer, cardiovascular disease (CVD), stroke, congestive heart failure (CHF), diabetes, hypertension, chronic obstructive pulmonary disease (COPD), psychiatric disorder and peripheral vascular disease (PVD).

Since hospitalization and death times are recorded in days, t represents day (i.e, day post DOPPS entry) in our analysis. The mean number of hospital admissions was 0.595 per patient, while the median length of stay per visit was 5 days. Observed follow-up time had a median of 326 days. Approximately 3% of subjects (154 subjects) were observed to die.

We fitted the model as described in Section $3.2.1$ to the DOPPS data. Specifically, Gauss-Hermite quadrature with 15 nodes is used to approximate the integral of likelihood function. Covariate effects for the out-of-hospital given alive (i.e. prevalence) process and the death hazard are summarized in Table 3.4. For the prevalence process, coronary artery disease ($p < 0.001$), cancer ($p = 0.006$), cardiovascular disease ($p = 0.004$), stroke ($p = 0.045$), chronic obstructive

pulmonary disease ($p = 0.004$), psychiatric disorder ($p = 0.011$), and peripheral vascular disease ($p = 0.048$) have significant negative effects on the probability of being out-of-hospital given survival. Patients from Belgium ($p < 0.001$), Canada ($p < 0.001$), China ($p = 0.002$), Gulf Coast Consortium ($p = 0.001$), Germany ($p < 0.001$), Italy ($p < 0.001$), Japan ($p < 0.001$), Spain ($p = 0.002$), Sweden ($p < 0.001$), and the United Kingdom ($p < 0.001$) have significantly lower out-of-hospital probability than patients from the U.S. (reference). Time on dialysis (coded in year) has a significantly negative effect ($p = 0.003$) on prevalence. For the mortality hazard, coronary artery disease ($p = 0.009$) , cancer ($p = 0.001$), and chronic obstructive pulmonary disease ($p = 0.030$) are associated with significantly higher death hazard. Patients from Italy ($p = 0.042$) have a significant higher mortality rate than patents from the U.S. Increasing age ($p = 0.003$) is also associated with a higher death hazard.

The frailty variance is estimated at $\widehat{\sigma}_\nu^2 = 4.546$ with standard error 0.158, indicating that there is significant within-patient correlation given covariates. The estimated scale parameter $\widehat{\gamma} = -0.350$ with standard error 0.037. This indicates that, given the covariates, the out-of-hospital process and death hazard are negative correlated, which makes sense intuitively; i.e., patients that are more ill (than accounted for their covariate pattern) could be less likely to be out-of-hospital while alive, and more likely to die. Moreover, the frailty effects are stronger in the prevalence process than that in the mortality process. We also compare the frailty estimates from the GLMM with logit link using only the prevalence process. Figure 3.1 shows relatively high concordance between the frailty estimates from the two models (Kendall's tau correlation: 0.885, Pearson correlation: 0.977, and Spearman's rank correlation: 0.971). The majority of the absolute difference of frailty terms between two models are below 0.7 (Figure 3.2).

3.5 Discussion

In this manuscript, we propose a method to jointly model the prevalence process and the mortality hazard using a shared frailty model, allowing a scale parameter in the death process to be estimated.

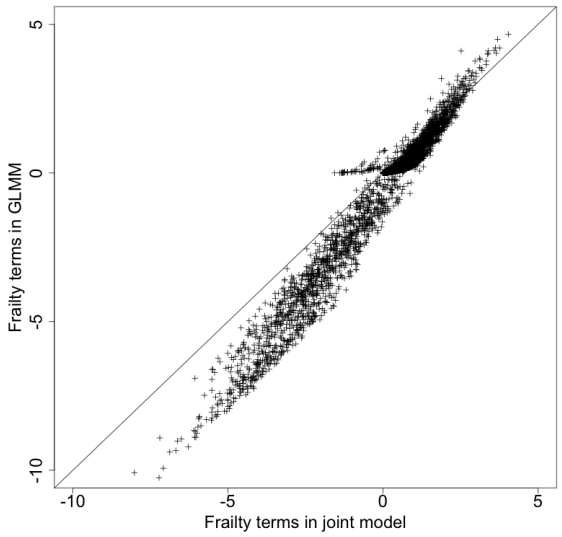

Figure 3.1: Scatterplots of Frailty Terms

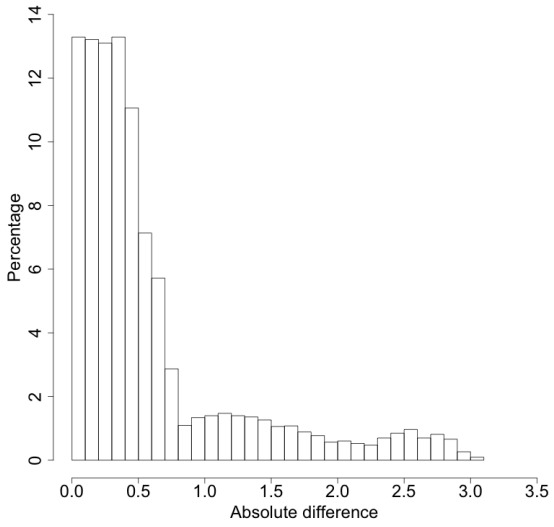

Figure 3.2: Histogram of Absolute Difference of Frailty Terms

The frailty terms are treated as fixed parameters in the iterative estimating procedures. We also utilize a resampling method to obtain a robust variance estimator for the regression parameter estimator.

In the application of the proposed methods to the DOPPS data, the scale parameter was estimated at $\widehat{\gamma} = -0.350$. This means that, given covariates, the out-of-hospital and mortality process are negatively correlated, perhaps due to unmeasured confounders. Since $|\widehat{\gamma}| < 1$, then the prevalence process appears to have larger frailty variance than the mortality process. This finding is consistent with the DOPPS data, since the contents of the covariate set for which data are obtained are choosen primarily based on the perceived effects on mortality. From this perspective, it is possible that some essential predictors for the prevalence model are not included in the analysis.

In Section 3.2.1, we described our iterative estimating procedures where frailty terms are updated in each iteration. An alternative method would be to estimate frailty terms and variance σ_ν^2 using only the prevalence model, similar to a GLMM. Fixing each frailty term and random variance at their estimated values, the algorithm will converge faster. However, this alternative method would lose some efficiency, because mortality information is abandoned. Moreover, it would likely introduce more biases to the frailty estimates for subjects with small death time.

The advantage of estimating frailty terms at each iteration is the allowance of using standard softwares, for example R [R Development Core Team, 2008] and SAS [SAS Institute Inc., 2013]. Having $\widehat{\boldsymbol{\nu}}$ and $\widehat{\gamma}$, one could use the analog to logistic regression and Cox model to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$ as described before. Similar to the estimation of γ and ${\sigma_\nu}^2$, numerical approximation approached could also be used to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

Table 3.4: Analysis of DOPPS Data: Covariate Effects on the Out-of-Hospital given Alive Process and Mortality Process

	Prevalence model			Mortality model			
	$\widehat{\beta}$	$\widehat{\text{SE}}(\widehat{\beta})$	p	$\widehat{\alpha}$	$\widehat{\text{SE}}(\widehat{\alpha})$	p	exp($\widehat{\alpha}$)
Intercept	7.510	0.204	<0.001*	-	-	-	-
Female	-0.137	0.194	0.481	-0.193	0.223	0.388	0.825
CAD[a]	-0.886	0.184	<0.001*	0.561	0.213	0.009*	1.752
Cancer	-0.618	0.224	0.006*	0.683	0.211	0.001*	1.980
CVD[b]	-0.582	0.203	0.004*	0.419	0.227	0.065	1.520
Stroke	-0.460	0.230	0.045*	-0.035	0.284	0.901	0.965
CHF[c]	-0.099	0.184	0.590	0.222	0.203	0.274	1.248
Diabetes	-0.060	0.163	0.712	-0.284	0.178	0.109	0.752
Hypertension	0.243	0.207	0.241	-0.092	0.216	0.670	0.912
COPD[d]	-0.654	0.225	0.004*	0.520	0.240	0.030*	1.682
Psychiatric Disorder	-0.568	0.225	0.011*	0.460	0.264	0.082	1.584
PVD[e]	-0.362	0.183	0.048*	0.164	0.215	0.445	1.178
Belgium	-1.968	0.323	<0.001*	-0.279	0.525	0.596	0.757
Canada	-1.406	0.312	<0.001*	-0.216	0.386	0.576	0.806
China	-1.325	0.434	0.002*	-0.401	1.017	0.693	0.670
Gulf	-1.215	0.365	0.001*	-0.025	0.531	0.963	0.976
Germany	-2.179	0.245	<0.001*	-0.051	0.324	0.874	0.950
Italy	-1.195	0.288	<0.001*	0.582	0.287	0.042*	1.789
Japan	-1.455	0.286	<0.001*	-0.647	0.366	0.077	0.524
Spain	-1.044	0.334	0.002*	0.028	0.374	0.941	1.028
Sweden	-2.094	0.331	<0.001*	0.082	0.527	0.876	1.086
UK	-1.378	0.380	<0.001*	-0.179	0.520	0.730	0.836
Age (per 5 years)	0.018	0.028	0.537	0.103	0.034	0.003*	1.108
Time on dialysis (years)	-3.092	1.050	0.003*	0.709	1.093	0.517	2.031
Height (per 10 cm)	0.064	0.103	0.536	-0.116	0.112	0.299	0.890

a: Coronary artery disease
b: Cardiovascular disease
c: Congestive heart failure
d: Chronic obstructive pulmonary disease
e: Peripheral vascular disease

APPENDIX A

Supplemental Materials to Chapter 1

A.1 Notations and Assumptions

We first list the notations and assumptions that are presented in the main paper.

$A_i^0(t) = I\{H_i(t) = 0, D_i > t\}$: Survival-out-of-hospital indicator for subject i at time t.

$\pi_i(t) \equiv E[A_i^0(t)|\boldsymbol{Z}_i(t)]$: Survival-out-of-hospital probability for subject i at time t.

$dM_i^*(t;\boldsymbol{\beta}) = A_i^0(t)dt - \pi_0(t)\exp[\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)]dt$: Survival-out-of-hospital Martingale increment for subject i at time t.

$dM_i(t;\boldsymbol{\beta}) = dM_i^*(t;\boldsymbol{\beta})I(C_i \geq t)$: Observed survival-out-of-hospital Martingale increment for subject i at time t.

$N_i^C(t) = I(C_i \leq t \wedge D_i)$: Observed censoring time counting process for subject i at time t.

$dN_i^C(t) = N_i^C(t^- + dt) - N_i^C(t^-)$: Censoring time counting process increment for subject i at time t.

$X_i = D_i \wedge C_i$: Observed terminal event for subject i.

$Y_i(t) = I(X_i \geq t)$: At-risk process for censoring time for subject i at time t.

$I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})$: At-risk process when $C_i^{\langle m\rangle}$ is imputed from $\widehat{G}(t;\widehat{\boldsymbol{\theta}})$ for subject i at time t.

$I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)$: At-risk process when $C_i^{\langle m\rangle}$ is imputed from $G(t;\boldsymbol{\theta}_0)$ for subject i at time t.

$dM_i^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = dM_i^*(t;\boldsymbol{\beta})I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta})$: Observed survival-out-of-hospital Martingale increment for imputation m for subject i at time t.

Let $\boldsymbol{S}^{(k)}(t;\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i(t)^{\otimes k}I(C_i \geq t)\text{exp}\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}$, and $\boldsymbol{s}^{(k)}(t;\boldsymbol{\beta}) = E[\boldsymbol{Z}_1(t)^{\otimes k}I(C_1 \geq t)\text{exp}\{\boldsymbol{\beta}^T\boldsymbol{Z}_1(t)\}]$ be the limiting value of $\boldsymbol{S}^{(k)}(t;\boldsymbol{\beta})$ for $k = 0, 1, 2$, where $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. Also let $\bar{\boldsymbol{Z}}(t;\boldsymbol{\beta}) = \boldsymbol{S}^{(1)}(t;\boldsymbol{\beta})/S^{(0)}(t;\boldsymbol{\beta})$, and denote its limiting value as $\bar{\boldsymbol{z}}(t;\boldsymbol{\beta}) = \boldsymbol{s}^{(1)}(t;\boldsymbol{\beta})/s^{(0)}(t;\boldsymbol{\beta})$. Moreover, define $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\int_0^\tau\{\boldsymbol{S}^{(2)}(t;\boldsymbol{\beta})/S^{(0)}(t;\boldsymbol{\beta}) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\beta})^{\otimes 2}\}A_i^0(t)I(C_i \geq t)dt$ and its limiting value $\boldsymbol{\Omega}(\boldsymbol{\beta}) = E[\int_0^\tau\{\boldsymbol{s}^{(2)}(t;\boldsymbol{\beta})/s^{(0)}(t;\boldsymbol{\beta}) - \bar{\boldsymbol{z}}(t;\boldsymbol{\beta})^{\otimes 2}\}A_1^0(t)I(C_1 \geq t)dt]$.

For mth imputed dataset, let $\boldsymbol{S}^{(k)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i(t)^{\otimes k}I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta})\text{exp}\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}$, and $\boldsymbol{s}^{(k)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = E[\boldsymbol{Z}_1(t)^{\otimes k}I(C_1^{\langle 1\rangle} \geq t;\boldsymbol{\theta})\text{exp}\{\boldsymbol{\beta}^T\boldsymbol{Z}_1(t)\}]$ be the limiting value of $\boldsymbol{S}^{(k)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$ for $k = 0, 1, 2$. Also let $\bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = \boldsymbol{S}^{(1)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})/S^{(0)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$, and denote its limiting value as $\bar{\boldsymbol{z}}^{\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) = \boldsymbol{s}^{(1)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})/s^{(0)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$. Moreover, define $\widehat{\boldsymbol{\Omega}}^{\langle m\rangle}(\boldsymbol{\beta},\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\int_0^\tau\{\boldsymbol{S}^{(2)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})/S^{(0)\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) - \bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})^{\otimes 2}\}A_i^0(t)I(C_i^{\langle m\rangle} \geq t)dt$ and its limiting value $\boldsymbol{\Omega}^{\langle 1\rangle}(\boldsymbol{\beta},\boldsymbol{\theta}) = E[\int_0^\tau\{\boldsymbol{s}^{(2)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})/s^{(0)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}) - \bar{\boldsymbol{z}}^{\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})^{\otimes 2}\}A_1^0(t)I(C_1^{\langle 1\rangle} \geq t)dt]$.

Proportional marginal rate model on the hospital-free survival event:

$$\pi_i(t) = \pi_0(t)\text{exp}\{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i(t)\} \tag{A.1}$$

Conditional independent censoring assumption:

$$E\{A_i^0(t)|\boldsymbol{Z}_i(t), I(C_i \geq t)\} = E\{A_i^0(t)|\boldsymbol{Z}_i(t)\} \tag{A.2}$$

Proportional hazard model on the censoring time:

$$\lambda_i^C(t) = \lambda_0^C(t)\text{exp}\{\boldsymbol{\theta}^T\boldsymbol{Z}_i(t)\} \tag{A.3}$$

A.2 Proof of Theorem 1

A.2.1 Consistency of $\widehat{\boldsymbol{\beta}}$

We first prove that $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$. Let

$$X_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} [(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \boldsymbol{Z}_i(t) - \log \frac{S^{(0)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}_0, t)}] A_i^0(t) I(C_i \geq t) dt$$

$$X(\boldsymbol{\beta}) = E[\int_0^{\tau} [(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \boldsymbol{Z}_1(t) - \log \frac{s^{(0)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)}] A_1^0(t) I(C_1 \geq t) dt]$$

Note that $\partial X_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = U_n(\boldsymbol{\beta})$.

For an arbitrary $\delta > 0$, consider a compact set $\mathscr{B}_\delta = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \delta\}$. Evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\partial X(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = E[\int_0^{\tau} \{\boldsymbol{Z}_1(t) - \bar{z}(t; \boldsymbol{\beta}_0)\} A_1^0(t) I(C_1 \geq t) dt] = 0$ by model assumptions, while $-\partial^2 X(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T = \boldsymbol{\Omega}(\boldsymbol{\beta}_0)$, which is positive definite by condition (c). Therefore, $X(\boldsymbol{\beta})$ is concave and has a local maximum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. We have $X(\boldsymbol{\beta}_0) \geq X(\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in \mathscr{B}_\delta$, and $X(\boldsymbol{\beta}_0) > X(\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in \partial \mathscr{B}_\delta$, where $\partial \mathscr{B}_\delta = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| = \delta\}$. For $\boldsymbol{\beta} \in \mathscr{B}_\delta$, by SLLN, $n^{-1} X_n(\boldsymbol{\beta}) \stackrel{a.s.}{\to} X(\boldsymbol{\beta})$. By continuous mapping theorem,

$$\|n^{-1} X_n(\boldsymbol{\beta}) - n^{-1} X_n(\boldsymbol{\beta}_0)\| \stackrel{a.s.}{\to} \|X(\boldsymbol{\beta}) - X(\boldsymbol{\beta}_0)\|$$

where $\|a\| = (a^T a)^{1/2}$. Therefore as $n \to \infty$, $X_n(\boldsymbol{\beta}_0) \geq X_n(\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in \mathscr{B}_\delta$, and $X_n(\boldsymbol{\beta}_0) > X_n(\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in \partial \mathscr{B}_\delta$. In this case, when $n \to \infty$, $X_n(\boldsymbol{\beta})$ has a maximum $\boldsymbol{\beta}_0$ in $\mathscr{B}_\delta$. Moreover, evaluated at $\widehat{\boldsymbol{\beta}}$, $\partial X_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = U(\boldsymbol{\beta}) = 0$, meaning that $\widehat{\boldsymbol{\beta}}$ is the local maximum. Taking δ arbitrary small, we would have $\widehat{\boldsymbol{\beta}} \stackrel{a.s.}{\to} \boldsymbol{\beta}_0$.

A.2.2 $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$

To show the asymptotic distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, we take the first-order Taylor expansion of $\boldsymbol{U}_n(\boldsymbol{\beta}_0)$ around $\widehat{\boldsymbol{\beta}}$ to get $n^{-1/2} \boldsymbol{U}_n(\boldsymbol{\beta}_0) = n^{1/2} \widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, where $\boldsymbol{\beta}^*$ is on the line segments between $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$. The consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta}_0)$ for $\boldsymbol{\beta}_0$ and $\boldsymbol{\Omega}(\boldsymbol{\beta}_0)$, along with the positive

definite of $\boldsymbol{\Omega}(\boldsymbol{\beta}_0)$ imply that

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = n^{-1/2}\boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{U}_n(\boldsymbol{\beta}_0) + o_p(1) \tag{A.4}$$

Then we express the estimating equation $\boldsymbol{U}_n(\boldsymbol{\beta}_0)$ as

$$\boldsymbol{U}_n(\boldsymbol{\beta}_0) = \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\beta}_0)\} dM_i(t;\boldsymbol{\beta}_0) = 0 \tag{A.5}$$

where $dM_i(t;\boldsymbol{\beta}) = A_i^0(t)I(C_i \geq t)dt - I(C_i \geq t)\exp[\boldsymbol{\beta}^T(t)\boldsymbol{Z}_i(t)]\pi_0(t)dt$. Without lose of generality, we assume that $\boldsymbol{Z}_i(t) \geq 0$. Therefore, $\boldsymbol{S}^{(1)}(t;\boldsymbol{\beta}_0)$ and $S^{(0)}(t;\boldsymbol{\beta}_0)$ are monotone functions in t and converges almost surely to $\boldsymbol{s}^{(1)}(t;\boldsymbol{\beta}_0)$ and $s^{(0)}(t;\boldsymbol{\beta}_0)$ by SLLN. Then applying Lemma 1 of Lin et al. (2000), we have

$$n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\bar{\boldsymbol{Z}}(t;\boldsymbol{\beta}_0) - \bar{\boldsymbol{z}}(t;\boldsymbol{\beta}_0)\} dM_i(t;\boldsymbol{\beta}_0) = o_p(1) \tag{A.6}$$

Combining (A.4), (A.5) and (A.6), we could represent $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ as a sum of independent and identically distributed random variables,

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) + o_p(1) \tag{A.7}$$

where $\boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \boldsymbol{\Omega}(\boldsymbol{\beta})^{-1}\boldsymbol{u}_i(\boldsymbol{\beta})$ and $\boldsymbol{u}_i(\boldsymbol{\beta}) = \int_0^{\tau}\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{z}}(t;\boldsymbol{\beta})\}dM_i(t;\boldsymbol{\beta})$. Since $\boldsymbol{S}^{(k)}(t;\boldsymbol{\beta}_0)$ $(k = 0, 1, 2)$ and $\boldsymbol{Z}_i(t)$ are totally bounded by condition (c), then Theorem 1 follows by Multivariate Central Limit Theorems. The consistency of $\widehat{\Sigma}(\widehat{\boldsymbol{\beta}})$ could be proved by iteratively applying SLLN, Lemma 1 of Lin et al (2000) and the strong consistency of $\widehat{\boldsymbol{\beta}}$.

A.3 Proof of Theorem 2

For $t \in [0, \tau]$, We normalize $\widehat{\pi}_0(t)$ and $\pi_0(t)$ to a proper scale for the proof of asymptotic properties:

$$\widehat{\pi}_0(t) = \frac{n^{-1/2} \sum_{i=1}^n A_i^0(t) I(C_i \geq t)}{n^{-1/2} \sum_{i=1}^n I(C_i \geq t) \text{exp}[\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)]}$$

$$\pi_0(t) = \frac{n^{1/2} E[A_1^0(t) I(C_1 \geq t)]}{n^{1/2} E\{I(C_1 \geq t) \text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)]\}}$$

We introduce some empirical process notations for the ease for derivations. Let $\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f_i(F_i)$ and $Pf = E[f_1(F_1)]$. The empirical process $\{\mathbb{G}_n f(t) : t \in [0, \tau]\}$, where $\mathbb{G}_n f(t) = n^{1/2}[\mathbb{P}_n f(t) - Pf(t)]$, can be identified as the empirical distribution of observations indexed by the function $f_i(t, F_i)$, for $t \in [0, \tau]$. In the following proof, we will drop the notation of data F_i in the parenthesis. The quantity in the numerator could be identified as

$$n^{1/2}\Big(n^{-1} \sum_{i=1}^n [A_i^0(t) I(C_i \geq t)] - E[A_1^0(t) I(C_1 \geq t)]\Big) = \mathbb{G}_n \tilde{f}^{\pi_1}(t)$$

where $\tilde{f}_i^{\pi_1}(t) = A_i^0(t) I(C_i \geq t)$. Consider the quantity $n^{1/2}\Big(n^{-1} \sum_{i=1}^n \{I(C_i \geq t) \text{exp}[\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)]\} - E\{I(C_1 \geq t) \text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)]\}\Big)$ in the denominator, which can be decomposed as

$$\begin{aligned}
& n^{1/2}[S^{(0)}(t; \widehat{\boldsymbol{\beta}}) - s^{(0)}(t; \boldsymbol{\beta}_0)] \\
=& n^{1/2}\Big(n^{-1} \sum_{i=1}^n \{I(C_i \geq t) \text{exp}[\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)]\} - n^{-1} \sum_{i=1}^n \{I(C_i \geq t) \text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)]\}\Big) \qquad \text{(A.8)} \\
+& n^{1/2}\Big(n^{-1} \sum_{i=1}^n \{I(C_i \geq t) \text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)]\} - E\{I(C_1 \geq t) \text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)]\}\Big) \qquad \text{(A.9)}
\end{aligned}$$

By Taylor expansion and (A.7), the first part (A.8) could be expressed as

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n} I(C_i \geq t)\{\exp[\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)] - \exp[\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)]\} \\
=& n^{-1/2}\sum_{i=1}^{n} I(C_i \geq t)\exp[\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)]\boldsymbol{Z}_i^T(t) n^{-1}\sum_{l=1}^{n} \boldsymbol{f}_l^{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) + o_p(1) \\
=& n^{-1/2}\sum_{i=1}^{n} \boldsymbol{S}^{(1)}(t;\boldsymbol{\beta}_0)^T \boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) + o_p(1) \\
=& n^{-1/2}\sum_{i=1}^{n} \boldsymbol{s}^{(1)}(t;\boldsymbol{\beta}_0)^T \boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) + o_p(1) \qquad \text{(A.10)}
\end{aligned}
$$

where (A.10) is by the strong consistency and monotonic property of $\boldsymbol{S}^{(1)}(t;\boldsymbol{\beta}_0)$. Combining (A.10) and (A.9) we have

$$
n^{1/2}\Big(n^{-1}\sum_{i=1}^{n}\{I(C_i \geq t)\exp[\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)]\} - E\{I(C_1 \geq t)\exp[\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)]\}\Big) = \mathbb{G}_n \tilde{f}^{\pi_2}(t;\boldsymbol{\beta}_0) + o_p(1) \qquad \text{(A.11)}
$$

where $\tilde{f}_i^{\pi_2}(t;\boldsymbol{\beta}) = \boldsymbol{s}^{(1)}(t;\boldsymbol{\beta}_0)^T \boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) + I(C_i \geq t)\exp[\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)]$.

By the empirical central limit theorem (van der Vaart and Wellner (1996), example 3.9.19, p.g. 383), we have:

$$
n^{1/2}\begin{pmatrix} \mathbb{P}_n \tilde{f}^{\pi_1} - P\tilde{f}^{\pi_1} \\ S^{(0)}(\widehat{\boldsymbol{\beta}}) - s^{(0)}(\boldsymbol{\beta}_0) \end{pmatrix} \Longrightarrow \begin{pmatrix} \mathbb{G}\tilde{f}^{\pi_1} \\ \mathbb{G}\tilde{f}^{\pi_2}[\boldsymbol{\beta}_0] \end{pmatrix}
$$

where $\mathbb{G}\tilde{f}^{\pi_1}$ and $\mathbb{G}\tilde{f}^{\pi_2}$ are tight Gaussian processes. Define a map

$$
(A, B) \mapsto \frac{A}{B}
$$

Since this composition map is Hadamard-differentiable, $1/B$ is bounded away from zero and A is of bounded variation, then its derivative map is given by

$$
(a, b) \mapsto \frac{a}{B} - \frac{Ab}{B^2}
$$

By Functional Delta method (van der Vaart (1998), p.g. 291)

$$
\begin{aligned}
n^{1/2}[\widehat{\pi}_0 - \pi_0] \Longrightarrow & \mathbb{G}\Big\{ \frac{\tilde{f}^{\pi_1}}{s^{(0)}(\boldsymbol{\beta}_0)} - \frac{[P\tilde{f}^{\pi_1}]\tilde{f}^{\pi_2}(\boldsymbol{\beta}_0)}{[s^{(0)}(\boldsymbol{\beta}_0)]^2} \Big\} \\
= & \mathbb{G}\Big\{ \frac{A^0(t)I(C \geq t) - I(C \geq t)\text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)]\pi_0(t) - \boldsymbol{s}^{(1)}(t;\boldsymbol{\beta}_0)^T \boldsymbol{f}^{\boldsymbol{\beta}}(\boldsymbol{\beta}_0)\pi_0(t)}{s^{(0)}(\boldsymbol{\beta}_0)} \Big\} \\
\equiv & \mathbb{G}\Big\{ \frac{f^{\pi_1} - f^{\pi_2}}{s^{(0)}(\boldsymbol{\beta}_0)} \Big\} \qquad \text{(A.12)}
\end{aligned}
$$

Therefore Theorem 2 follows, and by the consistency of $\widehat{\boldsymbol{\beta}}$ and $\mathbb{P}_n f$ to $\boldsymbol{\beta}_0$ and Pf, respectively, we have $\widehat{\sigma}(s,t)$ converges in probability to $\sigma(s,t)$.

A.4 Proof of Theorem 3

A.4.1 Consistency of $\widehat{\boldsymbol{\beta}}^M$

Before establishing the consistent property, we first show some useful results. When $C_i^{\langle m \rangle}$ is imputed with the true underlying $G(t;\boldsymbol{\theta}_0)$, $I(C_i^{\langle m \rangle} \geq t;\boldsymbol{\theta}_0)$ and $A_i^0(t)$ are independent condition on $\boldsymbol{Z}_i(t)$ by (A.2):

$$
E[A_i^0(t)I(C_i^{\langle m \rangle} \geq t;\boldsymbol{\theta}_0)|\boldsymbol{Z}_i(t)] = E[A_i^0(t)|\boldsymbol{Z}_i(t)]E[I(C_i^{\langle m \rangle} \geq t;\boldsymbol{\theta}_0)|\boldsymbol{Z}_i(t)] = 0
$$

$$
E[\{A_i^0(t)I(C_i^{\langle m \rangle} \geq t;\boldsymbol{\theta}_0)\}^2|\boldsymbol{Z}_i(t)] = E[\{A_i^0(t)\}^2|\boldsymbol{Z}_i(t)]E[\{I(C_i^{\langle m \rangle} \geq t;\boldsymbol{\theta}_0)\}|\boldsymbol{Z}_i(t)]
$$

Next consider $I(C_i^{\langle m \rangle} \geq t;\widehat{\boldsymbol{\theta}})$, where $C_i^{\langle m \rangle}$ are imputed from $\widehat{G}(t;\widehat{\boldsymbol{\theta}})$ if $N_i^D(X_i) = 1$. The survival function of imputed censoring time given $\boldsymbol{Z}_i(t)$ and $N_i^D(X_i) = 1$ is given by

$$
E[I(C_i^{\langle m \rangle} \geq t;\widehat{\boldsymbol{\theta}}|\boldsymbol{Z}_i(t), N_i^D(X_i) = 1] = I(t \geq D_i)\text{exp}\{[\widehat{\Lambda}_i^C(D_i;\widehat{\boldsymbol{\theta}}) - \widehat{\Lambda}_i^C(t;\widehat{\boldsymbol{\theta}})]
$$

where $d\widehat{\Lambda}_i(t;\boldsymbol{\theta}) = d\widehat{\Lambda}_0(t;\boldsymbol{\theta})\exp[\boldsymbol{\theta}^T\boldsymbol{Z}_i(t)]$ and the Breslow estimator $d\widehat{\Lambda}_0^C(t;\boldsymbol{\theta})$ is given by

$$d\widehat{\Lambda}_0^C(t;\boldsymbol{\theta}) = \frac{\sum_{i=1}^n dN_i^C(t)}{\sum_{i=1}^n Y_i(t)\exp[\boldsymbol{\theta}^T\boldsymbol{Z}_i(t)]}$$

Using the strong consistency of $\widehat{\boldsymbol{\theta}}$ and SLLN, we have $d\widehat{\Lambda}_0^C(t;\widehat{\boldsymbol{\theta}}) - d\Lambda_0^C(t)$ converges almost surely to zero. Applying continuous mapping theorem, we have

$$E[I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})|\boldsymbol{Z}_i(t), N_i^D(X_i) = 1] \xrightarrow{a.s.} E[I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)|\boldsymbol{Z}_i(t), N_i^D(X_i) = 1] \tag{A.13}$$

which means that asymptotically, $C_i^{\langle m\rangle}$ are drawn from the true $G(t;\boldsymbol{\theta}_0)$ for subjects with $N_i^D(X_i) = 1$. Note that for subjects $N_i^D(X_i) = 0$, $I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0) = I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}}) = I(C_i \geq t)$, then combining (A.13) we have

$$E[I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})|\boldsymbol{Z}_i(t)] = E[I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)|\boldsymbol{Z}_i(t)] + o(1) \tag{A.14}$$

Therefore, $I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})$ and $A_i^0(t)$ are asymptotically independent condition on $\boldsymbol{Z}_i(t)$. More explicitly,

$$E[A_i^0(t)I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})|\boldsymbol{Z}_i(t)] = E[A_i^0(t)|\boldsymbol{Z}_i(t)]E[I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})|\boldsymbol{Z}_i(t)] + o(1) = o(1) \tag{A.15}$$

$$E[\{A_i^0(t)I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})\}^2|\boldsymbol{Z}_i(t)] = E[\{A_i^0(t)\}^2|\boldsymbol{Z}_i(t)]E[\{I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})|\boldsymbol{Z}_i(t)] + o(1) \tag{A.16}$$

Then we will demonstrate that the two processes $n^{-1/2}\sum_{i=1}^n \int_0^t dM_i^*(t;\boldsymbol{\beta})I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})$ and $n^{-1/2}\sum_{i=1}^n \int_0^t dM_i^*(t;\boldsymbol{\beta})I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)$ converge to the same limiting process. By Donsker's theorem (van der Vaart 2000), $n^{-1/2}\sum_{i=1}^n \int_0^t dM_i^*(t;\boldsymbol{\beta})I(C_i^{\langle m\rangle} \geq t;\boldsymbol{\theta}_0)$ converges to a Gaussian process. By (A.14), (A.15) and (A.16), we have that $n^{-1/2}\sum_{i=1}^n \int_0^t dM_i^*(t;\boldsymbol{\beta})I(C_i^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\theta}})$

converges to the same limiting Gaussian process, i.e.,

$$n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau} dM_i^*(t;\boldsymbol{\beta})I(C_i^{\langle m\rangle}\geq t;\widehat{\boldsymbol{\theta}}) = n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau} dM_i^*(t;\boldsymbol{\beta})I(C_i^{\langle m\rangle}\geq t;\boldsymbol{\theta}_0)+o_p(1) \tag{A.17}$$

Now we are ready to show the consistency. Within the mth imputed dataset, define

$$X_n^{\langle m\rangle}(\boldsymbol{\beta},\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\int_0^{\tau}[(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{Z}_i(t) - \log\frac{S^{(0)\langle m\rangle}(t;\boldsymbol{\beta},\widehat{\boldsymbol{\theta}})}{S^{(0)\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})}]A_i^0(t)I(C_i^{\langle m\rangle}\geq t;\widehat{\boldsymbol{\theta}})dt]$$

$$X^{\langle 1\rangle}(\boldsymbol{\beta},\boldsymbol{\theta}_0) = E[\int_0^{\tau}[(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{Z}_1(t) - \log\frac{s^{(0)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}_0)}{s^{(0)\langle 1\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)}]A_1^0(t)I(C_1^{\langle 1\rangle}\geq t;\boldsymbol{\theta}_0)dt]$$

Note that $A_i^0(t)I(C_i^{\langle m\rangle}\geq t;\boldsymbol{\theta}_0) = A_i^0(t)I(C_i^{\langle m\rangle}\geq t;\widehat{\boldsymbol{\theta}})$, because $A_i^0(t)=0$ for $t\in[D_i,\tau]$. We have $\boldsymbol{S}^{(k)\langle m\rangle}(t;\boldsymbol{\beta},\widehat{\boldsymbol{\theta}})$ converges almost surely to $\boldsymbol{s}^{(k)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta}_0)$ by SLLN, (A.14) and (A.15), for $k=0,1,2$. Then by SLLN and the bounded total variations of $\boldsymbol{Z}_i(t)$, we get $n^{-1}X_n^{\langle m\rangle}(\boldsymbol{\beta},\widehat{\boldsymbol{\theta}})\xrightarrow{a.s.} X^{\langle 1\rangle}(\boldsymbol{\beta},\boldsymbol{\theta}_0)$. By (A.2) and (A.3), $X^{\langle 1\rangle}(\boldsymbol{\beta},\boldsymbol{\theta}_0) = X(\boldsymbol{\beta})$. In this case, $n^{-1}X_n^{\langle m\rangle}(\boldsymbol{\beta},\widehat{\boldsymbol{\theta}})\xrightarrow{a.s.} X(\boldsymbol{\beta})$. Following the arguments in the proof of theorem 1, we have $\|\widehat{\boldsymbol{\beta}}^{\langle m\rangle}-\boldsymbol{\beta}_0\|\xrightarrow{a.s.}0$. Since $\|\tilde{\boldsymbol{\beta}}^M-\boldsymbol{\beta}_0\| = M^{-1}\sum_{m=1}^{M}\|\widehat{\boldsymbol{\beta}}^{\langle m\rangle}-\boldsymbol{\beta}_0\|$, then by triangle inequality, $\|\tilde{\boldsymbol{\beta}}^M-\boldsymbol{\beta}_0\|\xrightarrow{a.s.}0$. Therefore, both $\widehat{\boldsymbol{\beta}}^{\langle m\rangle}$ and $\tilde{\boldsymbol{\beta}}^M$ converges almost surely to $\boldsymbol{\beta}_0$.

A.4.2 $n^{1/2}(\widehat{\boldsymbol{\beta}}^M-\boldsymbol{\beta}_0)$

Having established the strong consistency of $\widehat{\boldsymbol{\beta}}^{\langle m\rangle}$, we could apply first-order Taylor expansion on $U_n^{\langle m\rangle}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})$ around $\widehat{\boldsymbol{\beta}}$ and A3 in Lin et al. (2000) to get

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^{\langle m\rangle}-\boldsymbol{\beta}_0) = n^{-1/2}[\widehat{\boldsymbol{\Omega}}^{\langle m\rangle}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})]^{-1}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})\}dM_i^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})+o_p(1) \tag{A.18}$$

Iteratively applying WLLN, and by (A.2) (A.14) we have $\widehat{\boldsymbol{\Omega}}^{\langle m\rangle}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})\xrightarrow{p}\boldsymbol{\Omega}(\boldsymbol{\beta}_0)$. Then consider the quantity $n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})\}dM_i^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})$:

$$n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})\}dM_i^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})$$

$$=n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})\}[dM_i^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})-dM_i^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)] \tag{A.19}$$

$$+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{z}}^{\langle 1\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\}dM_i^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) \tag{A.20}$$

$$+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\bar{\boldsymbol{z}}^{\langle 1\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)-\bar{\boldsymbol{Z}}^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})\}dM_i^{\langle m\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) \tag{A.21}$$

Since $\boldsymbol{Z}_i(t)$ has bounded total variations, then by (A.17), we could show that (A.19) equals to $o_p(1)$. Without lose of generality we assume that $\boldsymbol{Z}_i(t)\geq 0$. Therefore, $\boldsymbol{S}^{(1)\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})$ and $S^{(0)\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})$ are monotone functions in t. By Lemma 1 in LWYY 2000, (A.17), the strong consistency of $\boldsymbol{S}^{(1)\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})$ and $S^{(0)\langle m\rangle}(t;\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}})$, we have (A.21) equals to $o_p(1)$. Combining results we have

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^{\langle m\rangle}-\boldsymbol{\beta}_0)=n^{-1/2}[\boldsymbol{\Omega}(\boldsymbol{\beta}_0)]^{-1}\sum_{i=1}^{n}\boldsymbol{u}_i^{\langle m\rangle}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)+o_p(1)$$

where $\boldsymbol{u}_i^{\langle m\rangle}(\boldsymbol{\beta},\boldsymbol{\theta})=\int_0^{\tau}\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{z}}^{\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})\}dM_i^{\langle m\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})$. Since $\widehat{\boldsymbol{\beta}}^{M}=M^{-1}\sum_{m=1}^{M}\widehat{\boldsymbol{\beta}}^{\langle m\rangle}$, then

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^{M}-\boldsymbol{\beta}_0)=n^{-1/2}[\boldsymbol{\Omega}(\boldsymbol{\beta}_0)]^{-1}\sum_{i=1}^{n}M^{-1}\sum_{m=1}^{M}\boldsymbol{u}_i^{\langle m\rangle}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)+o_p(1)$$

By multivariate central limit theorem, $n^{1/2}(\widehat{\boldsymbol{\beta}}^{M}-\boldsymbol{\beta}_0)$ converges to a mean zero normal random variable, where the variance is given by
$\Sigma_M(\boldsymbol{\beta}_0)=[\boldsymbol{\Omega}(\boldsymbol{\beta}_0)]^{-1}E\{[M^{-1}\sum_{m=1}^{M}\boldsymbol{u}_1^{\langle m\rangle}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]^{\otimes 2}\}[\boldsymbol{\Omega}(\boldsymbol{\beta}_0)]^{-1}$. We could also express the target quantity as

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^{M}-\boldsymbol{\beta}_0)=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,M) \tag{A.22}$$

where $\boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\theta},M)=[\boldsymbol{\Omega}(\boldsymbol{\beta})]^{-1}M^{-1}\sum_{m=1}^{M}\boldsymbol{u}_i^{\langle m\rangle}(\boldsymbol{\beta},\boldsymbol{\theta})$. The consistency of $\widehat{\Sigma}_M(\widehat{\boldsymbol{\beta}}^{M})$ can be

proved in similar manners of Theorem 1.

A.5 Proof of Theorem 4

Since $\widehat{\pi}_0^M(t)$ and $\pi_0(t)$ are given by

$$\widehat{\pi}_0^M(t) = \frac{n^{-1/2}\sum_{i=1}^n M^{-1}\sum_{m=1}^M A_i^0(t) I(C_i^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\theta}})}{n^{-1/2}\sum_{i=1}^n M^{-1}\sum_{m=1}^M I(C_i^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\theta}})\text{exp}[\widehat{\boldsymbol{\beta}}^M \boldsymbol{Z}_i(t)]},$$

$$\pi_0(t) = \frac{E[A_1^0(t) I(C_1^{\langle 1\rangle} \geq t; \boldsymbol{\theta}_0)]}{E\{I(C_1^{\langle 1\rangle} \geq t; \boldsymbol{\theta}_0)\text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)]\}}$$

Since the numerator is automatically an empirical process indexed by function $A_i^0(t)M^{-1}\sum_{m=1}^M I(C_i^{\langle m\rangle} \geq t; \boldsymbol{\theta}_0)$, then we will consider the asymptotic property of the denominator.

A.5.1 $n^{1/2}[M^{-1}\sum_{m=1}^M S^{(0)\langle m\rangle}(t; \widehat{\boldsymbol{\beta}}^M, \widehat{\boldsymbol{\theta}}) - s^{(0)\langle 1\rangle}(t; \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)]$

We could decompose the target quantity as

$$\begin{aligned}
&n^{1/2}[M^{-1}\sum_{m=1}^M S^{(0)\langle m\rangle}(t; \widehat{\boldsymbol{\beta}}^M, \widehat{\boldsymbol{\theta}}) - s^{(0)\langle 1\rangle}(t; \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)] \\
=&n^{-1/2}\Big(\sum_{i=1}^n\{M^{-1}\sum_{m=1}^M I(C_i^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\theta}})\text{exp}[\boldsymbol{Z}(t)^T\widehat{\boldsymbol{\beta}}^M]\} - \sum_{i=1}^n\{M^{-1}\sum_{m=1}^M I(C_i^{\langle m\rangle} \geq t; \boldsymbol{\theta}_0)\text{exp}[\boldsymbol{Z}(t)^T\widehat{\boldsymbol{\beta}}^M]\}\Big) \qquad \text{(A.23)} \\
&+n^{-1/2}\Big(\sum_{i=1}^n\{M^{-1}\sum_{m=1}^M I(C_i^{\langle m\rangle} \geq t; \boldsymbol{\theta}_0)\text{exp}[\boldsymbol{Z}(t)^T\widehat{\boldsymbol{\beta}}^M]\} - \sum_{i=1}^n\{M^{-1}\sum_{m=1}^M I(C_i^{\langle m\rangle} \geq t; \boldsymbol{\theta}_0)\text{exp}[\boldsymbol{\beta}_0^T\boldsymbol{Z}_i(t)]\}\Big) \qquad \text{(A.24)} \\
&+n^{1/2}\Big(n^{-1}\sum_{i=1}^n\{M^{-1}\sum_{m=1}^M I(C_i^{\langle m\rangle} \geq t; \boldsymbol{\theta}_0)\text{exp}[\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)]\} - E\{I(C_1^{\langle 1\rangle} \geq t; \boldsymbol{\theta}_0)\text{exp}[\boldsymbol{\beta}_0^T\boldsymbol{Z}_1(t)]\}\Big) \qquad \text{(A.25)}
\end{aligned}$$

By (A.14), the first term (A.23) can be shown to be $o_p(1)$. By (A.22), the strong consistency of $S^{(2)\langle m\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ and Taylor expansion, we have (A.24) equals to

$$n^{-1/2}\sum_{i=1}^{n}\{\boldsymbol{s}^{(1)\langle 1\rangle}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)^T\boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,M)\}+o_p(1)$$

Combining results, we would have

$$\begin{aligned}&n^{1/2}\Big(n^{-1}\sum_{i=1}^{n}\{M^{-1}\sum_{m=1}^{M}I(C_i^{\langle m\rangle}\geq t;\widehat{\boldsymbol{\theta}})\text{exp}[\boldsymbol{Z}(t)^T\tilde{\boldsymbol{\beta}}^M]\}-E\{I(C_1^{\langle 1\rangle}\geq t;\boldsymbol{\theta}_0)\text{exp}[\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)]\}\Big)\\ =&\mathbb{G}_n\tilde{f}^{\pi_2}(t;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,M)+o_p(1)\end{aligned}\tag{A.26}$$

where $\tilde{f}_i^{\pi_2}(t;\boldsymbol{\beta},\boldsymbol{\theta},M)=M^{-1}\sum_{m=1}^{M}\{I(C_i^{\langle m\rangle}\geq t;\boldsymbol{\theta})\text{exp}[\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)]\}+\boldsymbol{s}^{(1)\langle 1\rangle}(t;\boldsymbol{\beta},\boldsymbol{\theta})^T\boldsymbol{f}_i^{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\theta},M)$. Following similar arguments in the Proof of Theorem 2, we have Theorem 4.

APPENDIX B

Supplemental Materials to Chapter 2

B.1 Notations

$\boldsymbol{Z}_i(t)$: External covariates for subject i at time t.

$\boldsymbol{X}_i(t)$: The history of internal covariates for subject i up to time t.

$\mathcal{E}_i(t)$: Indicator of the state of interest for subject i at time t.

$A_i(t) = I\{\mathcal{E}_i(t) = 1\}I(D_i > t)$: Indicator of being alive and staying in the state of interest for subject i at time t.

C_{1i}: Conditional independent censoring time for subject i.

C_{2i}: Dependent censoring time for subject i.

$C_i = C_{1i} \wedge C_{2i}$: Observed censoring time for subject i at time t.

$dM_i^*(t;\boldsymbol{\beta}) = A_i(t)dt - \pi_0(t)\text{exp}\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}dt$: Martingale increment of the target of interest for subject i at time t.

$dM_i(t;\boldsymbol{\beta}) = A_i(t)I(C_i \geq t)dt - \pi_0(t)\text{exp}\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}I(C_i \geq t)dt$: Observed Martingale increment of the target of interest for subject i at time t.

$X_i = C_i \wedge D_i$: Observed terminal event time for subject i at time t.

$N_i^{C_2}(t) = I(C_{2i} \leq t \wedge X_i)$: Observed counting process of dependent censoring for subject i at time t.

$dN_i^{C_2}(t) = N_i^{C_2}(t^- + dt) - N_i^{C_2}(t^-)$: Counting process increment of observed dependent censoring for subject i at time t.

$Y_i(t) = I(X_i \geq t)$: At-risk process for dependent censoring for subject i at time t.

$dM_i^{C_2}(t;\boldsymbol{\theta}) = dN_i^{C_2}(t) - [\lambda_0^{C_2}(t) + \boldsymbol{\theta}_0^T \boldsymbol{X}_i(t)]Y_i(t)dt$: Martingale increment of observed dependent censoring for subject i at time t.

$W_i^A(t;\boldsymbol{\theta}) = I(C_{2i} \geq t)\text{exp}\{\int_0^t [\lambda_0^{C_2}(s) + \boldsymbol{\theta}_0^T \boldsymbol{X}_i(s)]Y_i(s)ds\}$: Type A weights for subject i at time t.

$W_i^B(t;\boldsymbol{\theta}) = I(C_{2i} \geq t)\text{exp}\{\int_0^t \boldsymbol{\theta}_0^T \boldsymbol{X}_i(s)Y_i(s)ds\}$: Type B weights for subject i at time t.

Let $\boldsymbol{X}^{(k)}(t) = n^{-1}\sum_{i=1}^n \boldsymbol{X}_i(t)^{\otimes k} Y_i(t)$, and $\boldsymbol{x}^{(k)}(t) = E[\boldsymbol{X}_1(t)^{\otimes k} Y_1(t)]$ be the limiting value of $\boldsymbol{X}^{(k)}(t)$ for $k = 0, 1, 2$, where $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. Denote $\bar{\boldsymbol{X}}(t) = \boldsymbol{X}^{(1)}(t)/X^{(0)}(t)$, and its limiting value as $\bar{\boldsymbol{x}}(t) = \boldsymbol{x}^{(1)}(t)/x^{(0)}(t)$. Moreover, define $\widehat{\boldsymbol{\Omega}}^{C_2} = n^{-1}\sum_{i=1}^n \int_0^\tau \{\boldsymbol{X}_i(t) - \bar{\boldsymbol{X}}(t)\}Y_i(t)\boldsymbol{X}_i^T(t)dt$, and its limiting value
$\boldsymbol{\Omega}^{C_2} = E[\int_0^\tau \{\boldsymbol{X}_1(t) - \bar{\boldsymbol{x}}(t)\}Y_1(t)\boldsymbol{X}_1^T(t)dt]$.

Let $\boldsymbol{Z}^{(k)}(t;\boldsymbol{\beta}, W) = n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i(t)^{\otimes k} I(C_{1i} \geq t)W_i(t)\text{exp}\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}$,
and $\boldsymbol{z}^{(k)}(t;\boldsymbol{\beta}, W) = E[\boldsymbol{Z}_1(t)^{\otimes k} I(C_{11} \geq t)W_1(t)\text{exp}\{\boldsymbol{\beta}^T \boldsymbol{Z}_1(t)\}]$ be the limiting value of $\boldsymbol{Z}^{(k)}(t;\boldsymbol{\beta}, W)$ for $k = 0, 1, 2$. Denote $\bar{\boldsymbol{Z}}(t;\boldsymbol{\beta}, W) = \boldsymbol{Z}^{(1)}(t;\boldsymbol{\beta}, W)/Z^{(0)}(t;\boldsymbol{\beta}, W)$, and its limiting value as $\bar{\boldsymbol{z}}(t;\boldsymbol{\beta}, W) = \boldsymbol{z}^{(1)}(t;\boldsymbol{\beta}, W)/z^{(0)}(t;\boldsymbol{\beta}, W)$. Moreover, define
$\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta}, W) = n^{-1}\sum_{i=1}^n \int_0^\tau \{\boldsymbol{Z}^{(2)}(t;\boldsymbol{\beta}, W)/Z^{(0)}(t;\boldsymbol{\beta}, W) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\beta}, W)^{\otimes 2}\}A_i(t)I(C_{1i} \geq t)dt$ and its limiting value $\boldsymbol{\Omega}(\boldsymbol{\beta}, W) = E[\int_0^\tau \{\boldsymbol{z}^{(2)}(t;\boldsymbol{\beta}, W)/z^{(0)}(t;\boldsymbol{\beta}, W) - \bar{\boldsymbol{z}}(t;\boldsymbol{\beta}, W)^{\otimes 2}\}A_1(t)I(C_{11} \geq t)dt]$.

B.2 Assumptions

Let

$$\lambda_i^{C_2}\{t\} = \lim_{\delta \to 0} \delta^{-1} Pr[t \leq C_{2i} < t + \delta | D_i \geq t, C_{2i} \geq t, \widetilde{\boldsymbol{X}}_i(t)]$$

No unmeasured confounders assumption:

$$\lambda_i^{C_2}\{t|\widetilde{\boldsymbol{X}}_i(t)\} = \lambda_i^{C_2}\{t|\widetilde{\boldsymbol{X}}_i(t), C_{1i}, C_{1i} \geq t, D_i, D_i \geq t, \mathcal{E}_i(s), s \in (t, \tau]\} \tag{B.1}$$

Note that $\widetilde{\boldsymbol{X}}_i(t)$ includes $\boldsymbol{Z}_i(t)$.

Assumptions for independent censoring time C_1:

$$E[A_i(t)|\boldsymbol{Z}_i(t), C_{1i} \geq t, C_{1i}] = E[A_i(t)|\boldsymbol{Z}_i(t)] \tag{B.2}$$

Additive hazard model on dependent censoring C_{2i}:

$$\lambda_i^{C_2}(t) = \lambda_0^{C_2}(t) + \boldsymbol{\theta}^T \boldsymbol{X}_i(t) \tag{B.3}$$

Model assumptions on the target of interest:

$$\pi_i(t) = \pi_0(t)\text{exp}[\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)]$$

or

$$E[dM_i^*(t)|\boldsymbol{Z}_i(t)] = 0 \tag{B.4}$$

Estimating equations:

$$\sum_{i=1}^{n}[A_i(t) - \pi_i(t)]I(C_{1i} \geq t)W_i(t) = 0 \tag{B.5}$$

$$\sum_{i=1}^{n}\int_0^{\tau}[A_i(t) - \pi_i(t)]I(C_{1i} \geq t)W_i(t)dt = 0 \tag{B.6}$$

B.3 Regularity Conditions for C_{1i} and C_{2i}

1. $\{X_i, Y_i(t), \boldsymbol{X}_i(t), \boldsymbol{Z}_i(t), I(C_{1i} \leq D_i \wedge C_{2i}), I(C_{2i} \leq D_i \wedge C_{1i})\}$ for $t \in [0, X_i], i = 1, 2, ..., n$ are independent and identically distributed.

2. $Pr(C_i \geq \tau) > 0$ for $i = 1, 2, ..., n$, where τ is a pre-specified constant.

3. $N_i^{C_1}(\tau) < \eta_1 < \infty$ and $N_i^{C_2}(\tau) < \eta_2 < \infty$ almost surely.

4. $|\boldsymbol{Z}_{ij}(0)| + \int_0^\tau |d\boldsymbol{Z}_{ij}(t)| < c_{\boldsymbol{Z}} < \infty$ and $|\boldsymbol{X}_{ij}(0)| + \int_0^\tau |d\boldsymbol{X}_{ij}(t)| < c_{\boldsymbol{X}} < \infty$ almost surely for $i = 1, 2, ..., n$, $j = 1, 2, ..., p$, i.e., $\boldsymbol{Z}_i(t)$ and $\boldsymbol{X}_i(t)$ have bounded total variations.

5. Let $\boldsymbol{z}_{C_1}^{(k)}(t; \boldsymbol{\gamma}) = E[\boldsymbol{Z}_1(t)^{\otimes k} Y_1(t) \exp\{\boldsymbol{\gamma}^T \boldsymbol{Z}_1(t)\}]$ and denote $\bar{\boldsymbol{z}}_{C_1}(t; \boldsymbol{\gamma}) = \boldsymbol{z}_{C_1}^{(1)}(t; \boldsymbol{\gamma}) / z_{C_1}^{(0)}(t; \boldsymbol{\gamma})$. $\boldsymbol{\Omega}^{C_1}(\boldsymbol{\gamma}_0) = E[\int_0^\tau \{\boldsymbol{z}_{C_1}^{(2)}(t; \boldsymbol{\gamma}_0) / z_{C_1}^{(0)}(t; \boldsymbol{\gamma}_0) - \bar{\boldsymbol{z}}_{C_1}(t; \boldsymbol{\gamma}_0)^{\otimes 2}\} Y_1(t) dN_1^{C_1}(t)]$ and $\boldsymbol{\Omega}^{C_2}$ are positive definite.

6. For $\boldsymbol{\gamma} \in \mathscr{B}_{\delta_1}$ where $\mathscr{B}_{\delta_1}$ is a small neighborhood around $\boldsymbol{\gamma}_0$, $z_{C_1}^{(0)}(t; \boldsymbol{\gamma})$, $\boldsymbol{z}_{C_1}^{(1)}(t; \boldsymbol{\gamma})$, $x^{(0)}(t)$ and $\boldsymbol{x}^{(1)}(t)$ are bounded away from zero.

B.4 Proof for Known Censoring Time

B.4.1 $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$

By no unmeasured confounders assumption (B.1), $C_{1i} \wedge D_i$ is independent of C_{2i} conditional on $\boldsymbol{X}_i(t)$. By assumption (B.1), we could get the additive hazards estimating equation

$$\boldsymbol{U}^{C_2}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^\tau \{\boldsymbol{X}_i(t) - \bar{\boldsymbol{X}}(t)\} dM_i^{C_2}(t; \boldsymbol{\theta})$$

and the expectation of $\boldsymbol{U}^{C_2}(\boldsymbol{\theta}_0)$ would be zero (Lin and Ying 1994, Schaubel et al. 2006). Let $\widehat{\boldsymbol{\theta}}$ be the solution of $\boldsymbol{U}^{C_2}(\boldsymbol{\theta}) = \mathbf{0}$, then $\widehat{\boldsymbol{\theta}}$ is a strong consistent estimator for $\boldsymbol{\theta}$. Furthermore, by Taylor

expansion and the consistency of $\widehat{\boldsymbol{\theta}}$ we have

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = [\widehat{\boldsymbol{\Omega}}^{C_2}]^{-1} n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{X}_i(t) - \bar{\boldsymbol{X}}(t)\} dM_i^{C_2}(t; \boldsymbol{\theta}_0) + o_p(1) \tag{B.7}$$

Without lose of generality, we assume that $\boldsymbol{X}_i(t) \geq 0$, then $\boldsymbol{X}^{(1)}(t)$ and $X^{(0)}(t)$ are monotone function in t and converge almost surely to $\boldsymbol{x}^{(1)}(t)$ and $x^{(0)}(t)$ by SLLN. Iteratively applying Lemma 1 of Lin et al. 2000, we could express $n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{X}_i(t) - \bar{\boldsymbol{X}}(t)\} dM_i^{C_2}(t; \boldsymbol{\theta}_0)$ as a sum of i.i.d random variables, i.e.,

$$n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{X}_i(t) - \bar{\boldsymbol{X}}(t)\} dM_i^{C_2}(t; \boldsymbol{\theta}_0) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0) + o_p(1)$$

where

$$\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}) = \int_0^{\tau} \{\boldsymbol{X}_i(t) - \bar{\boldsymbol{x}}(t)\} dM_i^{C_2}(t; \boldsymbol{\theta})$$

By the consistency of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Omega}}^{C_2}$, we could re-write (B.7) as

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = [\boldsymbol{\Omega}^{C_2}]^{-1} n^{-1/2} \sum_{i=1}^{n} \boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0) + o_p(1) \tag{B.8}$$

Then by multivariate central limit theorem, we have $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to a mean zero normal random variable.

B.4.2 $n^{1/2}\{d\widehat{\Lambda}_0^{C_2}(t; \widehat{\boldsymbol{\theta}}) - d\Lambda_0^{C_2}(t)\}$

By Lin and Ying 1994, the estimator of baseline hazard increment $d\widehat{\Lambda}_0^{C_2}(t)$ is given by

$$d\widehat{\Lambda}_0^{C_2}(t; \widehat{\boldsymbol{\theta}}) = \frac{\sum_{i=1}^{n} \{dN_i^{C_2}(t) - Y_i(t) \widehat{\boldsymbol{\theta}}^T \boldsymbol{X}_i(t) dt\}}{\sum_{i=1}^{n} Y_i(t)}$$

Consider $n^{1/2}[d\widehat{\Lambda}_0^{C_2}(t;\widehat{\boldsymbol{\theta}}) - d\Lambda_0^{C_2}(t)]$, which can be decomposed as

$$n^{1/2}[d\widehat{\Lambda}_0^{C_2}(t;\widehat{\boldsymbol{\theta}}) - d\Lambda_0^{C_2}(t)] = n^{1/2}\{d\widehat{\Lambda}_0^{C_2}(t;\widehat{\boldsymbol{\theta}}) - d\widehat{\Lambda}_0^{C_2}(t;\boldsymbol{\theta}_0)\} + n^{1/2}\{d\widehat{\Lambda}_0^{C_2}(t;\boldsymbol{\theta}_0) - d\Lambda_0^{C_2}(t)\} \tag{B.9}$$

Denote

$$\widehat{\boldsymbol{h}}_{C_2}(t) = -\frac{\boldsymbol{X}^{(1)}(t)dt}{X^{(0)}(t)}$$

and its limiting value $\boldsymbol{h}_C(t)$ as

$$\boldsymbol{h}_{C_2}(t) = -\frac{\boldsymbol{x}^{(1)}(t)dt}{x^{(0)}(t)}$$

By WLLN and continuous mapping theorem, we have $\widehat{\boldsymbol{h}}_{C_2}(t)$ converges in probability to $\boldsymbol{h}_{C_2}(t)$. Then the first term in (B.9) could be written as

$$\begin{aligned} n^{1/2}\{d\widehat{\Lambda}_0^{C_2}(t;\widehat{\boldsymbol{\theta}}) - d\widehat{\Lambda}_0^{C_2}(t;\boldsymbol{\theta}_0)\} &= -n^{1/2}\frac{\sum_{i=1}^n Y_i(t)\boldsymbol{X}_i^T(t)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)dt}{\sum_{i=1}^n Y_i(t)} \\ &= \widehat{\boldsymbol{h}}_{C_2}^T(t)n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \boldsymbol{h}_{C_2}^T(t)[\boldsymbol{\Omega}^{C_2}]^{-1}n^{-1/2}\sum_{i=1}^n \boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0) + o_p(1) \end{aligned} \tag{B.10}$$

Next consider the second part:

$$\begin{aligned} n^{1/2}\{d\widehat{\Lambda}_0^{C_2}(t;\boldsymbol{\theta}_0) - d\Lambda_0^{C_2}(t;\boldsymbol{\theta}_0)\} &= n^{1/2}\{\frac{\sum_{i=1}^n [dN_i^{C_2}(t) - Y_i(t)\boldsymbol{\theta}_0^T\boldsymbol{X}_i(t)dt - Y_i(t)\lambda_0^{C_2}(t)dt]}{\sum_{i=1}^n Y_i(t)}\} \\ &= n^{-1/2}\sum_{i=1}^n X^{(0)}(t)^{-1}dM_i^{C_2}(t;\boldsymbol{\theta}_0) \\ &= n^{-1/2}\sum_{i=1}^n x^{(0)}(t)^{-1}dM_i^{C_2}(t;\boldsymbol{\theta}_0) + o_p(1) \end{aligned} \tag{B.11}$$

where the last equality follows from Lemma 1 of Lin et al. 2000. Combining (B.10) and (B.11), we have

$$n^{1/2}[d\widehat{\Lambda}_0^{C_2}(t;\widehat{\boldsymbol{\theta}}) - d\Lambda_0^{C_2}(t)] = n^{-1/2}\sum_{i=1}^{n} f_i^{C_0}(t,\boldsymbol{\theta}_0) + o_p(1) \tag{B.12}$$

where

$$f_i^{C_0}(t,\boldsymbol{\theta}) = \boldsymbol{h}_{C_2}^T(t)[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}) + x^{(0)}(s)^{-1}dM_i^{C_2}(t;\boldsymbol{\theta})$$

By Central Limit Theorem, we have $n^{1/2}[d\widehat{\Lambda}_0^{C_2}(t;\widehat{\boldsymbol{\theta}}) - d\Lambda_0^{C_2}(t;\boldsymbol{\theta}_0)]$ converges in distribution to a mean zero normal random variable, with variance $E[f_1^{C_0}(t,\boldsymbol{\theta}_0)^2]$.

B.4.3 $n^{1/2}\{d\widehat{\Lambda}_i^{C_2}(t;\widehat{\boldsymbol{\theta}})Y_i(t) - d\Lambda_i^{C_2}(t;\boldsymbol{\theta}_0)Y_i(t)\}$

Similar to the previous section, we could decompose the target quantity as

$$\begin{aligned} n^{1/2}\{d\widehat{\Lambda}_i^{C_2}(t;\widehat{\boldsymbol{\theta}})Y_i(t) - d\Lambda_i^{C_2}(t;\boldsymbol{\theta}_0)Y_i(t)\} =& n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\boldsymbol{X}_i(t)Y_i(t)dt + & \text{(B.13)}\\ & n^{1/2}[d\widehat{\Lambda}_0^{C_2}(t;\widehat{\boldsymbol{\theta}}) - d\Lambda_0^{C_2}(t)]Y_i(t) & \text{(B.14)}\end{aligned}$$

By (B.8), the first term (B.13) could be written as

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\boldsymbol{X}_i(t)Y_i(t)dt = \boldsymbol{X}_i^T(t)[\boldsymbol{\Omega}^{C_2}]^{-1}Y_i(t)n^{-1/2}\sum_{l=1}^{n}\boldsymbol{u}_l^{C_2}(\boldsymbol{\theta}_0)dt + o_p(1) \tag{B.15}$$

Combining (B.15) and (B.12), we have $n^{1/2}[d\widehat{\Lambda}_i^{C_2}(t;\widehat{\boldsymbol{\theta}})Y_i(t) - d\Lambda_i^{C_2}(t)Y_i(t)]$ equals to $n^{-1/2}\sum_{l=1}^{n} f_l^{C_i}(t,\boldsymbol{\theta}_0) + o_p(1)$, where

$$f_l^{C_i}(t,\boldsymbol{\theta}) = [\boldsymbol{X}_i^T(t)dt + \boldsymbol{h}_{C_2}^T(t)][\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_l^{C_2}(\boldsymbol{\theta})Y_i(t) + x^{(0)}(t)^{-1}dM_l^{C_2}(t;\boldsymbol{\theta})Y_i(t)$$

By Central Limit Theorem again, we have $n^{1/2}[d\widehat{\Lambda}_i^{C_2}(t;\widehat{\boldsymbol{\theta}})Y_i(t) - d\Lambda_i^{C_2}(t;\boldsymbol{\theta}_0)Y_i(t)]$ converges in distribution to a mean zero normal random variable, with variance $E[f_1^{C_i}(t,\boldsymbol{\theta}_0)^2]$.

B.4.4 $n^{1/2}\{\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}}) - W_i^A(t;\boldsymbol{\theta}_0)\}$

Since the target quantity equals to $n^{1/2}I(C_{2i} \geq t)\{\exp[\int_0^t Y_i(s)d\widehat{\Lambda}_i^{C_2}(s;\widehat{\boldsymbol{\theta}})] - \exp[\int_0^t Y_i(s)d\Lambda_i^{C_2}(s;\boldsymbol{\theta}_0)]\}$, then by Taylor expansion and results of B.4.3 we have

$$n^{1/2}\{\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}}) - W_i^A(t;\boldsymbol{\theta}_0)\} = W_i^A(t;\boldsymbol{\theta}_0)n^{-1/2}\sum_{l=1}^{n}\int_0^t f_l^{C_i}(s,\boldsymbol{\theta}_0) + o_p(1) \tag{B.16}$$

B.4.5 $n^{1/2}\{\widehat{W}_i^B(t;\widehat{\boldsymbol{\theta}}) - W_i^B(t;\boldsymbol{\theta}_0)\}$

The target quantity is given by

$$\begin{aligned}
&n^{1/2}\{\widehat{W}_i^B(t;\widehat{\boldsymbol{\theta}}) - W_i^B(t;\boldsymbol{\theta}_0)\} \\
=&n^{1/2}I(C_{2i} \geq t)\{\exp[\int_0^t \widehat{\boldsymbol{\theta}}^T \boldsymbol{X}_i(s)Y_i(s)ds] - \exp[\int_0^t \boldsymbol{\theta}_0^T \boldsymbol{X}_i(s)Y_i(s)ds]\}
\end{aligned}$$

By Taylor expansion and the consistency of $\widehat{\boldsymbol{\theta}}$, the above quantity equals to

$$n^{1/2}W_i^B(t;\boldsymbol{\theta}_0)\int_0^t \boldsymbol{X}_i^T(s)Y_i(s)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)ds + o_p(1)$$

Combining results with (B.8) we have

$$n^{1/2}\{\widehat{W}_i^B(t;\widehat{\boldsymbol{\theta}}) - W_i^B(t;\boldsymbol{\theta}_0)\} = W_i^B(t;\boldsymbol{\theta}_0)n^{-1/2}\sum_{l=1}^{n}\int_0^t f_l^{X_i}(s,\boldsymbol{\theta}_0) + o_p(1) \tag{B.17}$$

where $f_l^{X_i}(t,\boldsymbol{\theta}) = [\boldsymbol{X}_i^T(t)Y_i(t)dt][\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_l^{C_2}(\boldsymbol{\theta})$.

B.4.6 Consistency of $\widehat{\boldsymbol{\beta}}_A$

Consider the two estimating equations

$$\sum_{i=1}^{n} W_i^A(t;\boldsymbol{\theta})dM_i(t;\boldsymbol{\beta}) = 0 \tag{B.18}$$

$$\sum_{i=1}^{n}\int_0^{\tau} \boldsymbol{Z}_i(t)W_i^A(t;\boldsymbol{\theta})dM_i(t;\boldsymbol{\beta}) = 0 \tag{B.19}$$

we will first show that they have expectation zero at $\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}_0$. Consider $W_1^A(t;\boldsymbol{\theta}_0)$:

$$\begin{aligned}
&W_1^A(t;\boldsymbol{\theta}_0)\\
=&\exp\left[\int_0^{t\wedge D_1} d\Lambda_1^{C_2}(t)\right] I(C_{21}\geq t)\\
=&\prod_{r\in(0,t\wedge D_1]}[1-d\Lambda_1^{C_2}(r)]^{-1} I(C_{21}\geq t)
\end{aligned}$$

Then consider the expectation of (B.18) condition on $\boldsymbol{X}_1(t)$. The first scenario is when $t \leq D_1$:

$$\begin{aligned}
&E\{W_1^A(t^-;\boldsymbol{\theta}_0)dM_1(t;\boldsymbol{\beta}_0)|\boldsymbol{X}_1(t), t\leq D_1\}\\
=&E\{e^{\Lambda_1^{C_2}(t^-)}dM_1^*(t;\boldsymbol{\beta}_0)I(C_{21}\geq t)I(C_{11}\geq t)|\boldsymbol{X}_1(t), t\leq D_1\}\\
=&e^{\Lambda_1^{C_2}(t^-)}E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{21}\geq t)I(C_{11}\geq t)|\boldsymbol{X}_1(t), t\leq D_1]\\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{21}> t^-)I(C_{11}\geq t)|C_{21}\geq t^-, \boldsymbol{X}_1(t), t\leq D_1][1-d\Lambda_i^{C_2}(t^-)]^{-1}\\
&\qquad\prod_{r\in(0,t^-)} E[C_{21}>r|C_{21}\geq r, \boldsymbol{X}_1(r)][1-d\Lambda_i^{C_2}(r)]^{-1}\\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{21}> t^-)I(C_{11}\geq t)|C_{21}\geq t^-, \boldsymbol{X}_1(t), t\leq D_1][1-d\Lambda_i^{C_2}(t^-)]^{-1}\\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{11}\geq t)|C_{21}\geq t^-, \boldsymbol{X}_1(t), t\leq D_1]\\
&\qquad E[I(C_{21}>t^-)|C_{21}\geq t^-, \boldsymbol{X}_1(t), t\leq D_1][1-d\Lambda_i^{C_2}(t^-)]^{-1}
\end{aligned} \tag{B.20}$$

$$\begin{aligned}
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{11}\geq t)|C_{21}\geq t^-, \boldsymbol{X}_1(t), t\leq D_1]\\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{11}\geq t)|\boldsymbol{X}_1(t), t\leq D_1]
\end{aligned} \tag{B.21}$$

where (B.20) and (B.21) are by no unmeasured confounders assumption (B.1). As for $t > D_1$, we have $dM_1^*(t;\boldsymbol{\beta}_0)$ is only a function of $\boldsymbol{Z}_1(t)$ because $A_1(t) = 0$. In this case

$$
\begin{aligned}
&E\{W_1^A(t^-;\boldsymbol{\theta}_0)dM_1(t;\boldsymbol{\beta}_0)|\boldsymbol{X}_1(t), t > D_1\} \\
=&E\{e^{\Lambda_1^{C_2}(D_1)}dM_1^*(t;\boldsymbol{\beta}_0)I(C_{21} \geq t)I(C_{11} \geq t)|\boldsymbol{X}_1(t), t > D_1\} \\
=&e^{\Lambda_1^{C_2}(D_1)}E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{21} \geq t)I(C_{11} \geq t)|\boldsymbol{X}_1(t), t > D_1] \\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{21} > D_1)I(C_{11} \geq t)|C_{21} \geq D_1, \boldsymbol{X}_1(t), t > D_1] \\
&\qquad\qquad \prod_{r\in(0,D_1]} E[C_{21} > r|C_{21} \geq r, \boldsymbol{X}_1(r)][1 - d\Lambda_i^{C_2}(r)]^{-1} \\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{11} \geq t)|C_{21} \geq t^-, \boldsymbol{X}_1(t), t > D_1] \qquad \text{(B.22)} \\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{11} \geq t)|\boldsymbol{X}_1(t), t > D_1]
\end{aligned}
$$

where (B.22) is by the fact that given $t > D_1$ (death is observed), C_{21} is infinity and therefore $C_{21} > D_1$. Combining the two scenarios, we have

$$E\{W_1^A(t^-;\boldsymbol{\theta}_0)dM_1(t;\boldsymbol{\beta}_0)|\boldsymbol{X}_1(t)\} = E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{11} \geq t)|\boldsymbol{X}_1(t)].$$

Since $\boldsymbol{X}_1(t)$ contains $\boldsymbol{Z}_1(t)$, then iterating expectations we have

$$
\begin{aligned}
&E\{W_1^A(t;\boldsymbol{\theta}_0)dM_1(t;\boldsymbol{\beta}_0)|\boldsymbol{Z}_1(t)\} \\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)I(C_{11} \geq t)|\boldsymbol{Z}_1(t)] \\
=&E[dM_1^*(t;\boldsymbol{\beta}_0)|\boldsymbol{Z}_1(t)]E[I(C_{11} \geq t)|\boldsymbol{Z}_1(t)] \qquad \text{(B.23)} \\
=&0 \qquad \text{(B.24)}
\end{aligned}
$$

where (B.23) is by conditional independent censoring assumption (B.2) and (B.23) is by the assumption (B.4). Similarly, we could show that (B.19) has expectation zero. Solving (B.18) and

(B.19), we get a mean zero estimating equation

$$\boldsymbol{U}[\boldsymbol{\beta}, W^A(\boldsymbol{\theta})] = \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t; \boldsymbol{\beta}, W^A(\boldsymbol{\theta})]\} W_i^A(t; \boldsymbol{\theta}) dM_i(t; \boldsymbol{\beta}) = 0$$

Consider the estimating equation with estimated weights $\widehat{W}^A(t; \widehat{\boldsymbol{\theta}})$, i.e.,

$$\boldsymbol{U}[\boldsymbol{\beta}, \widehat{W}^A(\widehat{\boldsymbol{\theta}})] = \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t; \boldsymbol{\beta}, \widehat{W}^A(\widehat{\boldsymbol{\theta}})]\} \widehat{W}_i^A(t; \widehat{\boldsymbol{\theta}}) dM_i(t; \boldsymbol{\beta}) = 0 \tag{B.25}$$

By SLLN and the strong consistency of $\widehat{W}_i^A(t; \boldsymbol{\theta})$ to $W_i^A(t; \boldsymbol{\theta})$ and $\widehat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$, we have that $n^{-1}\boldsymbol{U}[\boldsymbol{\beta}, \widehat{W}^A(\widehat{\boldsymbol{\theta}})]$ converges almost surely to zero. Then by the positive definite of $\boldsymbol{\Omega}[\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]$ and arguments in Lin et al. 2000 A1, the solution of (B.25) $\widehat{\boldsymbol{\beta}}_A$ is a consistent estimator of $\boldsymbol{\beta}_0$.

B.4.7 $n^{1/2}(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta})$

Since $\widehat{\boldsymbol{\beta}}_A$ is the solution of estimating equation

$$\boldsymbol{U}[\boldsymbol{\beta}, \widehat{W}^A(\widehat{\boldsymbol{\theta}})] = \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t; \boldsymbol{\beta}, \widehat{W}^A)\} \widehat{W}_i^A(t; \widehat{\boldsymbol{\theta}}) dM_i(t; \boldsymbol{\beta}) = 0$$

then by Taylor expansion and the consistency of $\widehat{\boldsymbol{\Omega}}[\boldsymbol{\beta}_0, \widehat{W}^A(\widehat{\boldsymbol{\theta}})]$ to $\boldsymbol{\Omega}[\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]$ we have

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_0) = \boldsymbol{\Omega}[\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]^{-1} n^{-1/2} \boldsymbol{U}[\boldsymbol{\beta}_0, \widehat{W}^A(\widehat{\boldsymbol{\theta}})] + o_p(1) \tag{B.26}$$

We could decompose $n^{-1/2}\boldsymbol{U}[\boldsymbol{\beta}_0, \widehat{W}^A(\widehat{\boldsymbol{\theta}})]$ as

$$
\begin{aligned}
& n^{-1/2}\boldsymbol{U}[\boldsymbol{\beta}_0, \widehat{W}^A(\widehat{\boldsymbol{\theta}})] \\
=& n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, \widehat{W}^A(\widehat{\boldsymbol{\theta}})]\}\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})dM_i(t;\boldsymbol{\beta}_0) \\
=& n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]\}W_i^A(t;\boldsymbol{\theta}_0)dM_i(t;\boldsymbol{\beta}_0) \qquad \text{(B.27)} \\
+& n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)] - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, \widehat{W}^A(\widehat{\boldsymbol{\theta}})]\}\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})dM_i(t;\boldsymbol{\beta}_0) \qquad \text{(B.28)} \\
+& n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]\}[\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}}) - W_i^A(t;\boldsymbol{\theta}_0)]dM_i(t;\boldsymbol{\beta}_0) \qquad \text{(B.29)}
\end{aligned}
$$

By the results of Lin et al. 2000 A2, (B.27) converges in probability to $n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{z}}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]\}W_i^A(t;\boldsymbol{\theta}_0)dM_i(t;\boldsymbol{\beta}_0) + o_p(1)$. Specifically, since $\boldsymbol{Z}_i(t)$ and $\boldsymbol{X}_i(t)$ are bounded, we assume without lose of generality that $\boldsymbol{Z}_i(t) \geq 0$ and $\boldsymbol{X}_i(t) \geq 0$. Therefore, $Z^{(0)}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]$ and $\boldsymbol{Z}^{(1)}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]$ are monotone functions in t and converges almost surely to $z^{(0)}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]$ and $\boldsymbol{z}^{(1)}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]$. Iteratively applying Lemma 1 in Lin et al. 2000, we establish the previous result. By Zhang and Schaubel 2011, the second term (B.28) equals $o_p(1)$. By (B.16), we have (B.29) equals to

$$
n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]\}\big[n^{-1}\sum_{l=1}^{n}\int_0^{t} f_l^{C_i}(s,\boldsymbol{\theta}_0)\big]W_i^A(t;\boldsymbol{\theta}_0)dM_i(t;\boldsymbol{\beta}_0) \qquad \text{(B.30)}
$$

Define

$$
\widehat{\boldsymbol{H}}(\boldsymbol{\beta},\boldsymbol{\theta},W) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta},W]\}[\int_0^{t}\{\boldsymbol{X}_i^T(s)ds + \boldsymbol{h}_{C_2}^T(s)\}Y_i(s)]W_i(t;\boldsymbol{\theta})dM_i(t;\boldsymbol{\beta})
$$

$$
\boldsymbol{H}(\boldsymbol{\beta},\boldsymbol{\theta},W) = E\Big[\int_0^{\tau}\{\boldsymbol{Z}_1(t) - \bar{\boldsymbol{z}}[t;\boldsymbol{\beta},W]\}[\int_0^{t}\{\boldsymbol{X}_1^T(s)ds + \boldsymbol{h}_{C_2}^T(s)\}Y_1(s)]W_1(t;\boldsymbol{\theta})dM_1(t;\boldsymbol{\beta})\Big]
$$

and

$$\widehat{\boldsymbol{J}}(t_1,t_2,\boldsymbol{\beta},\boldsymbol{\theta},W)=n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t_2}\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}[t;\boldsymbol{\beta},W]\}W_i(t;\boldsymbol{\theta})x^{(0)}(t)^{-1}dM_i(t;\boldsymbol{\beta})Y_i(t)$$

$$\boldsymbol{J}(t_1,t_2,\boldsymbol{\beta},\boldsymbol{\theta},W)=E\bigg[\int_{t_1}^{t_2}\{\boldsymbol{Z}_1(t)-\bar{\boldsymbol{z}}[t;\boldsymbol{\beta},W]\}W_1(t;\boldsymbol{\theta})x^{(0)}(t)^{-1}dM_1(t;\boldsymbol{\beta})Y_1(t)\bigg]$$

Switching the order of summation, we have (B.30) equals to

$$\begin{aligned}&\widehat{\boldsymbol{H}}[\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,W^A(\boldsymbol{\theta}_0)][\boldsymbol{\Omega}^{C_2}]^{-1}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0)+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\widehat{\boldsymbol{J}}[t,\tau,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,W^A(\boldsymbol{\theta}_0)]dM_i^{C_2}(t;\boldsymbol{\theta}_0)\\ =&n^{-1/2}\sum_{i=1}^{n}\bigg[\boldsymbol{H}[\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,W^A(\boldsymbol{\theta}_0)][\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0)+\int_0^{\tau}\boldsymbol{J}[t,\tau,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,W^A(\boldsymbol{\theta}_0)]dM_i^{C_2}(t;\boldsymbol{\theta}_0)\bigg]+o_p(1)\end{aligned}\tag{B.31}$$

where (B.31) is by iteratively applying SLLN. Combing (B.27) and (B.31) we have

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_A-\boldsymbol{\beta}_0)=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{f}_i^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)+o_p(1)\tag{B.32}$$

where $\boldsymbol{f}_i^{\boldsymbol{\beta}_A}(\boldsymbol{\theta},\boldsymbol{\beta})=\boldsymbol{\Omega}[\boldsymbol{\beta},W^A(\boldsymbol{\theta})]^{-1}\bigg[\int_0^{\tau}\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{z}}[t;\boldsymbol{\beta},W^A(\boldsymbol{\theta})]\}W_i^A(t;\boldsymbol{\theta})dM_i(t;\boldsymbol{\beta})+$
$\boldsymbol{H}[\boldsymbol{\beta},\boldsymbol{\theta},W^A(\boldsymbol{\theta})][\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta})+\int_0^{\tau}\boldsymbol{J}[t,\tau,\boldsymbol{\beta},\boldsymbol{\theta},W^A(\boldsymbol{\theta})]dM_i^{C_2}(t;\boldsymbol{\theta})\bigg]$. Therefore, $n^{1/2}(\widehat{\boldsymbol{\beta}}_A-\boldsymbol{\beta}_0)$ converges in distribution to a mean zero normal random variable, with variance covariance matrix $E[f_1^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)f_1^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)^T]$.

B.4.8 Consistency of $\widehat{\boldsymbol{\beta}}_B$

Similar to (B.24), we can show that $W_1^B(t;\boldsymbol{\theta}_0)dM_1(t;\boldsymbol{\beta}_0)$ have expectation zero condition on external covariates $\boldsymbol{Z}_1(t)$. Consider the first scenario where $t\le D_1$, then $W_1^B(t;\boldsymbol{\theta})=W_1^A(t;\boldsymbol{\theta})\exp[-\int_0^t d\Lambda_0^{C_2}(s;\boldsymbol{\theta})]$ and $\exp[-\int_0^t d\Lambda_0^{C_2}(s;\boldsymbol{\theta})]$ is just a function of t. For the second scenario where $t>D_1$, $W_1^B(t;\boldsymbol{\theta})=W_1^A(t;\boldsymbol{\theta})\exp[-\int_0^{D_1} d\Lambda_0^{C_2}(s;\boldsymbol{\theta})]$. Since D_1 is observed, then $\exp[-\int_0^{D_1} d\Lambda_0^{C_2}(s;\boldsymbol{\theta})]$ is also a function of t. Therefore combing two scenarios and iterating the expectation on $\boldsymbol{Z}_i(t)$, we also get two

mean zero estimating equations

$$\sum_{i=1}^{n} W_i^B(t;\boldsymbol{\theta}) dM_i(t;\boldsymbol{\beta}) = 0 \tag{B.33}$$

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i(t) W_i^B(t;\boldsymbol{\theta}) dM_i(t;\boldsymbol{\beta}) = 0 \tag{B.34}$$

Then the consistency of $\widehat{\boldsymbol{\beta}}_B$ could be proved in similar manners as of $\widehat{\boldsymbol{\beta}}_A$.

B.4.9 $n^{1/2}(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0)$

We have $\widehat{\boldsymbol{\beta}}_B$ is the solution of estimating equation

$$\boldsymbol{U}[\boldsymbol{\beta}, \widehat{W}^B(\widehat{\boldsymbol{\theta}})] = \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}, \widehat{W}^B(\widehat{\boldsymbol{\theta}})]\} \widehat{W}_i^B(t;\widehat{\boldsymbol{\theta}}) dM_i(t;\boldsymbol{\beta}) = 0$$

and by Taylor expansion and the strong consistency of $\widehat{\boldsymbol{\Omega}}[\boldsymbol{\beta}_0, \widehat{W}^B(\widehat{\boldsymbol{\theta}})]$ to $\boldsymbol{\Omega}[\boldsymbol{\beta}_0, W^B(\boldsymbol{\theta}_0)]$ we have

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0) = \boldsymbol{\Omega}[\boldsymbol{\beta}_0, W^B(\boldsymbol{\theta}_0)]^{-1} n^{-1/2} \boldsymbol{U}[\boldsymbol{\beta}_0, \widehat{W}^B(\widehat{\boldsymbol{\theta}})] + o_p(1) \tag{B.35}$$

We could decompose $n^{-1/2}\boldsymbol{U}[\boldsymbol{\beta}_0, \widehat{W}^B(\widehat{\boldsymbol{\theta}})]$ as

$$\begin{aligned}
& n^{-1/2}\boldsymbol{U}[\boldsymbol{\beta}_0, \widehat{W}^B(\widehat{\boldsymbol{\theta}})] \\
=& n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, \widehat{W}^B(\widehat{\boldsymbol{\theta}})]\} \widehat{W}_i^B(t;\widehat{\boldsymbol{\theta}}) dM_i(t;\boldsymbol{\beta}_0) \\
=& n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, W^B(\boldsymbol{\theta}_0)]\} W_i^B(t;\boldsymbol{\theta}_0) dM_i(t;\boldsymbol{\beta}_0) && \text{(B.36)} \\
+& n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, W^B(\boldsymbol{\theta}_0)] - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, \widehat{W}^B(\widehat{\boldsymbol{\theta}})]\} \widehat{W}_i^B(t;\widehat{\boldsymbol{\theta}}) dM_i(t;\boldsymbol{\beta}_0) && \text{(B.37)} \\
+& n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}[t;\boldsymbol{\beta}_0, W^B(\boldsymbol{\theta}_0)]\} [\widehat{W}_i^B(t;\widehat{\boldsymbol{\theta}}) - W_i^B(t;\boldsymbol{\theta}_0)] dM_i(t;\boldsymbol{\beta}_0) && \text{(B.38)}
\end{aligned}$$

Following the arguments in section (B.4.7), we have (B.36) converges in probability to $n^{-1/2}\sum_{i=1}^{n}\int_0^\tau\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{z}}[t;\boldsymbol{\beta}_0,W^B(\boldsymbol{\theta}_0)]\}W_i^B(t;\boldsymbol{\theta}_0)dM_i(t;\boldsymbol{\beta}_0)+o_p(1)$, the second term (B.37) equals to $o_p(1)$. Switching the order of summation, we have the third term (B.38) equals to

$$\begin{aligned}&\widehat{\boldsymbol{H}}^B[\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,W^B(t)][\boldsymbol{\Omega}^{C_2}]^{-1}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0)+o_p(1)\\ =&\boldsymbol{H}^B[\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,W^B(t)][\boldsymbol{\Omega}^{C_2}]^{-1}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0)+o_p(1)\end{aligned}\tag{B.39}$$

where

$$\widehat{\boldsymbol{H}}^B(\boldsymbol{\beta},\boldsymbol{\theta},W)=n^{-1}\sum_{i=1}^{n}\int_0^\tau\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{Z}}[t;\boldsymbol{\beta},W]\}[\int_0^t\boldsymbol{X}_i^T(s)Y_i(s)ds]W_i(t)dM_i(t;\boldsymbol{\beta})$$

$$\boldsymbol{H}^B(\boldsymbol{\beta},\boldsymbol{\theta},W)=E\bigg[\int_0^\tau\{\boldsymbol{Z}_1(t)-\bar{\boldsymbol{z}}[t;\boldsymbol{\beta},W]\}[\int_0^t\boldsymbol{X}_1^T(s)Y_1(s)ds]W_1(t)dM_1(t;\boldsymbol{\beta})\bigg]$$

Combing results we have

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_B-\boldsymbol{\beta}_0)=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{f}_i^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)+o_p(1)\tag{B.40}$$

where $\boldsymbol{f}_i^{\boldsymbol{\beta}_B}(\boldsymbol{\theta},\boldsymbol{\beta})=\boldsymbol{\Omega}[\boldsymbol{\beta},W^B(\boldsymbol{\theta})]^{-1}\bigg[\int_0^\tau\{\boldsymbol{Z}_i(t)-\bar{\boldsymbol{z}}[t;\boldsymbol{\beta},W^B(\boldsymbol{\theta})]\}W_i^B(t;\boldsymbol{\theta})dM_i(t;\boldsymbol{\beta})+$ $\boldsymbol{H}^B[\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,W^B(\boldsymbol{\theta})][\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta})\bigg]$. Therefore, $n^{1/2}(\widehat{\boldsymbol{\beta}}_B-\boldsymbol{\beta}_0)$ converges in distribution to a mean zero normal random variable, with variance covariance matrix equals to $E[\boldsymbol{f}_1^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)\boldsymbol{f}_1^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)^T]$.

B.4.10 $n^{1/2}\{\widehat{\pi}_0^A(t)-\pi_0(t)\}$

We consider the quantity in the numerator and the denominator separately.

B.4.10.1 $n^{1/2}\{n^{-1}\sum_{i=1}^{n}[I(C_i \geq t)A_i(t)\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})] - E[I(C_1 \geq t)A_1(t)W_1^A(t;\boldsymbol{\theta}_0)]\}$

We could decompose the target quantity as

$$
\begin{aligned}
& n^{1/2}\{n^{-1}\sum_{i=1}^{n}[I(C_i \geq t)A_i(t)\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})] - E[I(C_1 \geq t)A_1(t)W_1^A(t;\boldsymbol{\theta}_0)]\} \\
=& n^{1/2}\{n^{-1}\sum_{i=1}^{n}[I(C_i \geq t)A_i(t)\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})] - n^{-1}\sum_{i=1}^{n}[I(C_i \geq t)A_i(t)W_i^A(t;\boldsymbol{\theta}_0)]\} \qquad \text{(B.41)} \\
+& n^{1/2}\{n^{-1}\sum_{i=1}^{n}[I(C_i \geq t)A_i(t)W_i^A(t;\boldsymbol{\theta}_0)] - E[I(C_1 \geq t)A_1(t)W_1^A(t;\boldsymbol{\theta}_0)] \qquad \text{(B.42)}
\end{aligned}
$$

By (B.16), the first part (B.41) could be expressed as

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n} I(C_i \geq t)A_i(t)W_i^A(t;\boldsymbol{\theta}_0)n^{-1}\sum_{l=1}^{n}\Bigg[\int_0^t\{\boldsymbol{X}_i^T(s)ds + \boldsymbol{h}_{C_2}^T(s)\}Y_i(s)[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_l^{C_2}(\boldsymbol{\theta}_0) + \\
& \qquad\qquad \int_0^t x^{(0)}(s)^{-1}Y_i(s)dM_l^{C_2}(s;\boldsymbol{\theta}_0)\Bigg] \\
=& n^{-1/2}\sum_{i=1}^{n}\Bigg[n^{-1}\sum_{l=1}^{n} I(C_l \geq t)A_l(t)W_l^A(t;\boldsymbol{\theta}_0)\int_0^t\{\boldsymbol{X}_l^T(s)ds + \boldsymbol{h}_{C_2}^T(s)\}Y_l(s)[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0) \\
& \qquad + n^{-1}\sum_{l=1}^{n} I(C_l \geq t)A_l(t)W_l^A(t;\boldsymbol{\theta}_0)\int_0^t x^{(0)}(s)^{-1}Y_l(s)dM_i^{C_2}(s;\boldsymbol{\theta}_0)\Bigg] \\
=& n^{-1/2}\sum_{i=1}^{n}\Bigg[E\Bigg\{I(C_1 \geq t)A_1(t)W_1^A(t;\boldsymbol{\theta}_0)\int_0^t\{\boldsymbol{X}_1^T(s)ds + \boldsymbol{h}_{C_2}^T(s)\}Y_1(s)\Bigg\}[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0) \\
& \qquad + E\Bigg\{I(C_1 \geq t)A_1(t)W_1^A(t;\boldsymbol{\theta}_0)\Bigg\}\int_0^t x^{(0)}(s)^{-1}E[Y_1(s)]dM_i^{C_2}(s;\boldsymbol{\theta}_0)\Bigg] + o_p(1) \\
\equiv& n^{-1/2}\sum_{i=1}^{n}\tilde{f}_i^{\pi_{11},A}(t) + o_p(1) \qquad \text{(B.43)}
\end{aligned}
$$

Define $\tilde{f}_i^{\pi_{12},A}(t) = I(C_i \geq t)A_i(t)W_i^A(t;\boldsymbol{\theta}_0)$. Combining (B.42) and (B.43), we have

$$
n^{1/2}\{n^{-1}\sum_{i=1}^{n}[I(C_i \geq t)A_i(t)\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})] - E[I(C_1 \geq t)A_1(t)W_1^A(t;\boldsymbol{\theta}_0)]\} = \mathbb{G}_n\tilde{f}^{\pi_1,A} + o_p(1) \qquad \text{(B.44)}
$$

where $\tilde{f}_i^{\pi_1,A}(t) = \tilde{f}^{\pi_{11},A}(t) + \tilde{f}^{\pi_{12},A}(t)$, and by Donsker's theorem, (B.44) converges to $\mathbb{G}\tilde{f}^{\pi_1,A}$.

B.4.10.2 $n^{1/2}\{n^{-1}\sum_{i=1}^{n}[I(C_i \geq t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)}\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})] - E[I(C_1 \geq t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)}W_1^A(t;\boldsymbol{\theta}_0)]\}$

The target quantity could be expressed as

$$
\begin{aligned}
& n^{1/2}\{n^{-1}\sum_{i=1}^{n}[I(C_i \geq t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)}\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})] - E[I(C_1 \geq t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)}W_1^A(t;\boldsymbol{\theta}_0)]\} \\
=& n^{-1/2}\sum_{i=1}^{n}[I(C_i \geq t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)}\widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}}) - I(C_i \geq t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)}W_i^A(t;\boldsymbol{\theta}_0)] \qquad \text{(B.45)} \\
+& n^{-1/2}\sum_{i=1}^{n}[I(C_i \geq t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)}W_i^A(t;\boldsymbol{\theta}_0) - I(C_i \geq t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)}W_i^A(t;\boldsymbol{\theta}_0)] \qquad \text{(B.46)} \\
+& n^{-1/2}\sum_{i=1}^{n}[I(C_i \geq t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)}W_i^A(t;\boldsymbol{\theta}_0)] - n^{1/2}E[I(C_1 \geq t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)}W_1^A(t;\boldsymbol{\theta}_0)] \qquad \text{(B.47)}
\end{aligned}
$$

Having (B.16), the first term (B.45) equals to

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n} I(C_i \geq t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)}W_i^A(t;\boldsymbol{\theta}_0) \\
& \qquad n^{-1}\sum_{l=1}^{n}\left[\int_0^t \{\boldsymbol{X}_i^T(s)ds + \boldsymbol{h}_{C_2}^T(s)\}Y_i(s)[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_l^{C_2}(\boldsymbol{\theta}) + \int_0^t x^{(0)}(s)^{-1}Y_i(s)dM_l^{C_2}(s;\boldsymbol{\theta}_0)\right] \\
=& n^{-1/2}\sum_{i=1}^{n}\left[n^{-1}\sum_{l=1}^{n} I(C_l \geq t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_l(t)}W_l^A(t;\boldsymbol{\theta}_0)\int_0^t [\boldsymbol{X}_l^T(s)ds + \boldsymbol{h}_{C_2}^T(s)]Y_l(s)[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0)\right. \\
& \qquad \left. + n^{-1}\sum_{l=1}^{n} I(C_l \geq t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_l(t)}W_l^A(t;\boldsymbol{\theta}_0)\int_0^t x^{(0)}(s)^{-1}Y_l(s)dM_i^{C_2}(s;\boldsymbol{\theta}_0)\right] \\
=& n^{-1/2}\sum_{i=1}^{n}\left[E\left\{I(C_1 \geq t)e^{{\boldsymbol{\beta}_0}^T \boldsymbol{Z}_1(t)}W_1^A(t;\boldsymbol{\theta}_0)\int_0^t \{\boldsymbol{X}_1^T(s)ds + \boldsymbol{h}_{C_2}^T(s)\}Y_1(s)\right\}[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0)\right. \\
& \qquad \left. + \boldsymbol{z}^{(0)}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]\int_0^t x^{(0)}(s)^{-1}E[Y_1(s)]dM_i^{C_2}(s;\boldsymbol{\theta}_0)\right] + o_p(1) \\
\equiv& n^{-1/2}\sum_{i=1}^{n}\tilde{f}_i^{\pi_{21},A}(t) + o_p(1) \qquad \text{(B.48)}
\end{aligned}
$$

By (B.32) and Taylor expansion, the second term could be expressed as

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n} I(C_i \geq t) W_i^A(t;\boldsymbol{\theta}_0) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)} \boldsymbol{Z}_i^T(t) n^{-1} \sum_{l=1}^{n} \boldsymbol{f}_l^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\
=& n^{-1/2}\sum_{i=1}^{n} \boldsymbol{Z}^{(1)}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]^T \boldsymbol{f}_i^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\
=& n^{-1/2}\sum_{i=1}^{n} \boldsymbol{z}^{(1)}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]^T \boldsymbol{f}_i^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) + o_p(1) \\
\equiv& n^{-1/2}\sum_{i=1}^{n} \tilde{f}_i^{\pi_{22},A}(t) + o_p(1)
\end{aligned} \tag{B.49}
$$

Define $\tilde{f}_i^{\pi_{23},A}(t) = I(C_i \geq t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)} W_i^A(t;\boldsymbol{\theta}_0)$. Combining (B.47), (B.48) and (B.49), we have

$$
n^{1/2}\{n^{-1}\sum_{i=1}^{n}[I(C_i \geq t) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)} \widehat{W}_i^A(t;\widehat{\boldsymbol{\theta}})] - E[I(C_1 \geq t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_1(t)} W_1^A(t;\boldsymbol{\theta}_0)]\} = \mathbb{G}_n \tilde{f}^{\pi_2,A} + o_p(1) \tag{B.50}
$$

where $\tilde{f}_i^{\pi_2,A}(t) = \tilde{f}_i^{\pi_{21},A}(t) + \tilde{f}_i^{\pi_{22},A}(t) + \tilde{f}_i^{\pi_{23},A}(t)$.

B.4.10.3 $n^{1/2}\{\widehat{\pi}_0^A(t) - \pi_0(t)\}$

By the Donsker's theorem, we have

$$
n^{1/2} \begin{pmatrix} \mathbb{P}_n \tilde{f}^{\pi_1,A} - P\tilde{f}^{\pi_1,A} \\ \mathbb{P}_n \tilde{f}^{\pi_2,A} - P\tilde{f}^{\pi_1,A} \end{pmatrix} \Longrightarrow \begin{pmatrix} \mathbb{G}\tilde{f}^{\pi_1,A} \\ \mathbb{G}\tilde{f}^{\pi_2,A} \end{pmatrix}
$$

where $\mathbb{G}\tilde{f}^{\pi_1,A}$ and $\mathbb{G}\tilde{f}^{\pi_2,A}$ are tight Gaussian process on $(0,\tau]$.

By Delta method we have

$$
n^{1/2}[\widehat{\pi}_0^A - \pi_0] \Longrightarrow \mathbb{G}\Big[\frac{\tilde{f}^{\pi_1,A} - (P\tilde{f}^{\pi_1,A})\tilde{f}^{\pi_2,A}}{P\tilde{f}^{\pi_2,A}}\Big] \tag{B.51}
$$

B.4.11 $n^{1/2}\{\widehat{\pi}_0^B(t) - \pi_0(t)\}$

Following similar arguments we have

$$n^{1/2}[\widehat{\pi}_0^B - \pi_0] \Longrightarrow \mathbb{G}\Big[\frac{\tilde{f}^{\pi_1,B} - (P\tilde{f}^{\pi_1,B})\tilde{f}^{\pi_2,B}}{P\tilde{f}^{\pi_2,B}}\Big] \tag{B.52}$$

where

$$\begin{aligned}
\tilde{f}_i^{\pi_1,B}(t) &= \tilde{f}_i^{\pi_{11},B}(t) + \tilde{f}_i^{\pi_{12},B}(t) \\
\tilde{f}_i^{\pi_{11},B}(t) &= E\Big\{I(C_1 \geq t)A_1(t)W_1^B(t;\boldsymbol{\theta}_0)\int_0^t \boldsymbol{X}_1^T(s)Y_1(s)ds\Big\}[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0) \\
\tilde{f}_i^{\pi_{12},B}(t) &= I(C_i \geq t)A_i(t)W_i^B(t;\boldsymbol{\theta}_0) \\
\tilde{f}_i^{\pi_2,B}(t) &= \tilde{f}_i^{\pi_{21},B}(t) + \tilde{f}_i^{\pi_{22},B}(t) + \tilde{f}_i^{\pi_{23},B}(t) \\
\tilde{f}_i^{\pi_{21},A}(t) &= E\Big\{I(C_1 \geq t)e^{\boldsymbol{\beta}_0{}^T\boldsymbol{Z}_1(t)}W_1^B(t;\boldsymbol{\theta}_0)\int_0^t \boldsymbol{X}_1^T(s)Y_1(s)ds\Big\}[\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}_0) \\
\tilde{f}_i^{\pi_{22},B}(t) &= \boldsymbol{z}^{(1)}[t;\boldsymbol{\beta}_0, W^B(\boldsymbol{\theta}_0)]^T\boldsymbol{f}_i^{\boldsymbol{\beta}_B}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0) \\
\tilde{f}_i^{\pi_{23},B}(t) &= I(C_i \geq t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i(t)}W_i^B(t;\boldsymbol{\theta}_0)
\end{aligned}$$

B.5 Proof for Random Censoring

Suppose that we have a total of M imputations. Define $\Delta_{1i} = I[C_{1i} \leq D_i \wedge C_{2i}]$, where $X_i = D_i \wedge C_{1i} \wedge C_{2i}$, $Y_i(t) = I(X_i \geq t)$. Note that $Y_i(t)$ are the same for each imputed datasets, and for the underlying knowing censoring dataset. Let $\widehat{\boldsymbol{\gamma}}$ be the estimator of $\boldsymbol{\gamma}_0$ from the observed C_1 data $\{X_i, \Delta_{1i}\}_{i=1}^n$. In finite samples, $C_{1i}^{\langle m\rangle}$ are imputed from the estimated conditional probabilities with parameters $\widehat{\boldsymbol{\gamma}}$, which converge almost surely to the true value. It can then be shown that $I(C_{1i}^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\gamma}})$ and $I(C_{1i}^{\langle m\rangle} \geq t;\boldsymbol{\gamma}_0)$ are asymptotically equivalent, or more explicitly

$$n^{-1/2}\sum_{i=1}^n I(C_{1i}^{\langle m\rangle} \geq t;\widehat{\boldsymbol{\gamma}}) = n^{-1/2}\sum_{i=1}^n I(C_{1i}^{\langle m\rangle} \geq t;\boldsymbol{\gamma}_0) + o_p(1)$$

Asymptotically, our imputed censoring time are drawn from the underlying true conditional probability with $\boldsymbol{\gamma}_0$. More detailed proofs could be referred to a previous manuscript "Semiparametric Temporal Process Regression of Survival-Out-of-Hospital".

For the large sample properties of $\widehat{\boldsymbol{\beta}}_A(M)$, we have

$$n^{1/2}[\widehat{\boldsymbol{\beta}}_A(M) - \boldsymbol{\beta}_0] = n^{-1/2}\sum_{i=1}^{n} \boldsymbol{f}_i^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{\beta}_0, M) + o_p(1) \tag{B.53}$$

where

$$\begin{aligned}\boldsymbol{f}_i^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\beta}, M) = \boldsymbol{\Omega}[\boldsymbol{\beta}, W^A(\boldsymbol{\theta})]^{-1} \\ \Big[\int_0^\tau \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{z}}[t; \boldsymbol{\beta}, W^A(\boldsymbol{\theta})]\} W_i^A(t; \boldsymbol{\theta}) \frac{1}{M}\sum_{m=1}^{M} dM_i^{\langle m\rangle}(t; \boldsymbol{\beta}, \boldsymbol{\gamma}) + \\ \boldsymbol{H}[\boldsymbol{\beta}, \boldsymbol{\theta}, W^A(\boldsymbol{\theta})][\boldsymbol{\Omega}^{C_2}]^{-1}\boldsymbol{u}_i^{C_2}(\boldsymbol{\theta}) + \\ \int_0^\tau \boldsymbol{J}[t, \tau, \boldsymbol{\beta}, \boldsymbol{\theta}, W^A(\boldsymbol{\theta})] dM_i^{C_2}(t; \boldsymbol{\theta})\Big]\end{aligned}$$

$$dM_i^{\langle m\rangle}(t; \boldsymbol{\beta}, \boldsymbol{\gamma}) = \left[A_i(t) - \pi_0(t)\text{exp}\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}\right] I(C_{1i}^{\langle m\rangle} \geq t; \boldsymbol{\gamma}) I(C_{2i} \geq t) dt$$

Similarly, we could develop the asymptotic properties of $\widehat{\boldsymbol{\beta}}_B(M)$.

As for $\widehat{\pi}_0^A(t, M)$, it is given by

$$\widehat{\pi}_0^A(t, M) = \frac{M^{-1}\sum_{m=1}^{M}\sum_{i=1}^{n} A_i(t) I(C_{1i}^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\gamma}}) I(C_{2i} \geq t) W_i^A(t; \boldsymbol{\theta}_0)}{M^{-1}\sum_{m=1}^{M}\sum_{i=1}^{n} I(C_{1i}^{\langle m\rangle} \geq t; \widehat{\boldsymbol{\gamma}}) I(C_{2i} \geq t) W_i^A(t; \boldsymbol{\theta}_0)\text{exp}[\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(t)]}$$

We could show that

$$n^{1/2}[\widehat{\pi}_0^A(M) - \pi_0] \Longrightarrow \mathbb{G}\Big[\frac{\tilde{f}^{\pi_1, A} - \{P\tilde{f}^{\pi_1, A}\}\tilde{f}^{\pi_2, A}(M)}{P\tilde{f}^{\pi_2, A}(M)}\Big]$$

where

$$\begin{aligned}
\tilde{f}^{\pi_2,A}(t,M) =& \tilde{f}^{\pi_{21},A}(t) + \tilde{f}^{\pi_{22},A}(t,M) + \tilde{f}^{\pi_{23},A}(t,M) \\
\tilde{f}^{\pi_{22},A}(t,M) =& \boldsymbol{z}^{(1)}[t;\boldsymbol{\beta}_0, W^A(\boldsymbol{\theta}_0)]^T \boldsymbol{f}_i^{\boldsymbol{\beta}_A}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,M) \\
\tilde{f}^{\pi_{23},A}(t,M) =& \frac{1}{M}\sum_{m=1}^{M} I(C_{1i}^{\langle m\rangle} \geq t;\boldsymbol{\gamma}_0) I(C_{2i} \geq t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)} W_i^A(t;\boldsymbol{\theta}_0)
\end{aligned}$$

The asymptotic properties of $\widehat{\pi}_0^B(t,M)$ could be obtained following parallel arguments.

Bibliography

P. S. Albert. A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, 56(2):602–608, 2000.

P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: a large sample study. *The annals of statistics*, 10:1100–1120, 1982.

P. K. Andersen, J. P. Klein, and S. Rosthøj. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, pages 15–27, 2003.

A.-C. Andrei and S. Murray. Regression models for the mean of the quality-of-life-adjusted restricted survival time using pseudo-observations. *Biometrics*, 63(2):398–404, 2007.

H. Bang and A. A. Tsiatis. Estimating medical costs with censored data. *Biometrika*, 87(2): 329–343, 2000.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

N. E. Breslow. Contribution to the discussion of the paper by D.R. Cox. *Journal of the Royal Statistical Society, Series B*, 34(2):216–217, 1972.

N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25, 1993.

R. C. Carson, M. Juszczak, A. Davenport, and A. Burns. Is maximum conservative management an equivalent treatment option to dialysis for elderly patients with significant comorbid disease? *Clinical Journal of the American Society of Nephrology*, 4(10):1611–1619, 2009.

Y.-Y. Chi and J. G. Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445, 2006.

B. F. Cole, R. D. Gelber, and A. Goldhirsch. Cox regression models for quality adjusted survival analysis. *Statistics in medicine*, 12(10):975–987, 1993.

R. J. Cook and J. F. Lawless. Marginal analysis of recurrent events and a terminating event. *Statistics in medicine*, 16(8):911–924, 1997.

R. J. Cook, J. F. Lawless, L. Lakhal-Chaieb, and K.-A. Lee. Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination:

application to skeletal complications in cancer metastatic to bone. *Journal of the American Statistical Association*, 104(485):60–75, 2009.

D. R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220, 1972a.

D. R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220, 1972b.

D. R. Cox. Partial likelihood. *Biometrika*, pages 269–276, 1975.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

J. P. Estes, D. V. Nguyen, L. S. Dalrymple, Y. Mu, and D. Şentürk. Time-varying effect modeling with longitudinal data truncated by death: conditional models, interpretations, and inference. *Statistics in medicine*, 2015.

J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509, 1999.

J. P. Fine, J. Yan, and M. R. Kosorok. Temporal process regression. *Biometrika*, 91(3):683–703, 2004.

R. D. Gelber, R. S. Gelman, and A. Goldhirsch. A quality-of-life-oriented endpoint for comparing therapies. *Biometrics*, pages 781–795, 1989.

D. Ghosh and D. Y. Lin. Marginal regression models for recurrent and terminal events. *Statistica Sinica*, pages 663–688, 2002.

P. Glasziou, R. Simes, and R. Gelber. Quality adjusted survival analysis. *Statistics in medicine*, 9 (11):1259–1276, 1990.

H. Goldstein. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73(1):43–56, 1986.

H. Goldstein. Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, pages 45–51, 1991.

Q. Gong and D. E. Schaubel. Partly conditional estimation of the effect of a time-dependent factor in the presence of dependent censoring. *Biometrics*, 69(2):338–347, 2013.

Q. Gong and D. E. Schaubel. Estimating the average treatment effect on survival based on observational data and using partly conditional modeling. *Biometrics*, 73(1):134–144, 2017.

M. K. Grand and H. Putter. Regression models for expected length of stay. *Statistics in medicine*, 35(7):1178–1192, 2016.

R. J. Gray. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of statistics*, pages 1141–1154, 1988.

I. D. Ha, Y. Lee, and J.-k. Song. Hierarchical likelihood approach for frailty models. *Biometrika*, 88(1):233–233, 2001.

M. Á. Hernán, B. Brumback, and J. M. Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology*, 11(5):561–570, 2000.

D. C. Holland and M. Lam. Predictors of hospitalization and death among pre-dialysis patients: a retrospective cohort study. *Nephrology Dialysis Transplantation*, 15(5):650–658, 2000.

X. J. Hu, M. Lorenzi, J. J. Spinelli, S. C. Ying, and M. L. McBride. Analysis of recurrent events with non-negligible event duration, with application to assessing hospital utilization. *Lifetime data analysis*, 17(2):215–233, 2011.

C.-Y. Huang and M.-C. Wang. Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, 99(468):1153–1165, 2004.

Y. Huang and T. A. Louis. Expressing estimators of expected quality adjusted survival as functions of nelson-aalen estimators. *Lifetime Data Analysis*, 5(3):199–212, 1999.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. New York: Wiley, 2 edition, 2002.

J. D. Kalbfleisch, D. E. Schaubel, Y. Ye, and Q. Gong. An estimating function approach to the analysis of recurrent and terminal events. *Biometrics*, 69(2):366–374, 2013.

M. J. Laan and A. Hubbard. Locally efficient estimation of the quality-adjusted lifetime distribution with right-censored data and covariates. *Biometrics*, 55(2):530–536, 1999.

K. Larsen. Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics*, 60(1):85–92, 2004.

Y. Lee and J. A. Nelder. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 619–678, 1996.

J. A. Lewis. Statistical principles for clinical trials (ich e9): an introductory note on an international guideline. *Statistics in medicine*, 18(15):1903–1942, 1999.

X. Li, Y. Chen, and R. Li. A frailty model for recurrent events during alternating restraint and non-restraint time periods. *Statistics in Medicine*, 36(4):643–654, 2017.

K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

D. Lin and Z. Ying. Semiparametric analysis of the additive risk model. *Biometrika*, 81(1):61–71, 1994.

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

D. Lin, E. Feuer, R. Etzioni, and Y. Wax. Estimating medical costs from incomplete follow-up data. *Biometrics*, pages 419–434, 1997.

D. Y. Lin, L. J. Wei, I. Yang, and Z. Ying. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730, 2000.

R. J. Little and D. B. Rubin. *Statistical analysis with missing data*. New York: Wiley, 2002.

L. Liu, R. A. Wolfe, and X. Huang. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3):747–756, 2004.

Q. Liu and D. A. Pierce. A note on gauss—hermite quadrature. *Biometrika*, 81(3):624–629, 1994.

L. Mao and D. Y. Lin. Semiparametric regression for the weighted composite endpoint of recurrent and terminal events. *Biostatistics*, 17(2):390, 2016.

C. E. McCulloch and J. M. Neuhaus. *Generalized linear mixed models*. Wiley Online Library, 2001.

S. Murray and B. Cole. Variance and sample size calculations in quality-of-life-adjusted survival analysis (q-twist). *Biometrics*, 56(1):173–182, 2000.

J. D. Neaton, G. Gray, B. D. Zuckerman, and M. A. Konstam. Key issues in end point selection for heart failure trials: composite end points. *Journal of cardiac failure*, 11(8):567–575, 2005.

Q. Pan and D. E. Schaubel. Proportional hazards models based on biased samples and estimated selection probabilities. *Canadian Journal of Statistics*, 36(1):111–127, 2008.

L. Peng and Y. Huang. Survival analysis with temporal covariate effects. *Biometrika*, 94(3):719–733, 2007.

M. A. Pfeffer, K. Swedberg, C. B. Granger, P. Held, J. J. McMurray, E. L. Michelson, B. Olofsson, J. Östergren, S. Yusuf, C. Investigators, Committees, et al. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the charm-overall programme. *The Lancet*, 362(9386):759–766, 2003.

R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson Jr, N. Flournoy, V. T. Farewell, and N. E. Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, pages 541–554, 1978.

E. P. Pulkstenis, T. R. Ten Have, and J. R. Landis. Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*, 93(442):438–450, 1998.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL http://www.R-project.org. ISBN 3-900051-07-0.

D. Rizopoulos, G. Verbeke, E. Lesaffre, and Y. Vanrenterghem. A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, 64(2):611–619, 2008.

J. M. Robins. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section, American Statistical Association*, volume 24, page 3, 1993.

J. M. Robins and D. M. Finkelstein. Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics*, 56(3):779–788, 2000.

J. M. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology*, pages 297–331, 1992.

V. Rondeau, S. Mathoulin-Pelissier, H. Jacqmin-Gadda, V. Brouste, and P. Soubeyran. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*, 8(4):708–721, 2007.

P. K. Ruan and R. J. Gray. Analyses of cumulative incidence functions via non-parametric multiple imputation. *Statistics in medicine*, 27(27):5709–5724, 2008.

D. B. Rubin. *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons, 2004.

J. J. Sands, G. D. Etheredge, A. Shankar, J. Graff, J. Loeper, M. McKendry, and R. Farrell. Predicting hospitalization and mortality in end-stage renal disease (esrd) patients using an index of coexisting disease (iced)-based risk stratification model. *Disease Management*, 9(4):224–235, 2006.

SAS Institute Inc. *SAS Software, Version 9.4*. Cary, NC, 2013. URL http://www.sas.com/.

P. Sasieni. Maximum weighted partial likelihood estimators for the cox model. *Journal of the American Statistical Association*, 88(421):144–152, 1993.

D. E. Schaubel and J. Cai. Multiple imputation methods for recurrent event data with missing event category. *Canadian Journal of Statistics*, 34(4):677–692, 2006.

D. E. Schaubel and M. Zhang. Estimating treatment effects on the marginal recurrent event mean in the presence of a terminating event. *Lifetime data analysis*, 16(4):451–477, 2010.

D. E. Schaubel, D. Zeng, and J. Cai. A semiparametric additive rates model for recurrent event data. *Lifetime Data Analysis*, 12(4):389–406, 2006.

T. H. Scheike and M.-J. Zhang. Direct modelling of regression effects for transition probabilities in multistate models. *Scandinavian Journal of Statistics*, 34(1):17–32, 2007.

K. E. Shirley, D. S. Small, K. G. Lynch, S. A. Maisto, and D. W. Oslin. Hidden markov models for alcoholism treatment trial data. *The Annals of Applied Statistics*, pages 366–395, 2010.

R. Thomas, T. Have, A. R. Kunselman, E. P. Pulkstenis, and J. R. Landis. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, pages 367–383, 1998.

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

N. Wang and J. M. Robins. Large-sample theory for parametric multiple imputation procedures. *Biometrika*, 85(4):935–948, 1998.

A. R. Willan, D. Lin, R. Cook, and E. Chen. Using inverse-weighting in cost-effectiveness analysis with censored data. *Statistical methods in medical research*, 11(6):539–551, 2002.

M. S. Wulfsohn and A. A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339, 1997.

J. Yan and J. P. Fine. Functional association models for multivariate survival processes. *Journal of the American Statistical Association*, 100(469):184–196, 2005.

J. Yan and J. Huang. Partly functional temporal process regression with semiparametric profile estimating functions. *Biometrics*, 65(2):431–440, 2009.

Y. Ye, J. D. Kalbfleisch, and D. E. Schaubel. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics*, 63(1):78–87, 2007.

E. W. Young, D. A. Goodkin, D. L. Mapes, F. K. Port, M. L. Keen, K. Chen, B. L. Maroni, R. A. Wolfe, and P. J. Held. The dialysis outcomes and practice patterns study (dopps): an international hemodialysis study. *Kidney International*, 57:S74–S81, 2000.

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

D. Zeng and J. Cai. A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika*, 97:699–712, 2010.

M. Zhang and D. E. Schaubel. Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics*, 67(3):740–749, 2011.

H. Zhao and A. A. Tsiatis. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika*, 84(2):339–348, 1997.

Y. Zhao and H. Wang. Empirical likelihood inference for the regression model of mean quality-adjusted lifetime with censored data. *Canadian Journal of Statistics*, 36(3):463–478, 2008.

L. Zhu, H. Zhao, J. Sun, and S. Pounds. A conditional approach for regression analysis of longitudinal data with informative observation time and non-negligible observation duration. *Communications in Statistics-Theory and Methods*, 43(23):4998–5011, 2014.