Methods for Direct Modeling of Restricted Mean Survival Time for General Censoring Mechanisms and Causal Inference

by

Xin Wang

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 Associate Professor Yun Li
Professor Susan Murray
Emeritus Professor Robert Mark Merion

Xin Wang
vivianwx@umich.edu
ORCID iD: 0000-0002-1846-9726

DEDICATION

To my family.

ACKNOWLEDGEMENTS

My work was supported in part by National Institutes of Health Grant 5R01 DK070869. Data analyzed in the first chapter were supplied by the Minneapolis Medical Research Foundation as the contractor for the Scientific Registry of Transplant Recipients. And data in the second and third chapter were supplied by Kidney Epidemiology and Cost Center as the contractor for United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the Scientific Registry of Transplant Recipients or the U.S. Government.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

APPENDIX

ABSTRACT

Restricted mean survival time (RMST) is often of great clinical interest in practice. Several existing methods involve explicitly projecting out patient-specific survival curves using parameters estimated through Cox regression. It is often be preferable to directly model the restricted mean, for convenience and to yield more directly interpretable covariate effects.

In the first chapter, we propose generalized estimating equation methods to model RMST as a function of baseline covariates. The proposed methods avoid potentially problematic distributional assumptions pertaining to restricted survival time. Unlike existing methods, we allow censoring to depend on both baseline and time-dependent factors. The methods are motivated by the end-stage liver disease (ESLD) setting and, in particular, consider survival in the absence of the preferred therapy, liver transplantation.

In the second chapter, we propose generalized estimating equation methods to fit RMST models with multiplicative covariate effects. The proposed methods are applicable to several frequently occurring set-ups not considered in Chapter 1, including clustered data and data with a high-dimensional categorical covariate (e.g., center). Our proposed methods are motivated by modeling RMST among End-stage Renal Disease (ESRD) patients, in the presence of a high-dimensional covariate (1 million patients from over 5,000 dialysis facility). Estimation proceeds through a computationally efficient two-stage algorithm. In addition to evaluating large- and finite-sample properties, we demonstrate the considerable computational advantages

of the proposed techniques.

The third chapter is motivated by estimating the causal treatment in the presence of unmeasured confounding. We propose two-stage Instrumental Variable techniques for censored data. In particular, we develop closed-form, two-stage estimators for the causal treatment effect using an additive RMST model. Large sample properties are derived, with simulation studies conducted to assess finite sample properties. We apply the proposed methods to estimate the causal effect of peritoneal dialysis (PD) versus hemodialysis (HD) among End-Stage Renal Disease (ESRD) patients.

Keywords: Restricted mean survival time, Dependent censoring, Center effect, Instrumental variable, End-Stage Liver Disease, End-Stage Renal Disease.

CHAPTER I

Modeling Restricted Mean Survival Time under General Censoring Mechanisms

1.1 Introduction

The Cox proportional hazards model (Cox, 1972, 1975) is the strong default for analyzing time to event data with covariate adjustment. A key motivation for the hazard ratio (HR) is its connection to the ordering of the survival functions, under the assumption of proportional hazards. However, when there are departures from proportional hazards, this connection is lost and it is then difficult to interpret the HR. A HR estimated by ignoring the non-proportionality will be a poorly specified mixture of the survival distribution and censoring distribution (Gillen and Emerson, 2007), such that the resulting inference may then differ for studies with identical survival time distributions but different censoring patterns. In the presence of non-proportionality, alternatives to the HR include the 'average effect' (Xu and O'Quigley, 2000), or applying a 'stopped Cox model' (Van Houwelingen, 2007; Van Houwelingen and Putter, 2015).

Covariate effects that are cumulative in nature are often of greater interest than instantaneous effects, especially in the presence of non-proportionality (Schaubel and Wei, 2011). In particular, the contrast in restricted mean survival time (RMST) is a useful alternative. The RMST is defined as the average survival time up to a fixed

point L and can be written as the area under the survival curve on $[0, L]$. RMST is an easily interpretable and clinically relevant measure for summarizing the mortality over a fixed follow-up time period of interest. Most existing methods estimate RMST indirectly through hazard regression (Zucker, 1998; Chen and Tsiatis, 2001; Zhang and Schaubel, 2011). These approaches start by estimating the regression parameters and baseline hazard from a Cox model, calculating the cumulative baseline hazard, transforming it to obtain the survival function and, and finally integrating the survival function to obtain the RMST. Such indirect RMST estimation is inconvenient and computationally cumbersome, even for obtaining a point estimate, let alone its corresponding asymptotic standard error. Hence, it may be preferable to directly model RMST itself (Andersen et al., 2004; Tian et al., 2014).

The majority of existing methods for directly modeling RMST require assumptions regarding the censoring mechanism, which are often untenable. Censoring may result from multiple sources in an observational study. The simplest type would be *covariate-independent* censoring, which occurs independently of the death time and all the covariates. When this is the only type of censoring present, one can conduct regression analysis of RMST using imputed event times based on pseudo-observation methods (Andersen et al., 2004), or one can construct estimating equations for RMST based on Inverse Probability of Censoring Weighting (IPCW) (Robins and Rotnitzky, 1992; Robins, 1993; Robins and Finkelstein, 2000) as in Tian et al. (2014). However, in observational studies, censoring will often depend on the covariate vector. Censoring can depend on baseline covariates, but be conditionally independent of the event time given such covariates; this is referred to as *covariate-dependent* censoring. For example, it is common to have a staggered entry in an observational study with a fixed calendar period, such that subjects who enter later would have a different

censoring distribution than those who enter earlier; e.g., registration date on the wait-list for a liver transplant. Since mortality is often subject to calendar time trends, covariate-dependent censoring would be expected to be a frequent occurrence in observational studies. Andersen and Perme (2009) and Binder et al. (2014) conducted simulation studies to examine the bias and efficiency of the pseudo-observations approach for competing risks, in the presence of covariate dependent censoring. A third type of censoring is *dependent censoring*, which is often correlated with the event time through a mutual association with time varying covariates. Covariate-dependent and dependent censoring have been overcome in many applications by IPCW. Through pseudo-observations, Xiang and Murray (2012) modeled a standard linear regression of restricted survival time on the logarithm scale and handled dependent censoring through IPCW. Specifically, we connect the RMST and covariate vector through a user-specified link function, while Xiang and Murray (2012) model log restricted survival time through linear regression. In addition, being based on a pseudo-observation approach, their work has no systematic procedure for evaluating the asymptotic properties. To our knowledge, there is no existing method to directly model RMST in the presence of dependent censoring, or even covariate-dependent censoring.

The setting which motivated the proposed methods involves mortality in the absence of liver transplantation among End-Stage Liver Disease (ESLD) patients. Since the number of patients in need of liver transplantation is much greater than the number of available deceased-donor livers, medically suitable ESLD patients are placed on a wait-list. Priority for transplantation is then determined by medical urgency, as quantified by the Model for End-Stage Liver Disease (MELD) score. This score is calculated using the bounded versions of serum bilirubin, serum creatinine, inter-

national normalized ratio for prothrombin time (INR), and dialysis status (Kamath et al., 2001; Wiesner et al., 2003). The MELD score has been shown to be strongly predictive of pre-transplant survival among chronic ESLD patients (Kamath et al., 2001). For a given ESLD patient, the MELD score is updated frequently, such that MELD constitutes a time-varying covariate. Since wait-listed patients are sequenced on the wait-list in decreasing order of current MELD score, MELD is strongly associated with transplant rate. As the organ assignment is correlated with pre-transplant mortality through its mutual association with time varying MELD score, dependent censoring occurs through the receipt of a liver transplant, which precludes the observation of pre-transplant death. We are interested in the effect on pre-transplant mortality of prognostic factors observed at the time of wait-listing, as such information would be useful to hepatologists and transplant surgeons for counseling patients.

We propose semi-parametric regression methods for directly modeling RMST given baseline covariates in the presence of both covariate-dependent and dependent censoring. The proposed methods can be used to evaluate the cumulative effect of baseline covariates and to quantify treatment effects in terms of contrast in RMST. Our proposed methods do not require any distributional assumption on the death variates and, analogous to generalized linear models, allow for different link functions.

The contribution of our proposed work, compared to Tian et al. (2014), is that the latter requires that censoring does not depend on the covariate vector. Although random censoring may be a reasonable assumption in clinical trials, it will often fail in observational studies. Our methods not only allow for covariate-dependent censoring, but also allow for dependent censoring (e.g., dependence between the death and censoring times not captured by the covariates used in the death model). In Tian et al. (2014), the weight function is the inverse of the Kaplan-Meier estimator. In

the methods we propose, we distinguish between covariate dependent and dependent censoring; in particular, a double inverse weight is required and estimated through separate Cox models for the two types of censoring.

The remainder of this article is organized as follows. In Section (1.2), we formulate the data structure, define the necessary assumptions and then describe the proposed methods. Asymptotic properties are given in Section (1.3). In Section (1.4), we conduct simulation studies to evaluate the accuracy of the proposed procedures in finite samples. In Section (1.5), we apply our methods to the motivating ESLD data to determine the effect on pre-transplant mortality of several clinically meaningful variables. We conclude this paper with a brief discussion in Section (1.6). Derivation of the asymptotic properties and additional results for ESLD data analysis are provided in the Appendix A.

1.2 Proposed Methods

We begin with the necessary notation. Let D_i be the treatment-free survival time for subject i from a cohort of sample size n. We consider two types of censoring. One potential censoring time, denoted as C_i, is independent of D_i conditional on the baseline covariates; this type of censoring includes loss to follow-up or administrative censoring on the day the database closes. The other potential censoring, denoted as T_i, is not conditionally independent of D_i given baseline covariates; one example would be treatment time, which may dependently censor pre-treatment mortality. The observation time for subject i is $Z_i = D_i \wedge T_i \wedge C_i$, where $a \wedge b = \min\{a, b\}$; and the indicators for at risk status, pre-transplant death, dependent and independent censoring are denoted by $R_i(t) = I(Z_i \geq t), \Delta_i^D = I(D_i \leq T_i \wedge C_i), \Delta_i^T = I(T_i < D_i \wedge C_i)$ and $\Delta_i^C = I(C_i < D_i \wedge T_i)$ respectively. We denote the covariates predicting D_i, T_i

and C_i by $\boldsymbol{Z}_i^D(t), \boldsymbol{Z}_i^T(t)$ and $\boldsymbol{Z}_i^C$ respectively. Although we have defined these notations to accommodate time varying covariates, some elements of each may be time constant; e.g., gender or race. In some practical studies, the investigators might want to use the same covariate set for censoring and death time; however, we will distinguish these covariate sets for the purpose of generality. Stacking these covariates together and removing any redundancy, we obtain $\boldsymbol{Z}_i(t)$ and the corresponding covariate history as $\widetilde{\boldsymbol{Z}}_i(t) = \{\boldsymbol{Z}_i(u) : 0 \leq u < t\}$. Our observed data are then given by $\{Z_i, \Delta_i^D, \Delta_i^T, \Delta_i^C, \widetilde{\boldsymbol{Z}}(Z_i) : i = 1, \ldots, n\}$.

Let L be a pre-specified time point of interest, before the maximum follow-up time $\tau = \max\{Z_i : i = 1, \ldots, n\}$. Denote the restricted observation time as $Y_i = Z_i \wedge L$ and its corresponding indicator $\Delta_i = I(D_i \wedge L \leq T_i \wedge C_i)$. We are interested in the average survival time up to L and will model this measure through baseline covariates $\boldsymbol{Z}_i^D(0)$:

$$\mu_i(L) := E\left\{D_i \wedge L | \boldsymbol{Z}_i^D(0)\right\}.$$

Analogous to a generalized linear model, we assume a direct relationship between this RMST and baseline covariates as follows:

$$g\left[\mu_i(L)\right] \equiv g\left[E\left\{D_i \wedge L | \boldsymbol{Z}_i^D(0)\right\}\right] = \boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0), \tag{1.1}$$

where g is a strictly monotone link function with a continuous derivative within an open neighborhood $\mathcal{B}_D$ of $\boldsymbol{\beta}_D$. Examples of link functions include $g(x) = x$ (identity link), $g(x) = \log(x)$ (log link) and $g(x) = \log(x/(L - x))$ (logistic link). We choose to model the impact of baseline covariates for many reasons. First, our intention is to develop a model useful for application at the start of follow-up. For example, modeling (1.1) such modeling may be used in counseling ESLD patients regarding their prognosis in the absence of liver transplantation, given the information

observed at the time of wait-list registration. Second, RMST prediction based on time-dependent covariates are difficult to interpret, at least for internal time varying factors (Kalbfleisch and Prentice, 2011). The role of $\boldsymbol{Z}_i^D(t)$ depends on the model being considered. In Eq. (1.1), only baseline values are used and we average over the time-varying process. However, time-dependent values are needed to accommodate dependent censoring, as explained in the paragraphs below.

The choice of L requires careful thought. One would normally choose a time point of clinical relevance or, at least, of particular interest to the investigators, respecting the bound at the maximum follow-up time. If too small an L value is selected, $D \wedge L = L$ for most subjects, leading to a largely uninformative analysis. Conversely, if too large an L value is selected, $\hat{S}(L) \approx 0$. Setting L too large or too small will generally result in attenuated covariate effects. The choice of link function also requires some consideration. The identity link is usually most interesting because of its straightforward interpretation. However, it has the problem of unbounded predicted values. From this perspective, the logistic link may be a better choice, at least for the purposes of prediction. In addition, practitioners can conduct sensitivity analyses and diagnosis procedures to test the performance of different link functions, as we demonstrate in Section (1.5).

Note that we do not make any assumption about the error structure, in the interest of flexibility and robustness. Although it might be difficult to envision an arbitrarily truncated variate having a well-behaved distribution, it is reasonable to assume that the corresponding mean has a convenient form. Framing the model in terms of $g[E(D_i \wedge L)]$ instead of $E[g(D_i \wedge L)]$ is very important in our settings. For example, if $\log\{E(D_i \wedge L)\} = \boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0)$, the parameters in $\boldsymbol{\beta}_D$ can be interpreted as the multiplicative effect on RMST per unit increase in the corresponding covariate. This

is quite different from the model assumption $E\{\log(D_i \wedge L)\} = \boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0)$, where $\boldsymbol{\beta}_D$ equals the average change in logarithm of restricted survival time per unit increase in $\boldsymbol{Z}_i^D(0)$. The latter interpretation is much less intuitive, as back-transforming is invalid in the light of Jensen's Inequality.

We now derive the estimating equation for the parameter of interest, $\boldsymbol{\beta}_D$. In absence of censoring, based on (1.1), $\boldsymbol{\beta}_D$ can be estimated via the following estimating equation:

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i^D(0)\left[D_i \wedge L - g^{-1}\left\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D(0)\right\}\right] = \mathbf{0}. \tag{1.2}$$

Although connected to generalized linear models, (1.2) is more accurately interpreted as a generalized estimating equation due to the absence of distributional assumptions on $D_i \wedge L$.

However, we will not observe D_i for all patients due to the occurrence of censoring. Instead we may observe either independent censoring time C_i or treatment time T_i. For independent censoring, it is reasonable to assume C_i is independent of D_i conditional on the baseline covariates $\boldsymbol{Z}_i(0)$ after we include a rich set of variables in $\boldsymbol{Z}_i(0)$. This assumption does not hold for T_i; however, we can assume that the dependence of T_i and D_i occurs through (and only through) the time varying process $\widetilde{\boldsymbol{Z}}_i(t)$, such that conditional on $\widetilde{\boldsymbol{Z}}_i(t)$ we assume "no unmeasured confounders", formulated as,

$$\begin{aligned} & \lim_{h\to 0} \frac{P\{Z_i \in [t, t+h), \Delta_i^T = 1 | Z_i \geq t, \widetilde{\boldsymbol{Z}}_i(t), D_i\}}{h} \\ = & \lim_{h\to 0} \frac{P\{Z_i \in [t, t+h), \Delta_i^T = 1 | Z_i \geq t, \widetilde{\boldsymbol{Z}}_i(t)\}}{h}. \end{aligned}$$

This essentially assumes that the hazard of being censored by T_i at t depends only on the observed covariate history up to t and not additionally on future data. For

example, based on the current liver allocation system, the receipt of a deceased-donor transplant depends only on the patient's current prognostic factors, and not the future disease pathology.

Denote the hazard functions for C_i and T_i at time t as $\lambda_i^C(t)$ and $\lambda_i^T(t)$, respectively; i.e.,

$$\lambda_i^C(t) = \lim_{h\to 0} \frac{P\{Z_i \in [t, t+h), \Delta_i^C = 1 | Z_i \geq t, \widetilde{\boldsymbol{Z}}_i(t)\}}{h},$$
$$\lambda_i^T(t) = \lim_{h\to 0} \frac{P\{Z_i \in [t, t+h), \Delta_i^T = 1 | Z_i \geq t, \widetilde{\boldsymbol{Z}}_i(t)\}}{h},$$

with corresponding cumulative hazards, $\Lambda_i^C(t) = \int_0^t \lambda_i^C(u)du$ and $\Lambda_i^T(t) = \int_0^t \lambda_i^T(u)du$. Although $E(\boldsymbol{Z}_i^D(0)[Y_i - g^{-1}\{\boldsymbol{\beta}_D'\boldsymbol{Z}_i^D(0)\}]) \neq \mathbf{0}$ in the presence of censoring, we can show that under our assumption the IPCW weighted expectation is still zero; i.e., $E(\boldsymbol{Z}_i^D(0)W_i(Y_i)\Delta_i[Y_i - g^{-1}\{\boldsymbol{\beta}_D'\boldsymbol{Z}_i^D(0)\}]) = \mathbf{0}$, where $W_i(t) = W_i^T(t)W_i^C(t)$, $W_i^T(t) = \exp\{\Lambda_i^T(t)\}$ and $W_i^C(t) = \exp\{\Lambda_i^C(t)\}$. Therefore, the following equation is unbiased for $\boldsymbol{\beta}_D$:

$$\boldsymbol{\Phi}^*(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i^D(0)W_i(Y_i)\Delta_i\left[Y_i - g^{-1}\left\{\boldsymbol{\beta}_D'\boldsymbol{Z}_i^D(0)\right\}\right] = \mathbf{0}, \tag{1.3}$$

provided that the weight function $W_i(t) = \exp\{\Lambda_i^T(t)\}\exp\{\Lambda_i^C(t)\}$ is known. However, $\Lambda_i^T(t)$ and $\Lambda_i^C(t)$ are rarely known in practice and therefore must be estimated from the observed data. For this purpose, we assume Cox models for $\lambda_i^T(t)$ and $\lambda_i^C(t)$. Cox regression is a natural choice for modeling censoring times since it is a well established approach, especially in the context of IPCW. Besides its computational convenience, Cox regression can flexibly accommodate both time constant and time varying covariates.We assume the following Cox models, for T_i through

time-dependent covariates $\boldsymbol{Z}_i^T(t)$, and for C_i based on covariates $\boldsymbol{Z}_i^C$:

$$\lambda_i^T(t) = \lambda_0^T(t) \exp\left\{\boldsymbol{\beta}_T' \boldsymbol{Z}_i^T(t)\right\}.$$
$$\lambda_i^C(t) = \lambda_0^C(t) \exp\left(\boldsymbol{\beta}_C' \boldsymbol{Z}_i^C\right).$$

Using partial likelihood (Cox, 1975) and the Breslow estimator (Breslow, 1972), we can estimate $\widehat{\Lambda}_0^T(t)$ and $\widehat{\boldsymbol{\beta}}_T$ from data $\{Z_i, \Delta_i^T, \boldsymbol{Z}_i^T(t) : t \in [0, Z_i), i = 1, \ldots, n\}$, and $\widehat{\Lambda}_0^C(t)$ and $\widehat{\boldsymbol{\beta}}_C$ from data $\{Z_i, \Delta_i^C, \boldsymbol{Z}_i^C : i = 1, \ldots, n\}$ respectively. Plugging $\widehat{\Lambda}_i^T(t), \widehat{\boldsymbol{\beta}}_T, \widehat{\Lambda}_i^C(t), \widehat{\boldsymbol{\beta}}_C$ into (1.3), we obtain the following estimating equation,

$$\boldsymbol{\Phi}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i^D(0) \widehat{W}_i(Y_i) \Delta_i \left[Y_i - g^{-1}\left\{\boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0)\right\}\right] = \mathbf{0}, \tag{1.4}$$

where $\widehat{W}_i(t) = \widehat{W}_i^T(t)\widehat{W}_i^C(t)$, $\widehat{W}_i^T(t) = \exp\{\widehat{\Lambda}_i^T(t)\}$ and $\widehat{W}_i^C(t) = \exp\{\widehat{\Lambda}_i^C(t)\}$. The solution to (1.4) provides for consistent estimation of $\boldsymbol{\beta}_D$, with its asymptotic properties discussed in Section (1.3). The use of a double inverse weight shares some similarity with Schaubel and Wei (2011). However, unlike our methods, the first weight in Schaubel and Wei (2011) is derived from Inverse Probability of Treatment Weighting (IPTW) and serves to balance treatment-specific covariate distributions.

1.3 Asymptotic Properties

Before presenting the asymptotic properties of our proposed estimators, we specify the following regularity conditions (1)-(7) for $i = 1, \ldots, n$.

[(a)]

1. $\{Z_i, \Delta_i^T, \Delta_i^C, \widetilde{\boldsymbol{Z}}_i(Z_i)\}$ are independently and identically distributed.
2. $P(R_i(t) = 1) > 0$ for $t \in (0, \tau]$.
3. $|Z_{ik}(0)| + \int_0^\tau d|Z_{ik}(t)| < M_Z < \infty$ for $i = 1, \ldots, n$, where $Z_{ik}(0), Z_{ik}(t)$ are the kth components of $\boldsymbol{Z}_i(0)$ and $\boldsymbol{Z}_i(t)$, respectively.

4. $\Lambda_i^T(\tau) < \infty, \Lambda_i^C(\tau) < \infty$ and $\Lambda_i^T(t), \Lambda_i^C(t)$ are absolutely continuous for $t \in (0, \tau]$.

5. There exist neighborhoods $\mathcal{B}_T$ of $\boldsymbol{\beta}_T$ and $\mathcal{B}_C$ of $\boldsymbol{\beta}_C$ such that for $k = 0, 1, 2$,

$$\sup_{t\in(0,\tau],\boldsymbol{\beta}\in\mathcal{B}_T} \left\| \frac{1}{n}\sum_{i=1}^{n} \exp\left\{\boldsymbol{\beta}'\boldsymbol{Z}_i^T(t)\right\} R_i(t)\boldsymbol{Z}_i^T(t)^{\otimes k} - \boldsymbol{r}_T^{(k)}(t;\boldsymbol{\beta}) \right\| \xrightarrow{p} 0,$$

$$\sup_{t\in(0,\tau],\boldsymbol{\beta}\in\mathcal{B}_C} \left\| \frac{1}{n}\sum_{i=1}^{n} \exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i^C\right) R_i(t)\boldsymbol{Z}_i^{C\otimes k} - \boldsymbol{r}_C^{(k)}(t;\boldsymbol{\beta}) \right\| \xrightarrow{p} 0,$$

where $\boldsymbol{v}^{\otimes 0} = 1, \boldsymbol{v}^{\otimes 1} = \boldsymbol{v}, \boldsymbol{v}^{\otimes 2} = \boldsymbol{v}'\boldsymbol{v}$ and

$$\boldsymbol{r}_T^{(k)}(t;\boldsymbol{\beta}) = E\left[\exp\left\{\boldsymbol{\beta}'\boldsymbol{Z}_i^T(t)\right\} R_i(t)\boldsymbol{Z}_i^T(t)^{\otimes k}\right],$$

$$\boldsymbol{r}_C^{(k)}(t;\boldsymbol{\beta}) = E\left\{\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i^C\right) R_i(t)\boldsymbol{Z}_i^{C\otimes k}\right\}.$$

6. Define $h(x) = \partial g^{-1}(x)/\partial x$, where h exists and is continuous in an open neighborhood $\mathcal{B}_D$ of $\boldsymbol{\beta}_D$.

7. The matrices $\boldsymbol{A}(\boldsymbol{\beta}_D), \boldsymbol{\Omega}_T(\boldsymbol{\beta}_T), \boldsymbol{\Omega}_C(\boldsymbol{\beta}_C)$ are each positive definite, where

$$\boldsymbol{A}(\boldsymbol{\beta}) = E\left[\boldsymbol{Z}_i^D(0)^{\otimes 2} h\left\{\boldsymbol{\beta}_D'\boldsymbol{Z}_i^D(0)\right\}\right],$$

$$\boldsymbol{\Omega}_T(\boldsymbol{\beta}) = E\left[\int_0^\tau \left\{\frac{\boldsymbol{r}_T^{(2)}(t;\boldsymbol{\beta})}{r_T^{(0)}(t;\boldsymbol{\beta})} - \overline{\boldsymbol{z}}_T(t;\boldsymbol{\beta})^{\otimes 2}\right\} dN_i^T(t)\right],$$

$$\boldsymbol{\Omega}_C(\boldsymbol{\beta}) = E\left[\int_0^\tau \left\{\frac{\boldsymbol{r}_C^{(2)}(t;\boldsymbol{\beta})}{r_C^{(0)}(t;\boldsymbol{\beta})} - \overline{\boldsymbol{z}}_C(t;\boldsymbol{\beta})^{\otimes 2}\right\} dN_i^C(t)\right],$$

and

$$\overline{\boldsymbol{z}}_T(t;\boldsymbol{\beta}) = \frac{\boldsymbol{r}_T^{(1)}(t;\boldsymbol{\beta})}{r_T^{(0)}(t;\boldsymbol{\beta})},$$

$$\overline{\boldsymbol{z}}_C(t;\boldsymbol{\beta}) = \frac{\boldsymbol{r}_C^{(1)}(t;\boldsymbol{\beta})}{r_C^{(0)}(t;\boldsymbol{\beta})}.$$

Condition (1) can be relaxed at the expense of additional technical development. Condition (2) is needed for the purpose of identifiability. Conditions (3)-(6) ensure the convergence of several stochastic integrals used in the proofs. The matrices

$\boldsymbol{A}(\boldsymbol{\beta}_D)$, $\boldsymbol{\Omega}_T(\boldsymbol{\beta}_T)$, $\boldsymbol{\Omega}_C(\boldsymbol{\beta}_C)$ in condition (7) are at least non-negative definite and will be positive-definite under any non-redundant specification of the respective covariate vectors. Our main asymptotic results are summarized in Theorems I.1 and I.2 below, with the proofs presented in Appendix A.

Theorem I.1. *Under regularity conditions (1)-(7), as $n \to \infty$, $\sqrt{n}\boldsymbol{\Phi}(\boldsymbol{\beta}_D)$ converges to a zero-mean Normal with variance $\boldsymbol{B}(\boldsymbol{\beta}_D) = E\{\boldsymbol{B}_i(\boldsymbol{\beta}_D)^{\otimes 2}\}$, for any subject $i = 1, \ldots, n$,*

$$\begin{aligned}\boldsymbol{B}_i(\boldsymbol{\beta}) = \boldsymbol{\epsilon}_i(\boldsymbol{\beta}) + \boldsymbol{K}_T(\boldsymbol{\beta})\boldsymbol{\Omega}_T(\boldsymbol{\beta}_T)^{-1}\boldsymbol{U}_i^T(\boldsymbol{\beta}_T) + \boldsymbol{K}_C(\boldsymbol{\beta})\boldsymbol{\Omega}_C\left(\boldsymbol{\beta}_C\right)^{-1}\boldsymbol{U}_i^C(\boldsymbol{\beta}_C) \\ + \int_0^\tau \boldsymbol{H}_T(u;\boldsymbol{\beta}) r_T^{(0)}\left(u;\boldsymbol{\beta}_T\right)^{-1} dM_i^T(u) + \int_0^\tau \boldsymbol{H}_C(u;\boldsymbol{\beta}) r_C^{(0)}\left(u;\boldsymbol{\beta}_C\right)^{-1} dM_i^C(u),\end{aligned}$$

where we define

$$\begin{aligned}\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) &= \Delta_i W_i(Y_i)[Y_i - g^{-1}\{\boldsymbol{\beta}'\boldsymbol{Z}_i^D(0)\}]\boldsymbol{Z}_i^D(0), \\ \boldsymbol{U}_i^T(t) &= \int_0^t \{\boldsymbol{Z}_i^T(u) - \overline{\boldsymbol{z}}_T(u;\boldsymbol{\beta}_T)\} dM_i^T(u), \\ \boldsymbol{U}_i^C(t) &= \int_0^t \{\boldsymbol{Z}_i^C - \overline{\boldsymbol{z}}_C(u;\boldsymbol{\beta}_C)\} dM_i^C(u), \\ \boldsymbol{K}_T(\boldsymbol{\beta}) &= E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta})\boldsymbol{D}_i^T(Y_i)'\}, \\ \boldsymbol{K}_C(\boldsymbol{\beta}) &= E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta})\boldsymbol{D}_i^C(Y_i)'\}, \\ \boldsymbol{H}_T(u;\boldsymbol{\beta}) &= E[\boldsymbol{\epsilon}_i(\boldsymbol{\beta})\exp\{\boldsymbol{\beta}_T'\boldsymbol{Z}_i^T(u)\}R_i(u)], \\ \boldsymbol{H}_C(u;\boldsymbol{\beta}) &= E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta})\exp(\boldsymbol{\beta}_C'\boldsymbol{Z}_i^C)R_i(u)\}, \\ \boldsymbol{D}_i^T(t) &= \int_0^t \{\boldsymbol{Z}_i^T(u) - \overline{\boldsymbol{z}}_T(u;\boldsymbol{\beta}_T)\} d\Lambda_i^T(u), \\ \boldsymbol{D}_i^C(t) &= \int_0^t \{\boldsymbol{Z}_i^C(u) - \overline{\boldsymbol{z}}_C(u;\boldsymbol{\beta}_C)\} d\Lambda_i^C(u),\end{aligned}$$

with $\boldsymbol{\Omega}_T(\boldsymbol{\beta}), \boldsymbol{\Omega}_C(\boldsymbol{\beta})$ defined in Condition (7).

Here we use the usual counting process notations, where $N_i^T(t) = I(Z_i \le t, \Delta_i^T = 1)$ and $N_i^C(t) = I(Z_i \le t, \Delta_i^C = 1)$ are observed counting processes for T_i and

C_i respectively, with $dM_i^T(t) = dN_i^T(t) - R_i(t)d\Lambda_i^T(t)$ and $dM_i^C(t) = dN_i^C(t) - R_i(t)d\Lambda_i^C(t)$ being the corresponding zero mean processes. The proof utilizes various results derived in Zhang and Schaubel (2011), primarily the asymptotic expression of the empirical weight in terms of the true weight. The main purpose of Theorem I.1 is to set up Theorem I.2.

Theorem I.2. *Under regularity conditions (1)-(7), as* $n \to \infty$, $\widehat{\boldsymbol{\beta}}_D$ *converges in probability to* $\boldsymbol{\beta}_D$ *and* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_D)$ *converges to a zero-mean Normal with variance* $\boldsymbol{A}(\boldsymbol{\beta}_D)^{-1}\boldsymbol{B}(\boldsymbol{\beta}_D)\boldsymbol{A}(\boldsymbol{\beta}_D)^{-1}$ *with* $\boldsymbol{A}(\boldsymbol{\beta})$ *defined in condition (g) and* $\boldsymbol{B}(\boldsymbol{\beta})$ *defined in Theorem 1.*

The proof of consistency of $\widehat{\boldsymbol{\beta}}_D$ holds by the Inverse Function Theorem (Foutz, 1977) while the proof of asymptotic normality follows through the combination of various Taylor series expansions and the Cramér-Wold Theorem.

We propose two versions of asymptotic standard error (ASE) estimators for our proposed estimator $\widehat{\boldsymbol{\beta}}_D$. The first ASE is derived from (1.3) and, as such, treats the IPCW weights as known:

$$\text{ASE}_1 = \sqrt{\frac{1}{n}\text{Diag}\left[\widehat{\boldsymbol{A}}(\boldsymbol{\beta}_D)^{-1}\widehat{\boldsymbol{B}^*}(\boldsymbol{\beta}_D)\widehat{\boldsymbol{A}}(\boldsymbol{\beta}_D)^{-1}\right]},$$

where $\widehat{\boldsymbol{B}^*}(\boldsymbol{\beta}) = \widehat{E}\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta})^{\otimes 2}\}$. The second ASE is based on (1.4) and derived in Theorem I.2:

$$\text{ASE}_2 = \sqrt{\frac{1}{n}\text{Diag}\left[\widehat{\boldsymbol{A}}(\boldsymbol{\beta}_D)^{-1}\widehat{\boldsymbol{B}}(\boldsymbol{\beta}_D)\widehat{\boldsymbol{A}}(\boldsymbol{\beta}_D)^{-1}\right]},$$

where $\widehat{\boldsymbol{B}}(\boldsymbol{\beta}) = \widehat{E}\{\boldsymbol{B}_i(\boldsymbol{\beta})^{\otimes 2}\}$. These two ASEs can be obtained by plugging all the undetermined terms with their respective estimators. More detail about the calculation procedures is provided in Appendix A. These two versions of sandwich ASEs share the same second derivative matrix, $\boldsymbol{A}(\boldsymbol{\beta})$, but differ with two different middle matrices. The first version, ASE_1, which results from the weight function

being known as opposed to estimated, treats the weights as fixed and therefore its middle matrix involves $\boldsymbol{\epsilon}_i$ only. The second version, ASE_2, contains several extra terms in its middle matrix, in order to account for the variation due to estimating the weights. Although ASE_2 should be closer to the truth, it adds the complexity of middle matrix and is usually more complicated to calculate in practice than ASE_1. In contrast, ASE_1 can be easily computed with built-in commands from many statistical software packages (e.g., SAS, R) and therefore serves as a useful approximation of ASE_2. We evaluate the performance of both ASE_1 and ASE_2 through the simulations presented in the next section.

1.4 Simulation Study

We conducted simulations to evaluate the performance of the proposed methods in finite samples. Two different percentages of right censoring were considered, moderate and heavy censoring. For each simulated subject $i = 1, \ldots, n$, two baseline covariates Z_{i1}, Z_{i2} were generated from Bernoulli(0.5) distributions. The death time, D_i, was generated from $D_i = g^{-1}(\alpha_0 + \alpha_1 Z_{i1} + \alpha_2 Z_{i2}) + \epsilon_{1i}$, where $\epsilon_{1i} \sim \text{Uniform}(-\sigma, \sigma)$, $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \alpha_2]'$ and σ were chosen in accordance with the particular link function. More specifically, for linear link $\alpha = [5.5, 0.25, 0.25]'$ was tested and for log link $\alpha = [-0.63, .08, .08]'$ was tested. This death generator was used to induce the same mean structure for D_i and $D_i \wedge L$, as the mean structure of the former, $g\{E(D_i|Z_{i1}, Z_{i2})\} = \alpha_0 + \alpha_1 Z_{i1} + \alpha_2 Z_{i2}$, is similar to that of the latter,

$$g\{E(D_i \wedge L|Z_{i1}, Z_{i2})\} = \beta_{D0} + \beta_{D1} Z_{i1} + \beta_{D2} Z_{i2} + \beta_{D3} Z_{i1} Z_{i2}. \tag{1.5}$$

Because the above model (1.5) is saturated due to an extra interaction term $Z_{i1}Z_{i2}$, the true values of $\boldsymbol{\beta}_D = [\beta_{D0}, \beta_{D1}, \beta_{D2}, \beta_{D3}]'$ can be determined computationally and are calculated using Monte Carlo methods with size 10 million. We set $L = 10$ to

yield a reasonable range of $P(D > L)$. We evaluated the linear and log links, since they would be the most popular choices in practice.

We generated the independent censoring time C_i from a Cox model with the following hazard,

$$\lambda_i^C(t) = \lambda_0^C \exp\left(\beta_{C1} Z_{i1}\right), \tag{1.6}$$

where λ_0^C ranged from 1/36 to 1/21 and β_{C1} ranged from $-\log(3)$ to $\log(2)$. We generated a time-dependent covariate which was correlated with death time D_i and treatment time T_i while conditional on baseline covariates Z_{i1}, Z_{i2}. First, let $V_i = -V_0 \log\{(\epsilon_{i1} + \sigma)/(2\sigma)\} + \epsilon_{i2}$, where V_0 is a constant ranging from 40 to 100 and $\epsilon_{i2} \sim \text{Uniform}(0, 1)$. Then define a time-dependent variable $V_i(t) = I(t \leq V_i)$. Thus $V_i(t)$ is still correlated with D_i through ϵ_{i1} even after conditional on Z_{i1}, Z_{i2}. Treatment time T_i was generated from a Cox model with the following hazard,

$$\lambda_i^T(t) = \lambda_0^T \exp\left\{\beta_{T1} Z_{i2} + \beta_{T2} V_i(t)\right\}, \tag{1.7}$$

where λ_0^T ranged from 1/35 to 1/18, β_{T1} ranged from $-\log(4)$ to $\log(3)$ and β_{T2} ranged from $\log(2)$ to $\log(3)$. Therefore T_i is correlated with D_i conditional on Z_{i1}, Z_{i2}, through a mutual unobserved variable ϵ_{i1}.

We present results for samples sizes $n = 250$ and $n = 500$, under moderate and heavy censoring. For the linear link, $P(D > 10) \approx 11\%$, approximately 10% C_i and 21% T_i are observed in the moderate censoring scenario, and 15% C_i and 36% T_i are observed in the heavy censoring scenario. For the log link, $P(D > 10) \approx 10\%$, approximately 8% C_i and 24% T_i are observed in the moderate censoring scenario, and 15% C_i and 35% T_i are observed in the heavy censoring scenario. As displayed in Tables 1.1 and 1.2, the estimates for $\boldsymbol{\beta}_D$ are approximately unbiased, with the magnitudes of any bias generally decreasing with increasing sample size.

Table 1.1 Simulation results: $g(x) = x$

Scenario	N	$\boldsymbol{\beta}_D$	True	BIAS	ESD	ASE_1	$\text{CP}_1(\%)$	ASE_2	$\text{CP}_2(\%)$
Moderate Censoring	250	β_0	5.452	-0.008	0.505	0.519	94	0.515	94
		β_1	0.23	-0.037	0.714	0.72	95	0.735	96
		β_2	0.227	0.008	0.67	0.686	95	0.711	96
		β_3	-0.014	0.033	0.928	0.947	95	0.99	96
	500	β_0	5.452	0.002	0.346	0.368	96	0.364	96
		β_1	0.23	-0.016	0.503	0.508	95	0.519	95
		β_2	0.227	-0.018	0.469	0.487	96	0.505	96
		β_3	-0.014	0.012	0.674	0.671	95	0.701	97
Heavy Censoring	250	β_0	5.452	-0.006	0.467	0.486	96	0.577	98
		β_1	0.23	-0.042	0.704	0.725	95	0.858	98
		β_2	0.227	-0.037	0.821	0.83	95	0.991	98
		β_3	-0.014	-0.021	1.271	1.249	94	1.493	96
	500	β_0	5.452	0.01	0.328	0.345	96	0.398	98
		β_1	0.23	-0.024	0.512	0.512	94	0.589	96
		β_2	0.227	-0.043	0.591	0.594	95	0.694	97
		β_3	-0.014	-0.017	0.924	0.887	93	1.039	97

The calculated ASE_1s and ASE_2s are very close to their corresponding empirical standard deviations (ESD), and therefore their empirical coverage probabilities CP_1 and CP_2 are close to 95%. The implication from our simulation studies is that, in moderate samples, the proposed methods result in unbiased estimation, and that the easily computed ASE_1 is a useful approximation to the more complicated ASE_2. In our real-data application presented in Section (1.5), $n >> 1,000$, the sample size is far more than 1,000, which would render finite-sample bias in ASE_1 much less of an issue than in our simulation studies.

Mis-specification of the censoring model is a common issue in IPCW methods, in which case bias will generally exist in the mortality model estimation. To evaluate how much bias is introduced by such mis-specification, we conducted additional numerical studies. In particular, we considered 3 different types of mis-specification:

(i) Model for independent censoring, C, is mis-specified: add a covariate, Z_{i2}, to the censoring hazard $\lambda_i^C(t)$; i.e., Z_{i2} was added to the generator, but not the

Table 1.2 Simulation results: $g(x) = \log(x)$

Scenario	N	$\boldsymbol{\beta}_D$	True	BIAS	ESD	ASE_1	$\text{CP}_1(\%)$	ASE_2	$\text{CP}_2(\%)$
Moderate Censoring	250	β_0	-0.634	-0.008	0.104	0.104	95	0.102	94
		β_1	0.075	-0.007	0.143	0.139	95	0.142	96
		β_2	0.074	0.005	0.13	0.129	95	0.133	96
		β_3	-0.004	0.008	0.173	0.172	95	0.179	95
	500	β_0	-0.634	-0.004	0.068	0.073	97	0.072	96
		β_1	0.075	-0.001	0.095	0.097	96	0.099	96
		β_2	0.074	-0.001	0.088	0.091	96	0.094	97
		β_3	-0.004	0.001	0.12	0.121	96	0.126	97
Heavy Censoring	250	β_0	-0.634	-0.013	0.124	0.125	94	0.125	95
		β_1	0.075	-0.01	0.183	0.171	94	0.177	94
		β_2	0.074	0.007	0.157	0.157	95	0.167	96
		β_3	-0.004	0.01	0.222	0.212	95	0.227	96
	500	β_0	-0.634	-0.008	0.083	0.089	96	0.088	96
		β_1	0.075	0.001	0.121	0.12	95	0.124	95
		β_2	0.074	0.001	0.109	0.111	95	0.118	97
		β_3	-0.004	0.001	0.151	0.15	95	0.159	96

model for C_i being fitted.

(ii) Model for dependent censoring, T, is mis-specified: add a covariate, Z_{i1}, to censoring hazard $\lambda_i^T(t)$, but not to the T_i model being fitted.

(iii) Models are mis-specified for both C and T simultaneously.

Figure 1.1 displays the bias for sample size $N = 250$ calculated with $1,000$ replications for the 4 scenarios in Table 1.1 and 1.2, after introducing mis-specification (i)-(iii). In general, bias exists when the censoring model is not correctly specified, and is more pronounced when both types of censoring are incorrectly modeled.

An interesting comparison is between our methods and those proposed by Tian et al. (2014). Their methods were developed in the context of covariate-independent censoring and, thus, proposed IPCW weights based on Kaplan-Meier estimators. When the censoring mechanism is more complicated than independent censoring (i.e., in presence of either covariate-dependent or dependent censoring), then our

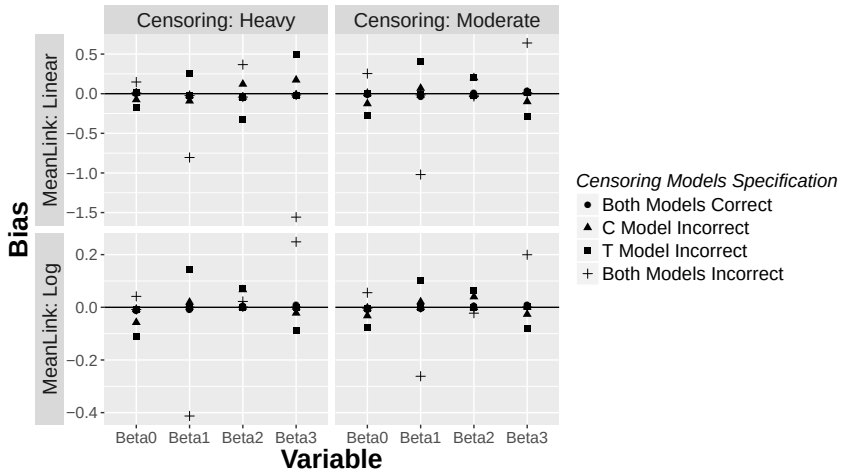


Figure 1.1 Bias of $[\beta_{D0}, \beta_{D1}, \beta_{D2}, \beta_{D3}]'$ when censoring models are mis-specified

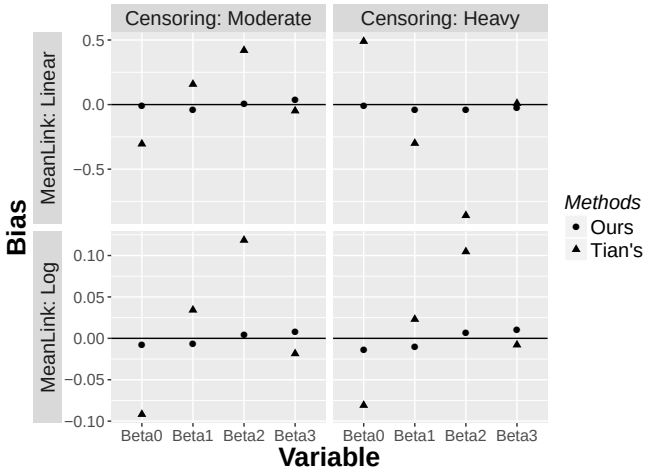


Figure 1.2 Bias comparison between our and Tian's methods in presence of dependent censoring

methods should perform better than Tian's methods. We illustrated this with additional simulations; we ran the simulation studies in Table 1.1 and 1.2 again, applied Tian et al. (2014), and then plotted the resulting bias along with the bias from our methods. As shown in Fig 1.2, Tian et al. (2014) has quite severe biases for $N = 250$ cases; similar results were obtained with $N = 500$. This is expected, since we are testing Tian et al. (2014) outside the set-up for which the methods were developed.

In presence of only covariate-independent censoring, both our methods and those of Tian et al. (2014) should apply. In evaluating our methods, we blindly added

Table 1.3 Efficiency comparison between our and Tian's methods in presence of only covariate-independent censoring: $g(x) = x$

N	$\boldsymbol{\beta}_D$	True	Bias_1	Bias_2	ESD_1	ESD_2	ERE
250	β_0	5.148	0.023	< 0.001	0.344	0.351	0.963
	β_1	0.406	-0.017	-0.002	0.404	0.401	1.012
	β_2	0.403	-0.009	0.005	0.403	0.402	1.004
500	β_0	5.148	0.012	0.001	0.244	0.25	0.953
	β_1	0.406	-0.007	< 0.001	0.286	0.286	1.003
	β_2	0.403	-0.009	-0.002	0.286	0.286	1.002

non-predictive covariates into the censoring model, in order to assess the degree of efficiency loss. We generated death time with $D_i = g^{-1}\left(5.25 + 0.5Z_{i1} + 0.5Z_{i2}\right) + \epsilon_{1i}$, where $Z_{i1}, Z_{i2} \sim \text{Bernoulli}(0.5)$ and $\epsilon_{1i} \sim \text{Uniform}(-5.25, 5.25)$. We chose $L = 9$ and introduced a simple censoring pattern (Exponential, rate 0.05). Approximately 79% D's are observed, with $P(D > L) \approx 19\%$. In Table 1.3, Bias_1 and ESD_1 are bias and empirical standard deviation (ESD) calculated by our methods with futile covariates Z_{i1}, Z_{i2} in the censoring model, and Bias_2 and ESD_2 are calculated using Tian et al. (2014) with a correctly specified non-parametric censoring model. The last column is Empirical Relative Efficiency (ERE) between Tian et al. (2014) and our methods, computed as the ratio of mean square error. Examining various replicates, the estimated regression coefficients for the C model tends to be very close to 0, with large p values. In practice, users would feel inclined to drop them from the model due to their negligible effect on C. Since these non-zero censoring coefficients are actually 0, our methods have a little bit greater bias than Tian et al. (2014). Regarding efficiency, our methods exhibited approximately the same efficiency as Tian et al. (2014), despite including the two irrelevant covariates in the C model.

1.5 Analysis of Liver Disease Data

We applied the proposed methods to directly estimate RMST among End-Stage Liver Disease (ESLD) patients. Of interest was survival in the absence of liver transplantation. We obtained data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donors, wait-listed candidates, and transplant recipients in the U.S., as submitted by the members of the Organ Procurement and Transplantation Network (OPTN), and has been described elsewhere. The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors.

The study population consisted of all chronic ESLD patients initially wait-listed for deceased-donor liver transplantation in U.S. at age ≥ 18 between January 1, 2005 and December 31, 2012. For each patient, the time origin ($t = 0$) is the date of wait-listing, with each patient followed from that date until earliest of death, receipt of a liver transplant, loss to follow-up, or the end of the observation period on 12/31/2012. Independent censoring occurs through loss to follow-up, administrative censoring, or receipt of a living-donor liver transplant. Note that living-donor transplantation is usually carried out with a liver segment donated by a family member or a close friend, such that the process is not systematically influenced by MELD score trajectory conditional on baseline covariates. As described in Section (1.1), dependent censoring occurs through the receipt of a deceased-donor transplant, which is correlated with pre-transplant mortality through its mutual association with time varying MELD score. A total of n =55,651 patients were included in our study population. Among them, 13,640 (25%) died before receipt of a transplant, 23,335 (42%) received a liver

transplant, and 18,676 (34%) were independently censored.

We constructed our independent censoring model and pre-transplant mortality model using baseline covariates historically reported to be important prognostic factors, including age, gender, race, blood type, United Network for Organ Sharing (UNOS) Region, calendar year of listing, underlying diagnosis, body mass index (BMI), dialysis, sodium, hospitalization status and MELD score at listing ($t = 0$). The liver transplant hazard model incorporated additional time-dependent covariates, including MELD score, dialysis, sodium, ascites and encephalopathy. UNOS has established 11 geographic Regions for administrative purposes. The availability of deceased-donor organs and the distribution of mortality is quite different across these 11 Regions. This therefore suggests the necessity of adjusting for UNOS Region in both our censoring and mortality models:

$$\lambda_{ij}^{T}(t) = A_i(t)\lambda_{0j}^{T}(t)\exp\left\{\boldsymbol{\beta}_T'\boldsymbol{Z}_i^{T}(t)\right\}, \tag{1.8}$$

$$\lambda_{ij}^{C}(t) = \lambda_{0j}^{C}(t)\exp\left(\boldsymbol{\beta}_C'\boldsymbol{Z}_i^{C}\right), \tag{1.9}$$

where the subscript $j = 1, 2, \ldots, 11$ stands for UNOS Region, while the indicator $A_i(t)$ records whether the patient is active and not removed from the wait-list at time t. Patients generally start follow-up as active, such that $A_i(0) = 1$, but may be made temporarily inactive due to illness ($A_i(t) = 0$), in which case the patient retains his/her position on the wait-list but cannot receive deceased-donor liver offers. A patient whose health condition declines to the point where liver transplantation is considered futile may be permanently removed from the wait-list ($A_i(u) = 0$ for any time point u after the time of removal). Therefore $A_i(t)$ serves as a time varying at-risk indicator for transplantation. Subintervals during which a given patient is inactive make no contribution to the fitting of model T. We

then compute cumulative hazard functions as $\widehat{\Lambda}_{ij}^{C}(t) = \exp\{-\int_0^t \exp(\widehat{\boldsymbol{\beta}}_C' \boldsymbol{Z}_i^C) d\widehat{\Lambda}_{0j}^{C}(u)\}$ and $\widehat{\Lambda}_{ij}^{T}(t) = \exp[-\int_0^t A_i(u) \exp\{\widehat{\boldsymbol{\beta}}_T' \boldsymbol{Z}_i^T(u)\} d\widehat{\Lambda}_{0j}^{T}(u)]$ and obtain the IPCW weight as $\widehat{W_{ij}}(t) = \exp\{\widehat{\Lambda}_{ij}^{C}(t)\} \exp\{\widehat{\Lambda}_{ij}^{T}(t)\}$. Since more than 99% of the estimated IPCW weights are below 10, we cap the weights by 10 in order to reduce variance.

We modeled restricted mean survival time at $L = 1$, $L = 3$, $L = 5$ years post wait-list registration, which are reasonable time windows in the ESLD setting. Overall crude survival probabilities are approximately 79%, 62% and 51% for the 3 time windows respectively. We present the parameter estimates for the pre-transplant mortality model using three link functions, including linear, log and logistic. As shown in Table 1.4, the covariate effects demonstrate the same trends across the different link functions. The average pre-transplant survival time out of the next 3 years is estimated as approximately 33, 35.8, and 33 months using the three link functions respectively, for a 'reference' patient; i.e., a white male wait-listed at age 50, registered in Region 5 (the Region with the largest population) during year 2005, diagnosed as none of the listed types, not hospitalized, not on dialysis, with blood Type O, BMI between 20 and 25, sodium level at 130 mmol/l and MELD score 6 (the minimum possible value). For another patient with the same profile but a different MELD score (e.g., MELD=30), the average pre-transplant survival time out of the next 3 years is estimated as approximately 11.6, 9.8, and 7.5 months respectively. This discrepancy in predicted RMST underscores the importance of the MELD score. The linear link does not always result in fitted RMST values within an admissible range $(0, L]$. This could perhaps be remedied by transforming various covariates, or by simply bounding estimated RMST. The remaining parameter estimates for the 1- and 5-year windows are provided in Appendix A.3.

Figure 1.3 plots fitted RMST values ($L = 3$ years) by MELD score (ranging from

Table 1.4 Estimated covariate effects on RMST in the absence of liver transplantation ($L = 36$ months)

	Linear			Log			Logistic		
$\boldsymbol{Z}_i^D(0)$	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p
Intercept	32.95	0.5	< 0.01	3.57	0.02	0.8	2.42	0.08	< 0.01
Year-2005	-1.02	0.05	< 0.01	-0.04	< 0.01	< 0.01	-0.16	0.01	< 0.01
Age-50 (Years)	-0.2	0.01	< 0.01	-0.01	< 0.01	< 0.01	-0.04	< 0.01	< 0.01
Sodium-130 (mmol/l)	0.47	0.02	< 0.01	0.02	< 0.01	< 0.01	0.08	< 0.01	< 0.01
MELD Score-6	-0.89	0.02	< 0.01	-0.05	< 0.01	< 0.01	-0.16	< 0.01	< 0.01
UNOS Region	*Reference Group: 5*								
1	-1.52	0.49	< 0.01	-0.06	0.02	< 0.01	-0.27	0.07	< 0.01
2	-1.35	0.38	< 0.01	-0.06	0.01	< 0.01	-0.22	0.06	< 0.01
3	-3.26	0.43	< 0.01	-0.11	0.02	< 0.01	-0.49	0.07	< 0.01
4	0.23	0.34	0.5	0.02	0.01	0.09	0.04	0.06	0.46
6	-0.97	0.54	0.07	< 0.01	0.02	0.95	-0.12	0.09	0.21
7	-0.59	0.42	0.16	-0.02	0.02	0.32	-0.1	0.07	0.14
8	-0.9	0.42	0.03	< 0.01	0.02	0.94	-0.13	0.07	0.06
9	-1.41	0.36	< 0.01	-0.07	0.01	< 0.01	-0.21	0.06	< 0.01
10	-2.24	0.48	< 0.01	-0.09	0.02	< 0.01	-0.39	0.07	< 0.01
11	-2.75	0.44	< 0.01	-0.1	0.02	< 0.01	-0.44	0.07	< 0.01
Gender	*Reference Group: Male*								
Female	0.49	0.22	0.02	< 0.01	0.01	0.95	0.06	0.03	0.1
Race	*Reference Group: White*								
lack	0.34	0.41	0.41	0.02	0.02	0.23	0.06	0.06	0.31
Hispanic	0.23	0.27	0.4	0.01	0.01	0.21	0.04	0.05	0.33
Asian	1.76	0.55	< 0.01	0.05	0.02	0.01	0.33	0.11	< 0.01
Others	-0.57	0.89	0.52	-0.01	0.04	0.74	-0.11	0.15	0.48
Blood Type	*Reference Group: O*								
A	-0.08	0.21	0.72	-0.01	0.01	0.26	-0.02	0.03	0.63
B	-0.02	0.36	0.95	-0.01	0.01	0.59	< 0.01	0.06	0.97
AB	-0.59	0.78	0.44	-0.05	0.03	0.08	-0.19	0.11	0.1
Diagnosis	*Reference Group: No or Yes*								
Hepatitis C	-1.33	0.33	< 0.01	-0.04	0.01	< 0.01	-0.26	0.05	< 0.01
Noncholestatic	0.76	0.33	0.02	0.06	0.01	< 0.01	0.12	0.05	0.02
Cholestatic	-1.26	0.45	0.01	-0.05	0.02	< 0.01	-0.28	0.07	< 0.01
Acute Hepatic Necrosis	2.27	0.81	0.01	0.06	0.03	0.04	0.56	0.18	< 0.01
Metastatic Disease	-2.95	0.66	< 0.01	-0.11	0.03	< 0.01	-0.53	0.11	< 0.01
Malignant Neoplasm	-7.24	0.37	< 0.01	-0.43	0.02	< 0.01	-1.23	0.06	< 0.01
BMI	*Reference Group:* $(20, 25]$								
$(0, 20]$	-2.05	0.47	< 0.01	-0.07	0.02	< 0.01	-0.28	0.07	< 0.01
$(25, 30]$	0.02	0.27	0.95	0.01	0.01	0.25	-0.01	0.04	0.8
> 30	-0.09	0.27	0.74	< 0.01	0.01	0.98	-0.02	0.04	0.57
Hospitalized	*Reference Group: Not Hospitalized*								
	-1.83	0.65	< 0.01	-0.54	0.08	< 0.01	-0.82	0.14	< 0.01
not ICU	-2.41	0.42	< 0.01	-0.33	0.04	< 0.01	-0.49	0.08	< 0.01
Dialysis	*Reference Group: No or Yes*								
Yes	1.69	0.54	< 0.01	0.13	0.04	< 0.01	0.49	0.09	< 0.01

An offset of $L = 36$ months is applied for log link.

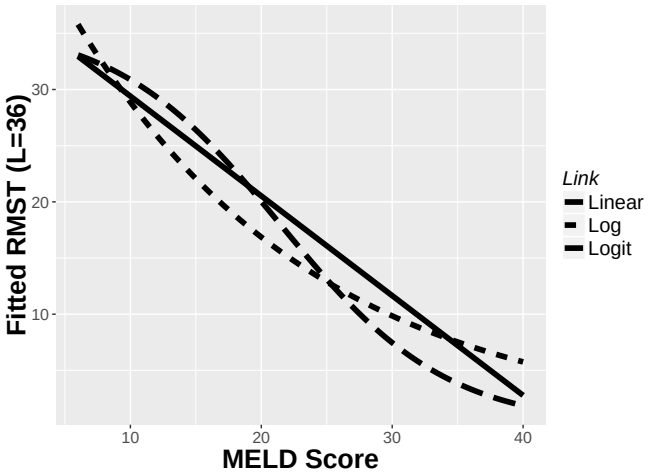


Figure 1.3 Fitted RMST ($L = 36$ months) by MELD score for a reference patient: white, male, age=50, Region=5, year=2005, not hospitalized, not on dialysis, blood Type=O, BMI $\in (20, 25]$, sodium=130

6 to 40; i.e., for all possible MELD scores), for the above-described reference patient. For all the three link functions, RMST decreases strongly with increasing MELD score, as anticipated. The fitted values based on all the three links result in fitted values which tail off at higher MELD scores. Among the three link functions, the linear link may be most appealing in terms of its straightforward interpretation. For example, for per unit increase in a patient's MELD score, the average survival time (capped at 3 years) will decrease approximately 0.9 months; for a 5-year increase in age at wait-listing, 3-year RMST decreases by approximately 1 month. Analogous trends are also observed in the models using the other two link functions. We will further compare the model adequacy using different link functions in terms of discrimination ability and prediction accuracy.

To evaluate each model's discrimination ability, we compute the Index of Concordance (IOC), also known as the C statistic (Harrell et al., 1996; Heagerty and Zheng, 2005; Uno et al., 2011), denoted by:

$$\text{IOC} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i\widehat{W}_i(Y_i)\widehat{W}_j(Y_i)I\left[Y_i < Y_j, g^{-1}\{\widehat{\boldsymbol{\beta}}_D'\boldsymbol{Z}_i^D(0)\} < g^{-1}\{\widehat{\boldsymbol{\beta}}_D'\boldsymbol{Z}_j^D(0)\}\right]}{\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i\widehat{W}_i(Y_i)\widehat{W}_j(Y_i)I\left(Y_i < Y_j\right)}.$$

This statistic converges to a censoring distribution free quantity, $P(\boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0) < \boldsymbol{\beta}_D' \boldsymbol{Z}_j^D(0) | D_i \wedge L < D_j \wedge L)$, measuring the agreement of predictions with observed failure order. Quantities frequently used to evaluate model prediction accuracy include the mean absolute deviation (MAD) and mean squared deviation (MSD) (Davison and Hinkley, 1997; Tian et al., 2007), formulated as,

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n} \Delta_i \widehat{W}_i(Y_i) \left| [Y_i - g^{-1}\{\widehat{\boldsymbol{\beta}}_D' \boldsymbol{Z}_i^D(0)\} \right|,$$
$$\text{MSD} = \frac{1}{n}\sum_{i=1}^{n} \Delta_i \widehat{W}_i(Y_i) \left[Y_i - g^{-1}\{\widehat{\boldsymbol{\beta}}_D' \boldsymbol{Z}_i^D(0)\} \right]^2,$$

which converge to $E|D_i \wedge L - g^{-1}\{\boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0)\}|$ and $E[D_i \wedge L - g^{-1}\{\boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0)\}]^2$ respectively, quantifying the "distance" between predicted and observed outcomes. Proof sketches of the convergence of IOC, MAD and MSD are provided in Appendix A.2.

In Table 1.5, we calculate these three statistics for the 9 models (3 link functions, with 3 values of L) through 2-fold cross validation. Specifically, we split our data by randomly selecting half the patients (n =27,825) into a "training" set (to which the models are fitted), and using the remaining half (n =27,826) as the "validation" set (to which the discrimination and predictive accuracy measures are applied). For $L = 1$, IOC=0.82 for all three link functions. For $L = 3$ and $L = 5$, IOC is largest for the log link, although not by a wide margin. As L increases, the IOCs decrease, for all link functions; this makes sense intuitively since covariate measurements at times more distant in the past should correspond to reduced discrimination. In terms of both MAD and MSD, the logistic link has the best prediction accuracy by a fair margin.

Table 1.5 Index of Concordance (IOC), Mean Absolute Deviation (MAD) and Mean Squared Deviation (MSD): Comparison of link functions by L (years)

Measure	L	Linear	Log	Logistic
IOC	1	0.82	0.82	0.82
	3	0.77	0.78	0.77
	5	0.72	0.75	0.74
MAD	1	1.52	1.67	1.36
	3	5.08	5.05	4.58
	5	7.19	6.56	6.06
MSD	1	6.66	6.97	6.42
	3	70.10	69.51	66.02
	5	153.25	145.78	139.75

1.6 Discussion

We have proposed methods to model restricted mean survival time as a function of baseline covariates, using techniques that are valid under a wide variety of censoring mechanisms. RMST is often of inherent interest to investigators, especially in settings where cumulative covariate effects are appealing. RMST is also an attractive alternative when the proportional hazards assumption does not hold. We have constructed double IPCW weights to simultaneously account for independent and dependent censoring. This general setup is frequently necessary in applications, and failing to account for either type of censoring may result in biased estimation of the mortality model. In studies with only one type of censoring, one would need to calculate the corresponding IPCW weight and set the other to 1. In the interests of flexibility and robustness, our proposed approach does not assume a model for the death time $D \wedge L$ but for its mean $E(D \wedge L)$. This, however, sacrifices some efficiency in settings wherein a (correctly specified) parametric model is assumed.

Stemming from our current work are several interesting directions worth exploring in the future. One is to consider 'residual' RMST (Grand and Putter, 2016),

i.e. conditional RMST given the patients still being alive at a later time point after being wait-listed. This could be achieved by applying our methods to landmark analysis, which incorporates time-varying covariates into the mortality model. The results from this landmark analysis should be of interest to those who are already wait-listed for a period of time and want to know, for instance, the effect of MELD on residual survival time. It would also be interesting to contrast RMST to the analogous number, expectancy of life lost before time L, which is the area under the curve of cumulative incidence functions rather than marginal survival functions. Andersen (2013) has discussed decomposition of number of life years lost using pseudo-observations in the context of competing risk. Assuming the absence of time-varying covariates (as needed in Andersen, 2013) and the use of a simple linear link, we expect the estimated covariate effects from modeling number of life years lost would remain the same magnitude but change the sign, and the intercept would change to L subtracted by the intercept in RMST model. Furthermore, one recent paper by Zhao et al. (2015) proposed to infer RMST curves as a function of L. Following this direction, extension of our methods to RMST curves is a very interesting idea worth further consideration.

We have applied our proposed method to ESLD data to study pre-transplant mortality. Such data requires consideration not only of independent censoring, but also of dependent censoring, where the receipt of a transplant precludes observation of wait-list mortality. This is the first paper to directly estimate RMST in the ESLD setting. The R code to implement our methods is available upon request to the first author.

CHAPTER II

Computationally Efficient Modeling of Restricted Mean Survival Time Based on Clustered Data

2.1 Introduction

Restricted mean survival time (RMST) is often of great clinical interest in practice and is gaining increased attention among biostatisticians. There are now several existing methods to model RMST, with the methods distinguished by their estimation approaches and assumptions on the censoring mechanism (Karrison, 1987; Chen and Tsiatis, 2001; Andersen et al., 2004; Andersen and Perme, 2009; Zhang and Schaubel, 2011; Tian et al., 2014; Wang and Schaubel, 2017). Compared to proceeding indirectly by transforming other pivotal functions into RMST, direct modeling of RMST is more appealing in terms of parameter interpretation and computational convenience. We propose generalized estimating equation (GEE) methods to model RMST as a function of baseline covariates. In the interest of robustness and flexibility, we avoid making any distributional assumptions on the underlying restricted survival time. To represent covariate effects in the mortality model, we make a regression assumption on RMST with baseline covariates, usually through a link function analogous to a Generalized Linear Model (GLM).

Although parametrization of the RMST is generally convenient, computational difficulties may arise if the dimension of the covariate vector is quite large. Standard

software packages (e.g., R and SAS) typically handle datasets with tens of thousands of subjects better than they handle several hundred covariates. Examples include clustered data, or data with a high-dimensional covariate. In this report, the terms "cluster", "facility", and "center" could be used interchangeably. In the interest of concreteness, we use the term "center" hereafter, in part due to its connection with the data which motivated our work in this chapter. A conventional way to adjust for fixed facility effects in a regression model is to code potentially very large number of center indicators; this can introduce a high covariate dimension, in turn greatly increasing the computational burden.

The case we consider in this paper involves data characterized by a large number of centers and collected in an environment where fixed cluster effects are desired (Kalbfleisch and Wolfe, 2013). The dataset which motivates our proposed methods consists of End-Stage Renal Disease (ESRD) patients receiving hemodialysis in the United States. The study population has over 1 million patients from over $5,000$ dialysis facilities. We are interested in evaluating the effect on survival of some variables historically reported to be important prognostic factors, including age, race, gender, underlying diagnosis, and comorbidity information. However, facility is reported as a strong predictor of ESRD patient survival, and it is strongly suspected that the distributions of these prognostic factors are unequal across the thousands of facilities. Hence, the potential for bias exists if our model does not adjust for facility.

The purpose of this chapter is not to contrast fixed versus random effects as methods for adjusting for center. Strong cases can be made in either direction, such that the choice of one over the other depends largely on the data structure at hand, along with the analytic objectives. In our cases, covariate effects are of interest, but so are the facility effects. In this particular framework, one reason to

prefer fixed versus random facility effects is that a key assumption of the random effect model (i.e., the independence of facility effect and patient characteristics) is unlikely to hold. Fixed cluster effects are more appropriate to avoid such confounding issues when we suspect the individual covariates are correlated with facility effects. Moreover, in addition to estimating covariate effects, an objective of our analysis of the ESRD data is to identify "unusual" or "interesting" centers with significantly below or above average performance. Such results are of great interest to various stakeholders (e.g., regulatory bodies, oversight committees, insurers).

the overseers and payers, fixed effect methods have been demonstrated to yield less biased estimates of the extreme responses with smaller mean squared error when the true facility effect is far from that of the average facility (Kalbfleisch and Wolfe, 2013; He et al., 2013). We continue discussion in contrasting fixed and random center effects in Section (2.5).

In the context of ESRD data, conventional methods would involve simultaneous estimation of the covariate effects and about $6,000$ facilities indicators. As will be demonstrated, the proposed techniques separate the estimation of center specific baseline RMST from the estimation of covariate effects. As will be detailed, we can exploit standard software for implementation, but the proposed yield much faster run times relative to those typically employed in the generalized linear model setting. In particular, we dissect the model structure and connect it to the estimation procedure of the stratified proportional hazards model (Cox, 1972, 1975; Boudreau and Lawless, 2006).

The novelty of the methods proposed in this report is primarily from two perspectives. First, to the best of our knowledge, no previous work has proposed methods for estimating center effects through a direct model of RMST. Second, no previous

report has addressed computational issues likely to arise in large data sets; we focus on center effects (as a frequently occurring instance of high dimensional categorical covariates) and propose techniques which greatly reduce computing time and that should be quite appealing to practitioners (e.g., can leverage standard software).

The remainder of this report is organized as follows. In Section (2.2), we formulate the data structure, describe the proposed methods, and then propose the estimation procedures. Large sample properties are derived in Section (2.3), with numerical studies conducted in Section (2.4) to assess the accuracy of the proposed procedures in finite samples. We illustrate our methods in Section (2.5) through application to the motivating ESRD data. Discussion and possible future directions are presented in Section (2.6).

2.2 Proposed Methods

2.2.1 Notation and assumptions

Let i denote the i'th patient ($i = 1, \ldots, n$) and g_i denote this patient's center, where $g_i = 1, \ldots, J$ and J is usually a relatively large number (e.g., $J = 1,000$). To simplify the notation, we create a vector variable $\boldsymbol{G}_i = (G_{i1}, \ldots, G_{iJ})$, where $G_{ij} = I(g_i = j)$, based on g_i, and this vector is all 0 except g_ith element equal to 1. Baseline covariates of interest are denoted by $\boldsymbol{Z}_i$, a vector of length p. Let D_i denote the mortality time, which is subject to right censoring time C_i. Due to the occurrence of censoring, we observe the minimum of death and censoring time, $X_i = D_i \wedge C_i$, and hence, we define the death indicator $\Delta_i^D = I(D_i \leq C_i)$. Suppose L is the pre-specified time point of interest; then define restricted observation time $Y_i = D_i \wedge L$ and its corresponding indicator $\Delta_i^Y = I(D_i \wedge L \leq C_i)$. Our observed data are then $\mathcal{O} = \{\mathcal{O}_i;\ i = 1, \ldots, n\}$, where $\mathcal{O}_i = \{\boldsymbol{Z}_i, \boldsymbol{G}_i, X_i, Y_i, \Delta_i^Y, \Delta_i^D\}$.

We are particularly interested in the average survival time up to L, i.e., RMST

at L. Since our intention is to develop a useful tool to evaluate survival based on the information available at the time origin, we model the RMST as a function of baseline (i.e., time 0) covariates. Analogous to a GLM with log link, we assume the following mortality model for the RMST, $\mu_{ij} = E\left(D_i \wedge L | \boldsymbol{Z}_i, g_i = j\right)$

$$\mu_{ij} \quad = \quad \mu_{0j} \exp\left(\boldsymbol{\beta}_0' \boldsymbol{Z}_i\right), \tag{2.1}$$

where $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0p})'$ is the covariate effect of interest, and $\boldsymbol{\mu}_0 = (\mu_{01}, \ldots, \mu_{0J})'$ is the center-specific baseline RMST. Model (2.1) has the same structure as a GLM with the log link. However, note that the variance structure is unspecified. The model is equivalent to model with centers represented by J indicator variables; i.e., $\exp\{\boldsymbol{\beta}_0' \boldsymbol{Z}_i + \boldsymbol{G}_i' \log(\boldsymbol{\mu}_0)\}$. For the data structure of our interest in this report, J is usually a large number, such that fitting the model requires careful consideration to avoid computational difficulties. In order to avoid estimating the J center effects simultaneously, we propose a two-stage procedure which allows us to separately estimate $\boldsymbol{\beta}_0$ and $\boldsymbol{\mu}_0$.

2.2.2 Censoring models

In the absence of censoring, $E[G_{ij} \boldsymbol{Z}_i \{D_i \wedge L - \mu_{0j} \exp(\boldsymbol{\beta}_0' \boldsymbol{Z}_i)\}] = \mathbf{0}$. This can serve as the basis for constructing estimating equations, but requires modification in the presence of censoring. To accommodate censoring, we employ a variant of Inverse Probability Censoring Weight (IPCW) (Robins and Rotnitzky, 1992; Robins and Finkelstein, 2000). In our context, IPCW re-weights the observed death (or at risk patients) such that the weighted uncensored data represent the $(D_i \wedge L)$ distribution that for the target population. We allow the censoring distribution to depend on the baseline covariates and to differ across centers. Note that covariate dependent censoring is quite common, such as the staggered entry in an observational study

with a fixed calendar period. In this case, subjects who enter later in the observation window would have a different censoring distribution than those who enter earlier.

Denote the hazard function for censoring time C by $\lambda_{ij}^C(t)$ for patient i from center j, with

$$\lambda_{ij}^C(t) = \lim_{h \to 0} \frac{P\left(X_i \in [t, t+h), \Delta_i^D = 0 | X_i \geq t, \boldsymbol{Z}_i, g_i = j\right)}{h},$$

and denote the corresponding cumulative hazard by $\Lambda_{ij}^C(t) = \int_0^t \lambda_{ij}^C(u)du$. The IPCW weight is given by can $W_{ij}^C(t) = \exp\{\Lambda_{ij}^C(t)\}$. Using the weight value at time Y_i, denoted by $W_i = \sum_{j=1}^J G_{ij} W_{ij}^C(Y_i)$, it can be shown that the following weighted expectation has mean zero:

$$E\left[G_{ij}\Delta_i^Y W_i \boldsymbol{Z}_i \left\{Y_i - \mu_{0j}\exp\left(\boldsymbol{\beta}_0'\boldsymbol{Z}_i\right)\right\}\right] = \mathbf{0}. \tag{2.2}$$

In practice, $\Lambda_{ij}^C(t)$ is rarely known and needs to be estimated from the observed data. For this purpose, we assume the following Cox model for censoring:

$$\lambda_{ij}^C(t) = \lambda_{0j}^C(t)\exp\left(\boldsymbol{\theta}_0'\boldsymbol{Z}_i\right). \tag{2.3}$$

The use of Cox regression is well-established in the context of IPCW, and the above censoring assumption can accommodate both covariate-independent and covariate-dependent censoring. After estimating $\widehat{\boldsymbol{\theta}}$ and $\widehat{\Lambda}_{0j}^C(t)$ through the partial likelihood (Cox, 1975) and Breslow (Breslow, 1972) estimators, respectively, we can estimate the IPCW weights at time t as $\widehat{W}_{ij}^C(t) = \exp\{\exp(\widehat{\boldsymbol{\theta}}'\boldsymbol{Z}_i)\widehat{\Lambda}_{0j}^C(t)\}$. With $\widehat{\boldsymbol{\theta}}$ and $\widehat{\Lambda}_{0j}^C(t)$ from censoring model (2.3), we can estimate W_i as $\widehat{W}_i = \exp\{\exp(\widehat{\boldsymbol{\theta}}'\boldsymbol{Z}_i)\widehat{\Lambda}_{0j}^C(Y_i)\}$. We then substitute the estimated weights $\widehat{\boldsymbol{W}} = (\widehat{W}_1, \ldots, \widehat{W}_n)'$ in place of their corresponding true values $\boldsymbol{W} = (W_1, \ldots, W_n)'$.

2.2.3 Estimating equations

Based on the zero-mean property (2.2), we construct the following estimating equations:

$$\begin{aligned}\sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}W_i\Delta_i^Y\left(Y_i-\mu_{ij}\right)\boldsymbol{Z}_i &= \mathbf{0}\\ \sum_{i=1}^{n} G_{ij}W_i\Delta_i^Y\left(Y_i-\mu_{ij}\right) &= 0, \quad j=1,\ldots,J.\end{aligned}$$

Substituting $\widehat{\boldsymbol{W}}$ for $\boldsymbol{W}$, we can estimate $\boldsymbol{\beta}_0$ and $\boldsymbol{\mu}_0$ from following $p+J$ working estimating equations:

$$\sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}\widehat{W}_i\Delta_i^Y\left\{Y_i-\mu_{0j}\exp\left(\boldsymbol{\beta}_0'\boldsymbol{Z}_i\right)\right\}\boldsymbol{Z}_i = \mathbf{0}, \tag{2.4}$$

$$\sum_{i=1}^{n} G_{ij}\widehat{W}_i\Delta_i^Y\left\{Y_i-\mu_{0j}\exp\left(\boldsymbol{\beta}_0'\boldsymbol{Z}_i\right)\right\} = 0, \quad j=1,\ldots,J. \tag{2.5}$$

Solving (2.4) and (2.5) simultaneously implies simultaneous estimation of $p+J$ parameters, which is subject to numerical instability when J is quite large. Instead, we propose estimating $\boldsymbol{\beta}_0$ first through iteration, and then estimating $\boldsymbol{\mu}_0$ through J separate closed-form expressions. Along these lines, we define:

$$\boldsymbol{S}_j^{(k)}\left(\boldsymbol{\beta},\boldsymbol{W}\right) = \frac{\sum_{i=1}^{n} G_{ij}W_i\Delta_i^Y\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}}{\sum_{i=1}^{n} G_{ij}}, \tag{2.6}$$

$$\overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta},\boldsymbol{W}\right) = \frac{\boldsymbol{S}_j^{(1)}\left(\boldsymbol{\beta},\boldsymbol{W}\right)}{S_j^{(0)}\left(\boldsymbol{\beta},\boldsymbol{W}\right)}, \tag{2.7}$$

for $j=1,\ldots,J, k=0,1,2$. For a vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 0}=1$, $\boldsymbol{a}^{\otimes 1}=\boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2}=\boldsymbol{a}'\boldsymbol{a}$. Using the defined $\boldsymbol{S}_j^{(k)}$'s and $\overline{\boldsymbol{S}}_j$'s, we can rewrite the estimating equations (2.4)-(2.5) as follows:

$$\sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}\left\{\boldsymbol{Z}_i-\overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta},\widehat{\boldsymbol{W}}\right)\right\}\widehat{W}_i\Delta_i^Y Y_i = \mathbf{0}, \tag{2.8}$$

$$\mu_{0j} = \frac{\sum_{i=1}^{n} G_{ij}\widehat{W}_i\Delta_i^Y Y_i}{\sum_{i=1}^{n} G_{ij}\widehat{W}_i\Delta_i^Y\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)}, \quad j=1,\ldots,J. \tag{2.9}$$

The algebra underlying the equivalence of (2.4)-(2.5) and (2.8)-(2.9) is provided in Appendix. Note that (2.8) is free of the center-specific parameters $\boldsymbol{\mu}_0$ and that (2.9) is a closed-form calculation of $\boldsymbol{\mu}_0$, allowing us to separately estimate $\boldsymbol{\beta}_0$ and $\boldsymbol{\mu}_0$.

2.2.4 Fitting proposed model using Cox regression software

We now demonstrate how (2.8)-(2.9) can be easily estimated using standard Cox regression software. Consider the following stratified Cox model,

$$\lambda_{ij}^{\dagger}(t) = \lambda_{0j}^{\dagger}(t) \exp\left(\boldsymbol{\gamma}' \boldsymbol{Z}_i\right),$$

for the death hazard of a patient $i \in \{1, \ldots, n\}$ from cluster $j \in \{1, \ldots, J\}$ with baseline covariate $\boldsymbol{Z}_i$. We set $N_i^{\dagger}(t)$ and $R_i^{\dagger}(t)$ as the counting process for death and at-risk indicator, respectively. As implemented by, for example, R and SAS, $\boldsymbol{\gamma}$ and $\lambda_{0j}^{\dagger}$ can be estimated from the estimating equations given below with weights $W_i^{\dagger}(t) = 1$. A variant of the regular estimating equation is well developed by weighting with $W_i^{\dagger}(t)$ as IPCW weights (Zhang and Schaubel, 2011).

Our goal is to coerce the software (e.g., *coxph* in R, *phreg* in SAS) to fit model (2.1) by solving the estimating equations (2.8)-(2.9). To do so using such software, we build connections between (2.8)-(2.9) and the IPCW version of the Cox score equations:

$$\sum_{j=1}^{J} \sum_{i=1}^{n} \int_0^{\tau} G_{ij} \widehat{W}_i^{\dagger}(u) \left\{ \boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j^{\dagger}\left(u; \boldsymbol{\gamma}, \boldsymbol{W}^{\dagger}\right) \right\} dN_i^{\dagger}(u) = \boldsymbol{0}, \tag{2.10}$$

$$\int_0^t \frac{\sum_{i=1}^n G_{ij} W_i^{\dagger}(u) dN_i^{\dagger}(u)}{\sum_{i=1}^n G_{ij} W_i^{\dagger}(u) \exp\left(\boldsymbol{\gamma}' \boldsymbol{Z}_i\right) R_i^{\dagger}(u)} = \Lambda_{0j}^{\dagger}(t), j = 1, \ldots, J, \tag{2.11}$$

where for $k = 0, 1, 2$,

$$\boldsymbol{S}_j^{(k)\dagger}\left(t; \boldsymbol{\gamma}, \boldsymbol{W}^{\dagger}\right) = \frac{\sum_{i=1}^n G_{ij} W_i^{\dagger}(t) \exp\left(\boldsymbol{\gamma}' \boldsymbol{Z}_i\right) R_i^{\dagger}(t) \boldsymbol{Z}_i^{\otimes k}}{\sum_{i=1}^n G_{ij} W_i^{\dagger}(t) \exp\left(\boldsymbol{\gamma}' \boldsymbol{Z}_i\right) R_i^{\dagger}(t)}, \tag{2.12}$$

$$\overline{\boldsymbol{S}}_j^{\dagger}\left(t; \boldsymbol{\gamma}, \boldsymbol{W}^{\dagger}\right) = \frac{\boldsymbol{S}_j^{(1)\dagger}\left(t; \boldsymbol{\gamma}, \boldsymbol{W}^{\dagger}\right)}{\boldsymbol{S}_j^{(0)\dagger}\left(t; \boldsymbol{\gamma}, \boldsymbol{W}^{\dagger}\right)}. \tag{2.13}$$

First, remove the integral signs from (2.10)-(2.11), such that only the increment at time u is considered. Next, replace $\widehat{W}_i^{\dagger}(u)$ with $\widehat{W}_i\Delta_i^Y Y_i$, then set $R_i^{\dagger}(u) = 1$ and $dN_i^{\dagger}(u) = 1$. By this point, it's clear that u is arbitrary. To conform with the software, we can set u equal to any positive number; here, we set $u = 1$. Then (2.10)-(2.11) are equivalent to our (2.8)-(2.9) after adding an offset $-\log(Y_i)$ to the linear predictor.

Combining the above information implies that our proposed model can be fitted using software for a standard Cox regression with the data set augmented such that: (a) observation time set to 1 for each subject; (b) $\widehat{W}_i\Delta_i^Y Y_i$ used for a weight; (c) $-\log(Y_i)$ used for an offset; (d) center serving as strata.

2.2.5 An efficient algorithm for our proposed methods

Based on the connection between our proposed estimating equations and the Cox analogous in Section (2.2.4), we propose the following computationally efficient estimation procedure:

(i) Estimate the censoring hazard $\widehat{\Lambda}_{ij}^C(t)$ from model (2.3) by unweighted partial likelihood and Breslow estimator; construct IPCW weights by $\widehat{W}_i = \exp\{\widehat{\Lambda}_{ij}^C(Y_i)\}$ for $i = 1, \ldots, n$.

(ii) Create a dataset wherein each patient is observed to die at time 1 and has the baseline covariate $\boldsymbol{Z}_i$.

(iii) Fit a stratified Cox model to the dataset create in Step (ii), with center serving as strata, covariate $\boldsymbol{Z}_i$, weight $\widehat{W}_i\Delta_i^Y Y_i$, and set the offset to $-\log(Y_i)$. Note that ties should be handled by Breslow option, which is default in SAS, but not R.

Step (iii) can be implemented by several statistical software packages (e.g., R, SAS).

The algorithm is quite fast, even in very large dataset, owing to the stratification. The resulting coefficient and baseline hazard serve as our proposed estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\mu}}$.

2.2.6 Center effects

Note that $\{\mu_{01}, \ldots \mu_{0J}\}$ represent the center-specific baseline RMST and, analogous to a center-specific intercept, do not represent center-specific contrasts. For settings where contrasts between centers are of interest, we propose rescaling μ_{0j} to $\eta_j = \mu_{0j}/\boldsymbol{w}'\boldsymbol{\mu}_0$, where $\boldsymbol{w} = (w_1, \ldots, w_J)'$ is a pre-specified weight vector with $\boldsymbol{w}'\mathbf{1} = 1$. An example of $\boldsymbol{w}$ would be $\boldsymbol{w} = (1, \ldots, 1)'/J$, i.e., equal weight across all J centers. The rescaled $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_J)$ represents covariate-adjusted contrasts between each center and a weighted average center. Note that the weighted average of the contrasts equals 1 (i.e., $\boldsymbol{w}'\widehat{\boldsymbol{\eta}} = 1$), which is a desirable property for interpretation purposes.

2.3 Asymptotic Properties

Before presenting the asymptotic properties of our proposed estimators, we specify the following regularity conditions for $i = 1, \ldots, n$ and $j = 1, 1, \ldots, J$.

(a) $\{\mathcal{O}_1, \ldots \mathcal{O}_n$ are independently and identically distributed.

(b) $P\{R_i(t) = 1\} > 0$ for $t \in (0, \tau]$, where $R_i(t) = I(X_i \geq t)$ is the at risk process.

(c) $|Z_{ik}| < M_Z < \infty$, where Z_{ik} is the kth component of $\boldsymbol{Z}_i$.

(d) $\Lambda_{0j}^C(\tau) < \infty$ and $\Lambda_{0j}^C(t)$ is absolutely continuous for $t \in (0, \tau]$.

(e) There exist neighborhoods $\mathcal{C}$ of $\boldsymbol{\theta}$ such that for $k = 0, 1, 2, j = 1, \ldots, J$,

$$\sup_{t \in (0,\tau], \boldsymbol{\theta} \in \mathcal{C}} \left\| \boldsymbol{R}_j^{(k)}(t; \boldsymbol{\theta}) - \boldsymbol{r}_j^{(k)}(t; \boldsymbol{\theta}) \right\| \xrightarrow{p} 0,$$

where

$$\boldsymbol{R}_j^{(k)}(t;\boldsymbol{\theta}) = \frac{\sum_{i=1}^n G_{ij}\exp\left(\boldsymbol{\theta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}R_i(t)}{\sum_{i=1}^n G_{ij}}, \tag{2.14}$$

$$\boldsymbol{r}_j^{(k)}(t;\boldsymbol{\theta}) = E\left\{G_{ij}\exp\left(\boldsymbol{\theta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}R_i(t)\right\}. \tag{2.15}$$

(f) There exist neighborhoods $\mathcal{B}$ of $\boldsymbol{\beta}_0$ such that for $k = 0, 1, 2, j = 1, \ldots, J$,

$$\sup_{t\in(0,\tau],\boldsymbol{\beta}\in\mathcal{B}}\left\|\boldsymbol{S}_j^{(k)}\left(\boldsymbol{\beta},\boldsymbol{W}\right) - \boldsymbol{s}_j^{(k)}(\boldsymbol{\beta})\right\| \xrightarrow{p} 0,$$

where

$$\boldsymbol{s}_j^{(k)}(\boldsymbol{\beta}) = E\left\{G_{ij}\Delta_i^Y W_i\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}\right\} = E\left\{G_{ij}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}\right\},$$

(g) The matrices $\boldsymbol{A}(\boldsymbol{\beta}_0), \boldsymbol{\Theta}(\boldsymbol{\theta})$ are each positive definite, where

$$\boldsymbol{A}(\boldsymbol{\beta}) = \sum_{j=1}^J E\left\{G_{ij}\Delta_i^Y W_i Y_i\left(\frac{\boldsymbol{s}_j^{(2)}(\boldsymbol{\beta})}{s_j^{(0)}(\boldsymbol{\beta})} - \overline{\boldsymbol{s}}_j(\boldsymbol{\beta})^{\otimes 2}\right)\right\},$$

$$\boldsymbol{\Theta}(\boldsymbol{\theta}) = \sum_{j=1}^J E\left\{G_{ij}\int_0^\tau\left(\frac{\boldsymbol{r}_j^{(2)}(t;\boldsymbol{\theta})}{r_j^{(0)}(t;\boldsymbol{\theta})} - \overline{\boldsymbol{r}}_j(t;\boldsymbol{\theta})^{\otimes 2}\right)\boldsymbol{r}_j^{(0)}(t;\boldsymbol{\theta})\lambda_{0j}^C(t)dt\right\}.$$

and $\overline{\boldsymbol{s}}_j(\boldsymbol{\beta}) = z_j^{(0)}(\boldsymbol{\beta})^{-1}\boldsymbol{z}_j^{(1)}(\boldsymbol{\beta})$, $\overline{\boldsymbol{r}}_j(t;\boldsymbol{\theta}) = r_j^{(0)}(t;\boldsymbol{\theta})^{-1}\boldsymbol{r}_j^{(1)}(t;\boldsymbol{\theta})$.

These conditions can be relaxed at the expense of additional technical development. Our main asymptotic results are summarized in the following three theorems, with the proofs presented in the Appendix.

Theorem II.1. *Under regularity conditions (a)-(g), as $n \to \infty$, $\widehat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a zero-mean Normal with variance $\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{B}(\boldsymbol{\beta}_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}$ with $\boldsymbol{A}(\boldsymbol{\beta}_0)$ defined in condition (g) and $\boldsymbol{B}(\boldsymbol{\beta}_0)$ defined as follows:*

$$\boldsymbol{B}\left(\boldsymbol{\beta},\boldsymbol{W}\right) = E\left[\sum_{j=1}^J G_{ij}\left\{\boldsymbol{b}_i\left(\boldsymbol{\beta},\boldsymbol{W}\right)\right\}^{\otimes 2}\right],$$

where

$$\begin{aligned}
\boldsymbol{b}_i(\boldsymbol{\beta}, \boldsymbol{W}) &= \left\{\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j(\boldsymbol{\beta}, \boldsymbol{W})\right\} W_i \Delta_i^Y (Y_i - \mu_{ij}) + \boldsymbol{K}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W}) \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \boldsymbol{U}_i(\boldsymbol{\theta}) \\
&+ \int_0^L \frac{\boldsymbol{H}_j(u; \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W})}{r_j^{(0)}(u; \boldsymbol{\theta})} dM_i^C(u), \\
\boldsymbol{K}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W}) &= \sum_{j=1}^{J} E\left\{G_{ij} \boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W}) \boldsymbol{D}_i(\boldsymbol{\theta})'\right\}, \\
\boldsymbol{U}_i(\boldsymbol{\theta}) &= G_{ij} \int_0^{\tau} \left\{\boldsymbol{Z}_i - \overline{\boldsymbol{r}}_j(u; \boldsymbol{\theta})\right\} M_i^C(u), \\
\boldsymbol{H}_j(t; \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W}) &= E\left\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W}) \exp(\boldsymbol{\theta}' \boldsymbol{Z}_i) R_i(t)\right\}.
\end{aligned}$$

and $M_i^C(t) = N_i^C(t) - \int_0^t R_i(u) \exp(\boldsymbol{\theta}' \boldsymbol{Z}_i) d\lambda_{0i}^C(u)$ *is the censoring martingale, with* $N_i^C(t) = I(X_i \leq t, \Delta_i^D = 0)$ *being the censoring counting process.*

The consistency of $\widehat{\boldsymbol{\beta}}$ holds by the Inverse Function Theorem (Foutz, 1977), while the proof of asymptotic normality follows through the combination of various Taylor series expansions and the Cramér-Wold Theorem. This sandwich variance with $\boldsymbol{B}$ as the middle matrix treats IPCW weights as estimated from the data, which well reflects the reality. However, the calculation of this variance could be complicated. A useful short cut involves replacing the middle matrix $\boldsymbol{B}$ with $\boldsymbol{B}^*$:

$$\boldsymbol{B}^*(\boldsymbol{\beta}, \boldsymbol{W}) = E\left[\sum_{j=1}^{J} G_{ij} \left\{\boldsymbol{b}_i^*(\boldsymbol{\beta}, \boldsymbol{W})\right\}^{\otimes 2}\right],$$

where $\boldsymbol{b}_i^*(\boldsymbol{\beta}, \boldsymbol{W}) = \boldsymbol{Z}_i - \sum_{j=1}^{J} \overline{\boldsymbol{S}}_j(\boldsymbol{\beta}, \boldsymbol{W}) W_i \Delta_i^Y (Y_i - \mu_{ij})$ is the first and primary component of the original $\boldsymbol{b}_i(\boldsymbol{\beta}, \boldsymbol{W})$. This short cut treats the IPCW weights as fixed rather than estimated. Although it does not fully reflect the actual estimating procedure, this short cut is much easier to calculate and should serve as a useful substitute for the more complicated variance estimator implied by Theorem II.1, particularly since the primary source of variation is still captured.

Theorem II.2. *Under regularity conditions (a)-(g), as* $n \to \infty$, $\widehat{\boldsymbol{\mu}}_0$ *converges in probability to* $\boldsymbol{\mu}_0$ *and* $\sqrt{n}(\widehat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0)$ *converges to a zero-mean Normal with variance*

$\boldsymbol{V}_\mu$, *where*

$$\boldsymbol{V}_\mu = E\left[\left\{\begin{pmatrix} \frac{n}{n_1 s_1^{(0)}(\boldsymbol{\beta})} G_{ij} W_i \Delta_i^Y \left(Y_i - \mu_{ij}\right) \\ \vdots \\ \frac{n}{n_J s_J^{(0)}(\boldsymbol{\beta})} G_{ij} W_i \Delta_i^Y \left(Y_i - \mu_{ij}\right) \end{pmatrix} - \sum_{j=1}^{J} G_{ij} \begin{pmatrix} \mu_{01} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mu_{0J} \end{pmatrix} \begin{pmatrix} \overline{\boldsymbol{s}}_1(\boldsymbol{\beta})' \\ \vdots \\ \overline{\boldsymbol{s}}_J(\boldsymbol{\beta})' \end{pmatrix} \boldsymbol{b}_i\left(\boldsymbol{\beta}, \boldsymbol{W}\right) \right\}^{\otimes 2}\right].$$

Theorem II.3. *Under regularity conditions (a)-(g), as* $n \to \infty$, $\widehat{\boldsymbol{\eta}}$ *converges in probability to* $\boldsymbol{\eta}_0$ *and* $\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ *converges to a zero-mean Normal with variance* $\boldsymbol{V}_\eta$, *where*

$$\boldsymbol{V}_\eta = \left(\boldsymbol{\mu}_0' \boldsymbol{w}\right)^{-4} \left(\boldsymbol{\mu}_0' \boldsymbol{w} \boldsymbol{I}_J - \boldsymbol{\mu}_0 \boldsymbol{w}'\right) \boldsymbol{V}_\mu \left(\boldsymbol{\mu}_0' \boldsymbol{w} \boldsymbol{I}_J - \boldsymbol{w} \boldsymbol{\mu}_0'\right).$$

The proofs of Theorems II.2 and II.3 proceed by applying the Delta Method to the results of Theorem II.1.

2.4 Simulation Study

We generated the number of patients across $J = 50$ centers from a multinomial distribution with equal weights $1/J$ and the total sample size n. Three total sample sizes are tested: $n = 2,500$, $n = 5,000$, $n = 10,000$ and $n = 20,000$.

Death times are generated from an Exponential with mean $1/\mu_{0j}^\dagger \exp(-\beta_1^\dagger Z_{1i} - \beta_2^\dagger Z_{2i})$, where Z_{1i} and Z_{2i} each follow independent Normal$(0, 1)$ distributions. We set $\beta_1^\dagger = 0.5, \beta_2^\dagger = 1$ and let $\mu_{01}^\dagger, \dots, \mu_{0J}^\dagger$ range from 0.158 to 0.550 with an equal increment. The true parameter values are determined computationally by Monte Carlo Methods with sample size 10 million. The censoring time also follows an Exponential distribution with hazard $\lambda_{0j}^C \exp(\theta_1 Z_{1i} + \theta_2 Z_{3i})$, where two censoring patterns are tested, resulting in ≈15% and ≈30% censoring. The first censoring pattern uses $\boldsymbol{\theta} = (0.4, 0.1)'$ and with $\lambda_{C01}, \dots, \lambda_{C0J}$ ranging from 0.0108 to 0.05 with an equal increment. The second censoring pattern uses $\boldsymbol{\theta} = (0.5, -0.5)'$ and lets $\lambda_{C01}, \dots, \lambda_{C0J}$ range from 0.712 to 0.810 with an equal increment. The performance of the proposed methods is evaluated at 5 different truncation points: $L = 0.18,\ 0.57,\ 1.8,\ 5.4,\ 13.4,$

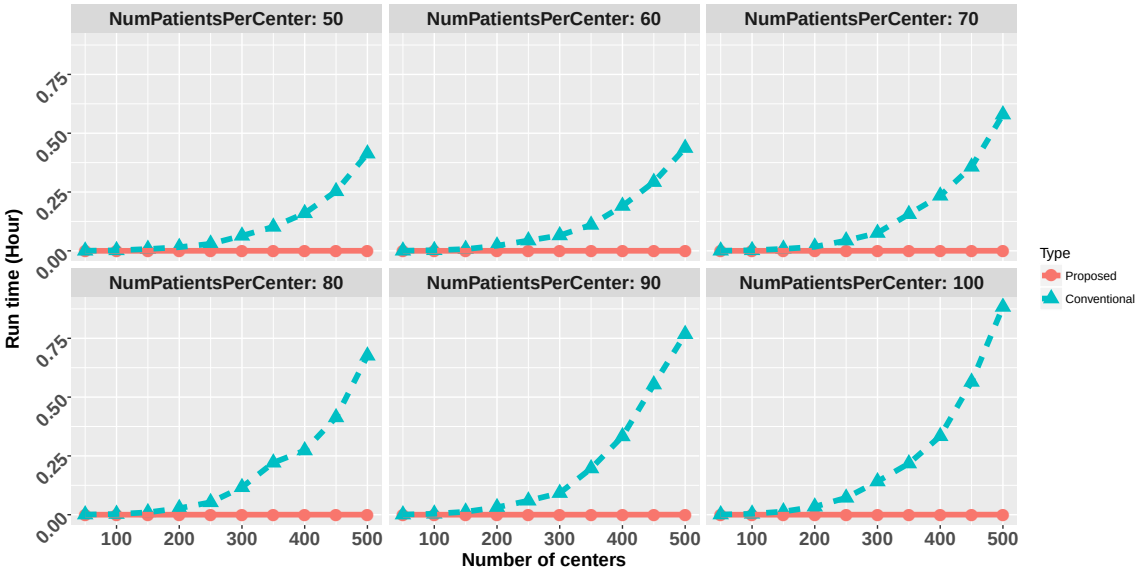
which represent approximately the 10th, 25th, 50th, 75th, and 90th percentiles, respectively, of the potentially censored death time distribution.

To illustrate the difference in the run time between our proposed methods and a standard weighted GLM approach (which would simultaneously estimate the covariate and center parameters), we choose $L = 1.8$ under the first censoring set-up aforedescribed. Run times are presented for $J = 25,\ 50,\ 100,\ 200,\ 400,\ 600,\ 800,\ 1000$ and on average $50,\ 60,\ 70,\ 80,\ 90,\ 100$ patients per center, respectively. Each run time is calculated using the average across 10 replicates. The conventional methods, which create dummy variables for each center and solve estimating equations (2.4)-(2.5) simultaneously, are implemented in R using the package geepack (Wang and Schaubel, 2017). Relative to the proposed methods, the conventional method runs much slower and result in approximately a $10-$ to $3000-$fold increase in run time required to estimate the model. This is depicted in Figure (2.1), when number of centers ranges from $J = 50$ to $J = 500$. The savings in computation time offered by our proposed algorithm increases rapidly with increasing J and also growing average number of patients per center.

Another disadvantage of using conventional methods to fit model (2.1) is that such an approach requires large storage to create the center indicators for a large data set. For example, the data set which motivated our methods (with $> 5,000$ centers and >1 million patients), requires R to allocate $\approx$ 50GB to create all the center indicators using the common data types, and about 10GB if special packages are used (e.g., *sparseMatrix*). In contrast, fitting model (2.1) through our proposed methods does not require the creation of center indicators, which greatly reduces the storage requirements.

For purposes of simplicity, we present the simulation results for $L = 1.8$ and

Figure 2.1 Computational time for our proposed and conventional methods with different J's and number of patients per center



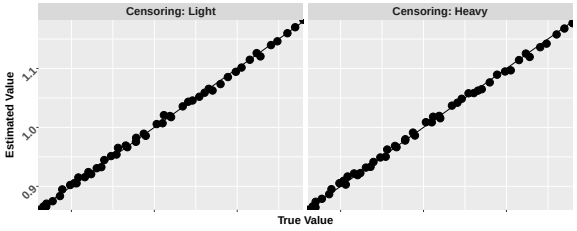
$L = 5.4$, and relegate those corresponding to the remaining L values to the Appendix. As shown in Table (2.1), the magnitude of the bias decreases generally as sample size increases. Average Standard Error (ASE) is calculated using the afore-mentioned short cut (which treats the estimated $\widehat{\boldsymbol{W}}$ as known), and is on average very close to empirical standard deviation (ESD). The coverage probability (CP) corresponding to ASE is quite close to 95%, except in a few scenarios under heavy censoring. We omit the simulation results corresponding to the standard error estimator derived from Theorem II.1, since the results are very similar to those presented for the short cut formula.

Figure (2.2) shows the true and estimated values, and (**??**) shows ESD and ASD, for $\widehat{\boldsymbol{\eta}}$ under light and heavy censoring for $L = 1.8$ and $L = 5.4$. As sample size increases, the distribution of the estimates shifts towards the true values and the variation decreases. The CPs are always around 95%.

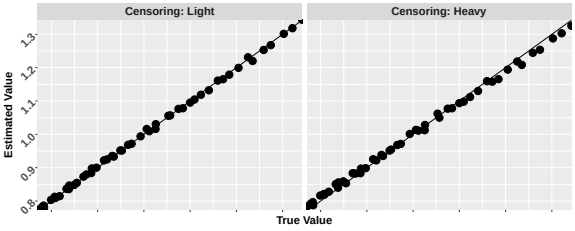
Table 2.1 Simulation results: $L = 1.8$ and $L = 5.4$ under light and heavy censoring

L	Censoring	Var (True)	n	Bias	ESD	ASE	CP(%)
1.8	Light	$\beta_1(-0.132)$	2500	0.002	0.01	0.009	94
			5000	0.001	0.007	0.007	93
			10000	0.001	0.005	0.005	95
		$\beta_2(-0.164)$	2500	0.002	0.01	0.01	93
			5000	0.001	0.007	0.007	95
			10000	0	0.005	0.005	94
	Heavy	$\beta_1(-0.132)$	2500	0.004	0.01	0.01	93
			5000	0.002	0.008	0.007	93
			10000	0.001	0.005	0.005	96
		$\beta_2(-0.164)$	2500	0.002	0.01	0.01	94
			5000	0.001	0.007	0.007	95
			10000	0.001	0.005	0.005	96
5.4	Light	$\beta_1(-0.132)$	2500	0.004	0.014	0.014	95
			5000	0.002	0.01	0.01	94
			10000	0.001	0.007	0.007	95
		$\beta_2(-0.164)$	2500	0.004	0.015	0.014	92
			5000	0.003	0.01	0.01	95
			10000	0.001	0.007	0.007	94
	Heavy	$\beta_1(-0.132)$	2500	0.016	0.018	0.018	86
			5000	0.01	0.013	0.013	90
			10000	0.005	0.009	0.01	92
		$\beta_2(-0.164)$	2500	0.006	0.017	0.016	93
			5000	0.003	0.012	0.012	94
			10000	0.002	0.008	0.008	95

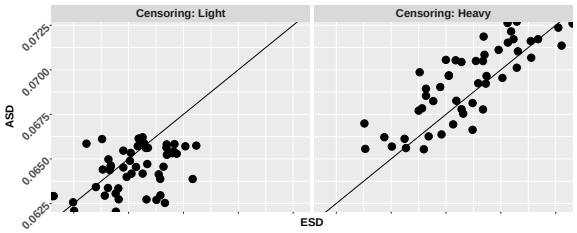
Figure 2.2 True and estimated values and standard deviation of $\widehat{\boldsymbol{\eta}}$ for $L = 1.8$ and $L = 5.4$



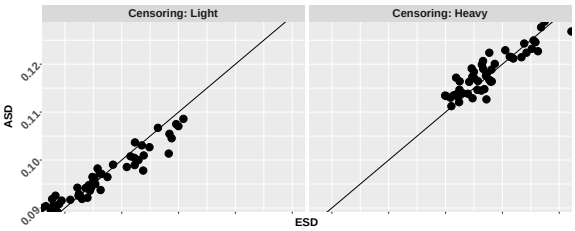
(a) Estimation for $L = 1.8$



(b) Estimation for $L = 5.4$



(c) Standard deviation for $L = 1.8$



(d) Standard deviation for $L = 5.4$

2.5 Application Data Analysis

We analyze survival for end-stage renal disease (ESRD) patients, using data obtained from the United States Renal Data System (USRDS). We include all patients initiating renal replacement therapy (RRT) on hemodialysis in the United States between January 1, 2004 and December 31, 2014. We excluded patients with a prior kidney transplant and patients aged <18 at the time of RRT-initiation. For each patient, follow-up starts at the date of RRT initiation and continues until the earliest of the following four events: death, transplantation, loss to follow-up, or 12/31/2014. The event of primary interest is death. We have $n = 1,061,403$ patients from $J = 5,301$ ESRD facilities. Approximately 64% of patients are observed to die. We choose $L = 5$ years as the truncation point. Out of $n = 1,061,403$ patients, 55% were observed to die before L, 27% were censored before L, and 18% were truncated at L.

Prognostic factors historically reported as being important and, hence, included in our analysis include: calendar year of RRT initiation (centered at 2004), age at RRT initiation (centered at 50 years and scaled by 5), gender, race (Caucasian, Asian, Black, and Other), ethnicity (Hispanic or not), primary renal diagnosis (glomerulonephritis (GN), diabetes, hypertension, and others), and 8 binary indicators of comorbidity conditions: cancer, diabetes, athlerosclerotic heart disease (ASHD), congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), cerebrovascular accident (CVA), peripheral vascular disease (PVD), elicit drug use, smoking status (current/former, non), and alcohol consumption. The RMST model of mortality includes the afore-mentioned covariates as predictors and $J = 5,301$ ESRD Network facilities as centers. The Cox model for censoring includes the same

set of covariates and is stratified by center. Estimated coefficients for the RMST model are displayed in Table (2.2).

The center effect is evaluated by both center-specific RMST μ_j and rescaled η_j. Figure (2.3) shows the histogram of the $J = 5,301$ center-specific $\widehat{\mu}_j$'s, the majority of which lie between 3.5 and 5. Figure (2.3) displays the point and interval estimates (95% confidence level) of rescaled η_j's. A total of 656 (12%) of facilities are significantly below average 5-year RMST, while 582 (11%) are significantly above average. There were 4,063 (77%) facilities that were not significantly different from the average 5-year RMST.

It took R approximately 11.33 minutes to calculate the IPCW weights; 2.65 minutes to estimate our proposed methods; then another 4.65 hours to calculate the standard error for $\widehat{\boldsymbol{\mu}}$. However, it requires R to allocate about 50GB memory to create the data needed for conventional methods, which is impossible for most of the local computers. Thus, in this particular example, storage considerations alone preclude a meaningful comparison between the proposed and conventional GLM procedures with respect to run times.

2.6 Discussion

We have developed a computationally attractive way to model RMST in the presence of a large number of centers. Computational advantages include great reductions in storage requirements and run times relative to conventional methods. We have demonstrated that our proposed methods have good finite-sample performance. The methods accommodate the estimation of fixed center effects through a normalized center effect measure.

We applied our methods to ESRD analysis and out of $J = 5,301$ ESRD facilities

Table 2.2 Estimated covariate effects on RMST ($L = 5$ years)

$\boldsymbol{Z}$	$\widehat{\boldsymbol{\beta}}$	SE	p
(Age-50)/5 (Years)	-0.057	0	< 0.001
Initiation year-2004	-0.041	0	< 0.001
Gender	*Reference: Male*		
female	-0.003	0.002	0.060
Ethnicity	*Reference: Not Hispanic*		
Hispanic	0.139	0.003	< 0.001
Race	*Reference: Caucasian*		
Asian	0.147	0.004	< 0.001
Black	0.11	0.002	< 0.001
Other	-0.041	0.008	< 0.001
PRD	*Reference: GN*		
Diabetes	-0.025	0.003	< 0.001
Hypertention	-0.016	0.003	< 0.001
Other	-0.117	0.004	< 0.001
CO	*Reference: No*		
ASHD	0.009	0.002	< 0.001
Cancer	-0.196	0.004	< 0.001
CHF	-0.15	0.002	< 0.001
COPD	-0.141	0.003	< 0.001
CVA	-0.08	0.003	< 0.001
Diabetes	-0.022	0.003	< 0.001
Drug use	-0.096	0.007	< 0.001
PVD	-0.102	0.003	< 0.001
Tabacoo use	-0.011	0.003	< 0.001
Alcohol use	-0.132	0.007	< 0.001

Figure 2.3 Histogram of estimated $J = 5,301$ center-specific RMST μ_j's

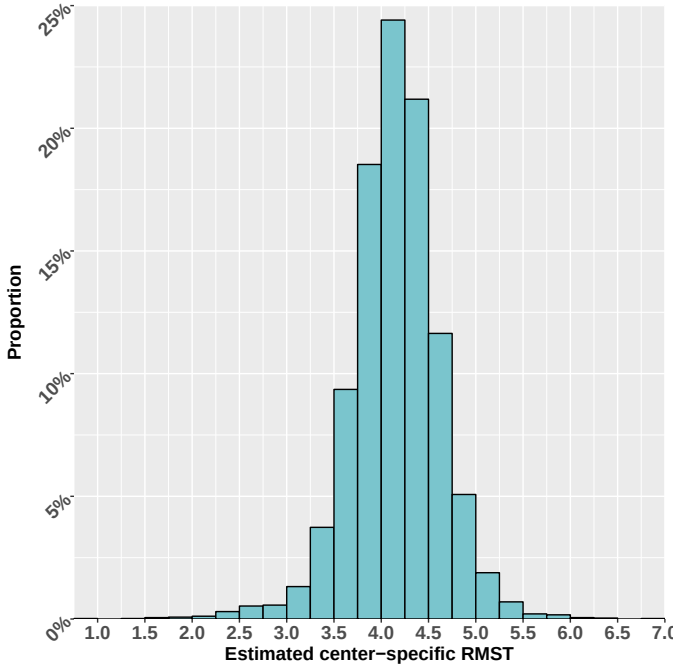
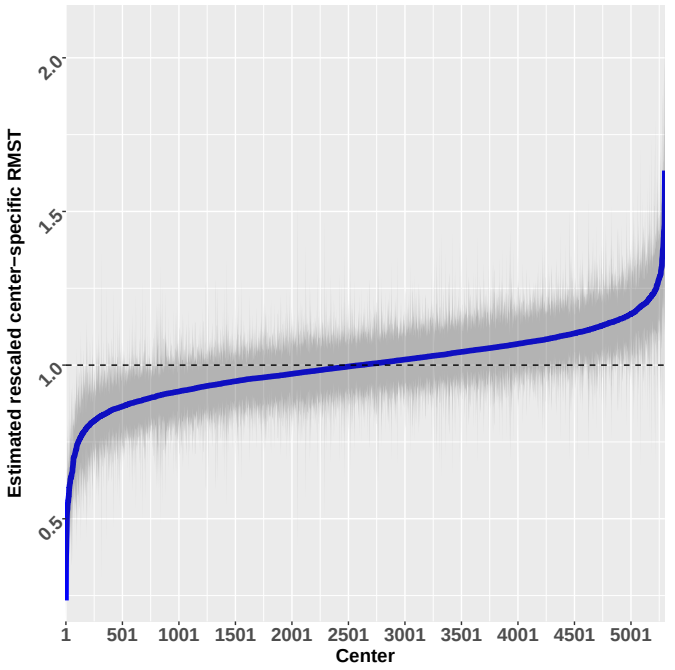


Figure 2.4 Point estimator and confidence interval of $J = 5,301$ rescaled η_j's



detected about 12% facilities significantly below and 11% significantly above average in terms of 5-year RMST. This proportion might be inflated and a more robust method empirical null is proposed by Efron (2004) and Kalbfleisch and Wolfe (2013). They propose to correct the normalized test statistics by the empirical distribution and avoid over flags of the significantly different centers. An application example is shown in (He et al., 2013). We expect fewer centers will be flagged as significantly different from average with empirical null methods employed.

The parameter choice of the weights depends on the research objective of the analysis. Our data analysis uses equal weight across all the center so that the resulted weighted average is not dominated by the large centers. This way it is easier to detect the centers with unusual performance. Some other reasonable choices include center-size proportional weights, resulting into the national average; this choice carries more interpretability but is driven by large centers.

CHAPTER III

Instrumental Variable Methods based on Restricted Mean Survival Time Models

3.1 Introduction

An important limitation of observational studies is that lack of randomization and the potential for unmeasured confounding generally results in a disconnect between association and causation. In non-experimental studies, where important confounding variables may be unobserved, the traditionally used covariate balancing approaches (e.g., matching, inverse probability weighting, stratification) do not suffice, since such approaches only control for measured confounders.

Instrumental variable (IV) methods are a popular approach for consistent estimation in the presence of unmeasured confounders. Unmeasured confounding that affects both treatment and response induces correlation between treatment and the error term in a regression model. Explanatory variables which suffer from such correlation issues are referred to as *endogenous*, which may include the treatment of interest in an observational study. Explanatory variables which do not suffer from such confounding issues are termed *exogenous* variables; i.e., the adjustment covariates. Fitting a regression model with only treatment and adjustment covariates and, hence, not accounting for unobserved confounders will often result in biased estimation. However, if an instrument is available, it may still be possible to obtain a

consistent estimator of the treatment effect. An instrument is a variable that does not directly impact the response, but is correlated with treatment and conditionally uncorrelated with error term, given the adjustment covariates. To control for unmeasured confounding, IV methods seek to find an instrument, then exploit a randomized experiment embedded in the observational study through which to estimate the causal treatment effect.

IV methods have a long history in economic applications, typically in the context of uncensored responses. The presence of censoring complicates the extension of IV methods to survival data and, correspondingly, very few methods have been developed in this vein. Existing IV methods for censored data can generally be classified as nonparametric, semiparametric or (fully) parametric. Non-parametric IV methods are usually used in the context of a binary instrument and treatment, and target the Local Average Treatment Effect (LATE), also known as Complier Average Causal Effect (CACE), which is the average causal treatment effect among the subpopulation of compliers (Baker, 1998; Baiocchi et al., 2014). For example, Nie et al. (2011) makes use of the mixture structure implied by the latent compliance, with estimation through empirical likelihood. Such nonparametric methods are generally robust, due to not making any distributional assumptions. However, such methods are of somewhat limited applicability, in the sense that they require the treatment variable to be binary and do not permit covariate adjustment. Parametric IV models which do permit covariate adjustment have been developed for censored outcomes (Tang and Lee, 1998; Chen et al., 2012). The treatment effect in a fully parametric model can be estimated by Maximum Likelihood or Expectation-Maximization methods. Kjaersgaard and Parner (2015) handled censoring with pseudo-observation methods, then modeled restricted mean survival time through a linear model. Para-

metric methods suffer the potential for substantial bias when the assumed model is incorrectly specified. Semiparametric approaches are generally more flexible, such as accelerated failure time models (Robins and Tsiatis, 1991; Loeys and Goetghebeur, 2003) and, more recently, additive hazards models (Li et al., 2015; Chan, 2016; Zheng et al., 2017). Martinussen et al. (2017)

Our proposed methods are semi-parametric, in the interests of both robustness and covariate adjustment. We propose to model the cumulative measure, Restricted Mean Survival Time (RMST) (Andersen and Perme, 2009; Tian et al., 2014). RMST is the average survival time up to a pre-specified time point, say L, and can be expressed as the area under the survival curve over the $[0, L]$ time interval; i.e., $\int_0^L S(t)dt$. RMST is a cumulative summary of survival from time 0 to L. Cumulative treatment effects are often of greater interest than instantaneous effects, especially in the presence of a treatment effect which changes direction over time (Schaubel and Wei, 2011). Mean survival time is generally not identifiable in the presence of censoring, except under a fully parametric model. Our proposed methods directly model RMST, which is much more computationally convenient and intuitively interpretable than indirect approaches (i.e., starting by estimating the hazard or survival function and then integrating the survival function to obtain the RMST). The assumed RMST model only makes assumptions regarding the mean structure and, in the interests of robustness, leaves the variance unspecified . In addition, we allow censoring to depend on the observed covariates, a property shared with few existing IV methods.

IV models are usually estimated by two-stage least squares when the response model is linear. For non-linear link functions, there are two variations, two-stage predicator substitution (2SPS) and two-stage residual inclusion (2SRI). The 2SPS

and 2SRI procedures use different decompositions of the treatment variable and error term to remove their mutual dependence. For censored data, the 2SPS and 2SRI procedures rely on different censoring assumptions (Chan, 2016).

Our motivating example involves the comparison between hemodialysis (HD) and peritoneal dialysis (PD) with respect to 5-year RMST among End-Stage Renal Disease (ESRD) patients. Although kidney transplantation is the preferred treatment for ESRD, most ESRD patients are placed on dialysis either while awaiting transplantation or as their only therapy (Fenton et al., 1997). HD uses a man-made membrane to filter waste and remove excess fluid from the blood. PD, a newer, less costly but much less employed method, uses the lining of the abdominal cavity and a solution to remove waste and excess fluid from the body. It has long been debated which dialytic method provides better survival. Some studies show that PD is associated with an initial survival advantage, but no significant difference afterwards (Fenton et al., 1997; Jaar et al., 2005; Heaf et al., 2002; Kumar et al., 2014). Others show that the mortality rate in PD patients is significantly higher than that in HD patients, especially in older patients (Kim et al., 2014; Weinhandl et al., 2010). Overall, results from the existing literature are conflicting, with a major concern being the potentially strong selection bias; e.g., PD patients tend to be younger and healthier. The majority of the above-cited studies only controlled for measured confounders, which leads naturally to the question we address. Which (if either) dialytic method emerges as superior in terms of patient survival if one accounts for unmeasured confounders? We apply IV methods in order to address this question.

Referring to the terminology familiar to the IV setting, the endogenous treatment in our ESRD data is a binary 0/1 indicator for taking PD rather than HD. We are interested in the survival difference at 5 years following dialysis initiation. A general

strategy to find an instrument for comparing treatments A and B is to look for naturally occurring variation in medical practice patterns (e.g., at the physician level). The instrument could then be defined as degree of PD usage; this IV represents a preference-based instrument (Baiocchi et al., 2014). In our motivating example, a possible preference-based instrument would be facility-level mean PD usage (e.g., fraction of patients initiating dialysis on PD). This IV may well have a strong influence on the individual treatment preference, without impacting directly on individual patient survival.

In the sections that follow, we first formulate the notations and data structure, describe the proposed methods and estimating procedure, and then derive the asymptotic properties in Section (3.2). We conduct simulation studies to evaluate the accuracy of the proposed procedures in finite samples in Section (3.3), and then apply our methods to ESRD data in Section (3.4). We close out the chapter with a discussion in Section (3.5).

3.2 Proposed Methods

3.2.1 Notations and Assumptions

We denote the treatment variable by A, observed adjustment covariates by $\boldsymbol{Z}$ (a vector of length p), and unobserved variables by U. We consider only one treatment variable for simplicity; generalization to more than one dimension is straightforward. Let time to event be represented by D (i.e., time of death) and let the pre-specified truncation time be $L \leq \tau$, where τ is the maximum censoring time. The death time D is subject to independent right censoring time, C. Let $X = D \wedge C$ denote the observed follow-up time, where $a \wedge b = \min(a, b)$, and we let $\Delta_D = I(D \leq C)$ be the observed-death indicator. Let $Y = X \wedge L$ denote the observed restricted survival time, and let $\Delta_Y = I(D \wedge L \leq C)$ be its corresponding event indicator.

We assume that the underlying RMST follows the following model,

$$E(D \wedge L|\boldsymbol{Z}, A, U) = \boldsymbol{\beta}_Z'\boldsymbol{Z} + \beta_A A + \beta_U U, \tag{3.1}$$

where U is unobserved and correlated with A, and the underlying error is independent of any other variable. The parameter vector $(\boldsymbol{\beta}_Z, \beta_A)$ is of primary interest. Fitting any regression model based on (3.1) with $\boldsymbol{Z}$ and A, but without U, generally leads to biased estimation of $(\boldsymbol{\beta}_Z, \beta_A)$, except when either $[U \perp (\boldsymbol{Z}, A)]$ or $[U \perp D|\boldsymbol{Z}, A]$ holds. Note that we do not assume that either condition holds.

To use the IV approach with treatment A, we need to find an observable variate I, not represented in Eq. (3.1), that satisfies the following two conditions:

Valid: I is uncorrelated with U: $\text{Cov}(I, U) = 0$.

Informative: I is correlated with A:

$$A = \boldsymbol{\alpha}_Z'\boldsymbol{Z} + \alpha_I I + \alpha_U U + \epsilon, \tag{3.2}$$

where $\alpha_I \neq 0$, $E(\epsilon) = 0$ and $\epsilon \perp (\boldsymbol{Z}, I)$.

When I satisfies both the validity and informativeness properties, it is referred to as an instrumental variable (or instrument), for A. We do not put any restriction on the distribution of I or A. They can be both continuous, or both discrete, or having continuous and discrete characteristics at the same time, as long as the second moments of all variables are finite (Wooldridge, 2010).

IV methods estimate the regression parameters from a reduced form of the response variate by plugging (3.2) to (3.1) and rearranging as follows:

$$\begin{aligned} E\left(D \wedge L|\boldsymbol{Z}, I\right) &= E\left[\left(\boldsymbol{\beta}_Z + \beta_A\alpha_Z\right)'\boldsymbol{Z} + \beta_A\alpha_I I + \left(\beta_U + \beta_A\alpha_U\right)U + \beta_A\epsilon|\boldsymbol{Z}, I\right] \\ &= \left(\boldsymbol{\beta}_Z + \beta_A\alpha_Z\right)'\boldsymbol{Z} + \beta_A\alpha_I I + E\left[\left(\beta_U + \beta_A\alpha_U\right)U + \beta_A\epsilon|\boldsymbol{Z}, I\right] \\ &= \left(\boldsymbol{\beta}_Z + \beta_A\alpha_Z\right)'\boldsymbol{Z} + \beta_A\alpha_I I, \end{aligned}$$

where $E(U|\boldsymbol{Z}, I) = 0$ holds by the definition of observed adjustment and instrumental variable, and $E(\epsilon|\boldsymbol{Z}, I) = 0$ follows from the informative property of instrument variable. If the parameters in model (3.2) are known, then we define $\widetilde{A} = \boldsymbol{\alpha}'_Z \boldsymbol{Z} + \alpha_I I$ and can obtain $(\boldsymbol{\beta}_Z, \beta_A)$ from the following model:

$$E\left(D \wedge L|\boldsymbol{Z}, \widetilde{A}\right) = \boldsymbol{\beta}'_Z \boldsymbol{Z} + \beta_A \widetilde{A}. \tag{3.3}$$

3.2.2 Estimation Procedure in the Absence of Censoring

Let i be a subject randomly drawn from the whole population of sample size n. The observed data are denoted by $\mathcal{O} = \{\boldsymbol{Z}_i, A_i, I_i, Y_i, \Delta_{Di} : i = 1, \ldots, n\}$.

In the absence of censoring, we can obtain $(\boldsymbol{\beta}_Z, \beta_A)$ by a two stage procedure. In the first stage, we would fit the linear model (3.2) through Ordinary Least Squares (OLS) by regressing A on $\boldsymbol{Z}, I$ to obtain $\widehat{\boldsymbol{\alpha}}_Z, \widehat{\alpha}_I$. We then calculate the predicted value of $\widetilde{A}$ as $\widehat{A} = \widehat{\boldsymbol{\alpha}}'_Z \boldsymbol{Z} + \widehat{\alpha}_I I$. At the second stage, we substitute the observed treatment variable A with its fitted value from the first stage, $\widehat{A}$. We then estimate $(\boldsymbol{\beta}_Z, \beta_A)$ by regressing $D \wedge L$ on $\boldsymbol{Z}, \widehat{A}$ via the following estimating equation,

$$\sum_{i=1}^{n} \left(D_i \wedge L - \begin{bmatrix} \boldsymbol{\beta}'_Z & \beta_A \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_i \\ \widehat{A}_i \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{Z}_i \\ \widehat{A}_i \end{bmatrix} = \mathbf{0}. \tag{3.4}$$

3.2.3 Estimation Procedure in the Presence of Censoring

The potential for censoring does not allow us to estimate the parameters in model (3.4) directly because we will not always observe the death time D. Denote the hazard function for right censoring time C by $\lambda_C(t)$, where

$$\lambda_C(t) = \lim_{h \to 0} \frac{P\{X \in [t, t+h), \Delta_D = 0 | X \geq t\}}{h}, \tag{3.5}$$

with corresponding cumulative hazards $\Lambda_C(t) = \int_0^t \lambda_C(u) du$. We propose to handle censoring by Inverse Probability of Censoring Weighting (IPCW) (Robins and

Rotnitzky, 1992; Robins, 1993; Robins and Finkelstein, 2000). The weight $W_C(t) = \exp\{\Lambda_C(t)\}$ reweights the observed death times by ghosting for the prognostically similar but censored observations. With the help of this weight at time Y, denoted by $W = W_C(Y)$, it can be shown that the following weighted expectation holds,

$$E\left\{\Delta_Y W\left(D \wedge L\right) | \boldsymbol{Z}, \widetilde{A}\right\} = \boldsymbol{\beta}_Z' \boldsymbol{Z} + \beta_A \widetilde{A}. \tag{3.6}$$

Since $\Lambda_C(t)$ is rarely known in practice, we usually need to estimate it from the observed data. For this purpose, we assume a Cox model for censoring. Cox regression is a well-established approach, especially in the context of IPCW. Since our model (3.3) holds conditional on $\boldsymbol{Z}$ and I, we allow censoring to depend on these variates though the model,

$$\lambda_C(t) = \lambda_{C0}(t) \exp\left\{\boldsymbol{\gamma}_Z' \boldsymbol{Z} + \gamma_I I\right\}. \tag{3.7}$$

Using partial likelihood (Cox, 1975) and the Breslow estimator (Breslow, 1972), we can estimate $\Lambda_{C0}(t)$ and $(\boldsymbol{\gamma}, \gamma_I)$. Plugging $\widehat{\Lambda}_{C0}(t)$ and $\widehat{\boldsymbol{\gamma}}_Z, \widehat{\gamma}_I$ into $\widehat{W} = \exp\{\int_0^Y d\widehat{\Lambda}_C(u)\}$, we obtain the following estimating equation:

$$\sum_{i=1}^{n} \Delta_i \widehat{W}_i \left(Y_i - \begin{bmatrix} \boldsymbol{\beta}_Z' & \beta_A \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_i \\ \widehat{A}_i \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{Z}_i \\ \widehat{A}_i \end{bmatrix} = \mathbf{0}. \tag{3.8}$$

3.2.4 Variance Estimation

A naive variance estimator can be obtained by ignoring the variability associated with $\widehat{A}$. Since this short-cut would generally result in variance estimation that is at least somewhat inaccurate, we instead derive the asymptotic properties and propose a more accurate sandwich variance estimator. To begin, we stack the estimating

equations from the two stages together and solve the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$.

$$\begin{aligned}
\boldsymbol{\phi}_1(\boldsymbol{\theta}; \mathcal{O}) &= \frac{1}{n}\sum_{i=1}^{n} \epsilon_{Yi} \begin{bmatrix} \boldsymbol{Z}_i \\ \widehat{A}_i \end{bmatrix} = \mathbf{0}, \\
\boldsymbol{\phi}_2(\boldsymbol{\theta}; \mathcal{O}) &= \frac{1}{n}\sum_{i=1}^{n} \epsilon_{Ai} \begin{bmatrix} \boldsymbol{Z}_i \\ I_i \end{bmatrix} = \mathbf{0},
\end{aligned}$$

where $\epsilon_{Yi} = \Delta_i \widehat{W}_i \left\{ Y_i - \left(\boldsymbol{\beta}_Z' \boldsymbol{Z}_i + \beta_A \widehat{A}_i \right) \right\}$ and $\epsilon_{Ai} = A_i - \left(\boldsymbol{\alpha}_Z' \boldsymbol{Z}_i + \alpha_I \widehat{I}_i \right)$. Let the true parameter value be $\boldsymbol{\theta}$, and the solution be our estimator $\widehat{\boldsymbol{\theta}}$. The Taylor expansion of $\boldsymbol{\phi} = (\boldsymbol{\phi}_1', \boldsymbol{\phi}_2')'$ at $\boldsymbol{\theta}$ around $\widehat{\boldsymbol{\theta}}$ is:

$$\begin{aligned}
\boldsymbol{\phi}(\boldsymbol{\theta}) &= 0 + \frac{\partial \boldsymbol{\phi}\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}} \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}} \right) + o\left(\left\| \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}} \right\| \right) \\
\Rightarrow \quad \sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) &= \left(-\frac{\partial \boldsymbol{\phi}\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}} \right)^{-1} \sqrt{n} \boldsymbol{\phi}(\boldsymbol{\theta}) + o\left(\left\| \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}} \right\| \right).
\end{aligned}$$

For purposes of simplicity, we treat the IPCW weights as fixed in this chapter, exploiting the findings from Chapter 1 that very little accuracy is lost. The derivation of the large-sample distribution follows:

$$-\frac{\partial \boldsymbol{\phi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{M}^A = \begin{bmatrix} \boldsymbol{M}_{11}^A & \boldsymbol{M}_{12}^A \\ \mathbf{0} & \boldsymbol{M}_{22}^A \end{bmatrix},$$

where

$$\begin{aligned}
\boldsymbol{M}_{11}^A &= E\left\{ \Delta_i W_i \begin{bmatrix} \boldsymbol{Z}_i \\ \widehat{A}_i \end{bmatrix}^{\otimes 2} \right\}, \\
\boldsymbol{M}_{12}^A &= E\left\{ \Delta_i W_i \left(\beta_A \begin{bmatrix} \boldsymbol{Z}_i \\ \widehat{A}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_i \\ I_i \end{bmatrix}' - \epsilon_{Yi} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \boldsymbol{Z}_i' & I_i \end{bmatrix} \right) \right\}, \\
\boldsymbol{M}_{22}^A &= E\left(\begin{bmatrix} \boldsymbol{Z}_i \\ I_i \end{bmatrix}^{\otimes 2} \right),
\end{aligned}$$

and

$$\sqrt{n}\boldsymbol{\phi}\left(\boldsymbol{\theta}\right)=\left[\begin{array}{c}\frac{1}{\sqrt{n}}\sum_{i}^{n}\boldsymbol{m}_{1}^{B}\\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{m}_{2}^{B}\end{array}\right]\overset{D}{\rightarrow}\text{Normal}\left(\left[\begin{array}{c}\mathbf{0}\\ \mathbf{0}\end{array}\right],\boldsymbol{M}^{B}\right),\boldsymbol{m}_{1}^{B}=\epsilon_{Yi}\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ \widehat{A}_{i}\end{array}\right],\boldsymbol{m}_{2}^{B}=\epsilon_{Ai}\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ I_{i}\end{array}\right],$$

where

$$\boldsymbol{M}^{B}=E\left(\left[\begin{array}{c}\boldsymbol{m}_{1}^{B}\\ \boldsymbol{m}_{2}^{B}\end{array}\right]^{\otimes2}\right)=E\left(\left[\begin{array}{cc}\epsilon_{Yi}^{2}\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ \widehat{A}_{i}\end{array}\right]^{\otimes2} & \epsilon_{Yi}\epsilon_{Ai}\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ \widehat{A}_{i}\end{array}\right]\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ I_{i}\end{array}\right]'\\ \epsilon_{Yi}\epsilon_{Ai}\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ I_{i}\end{array}\right]\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ \widehat{A}_{i}\end{array}\right]' & \epsilon_{Ai}^{2}\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ I_{i}\end{array}\right]^{\otimes2}\end{array}\right]\right).$$

Summarizing the above results, we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\overset{D}{\rightarrow}\text{Normal}\left(\left[\begin{array}{c}\mathbf{0}\\ \mathbf{0}\end{array}\right],\left(\boldsymbol{M}^{A}\right)^{-1}\boldsymbol{M}^{B}\left(\boldsymbol{M}^{A}\right)^{\prime-1}\right),$$

where the variance matrix $\left(\boldsymbol{M}^{A}\right)^{-1}\boldsymbol{M}^{B}\left(\boldsymbol{M}^{A}\right)^{-1}$ can be written as $E\left(\left[\begin{array}{cc}\boldsymbol{V}_{11} & \boldsymbol{V}_{12}\\ \boldsymbol{V}_{12}' & \boldsymbol{V}_{22}\end{array}\right]\right)$,

$$\begin{aligned}\boldsymbol{V}_{11} &= \left\{\left(\boldsymbol{M}_{11}^{A}\right)^{-1}\boldsymbol{m}_{1}^{B}-\left(\boldsymbol{M}_{11}^{A}\right)^{-1}\left(\boldsymbol{M}_{12}^{A}\right)\left(\boldsymbol{M}_{22}^{A}\right)^{-1}\boldsymbol{m}_{2}^{B}\right\}^{\otimes2}\\ \boldsymbol{V}_{12} &= \left(\boldsymbol{M}_{11}^{A}\right)^{-1}\left(\boldsymbol{m}_{1}^{B}\right)\left(\boldsymbol{m}_{2}^{B}\right)'\left(\boldsymbol{M}_{22}^{A}\right)^{-1}-\left(\boldsymbol{M}_{11}^{A}\right)^{-1}\left(\boldsymbol{M}_{12}^{A}\right)\left(\boldsymbol{M}_{22}^{A}\right)^{-1}\left(\boldsymbol{m}_{2}^{B}\right)^{\otimes2}\left(\boldsymbol{M}_{22}^{A}\right)^{-1}\\ \boldsymbol{V}_{22} &= \left(\boldsymbol{M}_{22}^{A}\right)^{-1}\left(\boldsymbol{m}_{2}^{B}\right)^{\otimes2}\left(\boldsymbol{M}_{22}^{A}\right)^{-1}.\end{aligned}$$

The asymptotic variance of $\widehat{\boldsymbol{\beta}}$ is then $\boldsymbol{V}_{11}=E\left\{\left(\boldsymbol{M}_{11}^{A}\right)^{-1}\boldsymbol{m}_{1}^{B}-\left(\boldsymbol{M}_{11}^{A}\right)^{-1}\left(\boldsymbol{M}_{12}^{A}\right)\left(\boldsymbol{M}_{22}^{A}\right)^{-1}\boldsymbol{m}_{2}^{B}\right\}^{\otimes2}$ and, therefore, the finite-sample variance of $\widehat{\boldsymbol{\beta}}$ can be estimated as

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\epsilon_{Yi}\left(\widehat{\boldsymbol{M}}_{11}^{A}\right)^{-1}\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ \widehat{A}_{i}\end{array}\right]-\epsilon_{Ai}\left(\boldsymbol{M}_{11}^{A}\right)^{-1}\left(\boldsymbol{M}_{12}^{A}\right)\left(\boldsymbol{M}_{22}^{A}\right)^{-1}\left[\begin{array}{c}\boldsymbol{Z}_{i}\\ I_{i}\end{array}\right]\right\}^{\otimes2}. \tag{3.9}$$

3.3 Simulation Study

We generated each of Z, I and U though a standard Normal distribution, then generated the treatment variable A by:

$$A = 0.5 + Z + 0.5I + U + \zeta,$$

where $\zeta \sim N(0, 1)$. We then simulated the death time D from an Exponential distribution with mean,

$$E(D|A, Z, U) = 5 - 0.8A + 0.8Z + 0.8U.$$

We generated the censoring time C from a Cox model with the following hazard,

$$\lambda_C(t) = \lambda_{C0} \exp\left(\gamma_Z Z + \gamma_I I\right).$$

Three scenarios, each with different censoring patterns, are evaluated:

Case I: No censoring.

Case II: $D \perp C$: $\lambda_{C0} = \exp(-3)$, $\gamma_Z = \gamma_I = 0$, resulting in $\approx 19\%$ censoring.

Case III: $D \perp C|Z, I$: $\lambda_{C0} = \exp(-2)$, $\gamma_Z = \gamma_I = -0.5$, where $\approx$39% censoring occurs.

The true RMST model is given by

$$E\left(D \wedge L|Z, A, U\right) = \beta_0 + \beta_Z Z + \beta_A A + \beta_U U,$$

where the true limiting values of the regression coefficient are calculated using Monte Carlo Methods with sample size 10 million. We will look at RMST at 2 different time points: $L = 3$ and $L = 6$, representing approximately the median and third quartile respectively. The corresponding true values of β_A are -0.12 and -0.311 respectively.

Table 3.1 Simulation results: $L = 3$, $\beta_A = -0.12$

Scenario	n	Bias			ASE	CP(%)
		BM	NE	Proposed		
I	500	-0.001	0.052	-0.001	0.093	94.3
	1000	-0.002	0.052	-0.003	0.065	95.2
II	500	0	0.053	-0.002	0.096	95.4
	1000	-0.001	0.053	0	0.068	95
III	500	-0.003	0.051	0.002	0.113	95.2
	1000	0.001	0.053	0.008	0.079	95.3

We compare the proposed IV estimators with two other estimators. The first is the benchmark estimator (BM), which is based on estimating equations similar to (3.8) but incorporates the unobserved U,

$$\sum_{i=1}^{n} \widehat{W}_i \left(Y_i - \left[\beta_Z^{\mathrm{BM}}, \beta_A^{\mathrm{BM}}, \beta_U^{\mathrm{BM}}\right] \begin{bmatrix} Z_i \\ A_i \\ U_i \end{bmatrix} \right) \begin{bmatrix} Z_i \\ A_i \\ U_i \end{bmatrix} = \mathbf{0}.$$

This estimator serves as the gold standard and applies only if we observe U. The second comparator is the naive estimator (NE) using similar techniques but without adjusting for the unmeasured confounder,

$$\sum_{i=1}^{n} \widehat{W}_i \left(Y_i - \left[\beta_Z^{\mathrm{NE}}, \beta_A^{\mathrm{NE}}\right] \begin{bmatrix} Z_i \\ A_i \end{bmatrix} \right) \begin{bmatrix} Z_i \\ A_i \end{bmatrix} = \mathbf{0}. \tag{3.10}$$

This estimator does not correct for unmeasured confounding with IV at all and thus the estimated treatment effect is expected to be biased.

The estimated treatment effect β_A from three methods are displayed in Tables (3.1) and (3.2), based on $n = 1000$ replicates. The benchmark estimators had very small bias, as expected. The naive estimators have much larger bias, which does not go away with increasing sample size. The bias of our proposed methods is small and generally shrinks as the sample size increases. Average Standard Error (ASE) is the

Table 3.2 Simulation results: $L = 6$, $\beta_A = -0.311$

Scenario	n	Bias			ASE	CP(%)
		BM	NE	Proposed		
I	500	-0.003	0.136	-0.006	0.194	94.5
	1000	-0.003	0.137	-0.007	0.136	93.8
II	500	-0.003	0.136	-0.007	0.211	95.4
	1000	-0.002	0.138	-0.001	0.149	95.7
III	500	0.01	0.14	0.064	0.306	93.9
	1000	0.008	0.143	0.055	0.225	94.3

average of $n = 1000$ proposed standard error estimates. The ASEs are quite close to the empirical standard deviation, thus making coverage probability (CP) for our proposed estimators quite close to the target value, 95%.

3.4 Application Data Analysis

We apply our proposed methods to evaluate the effect of hemodialysis (HD) versus peritoneal dialysis (PD), using data obtained from the United States Renal Data System (USRDS). The existing literature is conflicting with respect to the survival advantage of these two dialytic modalities, suggesting the presence of unmeasured treatment-outcome confounding. From this perspective, IV analysis may provide useful insight.

The study population consists of adults initiating dialysis between January 1, 2009 and December 31, 2014. For our analysis, each patient is classified by dialysis type at the time of dialysis initiation. We restrict the analysis to patients from the 929 dialysis facilities with at least 10 PD and 50 total patients during the study period. The instrumental variable we employed was facility-level PD usage, defined as the fraction of patients initiating dialysis on PD. We determined this proportion for each facility during a historical period, 2006-2008, in order to avoid inducing

Table 3.3 Analysis of USRDS data: Description of the study population by first modality

Covariate	HD	PD	Std Diff
Proportion of death	53	36	-36.3
Age (Years))	63.6	58.1	-36.5
Primary Renal Diagnosis			
Diabetes	46	43	-5.9
Hypertention	28	26	-4.6
Glomerulonephritis	8	15	22.3
Other	17	15	-5.5
Comorbidities			
Alcohol use	2	1	-11.6
ASHD	21	13	-21.3
Cancer	8	5	-11.9
CHF	33	16	-39.5
COPD	10	4	-22.4
CVA	10	6	-13.7
Diabetes	11	7	-12.1
Drug use	1	0	-10.7
PVD	14	9	-17.4
Tabacoo use	7	6	-2.9

patient-level confounding between the instrument and unmeasured variables. The calculated mean PD usage varies from 1.8% to 54.6% with a mean of 14.2%. Note that the Pearson correlation between facility-level PD usage in 2006–2008 and 2009-2014 is 0.57, while the coefficient for facility level PD usage in model 3.2 is 0.77 with $p < 0.001$, suggesting the potential for a good instrument candidate.

Table (3.3) confirms that patients treated with PD are generally healthier than those treated with HD. They are on average 5 years younger and suffer fewer comorbidities. There are likely some other unmeasured variables that affect the treatment choice of PD over HD, and it is likely that a covariate important enough to impact on treatment selection would also have an effect on survival.

In addition to the treatment indicator, other prognostic factors historically reported as being important and included in our analysis are: year of ESRD incidence (centered at year 2009), age at dialysis initiation (centered at 50 years, then scaled

by 5), gender, race (Caucasian, Asian, Black, and others), ethnicity (Hispanic or not), primary renal diagnosis (glomerulonephritis (GN) diabetes, hypertension, and others), and 10 binary comorbidity indicators of causes, including cancer, diabetes, athlero-sclerotic heart disease (ASHD), congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), cerebrovascular accident (CVA), peripheral vascular disease (PVD), elicit drug use, tobacco use, and alcohol consumption. The censoring model is a Cox model stratified by incidence year and with all the afore-mentioned covariates except PD treatment indicator. The calculated IPCW weights are capped at 100 to stabilize the estimating procedure. The naive estimator is obtained from estimating equation (3.10) with all the afore-listed prognostic factors.

For our proposed estimator, at the first stage we regressed the PD treatment indicator on all the prognostic factors and the facility level PD usage. This yields a fitted treatment indicator, which is then included in the second stage with the adjustment covariates. As shown in Table (3.4), PD is significantly protective compared to HD. Parameter estimates for are in the same direction for the proposed and naive methods. Results based on the proposed methods results in an increase in 5-year mean lifetimes of 0.31 years ($p = 0.001$).

3.5 Discussion

We have developed methods for employing instrumental variables to control for unmeasured confounding when modeling survival in terms of RMST. We consider only the linear link in this chapter. Future work should include extensions to additional choices for the link function (e.g., log, logistic). For non-linear link functions, different techniques can be used to estimate the mortality coefficients (Wooldridge, 2010). Another interesting potential extension involves the generalization of the

Table 3.4 Estimated covariate effects on RMST ($L = 5$ years)

Covariates	Naïve estimator	Proposed estimator	SE	p
Intercept	4.103	4.09	0.039	< 0.001
PD	0.279	0.308	0.09	0.001
(Age-50)/5 (Years)	-0.125	-0.124	0.003	< 0.001
Initiation year-2009	-0.385	-0.386	0.003	< 0.001
Gender	*Reference Group: Male*			
female	-0.013	-0.015	0.012	0.211
Ethnicity	*Reference Group: Not Hispanic*			
Hispanic	0.257	0.26	0.019	< 0.001
Race	*Reference Group: Caucasian*			
Asian	0.373	0.373	0.028	< 0.001
Black	0.185	0.187	0.016	< 0.001
Other	0.058	0.058	0.061	0.338
Primary Renal Diagnosis	*Reference Group: Glomerulonephritis*			
Diabetes	-0.162	-0.154	0.028	< 0.001
Hypertention	-0.136	-0.131	0.028	< 0.001
Other	-0.343	-0.341	0.032	< 0.001
Comorbidity	*Reference Group: No or Yes*			
ASHD	-0.06	-0.06	0.014	< 0.001
Cancer	-0.35	-0.349	0.021	< 0.001
CHF	-0.357	-0.352	0.014	< 0.001
COPD	-0.267	-0.264	0.018	< 0.001
CVA	-0.153	-0.15	0.019	< 0.001
Diabetes	-0.095	-0.091	0.02	< 0.001
Drug use	-0.272	-0.265	0.061	< 0.001
PVD	-0.208	-0.207	0.016	< 0.001
Tabacoo use	-0.08	-0.077	0.023	0.001
Alcohol use	-0.43	-0.428	0.051	< 0.001

censoring model. This chapter allows the censoring time to depend on adjustment covariates and instrumental variable. Our estimating procedure is called Two Stage Predictor Substitution (2SPS), because in the second stage the treatment variable is substituted by the predicted value from the first stage. Another related estimating procedure, Two Stage Residual Inclusion (2SRI), is identical to 2SPS for linear models, but permits a more relaxed censoring assumption. If the censoring depends on treatment variable, 2SRI is likely a good alternative to 2SPS.

APPENDICES

APPENDIX A

Appendix for Chapter I

A.1 Asymptotic Properties of The Proposed Estimator

A.1.1 Notations

To begin with, we review the essential notations needed for further discussion:

i: subject index, $i \in \{1, \ldots, n\}$

D_i: treatment-free death time

T_i: dependent censoring time; e.g. treatment

C_i: independent censoring time; e.g. administrative censoring

τ: end of follow up time

L: per-specified time point of interest, $L \leq \tau$

$Z_i = D_i \wedge T_i \wedge C_i$: observation time

$Y_i = Z_i \wedge L$: restricted observation time by L

$\Delta_i = I(D_i \wedge L \leq T_i \wedge C_i)$: indicator for restricted survival time $D_i \wedge L$

$\Delta_i^D = I(D_i \leq T_i \wedge C_i)$: death indicator

$\Delta_i^T = I(T_i < D_i \wedge C_i)$: dependent censoring indicator

$\Delta_i^C = I(C_i < D_i \wedge T_i)$: independent censoring indicator

$\boldsymbol{Z}_i^D(t)$: time-dependent covariates that predict death D_i

$\boldsymbol{Z}_i^T(t)$: time-dependent covariates that predict dependent censoring T_i

$\boldsymbol{Z}_i^C$: baseline covariates that predict independent censoring C_i

$\boldsymbol{Z}_i(t)$: a covariate set that stacks $\boldsymbol{Z}_i^D(t)$,$\boldsymbol{Z}_i^T(t)$, $\boldsymbol{Z}_i^C$ together and removes redundancy

$\widetilde{\boldsymbol{Z}}_i(t) = \{\boldsymbol{Z}_i(u) : 0 \leq u \leq t\}$: observation history of all the covariates up to time t

$\lambda_i^T(t)$: hazard rate for dependent censoring T_i

$\lambda_i^C(t)$: hazard rate for independent censoring C_i

$\Lambda_i^T(t) = \int_0^t \lambda_i^T(u)du$: cumulative hazard rate for dependent censoring T_i

$\Lambda_i^C(t) = \int_0^t \lambda_i^C(u)du$: cumulative hazard rate for independent censoring C_i

$N_i^D(t) = I(Z_i \leq t, \Delta_i^D = 1)$: counting process for death

$N_i^T(t) = I(Z_i \leq t, \Delta_i^T = 1)$: counting process for dependent censoring

$N_i^C(t) = I(Z_i \leq t, \Delta_i^C = 1)$: counting process for independent censoring

$R_i(t) = I(Z_i \geq t)$: at risk process

$dM_i^T(t) = dN_i^T(t) - R_i(t)d\Lambda_i^T(t)$: zero mean process for dependent censoring

$dM_i^C(t) = dN_i^C(t) - R_i(t)d\Lambda_i^C(t)$: zero mean process for independent censoring

A.1.2 Model Assumptions

We have made these assumptions in our paper:

(a) Assume restricted mean lifetime conditional on baseline covariates $\mu_i(L) := E\{D_i \wedge L|\boldsymbol{Z}_i^D(0)\}$ follows the model structure as below,

$$g\left[\mu_i\left(L\right)\right] \equiv g\left[E\left\{D_i \wedge L|\boldsymbol{Z}_i^D\left(0\right)\right\}\right] = \boldsymbol{\beta}_D'\boldsymbol{Z}_i^D\left(0\right),$$

where $g(*)$ is a given smooth and strictly monotone link function and $\boldsymbol{\beta}_D$ is of our primary interest.

(b) Assume Cox proportional hazards model for dependent and independent censor-

ing time T_i and C_i:

$$\lambda_i^T(t) = \lambda_0^T(t) \exp\left\{\boldsymbol{\beta}_T' \boldsymbol{Z}_i^T(t)\right\},$$

$$\lambda_i^C(t) = \lambda_0^C(t) \exp\left(\boldsymbol{\beta}_C' \boldsymbol{Z}_i^C\right).$$

(c) Assume no unmeasured confounders for dependent censoring T_i: for any $t > 0$,

$$\lim_{h\to 0} \frac{P\left\{Z_i \in [t, t+h), \Delta_i^T = 1 | Z_i \geq t, \widetilde{\boldsymbol{Z}}_i(t), D_i\right\}}{h} = \lim_{h\to 0} \frac{P\left\{Z_i \in [t, t+h), \Delta_i^T = 1 | Z_i \geq t, \widetilde{\boldsymbol{Z}}_i(t)\right\}}{h}.$$

(d) Assume independent censoring time is independent of either death time or dependent censoring time given baseline covariates; i.e.,

$$C_i \perp T_i | \boldsymbol{Z}_i(0), C_i \perp D_i | \boldsymbol{Z}_i(0).$$

A.1.3 Regularity Conditions

We specify the necessary regularity conditions (i)-(vii) as below.

(i) $\{Z_i, \Delta_i^D, \Delta_i^T, \Delta_i^C, \widetilde{\boldsymbol{Z}}_i(Z_i)\}, i = 1, \ldots, n$ are independently and identically distributed.

(ii) $P(R_i(t) = 1) > 0$ for $t \in (0, \tau], i = 1, \ldots, n$.

(iii) $|Z_{ik}(0)| + \int_0^\tau d|Z_{ik}(t)| < M_Z < \infty$ for $i = 1, \ldots, n$, where $Z_{ik}(t)$ are the kth components of $\boldsymbol{Z}_i(t)$.

(iv) $\Lambda_i^T(\tau) < \infty, \Lambda_i^C(\tau) < \infty$ and $\Lambda_i^T(t), \Lambda_i^C(t)$ are absolutely continuous for $t \in (0, \tau]$.

(v) There exist neighborhoods $\mathcal{B}_T$ of $\boldsymbol{\beta}_T$ and $\mathcal{B}_C$ of $\boldsymbol{\beta}_C$ such that for $k = 0, 1, 2$,

$$\sup_{t\in(0,\tau], \boldsymbol{\beta}\in\mathcal{B}_T} \left\| \frac{1}{n} \sum_{i=1}^n \exp\left\{\boldsymbol{\beta}' \boldsymbol{Z}_i^T(t)\right\} R_i(t) \boldsymbol{Z}_i^T(t)^{\otimes k} - \boldsymbol{r}_T^{(k)}(t; \boldsymbol{\beta}) \right\| \xrightarrow{p} 0,$$

$$\sup_{t\in(0,\tau], \boldsymbol{\beta}\in\mathcal{B}_C} \left\| \frac{1}{n} \sum_{i=1}^n \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i^C\right) R_i(t) \boldsymbol{Z}_i^{C\otimes k} - \boldsymbol{r}_C^{(k)}(t; \boldsymbol{\beta}) \right\| \xrightarrow{p} 0,$$

where $\boldsymbol{v}^{\otimes 0} = 1, \boldsymbol{v}^{\otimes 1} = \boldsymbol{v}, \boldsymbol{v}^{\otimes 2} = \boldsymbol{v}'\boldsymbol{v}$ and

$$\boldsymbol{r}_T^{(k)}(t; \boldsymbol{\beta}) = E\left[\exp\left\{\boldsymbol{\beta}' \boldsymbol{Z}_i^T(t)\right\} R_i(t) \boldsymbol{Z}_i^T(t)^{\otimes k}\right], \tag{A.1}$$

$$\boldsymbol{r}_C^{(k)}(t; \boldsymbol{\beta}) = E\left\{\exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i^C\right) R_i(t) \boldsymbol{Z}_i^{C\otimes k}\right\}. \tag{A.2}$$

(vi) Define $h(x) = \partial g^{-1}(x)/\partial x$, where h exists and is continuous in an open neighborhood $\mathcal{B}_D$ of $\boldsymbol{\beta}_D$.

(vii) The matrices $\boldsymbol{A}(\boldsymbol{\beta}_D), \boldsymbol{\Omega}_T(\boldsymbol{\beta}_T), \boldsymbol{\Omega}_C(\boldsymbol{\beta}_C)$ are each positive definite, where

$$\boldsymbol{A}(\boldsymbol{\beta}) = E\left[\boldsymbol{Z}_i^D(0)^{\otimes 2} h\left\{\boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0)\right\}\right], \tag{A.3}$$

$$\boldsymbol{\Omega}_T(\boldsymbol{\beta}) = E\left[\int_0^\tau \left\{\frac{\boldsymbol{r}_T^{(2)}(t;\boldsymbol{\beta})}{r_T^{(0)}(t;\boldsymbol{\beta})} - \overline{\boldsymbol{z}}_T(t;\boldsymbol{\beta})^{\otimes 2}\right\} dN_i^T(t)\right], \tag{A.4}$$

$$\boldsymbol{\Omega}_C(\boldsymbol{\beta}) = E\left[\int_0^\tau \left\{\frac{\boldsymbol{r}_C^{(2)}(t;\boldsymbol{\beta})}{r_C^{(0)}(t;\boldsymbol{\beta})} - \overline{\boldsymbol{z}}_C(t;\boldsymbol{\beta})^{\otimes 2}\right\} dN_i^C(t)\right], \tag{A.5}$$

and

$$\overline{\boldsymbol{z}}_T(t;\boldsymbol{\beta}) = \frac{\boldsymbol{r}_T^{(1)}(t;\boldsymbol{\beta})}{r_T^{(0)}(t;\boldsymbol{\beta})}, \tag{A.6}$$

$$\overline{\boldsymbol{z}}_C(t;\boldsymbol{\beta}) = \frac{\boldsymbol{r}_C^{(1)}(t;\boldsymbol{\beta})}{r_C^{(0)}(t;\boldsymbol{\beta})}. \tag{A.7}$$

A.1.4 Outline of Derivation

Two estimating equation mentioned in our paper are

$$\boldsymbol{\Phi}^*(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Phi}_i^*(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^n \Delta_i W_i(Y_i)\left[Y_i - g^{-1}\left\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D(0)\right\}\right] \boldsymbol{Z}_i^D(0) = \boldsymbol{0}, \tag{A.8}$$

where $W_i(t) = W_i^T(t) W_i^C(t)$, $W_i^T(t) = \exp\{\Lambda_i^T(t)\}$ and $W_i^C(t) = \exp\{\Lambda_i^C(t)\}$, and

$$\boldsymbol{\Phi}(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Phi}_i(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^n \Delta_i \widehat{W}_i(Y_i)\left[Y_i - g^{-1}\left\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D(0)\right\}\right] \boldsymbol{Z}_i^D(0) = \boldsymbol{0}, \tag{A.9}$$

where $\widehat{W}_i(t) = \widehat{W}_i^T(t)\widehat{W}_i^C(t)$, $\widehat{W}_i^T(t) = \exp\{\widehat{\Lambda}_i^T(t)\}$ and $\widehat{W}_i^C(t) = \exp\{\widehat{\Lambda}_i^C(t)\}$.

We will first show (A.8) is unbiased, and then (A.9) satisfies that $\sqrt{n}\boldsymbol{\Phi}(\boldsymbol{\beta}_D)$

converges to a zero-mean Normal with variance $\boldsymbol{B}(\boldsymbol{\beta}_D) = E\{\boldsymbol{B}_i(\boldsymbol{\beta}_D)^{\otimes 2}\}$, where

(A.10)

$$\boldsymbol{B}\left(\boldsymbol{\beta}_D\right) = E\{\boldsymbol{B}_i\left(\boldsymbol{\beta}_D\right)^{\otimes 2}\},$$

$$\boldsymbol{B}_i\left(\boldsymbol{\beta}\right) = \boldsymbol{\epsilon}_i\left(\boldsymbol{\beta}\right) + \boldsymbol{K}_T(\boldsymbol{\beta})\boldsymbol{\Omega}_T(\boldsymbol{\beta}_T)^{-1}\boldsymbol{U}_i^T\left(\boldsymbol{\beta}_T\right) + \int_0^L \boldsymbol{H}_T(u;\boldsymbol{\beta})r_T^{(0)}(u;\boldsymbol{\beta}_T)^{-1}dM_i^T\left(u\right)$$

(A.11) $$+ \boldsymbol{K}_C(\boldsymbol{\beta})\boldsymbol{\Omega}_C(\boldsymbol{\beta}_C)^{-1}\boldsymbol{U}_i^C\left(\boldsymbol{\beta}_C\right) + \int_0^L \boldsymbol{H}_C(u;\boldsymbol{\beta})r_C^{(0)}(u;\boldsymbol{\beta}_C)^{-1}dM_i^C\left(u\right),$$

(A.12) $$\boldsymbol{\epsilon}_i\left(\boldsymbol{\beta}\right) = \Delta_i W_i\left(Y_i\right)\left[Y_i - g^{-1}\{\boldsymbol{\beta}'\boldsymbol{Z}_i^D\left(0\right)\}\right]\boldsymbol{Z}_i^D\left(0\right),$$

(A.13) $$\boldsymbol{U}_i^T\left(\boldsymbol{\beta}_T\right) = \int_0^t \{\boldsymbol{Z}_i^T\left(u\right) - \overline{\boldsymbol{z}}_T(u;\boldsymbol{\beta}_T)\}dM_i^T\left(u\right),$$

(A.14) $$\boldsymbol{U}_i^C\left(\boldsymbol{\beta}_C\right) = \int_0^t \{\boldsymbol{Z}_i^C - \overline{\boldsymbol{z}}_C(u;\boldsymbol{\beta}_C)\}dM_i^C\left(u\right),$$

(A.15) $$\boldsymbol{K}_T\left(\boldsymbol{\beta}\right) = E\{\boldsymbol{\epsilon}_j\left(\boldsymbol{\beta}\right)\boldsymbol{D}_i^T\left(Y_i\right)'\},$$

(A.16) $$\boldsymbol{K}_C\left(\boldsymbol{\beta}\right) = E\{\boldsymbol{\epsilon}_j\left(\boldsymbol{\beta}\right)\boldsymbol{D}_i^C\left(Y_i\right)'\},$$

(A.17) $$\boldsymbol{H}_T\left(t;\boldsymbol{\beta}\right) = E[\boldsymbol{\epsilon}_j\left(\boldsymbol{\beta}\right)\exp\{\boldsymbol{\beta}_T'\boldsymbol{Z}_i^T\left(t\right)\}R_i\left(t\right)],$$

(A.18) $$\boldsymbol{H}_C\left(t;\boldsymbol{\beta}\right) = E\{\boldsymbol{\epsilon}_j\left(\boldsymbol{\beta}\right)\exp(\boldsymbol{\beta}_C'\boldsymbol{Z}_i^C)R_i\left(t\right)\},$$

(A.19) $$\boldsymbol{D}_i^T\left(t\right) = \int_0^t \{\boldsymbol{Z}_i^T\left(u\right) - \overline{\boldsymbol{z}}_T\left(u;\boldsymbol{\beta}_T\right)\}d\Lambda_i^T\left(u\right),$$

(A.20) $$\boldsymbol{D}_i^C\left(t\right) = \int_0^t \{\boldsymbol{Z}_i^C - \overline{\boldsymbol{z}}_C\left(u;\boldsymbol{\beta}_C\right)\}d\Lambda_i^C\left(u\right),$$

for any subject $i = 1, \ldots, n$, and $\boldsymbol{\Omega}_T(\boldsymbol{\beta}), \boldsymbol{\Omega}_C(\boldsymbol{\beta})$ are already defined in (A.4) and (A.5).

Let $\widehat{\boldsymbol{\beta}}_D$ denote the solution to (A.9). We will show that

(a) **(Consistency)** as $n \to \infty$, $\widehat{\boldsymbol{\beta}}_D$ converges in probability to $\boldsymbol{\beta}_D$.

(b) **(Asymptotic Properties)** as $n \to \infty$, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_D)$ converges to a zero-mean Normal with variance $\boldsymbol{A}(\boldsymbol{\beta}_D)^{-1}\boldsymbol{B}(\boldsymbol{\beta}_D)\boldsymbol{A}(\boldsymbol{\beta}_D)^{-1}$ with $\boldsymbol{A}(\boldsymbol{\beta})$ and $\boldsymbol{B}(\boldsymbol{\beta})$ defined in (A.3) and (A.10).

A.1.5 Unbiased Estimating Equation

Theorem A.1. *Under regularity conditions (i)-(vii), the estimating equation (A.8) is unbiased at the true value of $\boldsymbol{\beta}_D$; i.e. $E\{\boldsymbol{\Phi}^*(\boldsymbol{\beta}_D)\} = 0$.*

Proof. As defined in our paper, the i_{th} error term in (A.8) are independently and identically distributed. It would be enough to show that $E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta}_D)\} = \mathbf{0}$. This holds because the conditional expectation on $\boldsymbol{Z}_i^D(0)$ is unbiased:

$$
\begin{aligned}
&E\left\{\boldsymbol{\epsilon}_i\left(\boldsymbol{\beta}_D\right)|\boldsymbol{Z}_i^D\left(0\right)\right\} \\
&= \boldsymbol{Z}_i^D\left(0\right) E\left\{W_i\left(Y_i\right)\Delta_i Y_i|\boldsymbol{Z}_i^D\left(0\right)\right\} - \boldsymbol{Z}_i^D\left(0\right) g^{-1}\left\{\boldsymbol{\beta}_D'\boldsymbol{Z}_i^D\left(0\right)\right\} E\left\{W_i\left(Y_i\right)\Delta_i|\boldsymbol{Z}_i^D\left(0\right)\right\} \\
&= \boldsymbol{Z}_i^D\left(0\right) E\left[E\left\{W_i\left(Y_i\right)\Delta_i Y_i|D_i\right\}|\boldsymbol{Z}_i^D\left(0\right)\right] - \\
&\boldsymbol{Z}_i^D\left(0\right) g^{-1}\left\{\boldsymbol{\beta}_D'\boldsymbol{Z}_i^D\left(0\right)\right\} E\left[E\left\{W_i\left(Y_i\right)\Delta_i|D_i\right\}|\boldsymbol{Z}_i^D\left(0\right)\right] \\
&= \boldsymbol{Z}_i^D\left(0\right) E\left[E\left\{\frac{I\left(T_i \geq D_i \wedge L, C_i \geq D_i \wedge L\right)}{P\left(T_i \geq D_i \wedge L, C_i \geq D_i \wedge L\right)}\left(D_i \wedge L\right)|D_i\right\}|\boldsymbol{Z}_i^D\left(0\right)\right] \\
&- \boldsymbol{Z}_i^D\left(0\right) g^{-1}\left\{\boldsymbol{\beta}_D'\boldsymbol{Z}_i^D\left(0\right)\right\} E\left[E\left\{\frac{I\left(T_i \geq D_i \wedge L, C_i \geq D_i \wedge L\right)}{P\left(T_i \geq D_i \wedge L, C_i \geq D_i \wedge L\right)}|D_i\right\}|\boldsymbol{Z}_i^D\left(0\right)\right] \\
&= \boldsymbol{Z}_i^D\left(0\right) E\left\{D_i \wedge L|\boldsymbol{Z}_i^D\left(0\right)\right\} - \boldsymbol{Z}_i^D\left(0\right) g^{-1}\left\{\boldsymbol{\beta}_D'\boldsymbol{Z}_i^D\left(0\right)\right\} \\
&= \mathbf{0}.
\end{aligned}
$$

Then averaging over the baseline covariates, $E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta}_D)\}$ and therefore $E\{\boldsymbol{\Phi}_i^*(\boldsymbol{\beta}_D)\}$ will be $\mathbf{0}$. □

Theorem A.2. *Under regularity conditions (i)-(vii), as $n \to \infty$, $\sqrt{n}\boldsymbol{\Phi}(\boldsymbol{\beta}_D)$ converges to a zero-mean Normal with variance $\boldsymbol{B}(\boldsymbol{\beta}_D)$ defined in (A.10).*

Proof. As shown in Zhang and Schaubel (2011), the weight involved with dependent censoring time T_i can be written as

$$
\sqrt{n}\left\{\widehat{W}_i^T\left(t\right) - W_i^T\left(t\right)\right\} = \frac{1}{\sqrt{n}} W_i^T\left(t\right)\left\{\boldsymbol{D}_i^T\left(t\right)'\boldsymbol{\Omega}_T\left(\boldsymbol{\beta}_T\right)^{-1}\sum_{j=1}^{n}\boldsymbol{U}_j^T\left(\boldsymbol{\beta}_T\right) + \sum_{j=1}^{n} J_{ij}^T\left(t\right)\right\} + o_p\left(1\right),
$$

with defined $\boldsymbol{D}_i^T(t), \boldsymbol{U}_i^T(\boldsymbol{\beta}_T), \boldsymbol{\Omega}_T(\boldsymbol{\beta}), \boldsymbol{r}_T^{(k)}(t), \overline{\boldsymbol{z}}_T(t;\boldsymbol{\beta})$ in (A.19), (A.13), (A.4) , (A.1), (A.6) and

$$J_{ij}^T(t) = \int_0^t \exp\left\{\boldsymbol{\beta}_T' \boldsymbol{Z}_i^T(u)\right\} R_i(u)\, r_T^{(0)}(u;\boldsymbol{\beta}_T)^{-1}\, dM_j^T(u).$$

And we can derive the similar formula for independent censoring time C_i,

$$\sqrt{n}\left\{\widehat{W}_i^C(t) - W_i^C(t)\right\} = \frac{1}{\sqrt{n}} W_i^C(t)\left\{\boldsymbol{D}_i^C(t)'\, \boldsymbol{\Omega}_C(\boldsymbol{\beta}_C)^{-1} \sum_{j=1}^n \boldsymbol{U}_j^C(\boldsymbol{\beta}_C) + \sum_{j=1}^n J_{ij}^C(t)\right\} + o_p(1),$$

with defined $\boldsymbol{D}_i^C(t), \boldsymbol{U}_i^C(\boldsymbol{\beta}_C), \boldsymbol{\Omega}_C(\boldsymbol{\beta}), \boldsymbol{r}_C^{(k)}(t), \overline{\boldsymbol{z}}_C(t;\boldsymbol{\beta})$ in (A.20), (A.14), (A.5) , (A.2), (A.7)and

$$J_{ij}^C(t) = \int_0^t \exp\left(\boldsymbol{\beta}_C' \boldsymbol{Z}_i^C\right) R_i(u)\, r_C^{(0)}(u;\boldsymbol{\beta}_C)^{-1}\, dM_j^C(u).$$

Rewrite the target vector as

$$\begin{aligned}
\sqrt{n}\boldsymbol{\Phi}(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Delta_i(Y_i)\left[Y_i - g^{-1}\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D(0)\}\right] \boldsymbol{Z}_i^D(0)\, \widehat{W}_i^T(Y_i)\, \widehat{W}_i^C(Y_i) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Delta_i(Y_i)\left[Y_i - g^{-1}\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D(0)\}\right] \boldsymbol{Z}_i^D(0)\left[W_i^T(Y_i)\, W_i^C(Y_i)\right. && \text{(A.21)} \\
&\quad + W_i^C(Y_i)\left\{\widehat{W}_i^T(Y_i) - W_i^T(Y_i)\right\} && \text{(A.22)} \\
&\quad + W_i^T(Y_i)\left\{\widehat{W}_i^C(Y_i) - W_i^C(Y_i)\right\} && \text{(A.23)} \\
&\quad \left. + \left\{\widehat{W}_i^C(Y_i) - W_i^C(Y_i)\right\}\left\{\widehat{W}_i^T(Y_i) - W_i^T(Y_i)\right\}\right] && \text{(A.24)}
\end{aligned}$$

- The first part (A.21) is just

$$(A.21) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\epsilon}_i(\boldsymbol{\beta})$$

  where $\boldsymbol{\epsilon}_i(\boldsymbol{\beta})$ was defined in (A.12).

- The second part (A.22) involves the difference between estimated and true

IPCW weights for T:

$$(A.22) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i (\boldsymbol{\beta}) \left\{ \widehat{W}_i^T (Y_i) - W_i^T (Y_i) \right\}$$
$$= \frac{1}{n^{1.5}} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i (\boldsymbol{\beta}) \left\{ \boldsymbol{D}_i^T (Y_i)' \boldsymbol{\Omega}_T (\boldsymbol{\beta}_T)^{-1} \sum_{j=1}^{n} \boldsymbol{U}_j^T (\boldsymbol{\beta}_T) + \sum_{j=1}^{n} J_{ij}^T (Y_i) \right\} + \boldsymbol{o}_p (1)$$

(A.25)
$$= \frac{1}{n^{1.5}} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{\epsilon}_i (\boldsymbol{\beta}) \boldsymbol{D}_i^T (Y_i)' \boldsymbol{\Omega}_T (\boldsymbol{\beta}_T)^{-1} \boldsymbol{U}_j^T (\boldsymbol{\beta}_T)$$

(A.26)
$$+ \frac{1}{n^{1.5}} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{\epsilon}_i (\boldsymbol{\beta}) J_{ij}^T (Y_i) + \boldsymbol{o}_p (1)$$

Eq. (A.25) is simplified as

$$(A.25) = \frac{1}{n^{1.5}} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{\epsilon}_i (\boldsymbol{\beta}) \boldsymbol{D}_i^T (Y_i)' \boldsymbol{\Omega}_T (\boldsymbol{\beta}_T)^{-1} \boldsymbol{U}_j^T (\boldsymbol{\beta}_T)$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i (\boldsymbol{\beta}) \boldsymbol{D}_i^T (Y_i)' \right\} \boldsymbol{\Omega}_T (\boldsymbol{\beta}_T)^{-1} \boldsymbol{U}_j^T (\boldsymbol{\beta}_T)$$

where $\boldsymbol{K}_T(\boldsymbol{\beta}) = E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta})\boldsymbol{D}_i^T(Y_i)'\}$ was defined in (A.15), then

$$(A.25) = \frac{1}{\sqrt{n}} \boldsymbol{K}_T(\boldsymbol{\beta}) \boldsymbol{\Omega}_T (\boldsymbol{\beta}_T)^{-1} \sum_{j=1}^{n} \boldsymbol{U}_j^T (\boldsymbol{\beta}_T)$$

Since $J_{ij}^T (Y_i)$ can be written as

$$J_{ij}^T (Y_i) = \int_0^{Y_i} \exp \left\{ \boldsymbol{\beta}_T' \boldsymbol{Z}_i^T (u) \right\} R_i (u) r_T^{(0)} (u; \boldsymbol{\beta}_T)^{-1} dM_j^T (u)$$
$$= \int_0^{L} \exp \left\{ \boldsymbol{\beta}_T' \boldsymbol{Z}_i^T (u) \right\} I (Z_i \geq u) I (Z_i \wedge L \geq u) r_T^{(0)} (u; \boldsymbol{\beta}_T)^{-1} dM_j^T (u)$$
$$= \int_0^{L} \exp \left\{ \boldsymbol{\beta}_T' \boldsymbol{Z}_i^T (u) \right\} I (Z_i \geq u) r_T^{(0)} (u; \boldsymbol{\beta}_T)^{-1} dM_j^T (u)$$
$$\equiv \int_0^{L} \exp \left\{ \boldsymbol{\beta}_T' \boldsymbol{Z}_i^T (u) \right\} R_i (u) r_T^{(0)} (u; \boldsymbol{\beta}_T)^{-1} dM_j^T (u) ,$$

Eq. (A.26) is simplified as

$$(A.26) = \frac{1}{n^{1.5}} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{\epsilon}_i (\boldsymbol{\beta}) \left[\int_0^L \exp \left\{ \boldsymbol{\beta}_T' \boldsymbol{Z}_i^T (u) \right\} R_i (u) \, r_T^{(0)} (u; \boldsymbol{\beta}_T)^{-1} \, dM_j^T (u) \right]$$
$$= \frac{1}{\sqrt{n}} \int_0^L \left[\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i (\boldsymbol{\beta}) \exp \left\{ \boldsymbol{\beta}_T' \boldsymbol{Z}_i^T (u) \right\} R_i (u) \right] r_T^{(0)} (u; \boldsymbol{\beta}_T)^{-1} \left\{ d \sum_{j=1}^{n} M_j^T (u) \right\}.$$

where $\boldsymbol{H}_T(u; \boldsymbol{\beta}) = E[\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) \exp\{\boldsymbol{\beta}_T' \boldsymbol{Z}_i^T(u)\} R_i(u)]$ was defined in (A.17), then

$$(A.26) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^L \boldsymbol{H}_T (u; \boldsymbol{\beta}) \, r_T^{(0)} (u; \boldsymbol{\beta}_T)^{-1} \, dM_i^T (u).$$

To sum up, (A.22) can be rewritten as:

$$(A.22) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\boldsymbol{K}_T(\boldsymbol{\beta}) \boldsymbol{\Omega}_T(\boldsymbol{\beta}_T)^{-1} \boldsymbol{U}_i^T (\boldsymbol{\beta}_T) + \int_0^L \boldsymbol{H}_T(u; \boldsymbol{\beta}) r_T^{(0)}(u; \boldsymbol{\beta}_T)^{-1} dM_i^T (u) \right] + \boldsymbol{o}_p (1).$$

- Similarly, (A.23) can be rewritten as:

$$(A.23) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\boldsymbol{K}_C(\boldsymbol{\beta}) \boldsymbol{\Omega}_C(\boldsymbol{\beta}_C)^{-1} \boldsymbol{U}_i^C (\boldsymbol{\beta}_C) + \int_0^L \boldsymbol{H}_C(u; \boldsymbol{\beta}) r_C^{(0)}(u; \boldsymbol{\beta}_C)^{-1} dM_i^C (u) \right] + \boldsymbol{o}_p (1).$$

where $\boldsymbol{K}_C(\boldsymbol{\beta}) = E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) \boldsymbol{D}_i^C(Y_i)'\}$ and $\boldsymbol{H}_C(u; \boldsymbol{\beta}) = E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) \exp(\boldsymbol{\beta}_C' \boldsymbol{Z}_i^C) R_i(u)\}$ were defined in (A.18).

- Eq. (A.24) can be rewritten as:

(A.24)

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta_i \left(Y_i\right) \left[Y_i - g^{-1}\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D \left(0\right)\}\right] \boldsymbol{Z}_i^D \left(0\right)$$

$$* \left\{\widehat{W}_i^C \left(Y_i\right) - W_i^C \left(Y_i\right)\right\} \left\{\widehat{W}_i^T \left(Y_i\right) - W_i^T \left(Y_i\right)\right\}$$

$$= \frac{1}{n^{2.5}} \sum_{i=1}^{n} \Delta_i \left(Y_i\right) \left[Y_i - g^{-1}\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D \left(0\right)\}\right] \boldsymbol{Z}_i^D \left(0\right)$$

$$* W_i^T \left(Y_i\right) \left\{\boldsymbol{D}_i^T \left(Y_i\right)' \boldsymbol{\Omega}_T \left(\boldsymbol{\beta}_T\right)^{-1} \sum_{j=1}^{n} \boldsymbol{U}_j^T \left(\boldsymbol{\beta}_T\right) + \sum_{j=1}^{n} J_{ij}^T \left(Y_i\right) + \boldsymbol{o}_p \left(\sqrt{n}\right)\right\}$$

$$* W_i^C \left(Y_i\right) \left\{\boldsymbol{D}_i^C \left(Y_i\right)' \boldsymbol{\Omega}_C \left(\boldsymbol{\beta}_C\right)^{-1} \sum_{k=1}^{n} \boldsymbol{U}_k^C \left(\boldsymbol{\beta}_C\right) + \sum_{k=1}^{n} J_{ik}^C \left(Y_i\right) + \boldsymbol{o}_p \left(\sqrt{n}\right)\right\}$$

$$= \frac{1}{n^{2.5}} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i \left(\boldsymbol{\beta}\right) \left\{\boldsymbol{D}_i^T \left(Y_i\right)' \boldsymbol{\Omega}_T \left(\boldsymbol{\beta}_T\right)^{-1} \sum_{j=1}^{n} \boldsymbol{U}_j^T \left(\boldsymbol{\beta}_T\right) + \sum_{j=1}^{n} J_{ij}^T \left(Y_i\right) + \boldsymbol{o}_p \left(\sqrt{n}\right)\right\}$$

$$* \left\{\boldsymbol{D}_i^C \left(Y_i\right)' \boldsymbol{\Omega}_C \left(\boldsymbol{\beta}_C\right)^{-1} \sum_{k=1}^{n} \boldsymbol{U}_k^C \left(\boldsymbol{\beta}_C\right) + \sum_{k=1}^{n} J_{ik}^C \left(Y_i\right) + \boldsymbol{o}_p \left(\sqrt{n}\right)\right\}$$

$$= \frac{1}{n^{2.5}} \sum_{i=1}^{n} \boldsymbol{\epsilon}_i \left(\boldsymbol{\beta}\right) \left\{\boldsymbol{D}_i^T \left(Y_i\right)' \boldsymbol{\Omega}_T \left(\boldsymbol{\beta}_T\right)^{-1} \sum_{j=1}^{n} \boldsymbol{U}_j^T \left(\boldsymbol{\beta}_T\right) + \sum_{j=1}^{n} J_{ij}^T \left(Y_i\right)\right\}$$

$$* \left\{\boldsymbol{D}_i^C \left(Y_i\right)' \boldsymbol{\Omega}_C \left(\boldsymbol{\beta}_C\right)^{-1} \sum_{k=1}^{n} \boldsymbol{U}_k^C \left(\boldsymbol{\beta}_C\right) + \sum_{k=1}^{n} J_{ik}^C \left(Y_i\right)\right\} + \boldsymbol{o}_p \left(1\right)$$

(A.27)

$$= \frac{1}{n^{2.5}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \boldsymbol{\epsilon}_i \left(\boldsymbol{\beta}\right) \boldsymbol{D}_i^T \left(Y_i\right)' \boldsymbol{\Omega}_T \left(\boldsymbol{\beta}_T\right)^{-1} \boldsymbol{U}_j^T \left(\boldsymbol{\beta}_T\right) \boldsymbol{D}_i^C \left(Y_i\right)' \boldsymbol{\Omega}_C \left(\boldsymbol{\beta}_C\right)^{-1} \boldsymbol{U}_k^C \left(\boldsymbol{\beta}_C\right)$$

(A.28)

$$+ \frac{1}{n^{2.5}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \boldsymbol{\epsilon}_i \left(\boldsymbol{\beta}\right) J_{ij}^T \left(Y_i\right) \boldsymbol{D}_i^C \left(Y_i\right)' \boldsymbol{\Omega}_C \left(\boldsymbol{\beta}_C\right)^{-1} \boldsymbol{U}_k^C \left(\boldsymbol{\beta}_C\right)$$

(A.29)

$$+ \frac{1}{n^{2.5}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \boldsymbol{\epsilon}_i \left(\boldsymbol{\beta}\right) \boldsymbol{D}_i^T \left(Y_i\right)' \boldsymbol{\Omega}_T \left(\boldsymbol{\beta}_T\right)^{-1} \boldsymbol{U}_j^T \left(\boldsymbol{\beta}_T\right) J_{ik}^C \left(\boldsymbol{\beta}_C\right)$$

(A.30)

$$+ \frac{1}{n^{2.5}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \boldsymbol{\epsilon}_i \left(\boldsymbol{\beta}\right) J_{ij}^T \left(Y_i\right) J_{ik}^C \left(Y_i\right) + \boldsymbol{o}_p \left(1\right)$$

Eq. (A.27)-(A.30) can be shown to be negligible.

To sum up, we can rewrite $\sqrt{n}\boldsymbol{\Phi}(\boldsymbol{\beta})$ as

$$n\boldsymbol{\Phi}\left(\boldsymbol{\beta}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{B}_i\left(\boldsymbol{\beta}\right) + \boldsymbol{o}_p\left(1\right),$$

where as defined in (A.11), Since we have defined $\boldsymbol{B}(\boldsymbol{\beta}) = E\{\boldsymbol{B}_i(\boldsymbol{\beta})^{\otimes 2}\}$ in (A.10), then we have proven that

$$\sqrt{n}\boldsymbol{\Phi}\left(\boldsymbol{\beta}_D\right) \xrightarrow{D} \text{Normal}\left(\mathbf{0}, \boldsymbol{B}\left(\boldsymbol{\beta}_D\right)\right),$$

following that the mean of each term in the summation above is $\mathbf{0}$ at $\boldsymbol{\beta}_D$.

□

A.1.6 Consistency

Theorem A.3. *Under regularity conditions (i)-(vii), as $n \to \infty$, $\widehat{\boldsymbol{\beta}}_D \xrightarrow{p} \boldsymbol{\beta}_D$.*

Proof. We use the Inverse Function Theorem (Foutz, 1977) by verifying the following conditions:

- $\partial\boldsymbol{\Phi}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$ exists and is continuous in an open neighborhood $\mathcal{B}_D$ of $\boldsymbol{\beta}_D$.
- $-n^{-1}\partial\boldsymbol{\Phi}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'|_{\boldsymbol{\beta}=\boldsymbol{\beta}_D}$ is positive definite with probability 1 as $n \to \infty$.
- $-n^{-1}\partial\boldsymbol{\Phi}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$ converges in probability to a fixed function uniformly in an open neighborhood $\mathcal{B}_D$ of $\boldsymbol{\beta}_D$.
- Asymptotic unbiasedness of the estimating function:$-\boldsymbol{\Phi}(\boldsymbol{\beta}_D)/n \xrightarrow{p} \mathbf{0}$.

We know that

$$\frac{\partial\boldsymbol{\Phi}\left(\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}'} = -\sum_{i=1}^{n} \Delta_i \widehat{W}_i^T\left(Y_i\right) \widehat{W}_i^C\left(Y_i\right) h\left\{\boldsymbol{\beta}'\boldsymbol{Z}_i^D\left(0\right)\right\} \boldsymbol{Z}_i^D\left(0\right)^{\otimes 2}.$$

where $h(x) = \partial g^{-1}(x)/\partial x$. We will show that this derivative vector satisfies all the necessary conditions above.

- The first condition here holds because of the regularity condition (iv), which states that h exists and is continuous in an open neighborhood $\mathcal{B}_D$ of $\boldsymbol{\beta}_D$.

- As to the second condition here, we know

$$
\begin{aligned}
& -\frac{1}{n}\frac{\partial \boldsymbol{\Phi}\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_D} \\
& = E\left[\Delta_i W_i^T\left(Y_i\right) W_i^C\left(Y_i\right) h\left\{\boldsymbol{\beta}_D' \boldsymbol{Z}_i^D(0)\right\} \boldsymbol{Z}_i^D(0)^{\otimes 2}\right]+\boldsymbol{o}_p\left(1\right) \\
& = E\left[E\left\{\frac{I\left(T_i \wedge C_i \geq D_i \wedge L\right)}{P\left(T_i \geq D_i \wedge L\right) P\left(C_i \geq D_i \wedge L\right)}|D_i, \boldsymbol{Z}_i^D\left(0\right)\right\} h\left\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D\left(0\right)\right\} \boldsymbol{Z}_i^D\left(0\right)^{\otimes 2}\right]+\boldsymbol{o}_p\left(1\right) \\
& = E\left[E\left\{\frac{I\left(T_i \geq D_i \wedge L\right) I\left(C_i \geq D_i \wedge L\right)}{P\left(T_i \geq D_i \wedge L\right) P\left(C_i \geq D_i \wedge L\right)}|D_i, \boldsymbol{Z}_i^D\left(0\right)\right\} h\left\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D\left(0\right)\right\} \boldsymbol{Z}_i^D\left(0\right)^{\otimes 2}\right]+\boldsymbol{o}_p\left(1\right) \\
& = E\left[h\left\{\boldsymbol{\beta}' \boldsymbol{Z}_i^D\left(0\right)\right\} \boldsymbol{Z}_i^D\left(0\right)^{\otimes 2}\right]+\boldsymbol{o}_p\left(1\right) \\
& \equiv \boldsymbol{A}\left(\boldsymbol{\beta}\right).
\end{aligned}
$$

  where $\boldsymbol{A}(\boldsymbol{\beta})$ is defined as (A.3). Since we have assumed $\boldsymbol{A}(\boldsymbol{\beta}_D)$ is positive definite, the second condition holds here too.

- The third condition holds by the law of large numbers.

- Finally, since we have proven that

$$\sqrt{n}\boldsymbol{\Phi}(\boldsymbol{\beta}_D) \xrightarrow{D} \text{Normal}\left(0, \boldsymbol{B}(\boldsymbol{\beta}_D)\right).$$

  The last condition holds by Chebyshev's inequality.

Having verified all the four conditions, we can argue that $\widehat{\boldsymbol{\beta}}_D \xrightarrow{p} \boldsymbol{\beta}_D$ follows from Inverse Function Theorem. □

A.1.7 Asymptotic Distribution

Theorem A.4. *Under regularity conditions (i)-(vii), as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_D\right) \xrightarrow{D} \textit{Normal}\left(\mathbf{0}, \boldsymbol{A}(\boldsymbol{\beta}_D)^{-1}\boldsymbol{B}(\boldsymbol{\beta}_D)\boldsymbol{A}(\boldsymbol{\beta}_D)^{-1}\right).$$

Proof. Taylor expansion of $\boldsymbol{\Phi}\left(\widehat{\boldsymbol{\beta}}_D\right)$ around $\boldsymbol{\beta}_D$ is:

$$\mathbf{0} = \boldsymbol{\Phi}\left(\widehat{\boldsymbol{\beta}}_D\right) = \boldsymbol{\Phi}\left(\boldsymbol{\beta}_D\right) + \frac{\partial \boldsymbol{\Phi}\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}\left(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_D\right),$$

where $\widetilde{\boldsymbol{\beta}}$ lies between $\widehat{\boldsymbol{\beta}}_D$ and $\boldsymbol{\beta}_D$. So

$$\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_D\right) &= -\left\{\frac{\partial \boldsymbol{\Phi}\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}\right\}^{-1}\sqrt{n}\boldsymbol{\Phi}\left(\boldsymbol{\beta}_D\right) \\
&= \left[-\frac{1}{n}\sum_{i=1}^{n}\Delta_i \widehat{W}_i\left(Y_i\right)\boldsymbol{Z}_i^D\left(0\right)^{\otimes 2} h\left\{\widetilde{\boldsymbol{\beta}}'\boldsymbol{Z}_i^D\left(0\right)\right\}\right]^{-1}\sqrt{n}\boldsymbol{\Phi}\left(\boldsymbol{\beta}_D\right) \\
&= \boldsymbol{A}\left(\boldsymbol{\beta}_D\right)^{-1}\sqrt{n}\boldsymbol{\Phi}\left(\boldsymbol{\beta}_D\right) + \boldsymbol{o}_p\left(1\right).
\end{aligned}$$

Following Theorem A.2, it holds that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_D\right) \xrightarrow{D} \text{Normal}\left(\mathbf{0}, \boldsymbol{A}\left(\boldsymbol{\beta}_D\right)^{-1}\boldsymbol{B}\left(\boldsymbol{\beta}_D\right)\boldsymbol{A}\left(\boldsymbol{\beta}_D\right)^{-1}\right).$$

□

A.2 Model Selection Criteria

We suggest using Concordance Statistics (IOC), Mean Absolute Deviation (MAD) and Mean Squared Deviation (MSD) to select the proper link function. To simplify the notation, we denote $D_i^L = D_i \wedge L$ and its predicted value as $\widehat{D_i^L} = g^{-1}\{\boldsymbol{\beta}'_D\boldsymbol{Z}_i^D(0)\}$. Due to the occurrence of censoring, we observe $Z_i = D_i^L \wedge T_i \wedge C_i$ for subject i.

Our version of IOC is adapted from Frank Harrell's formula of concordance (Harrell, 1996; Heagerty, 2005, Uno et al., 2011):

$$\text{IOC} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i \widehat{W}_i\left(Y_i\right)\widehat{W}_j\left(Y_i\right) I\left(Y_i < Y_j, \widehat{D_i^L} < \widehat{D_j^L}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i \widehat{W}_i\left(Y_i\right)\widehat{W}_j\left(Y_i\right) I\left(Y_i < Y_j\right)}.$$

It converges to a censoring distribution free quantity $P(\widehat{D_i^L} < \widehat{D_j^L} | D_i^L < D_j^L)$ because

(i) as to the numerator,

$$
\begin{aligned}
&\frac{1}{n^2}\Delta_i \widehat{W}_i\left(Y_i\right)\widehat{W}_j\left(Y_i\right) I\left(Y_i < Y_j, \widehat{D_i^L} < \widehat{D_j^L}\right) \\
&\xrightarrow{p} E\left\{I\left(D_i^L \le T_i \wedge C_i\right) W_i\left(Y_i\right) W_j\left(Y_i\right) I\left(Y_i < Y_j\right) I\left(\widehat{D_i^L} < \widehat{D_j^L}\right)\right\} \\
&= E\left\{I\left(D_i^L \le T_i \wedge C_i\right) W_i\left(D_i^L\right) W_j\left(D_i^L\right) I\left(T_j \wedge C_j > D_i^L\right) I\left(D_j^L > D_i^L\right)\right. \\
&\left.* \; I\left(\widehat{D_i^L} < \widehat{D_j^L}\right)\right\} \\
&= E\left[E\left\{\frac{I\left(T_i \wedge C_i \ge D_i \wedge L\right) I\left(T_j \wedge C_j > D_i \wedge L\right)}{P\left(T_i \wedge C_i > D_i \wedge L\right) P\left(T_j \wedge C_j > D_i \wedge L\right)}\right.\right. \\
&\left.\left.* \; I\left(D_i^L < D_j^L, \widehat{D_i^L} < \widehat{D_j^L}\right) | \boldsymbol{Z}_i^D(0), D_i\right\}\right] \\
&\xrightarrow{p} P\left(D_i^L < D_j^L, \widehat{D_i^L} < \widehat{D_j^L}\right).
\end{aligned}
$$

(ii) Similarly, the denominator follows that

$$\frac{1}{n^2}\Delta_i \widehat{W}_i(Y_i)\widehat{W}_j(Y_i) I(Y_i < Y_j) \xrightarrow{p} P(D_i^L < D_j^L).$$

(iii) So

$$\text{IOC} \xrightarrow{p} P(\widehat{D_i^L} < \widehat{D_j^L} | D_i^L < D_j^L).$$

We can also use the similar trick to prove that

$$
\begin{aligned}
\text{MAD} &:= \frac{1}{n}\sum_{i=1}^{n} \Delta_i \widehat{W}_i\left(Y_i\right)\left|Y_i - g^{-1}\left\{\widehat{\boldsymbol{\beta}}_D' \boldsymbol{Z}_i^D\left(0\right)\right\}\right| \xrightarrow{p} E\left|D_i^L - \widehat{D_i^L}\right|, \\
\text{MSD} &:= \frac{1}{n}\sum_{i=1}^{n} \Delta_i \widehat{W}_i\left(Y_i\right)\left[Y_i - g^{-1}\left\{\widehat{\boldsymbol{\beta}}_D' \boldsymbol{Z}_i^D\left(0\right)\right\}\right]^2 \xrightarrow{p} E\left[D_i^L - \widehat{D_i^L}\right]^2.
\end{aligned}
$$

A.3 More Results in Application Data Analysis

Below are the plots of RMST within 1 year and 5 years post wait-list for chronic ESLD patients with different MELD scores.

Below are the estimated effects of prognostic factors on pre-transplant survival time within 1 year and 5 year post wait-list for chronic ESLD patients.

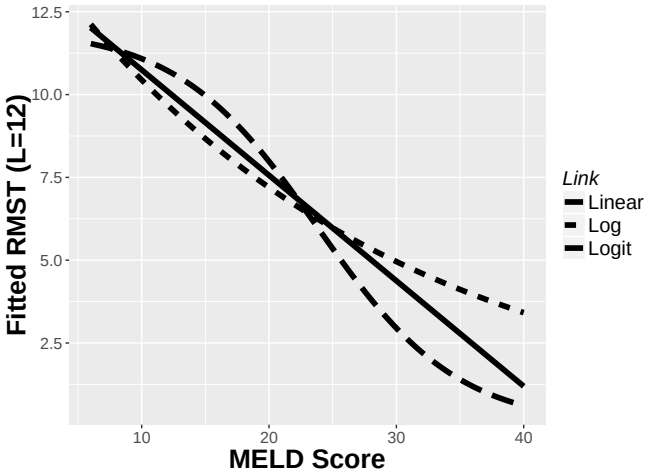


Figure A.1 Fitted RMST ($L = 12$ months) by MELD score for a reference patient: white, male, age=50, Region=5, year=2005, not hospitalized, not on dialysis, blood Type=O, BMI $\in (20, 25]$, sodium=130

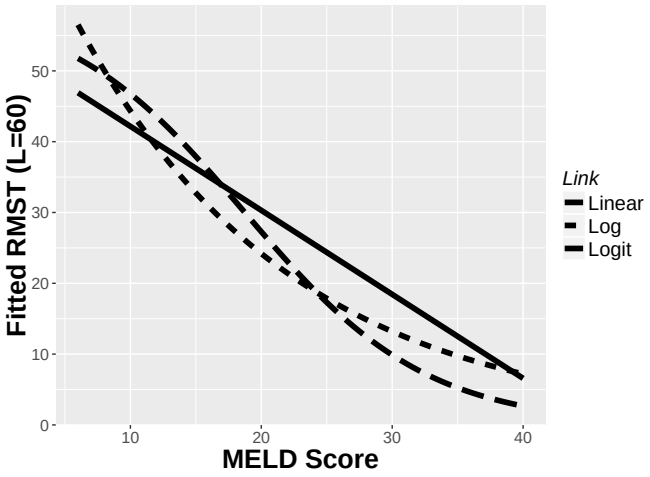


Figure A.2 Fitted RMST ($L = 60$ months) by MELD score for a reference patient: white, male, age=50, Region=5, year=2005, not hospitalized, not on dialysis, blood Type=O, BMI $\in (20, 25]$, sodium=130

Table A.1 Estimated covariate effects on RMST in the absence of liver transplantation ($L = 12$ months)

	Linear			Log			Logistic		
$\boldsymbol{Z}_i^D(0)$	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p
Intercept	12.02	0.14	< 0.01	2.49	0.01	0.44	3.22	0.1	< 0.01
Year-2005	0.1	0.01	< 0.01	0.01	< 0.01	< 0.01	0.06	0.01	< 0.01
Age-50 (Years)	-0.05	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	-0.04	< 0.01	< 0.01
Sodium-130 (mmol/l)	0.13	0.01	< 0.01	0.01	< 0.01	< 0.01	0.08	< 0.01	< 0.01
MELD Score-6	-0.32	< 0.01	< 0.01	-0.04	< 0.01	< 0.01	-0.18	< 0.01	< 0.01
UNOS Region				*Reference Group: 5*					
1	-0.32	0.12	0.01	-0.02	0.01	0.02	-0.31	0.09	< 0.01
2	-0.52	0.09	< 0.01	-0.04	0.01	< 0.01	-0.41	0.06	< 0.01
3	-0.63	0.12	< 0.01	-0.03	0.01	< 0.01	-0.54	0.08	< 0.01
4	-0.08	0.08	0.29	< 0.01	0.01	0.75	-0.09	0.07	0.18
6	-0.01	0.14	0.93	0.03	0.01	0.02	-0.2	0.1	0.05
7	0.06	0.11	0.59	0.02	0.01	0.04	0.03	0.08	0.69
8	-0.15	0.1	0.15	0.01	0.01	0.49	-0.17	0.08	0.04
9	-0.45	0.09	< 0.01	-0.04	0.01	< 0.01	-0.27	0.08	< 0.01
10	-0.51	0.12	< 0.01	-0.04	0.01	< 0.01	-0.47	0.09	< 0.01
11	-0.79	0.11	< 0.01	-0.06	0.01	< 0.01	-0.59	0.08	< 0.01
Gender				*Reference Group: Male*					
Female	0.01	0.05	0.88	-0.01	< 0.01	0.03	0.03	0.04	0.47
Race				*Reference Group: White*					
Black	0.19	0.11	0.07	0.02	0.01	0.02	0.05	0.07	0.5
Hispanic	-0.02	0.07	0.76	< 0.01	0.01	0.44	-0.05	0.05	0.34
Asian	0.2	0.12	0.1	0.02	0.01	0.11	0.21	0.11	0.05
Others	-0.29	0.23	0.22	-0.02	0.02	0.28	-0.1	0.18	0.57
Blood Type				*Reference Group: O*					
A	-0.05	0.05	0.3	< 0.01	< 0.01	0.44	< 0.01	0.04	0.96
B	-0.05	0.09	0.57	-0.01	0.01	0.46	0.01	0.06	0.89
AB	-0.28	0.21	0.19	-0.03	0.01	0.04	-0.23	0.15	0.13
Diagnosis				*Reference Group: No or Yes*					
Hepatitis C	-0.09	0.09	0.32	< 0.01	0.01	0.83	-0.16	0.06	0.01
Noncholestatic	0.29	0.09	< 0.01	0.04	0.01	< 0.01	0.13	0.06	0.04
Cholestatic	-0.05	0.12	0.68	-0.01	0.01	0.43	-0.11	0.09	0.25
Acute Hepatic Necrosis	0.9	0.22	< 0.01	0.05	0.02	0.01	0.86	0.17	< 0.01
Metastatic Disease	-0.45	0.19	0.02	-0.04	0.02	0.06	-0.27	0.13	0.04
Malignant Neoplasm	-1.64	0.1	< 0.01	-0.19	0.01	< 0.01	-1.02	0.07	< 0.01
BMI				*Reference Group:* $(20, 25]$					
$(0, 20]$	-0.46	0.12	< 0.01	-0.04	0.01	< 0.01	-0.24	0.08	< 0.01
$(25, 30]$	0.08	0.07	0.21	0.01	0.01	0.02	0.03	0.05	0.48
> 30	0.08	0.07	0.23	0.01	0.01	0.06	0.04	0.05	0.47
Hospitalized				*Reference Group: Not Hospitalized*					
ICU	-1.63	0.18	< 0.01	-0.57	0.05	< 0.01	-0.98	0.1	< 0.01
not ICU	-1.43	0.13	< 0.01	-0.3	0.02	< 0.01	-0.56	0.06	< 0.01
Dialysis				*Reference Group: No or Yes*					
Yes	0.81	0.15	< 0.01	0.11	0.02	< 0.01	0.56	0.08	< 0.01

An offset of $L = 12$ months is applied for log link.

Table A.2 Estimated covariate effects on RMST in the absence of liver transplantation ($L = 60$ months)

	Linear			Log			Logistic		
$\boldsymbol{Z}_i^D(0)$	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p	$\widehat{\boldsymbol{\beta}}_D$	ASE_1	p
Intercept	46.91	0.88	< 0.01	4.03	0.03	0.06	1.84	0.09	< 0.01
Year-2005	-3.2	0.07	< 0.01	-0.13	< 0.01	< 0.01	-0.32	0.01	< 0.01
Age-50 (Years)	-0.3	0.02	< 0.01	-0.01	< 0.01	< 0.01	-0.03	< 0.01	< 0.01
Sodium-130 (mmol/l)	0.66	0.04	< 0.01	0.03	< 0.01	< 0.01	0.07	< 0.01	< 0.01
MELD Score-6	-1.19	0.02	< 0.01	-0.06	< 0.01	< 0.01	-0.14	< 0.01	< 0.01
UNOS Region	*Reference Group: 5*								
1	-2.37	0.78	< 0.01	-0.07	0.03	0.02	-0.15	0.08	0.07
2	-1.86	0.64	< 0.01	-0.06	0.02	0.01	-0.14	0.06	0.03
3	-4.36	0.76	< 0.01	-0.13	0.03	< 0.01	-0.34	0.08	< 0.01
4	0	0.63	1	0.02	0.02	0.3	0.04	0.06	0.57
6	0.66	0.94	0.48	0.03	0.03	0.28	0.1	0.1	0.34
7	-0.55	0.68	0.41	-0.02	0.03	0.34	-0.01	0.07	0.86
8	-1.59	0.83	0.05	-0.01	0.03	0.82	-0.07	0.08	0.4
9	-1.05	0.65	0.1	-0.06	0.02	< 0.01	-0.1	0.07	0.15
10	-3.94	0.83	< 0.01	-0.14	0.03	< 0.01	-0.37	0.08	< 0.01
11	-3.46	0.77	< 0.01	-0.13	0.03	< 0.01	-0.31	0.08	< 0.01
Gender	*Reference Group: Male*								
Female	0.54	0.36	0.14	< 0.01	0.01	0.81	0.03	0.04	0.38
Race	*Reference Group: White*								
Black	-0.99	0.61	0.1	-0.04	0.03	0.16	-0.06	0.07	0.37
Hispanic	0.5	0.49	0.3	0.01	0.02	0.52	0.04	0.05	0.39
Asian	1.63	0.87	0.06	0.02	0.03	0.55	0.14	0.1	0.17
Others	-0.86	1.56	0.58	-0.01	0.06	0.89	-0.02	0.16	0.89
Blood Type	*Reference Group: O*								
A	0.06	0.36	0.86	0.01	0.01	0.61	0.03	0.04	0.47
B	-0.51	0.59	0.38	-0.01	0.02	0.44	-0.04	0.06	0.54
AB	0.55	1.07	0.61	0.01	0.04	0.8	0.03	0.13	0.81
Diagnosis	*Reference Group: No or Yes*								
Hepatitis C	-1.54	0.54	< 0.01	-0.05	0.02	0.01	-0.21	0.06	< 0.01
Noncholestatic	2.34	0.55	< 0.01	0.1	0.02	< 0.01	0.23	0.06	< 0.01
Cholestatic	-0.98	0.77	0.2	-0.04	0.03	0.08	-0.15	0.08	0.07
Acute Hepatic Necrosis	2.4	1.5	0.11	0.01	0.04	0.84	0.1	0.15	0.51
Metastatic Disease	-2.58	1.41	0.07	-0.07	0.06	0.25	-0.26	0.17	0.12
Malignant Neoplasm	-8.59	0.69	< 0.01	-0.44	0.04	< 0.01	-0.99	0.08	< 0.01
BMI	*Reference Group:* $(20, 25]$								
$(0, 20]$	-2.64	0.86	< 0.01	-0.09	0.03	< 0.01	-0.22	0.09	0.02
$(25, 30]$	-0.15	0.46	0.75	< 0.01	0.02	0.8	-0.03	0.05	0.59
> 30	-0.12	0.47	0.8	< 0.01	0.02	0.88	-0.05	0.05	0.33
Hospitalized	*Reference Group: Not Hospitalized*								
ICU	-1.87	0.75	0.01	-0.63	0.11	< 0.01	-0.83	0.16	< 0.01
not ICU	-1.76	0.59	< 0.01	-0.27	0.05	< 0.01	-0.36	0.09	< 0.01
Dialysis	*Reference Group: No or Yes*								
Yes	2.01	0.77	0.01	0.13	0.05	0.01	0.47	0.11	< 0.01

An offset of $L = 60$ months is applied for log link.

APPENDIX B

Appendix for Chapter II

B.1 Notations and Assumptions from Manuscript

- The true underlying mortality model for RMST at L is, for $i = 1, \ldots, n, j = 1, \ldots, J$,

$$\mu_{ij} = E\left(D_i \wedge L | \boldsymbol{Z}_i, g_i = j\right) = \mu_{0j} \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i\right).$$

- The censoring model follows proportional hazards assumption:

$$\lambda_{ij}^C(t) = \lambda_{0j}^C(t) \exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i\right).$$

- The working estimating equations are:

$$\text{(B.1)} \boldsymbol{\phi}_1\left(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}}\right) = \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_i \Delta_i^Y \left\{Y_i - \mu_{0j} \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i\right)\right\} \boldsymbol{Z}_i = \mathbf{0},$$

$$\text{(B.2)} \boldsymbol{\phi}_2\left(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}}\right) = \begin{pmatrix} \sum_{i=1}^{n} G_{i1} \widehat{W}_i \Delta_i^Y \left\{Y_i - \mu_{0j} \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i\right)\right\} \\ \vdots \\ \sum_{i=1}^{n} G_{iJ} \widehat{W}_i \Delta_i^Y \left\{Y_i - \mu_{0j} \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i\right)\right\} \end{pmatrix} = \mathbf{0}.$$

- With defined notations,

$$\text{(B.3)} \quad \boldsymbol{S}_j^{(k)}\left(\boldsymbol{\beta}, \boldsymbol{W}\right) = \sum_{i=1}^{n} G_{ij} W_i \Delta_i^Y \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i\right) \boldsymbol{Z}_i^{\otimes k},$$

$$\text{(B.4)} \quad \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \boldsymbol{W}\right) = \frac{\boldsymbol{S}_j^{(1)}\left(\boldsymbol{\beta}, \boldsymbol{W}\right)}{S_j^{(0)}\left(\boldsymbol{\beta}, \boldsymbol{W}\right)},$$

(2.4)-(2.5) are equivalent to

$$\text{(B.5)} \qquad \boldsymbol{\psi}_1\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right) = \sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}\left\{\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)\right\}\widehat{W}_i\Delta_i^Y Y_i = \mathbf{0},$$

$$\text{(B.6)} \quad \boldsymbol{\psi}_2\left(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}}\right) = \begin{pmatrix} \frac{\sum_{i=1}^n G_{i1}\widehat{W}_i\Delta_i^Y Y_i}{\sum_{i=1}^n G_{i1}\widehat{W}_i\Delta_i^Y \exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)} - \mu_{01} \\ \vdots \\ \frac{\sum_{i=1}^n G_{iJ}\widehat{W}_i\Delta_i^Y Y_i}{\sum_{i=1}^n G_{iJ}\widehat{W}_i\Delta_i^Y \exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)} - \mu_{0J} \end{pmatrix} = \mathbf{0}.$$

- Regularity conditions:

  (a) $\mathcal{O}_i$'s are independently and identically distributed.

  (b) $P(R_i(t) = 1) > 0$ for $t \in (0, \tau]$.

  (c) $|Z_{ik}| < M_Z < \infty$, where Z_{ik} is the kth component of $\boldsymbol{Z}_i$.

  (d) $\Lambda_{0j}^C(\tau) < \infty$ and $\Lambda_{0j}^C(t)$ is absolutely continuous for $t \in (0, \tau]$.

  (e) There exist neighborhoods $\mathcal{C}$ of $\boldsymbol{\theta}$ such that for $k = 0, 1, 2, j = 1, \ldots, J$,

$$\sup_{t\in(0,\tau],\boldsymbol{\theta}\in\mathcal{C}} \left\| \boldsymbol{R}_j^{(k)}(t;\boldsymbol{\theta}) - \boldsymbol{r}_j^{(k)}(t;\boldsymbol{\theta}) \right\| \xrightarrow{p} 0,$$

  where

$$\text{(B.7)} \qquad \boldsymbol{R}_j^{(k)}(t;\boldsymbol{\theta}) = \frac{\sum_{i=1}^n G_{ij}\exp\left(\boldsymbol{\theta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}R_i(t)}{\sum_{i=1}^n G_{ij}},$$

$$\text{(B.8)} \qquad \boldsymbol{r}_j^{(k)}(t;\boldsymbol{\theta}) = E\left\{G_{ij}\exp\left(\boldsymbol{\theta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}R_i(t)\right\}.$$

  (f) There exist neighborhoods $\mathcal{B}$ of $\boldsymbol{\beta}$ such that for $k = 0, 1, 2, j = 1, \ldots, J$,

$$\sup_{t\in(0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \left\| \boldsymbol{S}_j^{(k)}\left(\boldsymbol{\beta}, \boldsymbol{W}\right) - \boldsymbol{s}_j^{(k)}(\boldsymbol{\beta}) \right\| \xrightarrow{p} 0,$$

  where

$$\boldsymbol{s}_j^{(k)}(\boldsymbol{\beta}) = E\left\{G_{ij}\Delta_i^Y W_i \exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}\right\} = E\left\{G_{ij}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\boldsymbol{Z}_i^{\otimes k}\right\},$$

(g) The matrices $\boldsymbol{A}(\boldsymbol{\beta}), \boldsymbol{\Theta}(\boldsymbol{\theta})$ are each positive definite, where

$$\begin{aligned}
(\text{B.4}) \quad \boldsymbol{A}(\boldsymbol{\beta}) &= \sum_{j=1}^{J} E\left\{ G_{ij} \Delta_i^Y W_i Y_i \left(\frac{\boldsymbol{s}_j^{(2)}(\boldsymbol{\beta})}{s_j^{(0)}(\boldsymbol{\beta})} - \overline{\boldsymbol{s}}_j(\boldsymbol{\beta})^{\otimes 2} \right) \right\}, \\
(\text{B.5}) \quad \boldsymbol{\Theta}(\boldsymbol{\theta}) &= \sum_{j=1}^{J} E\left\{ G_{ij} \int_0^{\tau} \left(\frac{\boldsymbol{r}_j^{(2)}(t;\boldsymbol{\theta})}{r_j^{(0)}(t;\boldsymbol{\theta})} - \overline{\boldsymbol{r}}_j(t;\boldsymbol{\theta})^{\otimes 2} \right) \boldsymbol{r}_j^{(0)}(t;\boldsymbol{\theta}) \lambda_{0j}^C(t) dt \right\}.
\end{aligned}$$

and

$$\begin{aligned}
\overline{\boldsymbol{s}}_j(\boldsymbol{\beta}) &= s_j^{(0)}(\boldsymbol{\beta})^{-1} \boldsymbol{s}_j^{(1)}(\boldsymbol{\beta}), \\
\overline{\boldsymbol{r}}_j(t;\boldsymbol{\theta}) &= r_j^{(0)}(t;\boldsymbol{\theta})^{-1} \boldsymbol{r}_j^{(1)}(t;\boldsymbol{\theta}).
\end{aligned}$$

B.2 Unbiased Estimating Equation

B.2.1 Equivalence of ϕ and ψ

Since

$$\boldsymbol{\phi}_2\left(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}}\right) = \mathbf{0} \Rightarrow \mu_{0j} = \frac{\sum_{i=1}^n G_{ij} \widehat{W}_i \Delta_i^Y Y_i}{\sum_{i=1}^n G_{ij} \widehat{W}_i \Delta_i^Y \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i\right)}, j = 1, \ldots, J,$$

which is the same as $\boldsymbol{\psi}_2(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}})$. Plug this into $\boldsymbol{\phi}_1(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}})$:

$$\begin{aligned}
& \boldsymbol{\phi}_1\left(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}}\right) \\
= & \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_i \Delta_i^Y Y_i \boldsymbol{Z}_i - \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_i \Delta_i^Y \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_i\right) \boldsymbol{Z}_i \frac{\sum_{k=1}^n G_{kj} \widehat{W}_k \Delta_{Yk} Y_k}{\sum_{k=1}^n G_{kj} \widehat{W}_k \Delta_{Yk} \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_k\right)} \\
= & \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_i \Delta_i^Y Y_i \boldsymbol{S}_i - \sum_{j=1}^{J} \sum_{k=1}^{n} G_{kj} \widehat{W}_k \Delta_{Yk} Y_k \frac{\boldsymbol{S}_j^{(1)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)}{S_j^{(0)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)} \\
= & \sum_{j=1}^{J} \sum_{i=1}^{n} G_{ij} \widehat{W}_i \Delta_i^Y Y_i \left\{ \boldsymbol{S}_i - \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right) \right\} \\
= & \boldsymbol{\psi}_1\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)
\end{aligned}$$

B.2.2 Unbiased Estimating Equation

The unbiasedness of the estimating equations $\boldsymbol{\phi}_1(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}})$ and $\boldsymbol{\phi}_2(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}})$ follow from the consistency of IPCW weights $\widehat{W}_i$ to W_i, which is proven in the next

section, and each item in $\boldsymbol{\phi}_1\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)$ is mean zero:

$$
\begin{aligned}
& E\left[G_{ij} W_i \Delta_i^Y \left\{Y_i - \mu_{0j}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\right\}\boldsymbol{Z}_i\right] \\
= \; & E\left(E\left[G_{ij}\frac{I\left(D_i \wedge L \leq C_i\right)}{P\left(C_i > Y_i\right)}\left\{Y_i - \mu_{0j}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\right\}\boldsymbol{Z}_i|\boldsymbol{Z}_i\right]\right) \\
= \; & E\left\{E\left(E\left[G_{ij}\frac{I\left(D_i \wedge L \leq C_i\right)}{P\left(C_i > Y_i\right)}\left\{Y_i - \mu_{0j}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\right\}\boldsymbol{Z}_i|\boldsymbol{Z}_i, D_i\right]|\boldsymbol{Z}_i\right)\right\} \\
= \; & E\left\{E\left(E\left[G_{ij}\frac{I\left(D_i \wedge L \leq C_i\right)}{P\left(C_i > D_i \wedge L\right)}\left\{D_i \wedge L - \mu_{0j}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\right\}\boldsymbol{Z}_i|\boldsymbol{Z}_i, D_i\right]|\boldsymbol{Z}_i\right)\right\} \\
= \; & E\left(E\left[G_{ij}\left\{D_i \wedge L - \mu_{0j}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\right\}\boldsymbol{Z}_i|\boldsymbol{Z}_i\right]\right) \\
= \; & E\left[G_{ij}\left\{E\left(D_i \wedge L|\boldsymbol{Z}_i\right) - \mu_{0j}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right)\right\}\boldsymbol{Z}_i\right] \\
= \; & \mathbf{0}.
\end{aligned}
$$

Since $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2$ are equivalent to $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2$, the unbiasedness of the latter holds too.

B.3 Consistency of IPCW Weights $\widehat{W}_i$

B.3.1 Asymptotic Distribution of $\widehat{\boldsymbol{\theta}}$

From definition (**??**) and (2.6), we define

$$\overline{\boldsymbol{R}}_j(t;\boldsymbol{\theta}) = \frac{\boldsymbol{R}_j^{(1)}(t;\boldsymbol{\theta})}{R_j^{(0)}(t;\boldsymbol{\theta})}.$$

and the score residual:

$$\boldsymbol{U}_i(\boldsymbol{\theta}) = G_{ij}\int_0^\tau \left\{\boldsymbol{Z}_i - \overline{\boldsymbol{r}}_j(u;\boldsymbol{\theta})\right\} dM_i^C(u).$$

Then the asymptotic properties of $\widehat{\boldsymbol{\theta}}$ indicates that:

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}\boldsymbol{U}_i(\boldsymbol{\theta}).$$

B.3.2 Asymptotic Distribution of $\widehat{\Lambda}_{0j}^C(t)$

The difference between $\widehat{\Lambda}_{0j}^C(t)$ and $\Lambda_{0j}^C(t)$ can be separated as

$$
\begin{aligned}
\sqrt{n_j}\left\{\Lambda_{0j}^C(t) - \Lambda_{0j}^C(t)\right\} &= \sqrt{n_j}\left\{\widehat{\Lambda}_{0j}^C(t;\widehat{\boldsymbol{\theta}}) - \widehat{\Lambda}_{0j}^C(t;\boldsymbol{\theta})\right\} && \text{(B.11)} \\
&+ \sqrt{n_j}\left\{\widehat{\Lambda}_{0j}^C(t;\boldsymbol{\theta}) - \Lambda_{0j}^C(t;\boldsymbol{\theta})\right\}. && \text{(B.12)}
\end{aligned}
$$

Since

$$\widehat{\Lambda}_{0j}^{C}(t;\boldsymbol{\theta}) = \int_0^t \frac{\sum_{i=1}^n G_{ij} dN_i^C(u)}{\sum_{i=1}^n G_{ij} R_i(u) \exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i\right)} = \sum_{i=1}^n G_{ij} \int_0^t \frac{1}{n_j R_j^{(0)}(u;\boldsymbol{\theta})} dN_i^C(u),$$

$$\Rightarrow \quad (B.11) = \sqrt{n_j} \sum_{i=1}^n G_{ij} \int_0^t \left\{ \frac{1}{R_j^{(0)}(u;\widehat{\boldsymbol{\theta}})} - \frac{1}{R_j^{(0)}(u;\boldsymbol{\theta})} \right\} dN_i^C(u).$$

Through a Taylor expansion,

$$\begin{aligned}
(B.11) &= \sqrt{\frac{n_j}{n}} \boldsymbol{h}_j(t;\boldsymbol{\theta})' \sqrt{n} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) + o_p(1), \\
&= -\sqrt{\frac{n_j}{n}} \left\{ \int_0^t \overline{\boldsymbol{r}}_j(u;\boldsymbol{\theta})' d\Lambda_{0j}^C(u;\boldsymbol{\theta}) \right\} \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \sum_{i=1}^n \sum_{l=1}^J G_{il} \boldsymbol{U}_i(\boldsymbol{\theta}) + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{h}_j(t;\boldsymbol{\theta}) &= -\frac{1}{n_j} \sum_{i=1}^n G_{ij} \int_0^t \frac{\overline{\boldsymbol{R}}_j(u;\boldsymbol{\theta})}{R_j^{(0)}(u;\boldsymbol{\theta})} dN_i^C(u) \\
&= -\frac{1}{n_j} \sum_{i=1}^n G_{ij} \int_0^t \overline{\boldsymbol{R}}_j(u;\boldsymbol{\theta}) d\widehat{\Lambda}_{0j}^C(u;\boldsymbol{\theta}) \\
&\xrightarrow{p} -\int_0^t \overline{\boldsymbol{r}}_j(u;\boldsymbol{\theta}) d\Lambda_{0j}^C(u;\boldsymbol{\theta}).
\end{aligned}$$

holds from Slutsky's Theorem. In addition, since $(B.12) = \frac{1}{\sqrt{n_j}} \sum_{i=1}^n G_{ij} \int_0^t r_j^{(0)}(u;\boldsymbol{\theta})^{-1} dM_i^C(u) +$ $o_p(1)$,

$$\begin{aligned}
\sqrt{n_j} \left\{ \widehat{\Lambda}_{0j}^C(t) - \Lambda_{0j}^C(t) \right\} &= -\frac{\sqrt{n_j}}{n} \sum_{i=1}^n \sum_{l=1}^J G_{il} \int_0^t \overline{\boldsymbol{r}}_j(u;\boldsymbol{\theta})' \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \boldsymbol{U}_i(\boldsymbol{\theta}) d\Lambda_{0j}^C(u;\boldsymbol{\theta}) \\
&\quad + \frac{1}{\sqrt{n_j}} \sum_{i=1}^n G_{ij} \int_0^t \frac{1}{r_j^{(0)}(u;\boldsymbol{\theta})} dM_i^C(u) + o_p(1).
\end{aligned}$$

B.3.3 Asymptotic Distribution of $\widehat{\Lambda}_{ij}^C(t)$

$$\begin{aligned}
&(B.13) \qquad \sqrt{n_j} \left\{ \widehat{\Lambda}_{ij}^C(t) - \Lambda_{ij}^C(t) \right\} \\
&(B.14) = \sqrt{n_j} \left\{ \int_0^t \exp\left(\widehat{\boldsymbol{\theta}}' \boldsymbol{Z}_i \right) R_i(u) d\widehat{\Lambda}_{0j}^C(u) - \int_0^t \exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i \right) R_i(u) d\widehat{\Lambda}_{0j}^C(u) \right\} \\
&(B.15) + \sqrt{n_j} \left\{ \int_0^t \exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i \right) R_i(u) d\widehat{\Lambda}_{0j}^C(u) - \int_0^t \exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i \right) R_i(u) d\Lambda_{0j}^C(u) \right\}.
\end{aligned}$$

Since (B.14) $= \sqrt{\frac{n_j}{n}} \int_0^t \sqrt{n} \left\{ \exp\left(\widehat{\boldsymbol{\theta}}' \boldsymbol{Z}_i\right) - \exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i\right) \right\} R_i(u) d\widehat{\Lambda}_{0j}^C(u) + o_p(1)$, by a Taylor expansion, we obtain

$$\begin{aligned}
(B.14) &= \sqrt{\frac{n_j}{n}} \exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i\right) \boldsymbol{Z}_i' \int_0^t R_i(u) d\widehat{\Lambda}_{0j}^C(u) \sqrt{n} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) + o_p(1) \\
&= \frac{\sqrt{n_j}}{n} \boldsymbol{Z}_i' \int_0^t d\widehat{\Lambda}_{ij}^C(u) \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \sum_{k=1}^n \sum_{l=1}^J G_{kl} \boldsymbol{U}_k(\boldsymbol{\theta}) + o_p(1) \\
&= \frac{\sqrt{n_j}}{n} \sum_{k=1}^n \sum_{l=1}^J G_{kl} \int_0^t \boldsymbol{Z}_i' d\widehat{\Lambda}_{ij}^C(u) \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \boldsymbol{U}_k(\boldsymbol{\theta}) + o_p(1).
\end{aligned}$$

Plug the result of the above asymptotic Distribution of $\widehat{\lambda}_{0j}^C(t)$ into (B.15) and obtain:

$$\begin{aligned}
(B.15) &= \sqrt{n_j} \int_0^t \exp\left(\widehat{\boldsymbol{\theta}}' \boldsymbol{Z}_i\right) R_i(u) d\left\{ \widehat{\Lambda}_{0j}^C(u) - \Lambda_{0j}^C(u) \right\} + o_p(1) \\
&= -\frac{\sqrt{n_j}}{n} \sum_{k=1}^n \sum_{l=1}^J G_{kl} \int_0^t \overline{\boldsymbol{r}}_j(u;\boldsymbol{\theta})' \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \boldsymbol{U}_k(\boldsymbol{\theta}) \exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i\right) R_i(u) d\Lambda_{0j}^C(u;\boldsymbol{\theta}) \\
&\quad + \frac{1}{\sqrt{n_j}} \sum_{k=1}^n G_{kj} \int_0^t \frac{\exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i\right) R_i(u)}{r_j^{(0)}(u;\boldsymbol{\theta})} dM_k^C(u) + o_p(1) \\
&= -\frac{\sqrt{n_j}}{n} \sum_{k=1}^n \sum_{l=1}^J G_{kl} \int_0^t \overline{\boldsymbol{r}}_j(u;\boldsymbol{\theta})' \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \boldsymbol{U}_k(\boldsymbol{\theta}) d\Lambda_{ij}^C(u;\boldsymbol{\theta}) \\
&\quad + \frac{1}{\sqrt{n_j}} \sum_{k=1}^n G_{kj} \int_0^t \frac{\exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i\right) R_i(u)}{r_j^{(0)}(u;\boldsymbol{\theta})} dM_k^C(u) + o_p(1).
\end{aligned}$$

Combining the above re-expressions for (B.14) and (B.15) leads to

$$\begin{aligned}
\sqrt{n_j} \left\{ \widehat{\Lambda}_{ij}^C(t) - \Lambda_{ij}^C(t) \right\} &= \frac{\sqrt{n_j}}{n} \sum_{k=1}^n \sum_{l=1}^J G_{kl} \int_0^t \left\{ \boldsymbol{Z}_i - \overline{\boldsymbol{r}}_j(u;\boldsymbol{\theta}) \right\}' \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \boldsymbol{U}_k(\boldsymbol{\theta}) d\Lambda_{ij}^C(u;\boldsymbol{\theta}) \\
&\quad + \frac{1}{\sqrt{n_j}} \sum_{k=1}^n G_{kj} \int_0^t \frac{\exp\left(\boldsymbol{\theta}' \boldsymbol{Z}_i\right) R_i(u)}{r_j^{(0)}(u;\boldsymbol{\theta})} dM_k^C(u) + o_p(1)
\end{aligned}$$

Furthermore, at time point Y_i, we have

$$\sqrt{n_j} \left\{ \widehat{\Lambda}_{ij}^C(Y_i) - \Lambda_{ij}^C(Y_i) \right\} = \frac{\sqrt{n_j}}{n} \sum_{k=1}^n \sum_{l=1}^J G_{kl} \boldsymbol{D}_i(\boldsymbol{\theta})' \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \boldsymbol{U}_k(\boldsymbol{\theta}) + \frac{1}{\sqrt{n_j}} \sum_{k=1}^n G_{kj} J_{ik}(\boldsymbol{\theta}) + o_p(1),$$

where

$$\begin{aligned}
\boldsymbol{D}_i(\boldsymbol{\theta}) &= \sum_{j=1}^{J} G_{ij} \int_0^L \{\boldsymbol{Z}_i - \overline{\boldsymbol{r}}_j(u;\boldsymbol{\theta})\} R_i(u) d\Lambda_{ij}^C(u), \\
J_{ik}(\boldsymbol{\theta}) &= \sum_{j=1}^{J} G_{ij} G_{kj} \int_0^L r_j^{(0)}(u;\boldsymbol{\theta})^{-1} \exp(\boldsymbol{\theta}'\boldsymbol{Z}_i) R_i(u) dM_k^C(u).
\end{aligned}$$

B.3.4 Asymptotic Distribution of $\widehat{W}_i$

By definition, $\widehat{W}_i - W_i = \exp\left\{\widehat{\Lambda}_{ij}^C(Y_i)\right\} - \exp\left\{\Lambda_{ij}^C(Y_i)\right\} = \exp\left\{\Lambda_{ij}^C(Y_i)\right\}\left\{\widehat{\Lambda}_{ij}^C(Y_i) - \Lambda_{ij}^C(Y_i)\right\}$. From the previous result of $\widehat{\Lambda}_{ij}^C(Y_i) - \Lambda_{ij}^C(Y_i)$ into this, we get

$$\begin{aligned}
\sqrt{n_j}\left(\widehat{W}_i - W_i\right) &= \sqrt{n_j} \exp\left\{\Lambda_{ij}^C(Y_i)\right\}\left\{\widehat{\Lambda}_{ij}^C(Y_i) - \Lambda_{ij}^C(Y_i)\right\} \\
&= \sqrt{n_j} W_i \left\{\frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{J} G_{kl} \boldsymbol{D}_i(\boldsymbol{\theta})' \boldsymbol{\Theta}(\boldsymbol{\theta})^{-1} \boldsymbol{U}_k(\boldsymbol{\theta}) + \frac{1}{n_j}\sum_{k=1}^{n} G_{kj} J_{ik}(\boldsymbol{\theta})\right\} \\
&\quad + o_p\left(\frac{1}{\sqrt{n_j}}\right).
\end{aligned}$$

B.4 Consistency of $\widehat{\boldsymbol{\beta}}$

Using Inverse Function Theorem Foutz (1977), we need to verify these conditions to conclude the consistency of $\widehat{\boldsymbol{\beta}}$.

(a) $\partial \boldsymbol{\phi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})/\partial \boldsymbol{\beta}'$ exists and is continuous in an open neighborhood of $\boldsymbol{\beta}$.

(b) $-n^{-1}\partial \boldsymbol{\phi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})/\partial \boldsymbol{\beta}'$ is positive definite at $\boldsymbol{\beta}$ with probability 1 as $n \to \infty$.

(c) $-n^{-1}\partial \boldsymbol{\phi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})/\partial \boldsymbol{\beta}'$ converges in probability to a fixed function uniformly in an open neighborhood of $\boldsymbol{\beta}$.

(d) Asymptotic unbiasedness of the estimating function: $-n^{-1}\boldsymbol{\phi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}) \xrightarrow{p} \mathbf{0}$.

The only quantity in $\boldsymbol{\phi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})$ involved with $\boldsymbol{\beta}$ is $\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})$, and

$$\begin{aligned}
& \frac{\partial \left\{ \boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j \left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}} \right) \right\}}{\partial \boldsymbol{\beta}'} \\
= & \frac{\partial}{\partial \boldsymbol{\beta}'} \frac{\sum_{k=1}^n G_{kj} \Delta_{Yk} \widehat{W}_k \boldsymbol{Z}_k \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_k\right)}{\sum_{k=1}^n G_{kj} \Delta_{Yk} \widehat{W}_k \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_k\right)} \\
= & \frac{\boldsymbol{S}_j^{(2)} \left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}} \right) S_j^{(0)} \left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}} \right) - \boldsymbol{S}_j^{(1)} \left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}} \right)^{\otimes 2}}{S_j^{(0)} \left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}} \right)^2} \\
= & \frac{\boldsymbol{s}_j^{(2)}(\boldsymbol{\beta}) s_j^{(0)}(\boldsymbol{\beta}) - \boldsymbol{s}_j^{(1)}(\boldsymbol{\beta})^{\otimes 2}}{s_j^{(0)}(\boldsymbol{\beta})^2} + o_p(1).
\end{aligned}$$

So the derivative of $\boldsymbol{\psi}_1$ is

$$\begin{aligned}
-\frac{1}{n} \frac{\boldsymbol{\psi}_1 \left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}} \right)}{\partial \boldsymbol{\beta}'} &= -\frac{1}{n} \sum_{j=1}^J \sum_{i=1}^n G_{ij} \Delta_i^Y \widehat{W}_i Y_i \frac{\partial \left\{ \boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j \left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}} \right) \right\}}{\partial \boldsymbol{\beta}'} \\
&= \frac{1}{n} \sum_{j=1}^J \sum_{i=1}^n G_{ij} \Delta_i^Y \widehat{W}_i Y_i \frac{\boldsymbol{s}_j^{(2)}(\boldsymbol{\beta}) s_j^{(0)}(\boldsymbol{\beta}) - \boldsymbol{s}_j^{(1)}(\boldsymbol{\beta})^{\otimes 2}}{s_j^{(0)}(\boldsymbol{\beta})^2} + o(1) \\
&= \boldsymbol{A}(\boldsymbol{\beta}) + o(1).
\end{aligned}$$

So condition (b) holds. And condition (d) holds by law of large numbers. The remaining conditions would hold under some certain regularity conditions. So it follows that

$$\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}.$$

B.5 Asymptotic Distribution of $\widehat{\boldsymbol{\beta}}$

B.5.1 Write $\boldsymbol{\psi}(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})$ as a sum of i.i.d. terms

We want to replace Y_i in $\boldsymbol{\psi}_1$ with $Y_i - \mu_{ij}$, and it works because

$$\begin{aligned}
& \sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}\left\{\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)\right\}\widehat{W}_i\Delta_i^Y\mu_{ij} \\
= & \sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}\left\{\boldsymbol{Z}_i - \frac{\boldsymbol{S}_j^{(1)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)}{S_j^{(0)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)}\right\}\widehat{W}_i\Delta_i^Y\mu_{0j}\exp\left(\boldsymbol{\beta}'\boldsymbol{Z}_i\right) \\
= & \sum_{j=1}^{J} n_j \frac{\boldsymbol{S}_j^{(1)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right) S_j^{(0)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right) - \boldsymbol{S}_j^{(1)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right) S_j^{(0)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)}{S_j^{(0)}\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)} \\
= & \mathbf{0} \\
\Rightarrow \quad & \boldsymbol{\psi}_1\left(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \widehat{\boldsymbol{W}}\right) = \sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}\left\{\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)\right\}\widehat{W}_i\Delta_i^Y Y_i \\
= & \sum_{j=1}^{J}\sum_{i=1}^{n} G_{ij}\left\{\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)\right\}\widehat{W}_i\Delta_i^Y\left(Y_i - \mu_{ij}\right)
\end{aligned}$$

The main estimating equation $\boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})$ can be further separated as the sum of the following three components,

$$\begin{aligned}
\text{(B.16)} \quad \boldsymbol{\psi}_1\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{J} G_{ij}\left\{\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \boldsymbol{W}\right)\right\}W_i\Delta_i^Y\left(Y_i - \mu_{ij}\right) \\
\text{(B.17)} \quad &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{J} G_{ij}\left\{\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \boldsymbol{W}\right)\right\}\left(\widehat{W}_i - W_i\right)\Delta_i^Y\left(Y_i - \mu_{ij}\right) \\
\text{(B.18)} \quad &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{J} G_{ij}\left\{\overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \boldsymbol{W}\right) - \overline{\boldsymbol{S}}_j\left(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}\right)\right\}\widehat{W}_i\Delta_i^Y\left(Y_i - \mu_{ij}\right).
\end{aligned}$$

(B.18) can be shown to converge in probability to 0 through the Functional Delta Method. Through techniques from empirical processes (Lin et al., 2000), it can be shown that (B.16) is asymptotically equivalent to

$$(B.16) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{J} G_{ij}\left\{\boldsymbol{Z}_i - \overline{\boldsymbol{s}}_j\left(\boldsymbol{\beta}, \boldsymbol{W}\right)\right\}W_i\Delta_i^Y\left(Y_i - \mu_{ij}\right).$$

Let $\boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W}) = G_{ij}\{\boldsymbol{Z}_i - \overline{\boldsymbol{S}}_j(\boldsymbol{\beta}, \boldsymbol{W})\}W_i\Delta_i^Y(Y_i - \mu_{ij})$, then (B.17) can be written as,

$$
\begin{aligned}
(B.17) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{J}\boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W})\frac{\widehat{W}_i - W_i}{W_i} \\
\text{(B.19)} &= \frac{1}{\sqrt{n^3}}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{n}\sum_{l=1}^{J}G_{ij}G_{kl}\boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W})\boldsymbol{D}_i(\boldsymbol{\theta})'\boldsymbol{\Theta}(\boldsymbol{\theta})^{-1}\boldsymbol{U}_k(\boldsymbol{\theta}) \\
\text{(B.20)} &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{n}\frac{1}{n_j}G_{ij}G_{kj}\boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W})J_{ik}(\boldsymbol{\theta}) + o_p(1).
\end{aligned}
$$

The first component can be simplified as $(B.19) = \frac{1}{\sqrt{n}}\boldsymbol{K}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W})\boldsymbol{\Theta}(\boldsymbol{\theta})^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}G_{ij}\boldsymbol{U}_i(\boldsymbol{\theta}) +$ $o_p(1)$, with the following additional notations:

$$
\boldsymbol{K}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W}) = \sum_{j=1}^{J}\boldsymbol{K}_j(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W}) = \sum_{j=1}^{J}E\left\{G_{ij}\boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W})\boldsymbol{D}_i(\boldsymbol{\theta})'\right\}
$$

Let $\boldsymbol{H}_j(t; \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W}) = E\{\boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W})\exp(\boldsymbol{\theta}'\boldsymbol{Z}_i)R_i(t)\}$, then (B.20) is equivalent to

$$
\begin{aligned}
(B.20) &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{j=1}^{J}G_{kj}\int_0^L\frac{\sum_{i=1}^{n}G_{ij}\boldsymbol{\epsilon}_i(\boldsymbol{\beta}, \boldsymbol{W})\exp\left(\boldsymbol{\theta}'\boldsymbol{Z}_i\right)R_i(u)}{n_j r_j^{(0)}(u; \boldsymbol{\theta})}dM_k^C(u) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{j=1}^{J}G_{kj}\int_0^L\frac{\boldsymbol{H}_j(u; \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W})}{r_j^{(0)}(u; \boldsymbol{\theta})}dM_k^C(u) + o_p(1).
\end{aligned}
$$

Combining (B.19)and (B.20) together, we can write (B.17) as

$$
\begin{aligned}
\frac{1}{\sqrt{n}}(B.17) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{J}G_{ij}\left\{\boldsymbol{K}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W})\boldsymbol{\Theta}(\boldsymbol{\theta})^{-1}\boldsymbol{U}_i(\boldsymbol{\theta}) + \int_0^L\frac{\boldsymbol{H}_j(u; \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W})}{r_j^{(0)}(u; \boldsymbol{\theta})}dM_i^C(u)\right\} \\
&\quad + o_p(1).
\end{aligned}
$$

To sum up, our estimating equation $\frac{1}{\sqrt{n}}\boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})$ can be written as a sum of n i.i.d. terms at a difference of $o_p(1)$: $\frac{1}{\sqrt{n}}\boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{J}G_{ij}\boldsymbol{b}_i(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})$, where

$$
\begin{aligned}
\boldsymbol{b}_i\left(\boldsymbol{\beta}, \boldsymbol{W}\right) &= \left\{\boldsymbol{Z}_i - \overline{\boldsymbol{s}}_j\left(\boldsymbol{\beta}, \boldsymbol{W}\right)\right\}W_i\Delta_i^Y\left(Y_i - \mu_{ij}\right) \\
&+ \boldsymbol{K}\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W}\right)\boldsymbol{\Theta}(\boldsymbol{\theta})^{-1}\boldsymbol{U}_i(\boldsymbol{\theta}) \\
&+ \int_0^L\frac{\boldsymbol{H}_j\left(u; \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{W}\right)}{r_j^{(0)}(u; \boldsymbol{\theta})}dM_i^C(u).
\end{aligned}
$$

B.5.2 Asymptotic Distribution of $\boldsymbol{\psi}_1$

$$\frac{1}{\sqrt{n}}\boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}) \xrightarrow{D} \text{Normal}\left(\mathbf{0}, \boldsymbol{B}(\boldsymbol{\beta}, \boldsymbol{W})\right),$$

where

$$\boldsymbol{B}\left(\boldsymbol{\beta}, \boldsymbol{W}\right) = E\left[\sum_{j=1}^{J} G_{ij}\left\{\boldsymbol{b}_i\left(\boldsymbol{\beta}, \boldsymbol{W}\right)\right\}^{\otimes 2}\right].$$

B.5.3 Asymptotic Distribution of $\widehat{\boldsymbol{\beta}}$

Taylor expansion of $\boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})$ around $\boldsymbol{\beta}$ is

$$\begin{aligned}
\mathbf{0} = \boldsymbol{\psi}_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{W}}) &\approx \boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}) + \frac{\partial \boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})}{\partial \boldsymbol{\beta}'}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \\
\Rightarrow \quad \sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) &= \left[\frac{1}{n}\frac{\partial \boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}})}{\partial \boldsymbol{\beta}'}\right]^{-1} \frac{1}{\sqrt{n}}\boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}) = \boldsymbol{A}(\boldsymbol{\beta})^{-1}\frac{1}{\sqrt{n}}\boldsymbol{\psi}_1(\boldsymbol{\beta}, \widehat{\boldsymbol{W}}) + o\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

By Delta methods,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{D} \text{Normal}\left(\mathbf{0}, \boldsymbol{A}(\boldsymbol{\beta})^{-1}\boldsymbol{B}(\boldsymbol{\beta}, \boldsymbol{W})\boldsymbol{A}(\boldsymbol{\beta})^{-1}\right).$$

B.6 Asymptotic Distribution of $\widehat{\boldsymbol{\mu}}_0$ and $\widehat{\boldsymbol{\eta}}$

B.6.1 From $\widehat{\boldsymbol{\theta}}$ to $\widehat{\boldsymbol{\mu}}_0$

Define functions

$$\overline{d}_j(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n} G_{ij}\widehat{W}_i\Delta_i^Y Y_i}{\sum_{i=1}^{n} G_{ij}\widehat{W}_i\Delta_i^Y \exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)},$$

$$d_j(\boldsymbol{\beta}) = \lim_{n\to 0}\overline{d}_j(\boldsymbol{\beta}) = \frac{E\left\{G_{ij}\widehat{W}_i\Delta_i^Y Y_i\right\}}{E\left\{G_{ij}\widehat{W}_i\Delta_i^Y \exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\right\}} = \frac{\mu_{0j}E\left\{\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\right\}}{E\left\{\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\right\}},$$

and $\overline{\boldsymbol{d}}(\boldsymbol{\beta}) = (\overline{d}_1(\boldsymbol{\beta}), \ldots, \overline{d}_J(\boldsymbol{\beta}))$, $\boldsymbol{d}(\boldsymbol{\beta}) = (d_1(\boldsymbol{\beta}), \ldots, d_J(\boldsymbol{\beta}))$. Then $\widehat{\boldsymbol{\mu}}_0 = \overline{\boldsymbol{d}}(\widehat{\boldsymbol{\beta}})$ and $\boldsymbol{\mu}_0 = \boldsymbol{d}(\boldsymbol{\beta})$. The difference between the estimated and true value of μ_{0j} can be written

as:

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0\right) &= \sqrt{n}\left\{\overline{\boldsymbol{d}}(\widehat{\boldsymbol{\beta}})-\overline{\boldsymbol{d}}(\boldsymbol{\beta})\right\}+\sqrt{n}\left\{\overline{\boldsymbol{d}}(\boldsymbol{\beta})-\boldsymbol{d}(\boldsymbol{\beta})\right\}\\ &= \frac{\partial\overline{\boldsymbol{d}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+\sqrt{n}\left\{\overline{\boldsymbol{d}}(\boldsymbol{\beta})-\boldsymbol{d}(\boldsymbol{\beta})\right\}+o_p(1),\end{aligned}$$

where $\frac{\partial\overline{\boldsymbol{d}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}$ is a $J*p$ matrix and its jth row is

$$\begin{aligned}\left(\frac{\partial\overline{\boldsymbol{d}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)_{j*} &= -\frac{\left\{\sum_{i=1}^n G_{ij}\widehat{W}_i\Delta_i^Y Y_i\right\}\left\{\sum_{i=1}^n G_{ij}\widehat{W}_i\Delta_i^Y\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\boldsymbol{Z}_i'\right\}}{\left\{\sum_{i=1}^n G_{ij}\widehat{W}_i\Delta_i^Y\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\right\}^2}\\ &= -\frac{\left[\mu_{0j}E\left\{G_{ij}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\right\}\right]\left[E\left\{G_{ij}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\boldsymbol{Z}_i'\right\}\right]}{\left[E\left\{G_{ij}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\right\}\right]^2}+o_p(1)\\ &= -\mu_{0j}\overline{\boldsymbol{s}}_j(\boldsymbol{\beta})'+o_p(1).\end{aligned}$$

So we can write $\frac{\partial\overline{\boldsymbol{d}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}$ as $-\boldsymbol{\mu}_M\overline{\boldsymbol{s}}(\boldsymbol{\beta})$, where

$$\begin{aligned}\boldsymbol{\mu}_M &= \begin{pmatrix}\mu_{01} & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \mu_{0J}\end{pmatrix},\\ \overline{\boldsymbol{s}}(\boldsymbol{\beta}) &= \begin{pmatrix}\overline{\boldsymbol{s}}_1(\boldsymbol{\beta})'\\ \vdots\\ \overline{\boldsymbol{s}}_J(\boldsymbol{\beta})'\end{pmatrix}.\end{aligned}$$

And the jth element the second part is

$$\begin{aligned}\sqrt{n}\left\{\overline{\boldsymbol{d}}(\boldsymbol{\beta})-\boldsymbol{d}(\boldsymbol{\beta})\right\}_j &= \frac{\sqrt{n}}{n_j}\frac{\sum_{i=1}^n G_{ij}\widehat{W}_i\Delta_i^Y\left\{Y_i-\mu_{0j}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\right\}}{\frac{1}{n_j}\sum_{i=1}^n G_{ij}\widehat{W}_i\Delta_i^Y\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)}\\ &= \frac{\sqrt{n}}{n_j}\sum_{i=1}^n\frac{G_{ij}\widehat{W}_i\Delta_i^Y}{s_j^{(0)}(\boldsymbol{\beta})}\left(Y_i-\mu_{ij}\right)\\ &\xrightarrow{p} \frac{1}{\sqrt{n}}\frac{n}{n_j s_j^{(0)}(\boldsymbol{\beta})}\sum_{i=1}^n G_{ij}W_i\Delta_i^Y\left(Y_i-\mu_{ij}\right).\end{aligned}$$

So

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\begin{pmatrix}\frac{n}{n_1 s_1^{(0)}(\boldsymbol{\beta})}G_{i1}W_i\Delta_i^Y\left(Y_i-\mu_{ij}\right)\\ \vdots \\ \frac{n}{n_J s_J^{(0)}(\boldsymbol{\beta})}G_{iJ}W_i\Delta_i^Y\left(Y_i-\mu_{ij}\right)\end{pmatrix}-\sum_{j=1}^{J}G_{ij}\boldsymbol{\mu}_M\overline{\boldsymbol{s}}(\boldsymbol{\beta})\boldsymbol{b}_i\left(\boldsymbol{\beta},\boldsymbol{W}\right)\right\}\\ &+ o_p(1).\end{aligned}$$

Let

$$\boldsymbol{v}_{\mu i}\left(\boldsymbol{\beta},\boldsymbol{W}\right)=\begin{pmatrix}\frac{n}{n_1 s_1^{(0)}(\boldsymbol{\beta})}G_{i1}W_i\Delta_i^Y\left(Y_i-\mu_{ij}\right)\\ \vdots \\ \frac{n}{n_J s_J^{(0)}(\boldsymbol{\beta})}G_{iJ}W_i\Delta_i^Y\left(Y_i-\mu_{ij}\right)\end{pmatrix}-\sum_{j=1}^{J}G_{ij}\boldsymbol{\mu}_M\overline{\boldsymbol{s}}(\boldsymbol{\beta})\boldsymbol{b}_i\left(\boldsymbol{\beta},\boldsymbol{W}\right),$$

then

$$\sqrt{n}\left(\widehat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0\right)\xrightarrow{D}\text{Normal}\left(0,E\left\{\boldsymbol{v}_{\mu i}\left(\boldsymbol{\beta},\boldsymbol{W}\right)^{\otimes 2}\right\}\right).$$

B.6.2 From $\widehat{\mu}_0$ to $\widehat{\eta}$

Rescale $\widehat{\boldsymbol{\mu}}_0$ to $\widehat{\boldsymbol{\eta}}=\boldsymbol{\mu}/(\boldsymbol{\mu}'\boldsymbol{w})$ with a pre-specified weight vector $\boldsymbol{w}$. Define a new function $\boldsymbol{\zeta}(\boldsymbol{\mu}_0)=\frac{\boldsymbol{\mu}_0}{\boldsymbol{\mu}_0'\boldsymbol{w}}$. Then $\boldsymbol{\eta}=\boldsymbol{\zeta}(\boldsymbol{\mu}_0)$ and $\widehat{\boldsymbol{\eta}}=\boldsymbol{\zeta}(\widehat{\boldsymbol{\mu}}_0)$. Since

$$\frac{\partial\boldsymbol{\zeta}(\boldsymbol{\mu}_0)}{\partial\boldsymbol{\mu}_0}=\frac{\boldsymbol{\mu}_0'\boldsymbol{w}\boldsymbol{I}_J-\boldsymbol{\mu}_0\boldsymbol{w}'}{\left(\boldsymbol{\mu}_0'\boldsymbol{w}\right)^2},$$

By Delta Methods,

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}\right) &= \sqrt{n}\left\{\boldsymbol{\zeta}(\widehat{\boldsymbol{\mu}}_0)-\boldsymbol{\zeta}(\boldsymbol{\mu}_0)\right\}\\ &\xrightarrow{D} \text{Normal}\left(\boldsymbol{0},\frac{\boldsymbol{\mu}_0'\boldsymbol{w}\boldsymbol{I}_J-\boldsymbol{\mu}_0\boldsymbol{w}'}{\left(\boldsymbol{\mu}_0'\boldsymbol{w}\right)^2}E\left\{\boldsymbol{v}_{\mu i}\left(\boldsymbol{\beta},\boldsymbol{W}\right)^{\otimes 2}\right\}\frac{\boldsymbol{\mu}_0'\boldsymbol{w}\boldsymbol{I}_J-\boldsymbol{w}\boldsymbol{\mu}_0'}{\left(\boldsymbol{\mu}_0'\boldsymbol{w}\right)^2}\right).\end{aligned}$$

.

B.7 Simulation Result

B.7.1 Bias and Standard Error for $\widehat{\beta}$ when $L = 0.18, 0.57$ and 13.4

B.7.2 Point and interval estimation for η $L = 0.18, 0.57$ and 13.4

Table B.1 Simulation results: $L = 0.18, 0.57$ and 13.4 under light and heavy censoring

L	Censoring	Var (True)	n	Bias	ESD	ASE	CP(%)
0.18	Light	$\beta_1(-0.025)$	2500	0	0.004	0.004	94
			5000	0	0.003	0.003	94
			10000	0	0.002	0.002	94
		$\beta_2(-0.05)$	2500	0	0.005	0.005	95
			5000	0	0.003	0.003	95
			10000	0	0.002	0.002	95
	Heavy	$\beta_1(-0.025)$	2500	0	0.004	0.004	94
			5000	0	0.003	0.003	94
			10000	0	0.002	0.002	95
		$\beta_2(-0.05)$	2500	0	0.005	0.005	93
			5000	0	0.003	0.003	94
			10000	0	0.002	0.002	95
0.57	Light	$\beta_1(-0.062)$	2500	0.001	0.006	0.006	95
			5000	0	0.005	0.004	92
			10000	0	0.003	0.003	96
		$\beta_2(-0.125)$	2500	0	0.007	0.007	94
			5000	0	0.005	0.005	95
			10000	0	0.003	0.003	95
	Heavy	$\beta_1(-0.062)$	2500	0.001	0.006	0.006	95
			5000	0.001	0.005	0.005	95
			10000	0.001	0.003	0.003	95
		$\beta_2(-0.125)$	2500	0	0.007	0.007	94
			5000	0	0.005	0.005	95
			10000	0	0.003	0.003	95
13.4	Light	$\beta_1(-0.316)$	2500	0.009	0.02	0.02	92
			5000	0.005	0.015	0.014	92
			10000	0.003	0.01	0.01	94
		$\beta_2(-0.633)$	2500	0.006	0.021	0.019	91
			5000	0.005	0.014	0.014	94
			10000	0.003	0.01	0.01	93
	Heavy	$\beta_1(-0.316)$	2500	0.045	0.03	0.026	58
			5000	0.032	0.022	0.02	63
			10000	0.021	0.017	0.016	72
		$\beta_2(-0.633)$	2500	-0.006	0.028	0.025	91
			5000	-0.006	0.021	0.019	91
			10000	-0.004	0.016	0.014	92

Figure B.1 Point and interval estimation of $\boldsymbol{\eta}$ under light and heavy censoring for $L = 0.18$

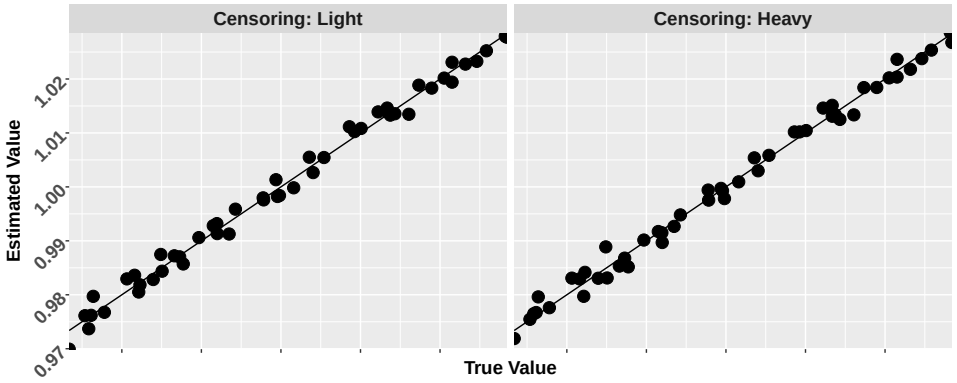


Figure B.2 Point and interval estimation of $\boldsymbol{\eta}$ under light and heavy censoring for $L = 0.57$

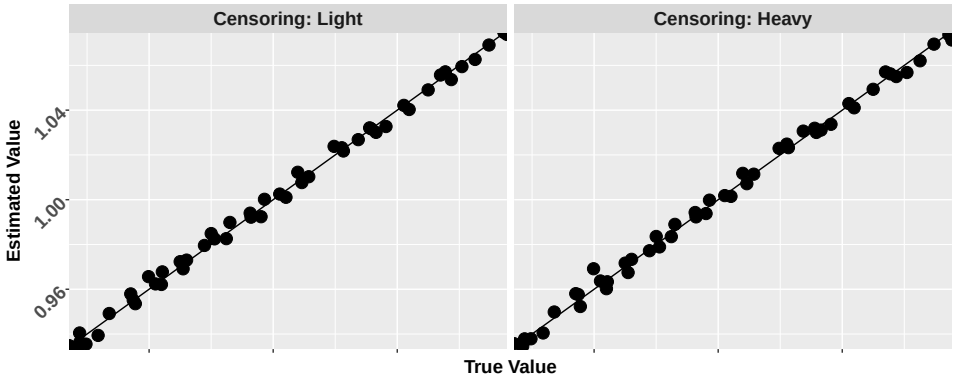
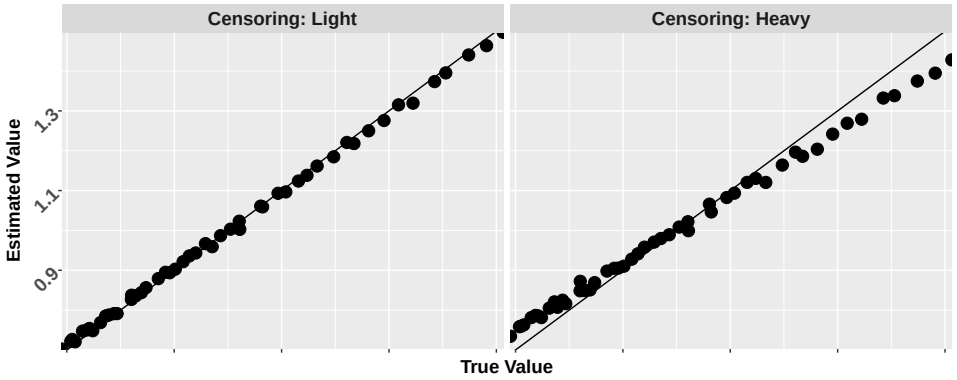


Figure B.3 Point and interval estimation of $\boldsymbol{\eta}$ under light and heavy censoring for $L = 13.4$



BIBLIOGRAPHY

BIBLIOGRAPHY

Per Kragh Andersen. Decomposition of number of life years lost according to causes of death. *Statistics in medicine*, 32(30):5278–5285, 2013.

Per Kragh Andersen and Maja Pohar Perme. Pseudo-observations in survival analysis. *Statistical methods in medical research*, 2009.

Per Kragh Andersen, Mette Gerster Hansen, and John P Klein. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime data analysis*, 10(4):335–350, 2004.

Michael Baiocchi, Jing Cheng, and Dylan S Small. Instrumental variable methods for causal inference. *Statistics in medicine*, 33(13):2297–2340, 2014.

Stuart G Baker. Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program. *Journal of the American Statistical Association*, 93(443):929–934, 1998.

Nadine Binder, Thomas A Gerds, and Per Kragh Andersen. Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime data analysis*, 20(2):303–315, 2014.

Christian Boudreau and Jerald F. Lawless. Survival analysis based on the proportional hazards model and survey data. *Canadian Journal of Statistics*, 34(2):203–216, 2006. ISSN 1708-945X.

Norman E Breslow. Contribution to the discussion of the paper by dr cox. *Journal of the Royal Statistical Society, Series B*, 34(2):216–217, 1972.

Kwun Chuen Gary Chan. Instrumental variable additive hazards models with exposure-dependent censoring. *Biometrics*, 2016.

Pei-Yun Chen and Anastasios A Tsiatis. Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics*, 57(4):1030–1038, 2001.

Songnian Chen, Cheng Hsiao, and Liqun Wang. Measurement errors and censored structural latent variables models. *Econometric Theory*, 28(03):696–703, 2012.

David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

David R Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

Anthony Christopher Davison and David Victor Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

Bradley Efron. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association*, 99(465):96–104, 2004.

Stanley SA Fenton, Douglas E Schaubel, Marie Desmeules, Howard I Morrison, Yang Mao, Pauline Copleston, John R Jeffery, and Carl M Kjellstrand. Hemodialysis versus peritoneal dialysis: a comparison of adjusted mortality rates. *American Journal of Kidney Diseases*, 30(3):334–342, 1997.

Robert V Foutz. On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, 72(357):147–148, 1977.

Daniel L Gillen and Scott S Emerson. Nontransitivity in a class of weighted logrank statistics under nonproportional hazards. *Statistics & probability letters*, 77(2):123–130, 2007.

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

Frank E Harrell, Kerry L Lee, and Daniel B Mark. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15:361–387, 1996.

Kevin He, Jack D Kalbfleisch, Yijiang Li, and Yi Li. Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. *Lifetime data analysis*, 19(4):490–512, 2013.

James G Heaf, Hans Løkkegaard, and Melvin Madsen. Initial survival advantage of peritoneal dialysis relative to haemodialysis. *Nephrology Dialysis Transplantation*, 17(1):112–117, 2002.

Patrick J Heagerty and Yingye Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61(1):92–105, 2005.

Bernard G Jaar, Josef Coresh, Laura C Plantinga, Nancy E Fink, Michael J Klag, Andrew S Levey, Nathan W Levin, John H Sadler, Alan Kliger, and Neil R Powe. Comparing the risk for death with peritoneal dialysis and hemodialysis in a national cohort of patients with chronic kidney disease. *Annals of internal medicine*, 143(3):174–183, 2005.

John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

John D Kalbfleisch and Robert A Wolfe. On monitoring outcomes of medical providers. *Statistics in Biosciences*, 5(2):286–302, 2013.

Patrick S Kamath, Russell H Wiesner, Michael Malinchoc, Walter Kremers, Terry M Therneau, Catherine L Kosberg, Gennaro D'Amico, E Rolland Dickson, and W Ray Kim. A model to predict survival in patients with end-stage liver disease. *Hepatology*, 33(2):464–470, 2001.

Theodore Karrison. Restricted mean life with adjustment for covariates. *Journal of the American Statistical Association*, 82(400):1169–1176, 1987.

Hyunwook Kim, Kyoung Hoon Kim, Kisoo Park, Shin-Wook Kang, Tae-Hyun Yoo, Song Vogue Ahn, Hyeong Sik Ahn, Hoo Jae Hann, Shina Lee, Jung-Hwa Ryu, et al. A population-based approach indicates an overall higher patient mortality with peritoneal dialysis compared to hemodialysis in korea. *Kidney international*, 86(5):991–1000, 2014.

Maiken IS Kjaersgaard and Erik T Parner. Instrumental variable method for time-to-event data using a pseudo-observation approach. *Biometrics*, 2015.

Victoria A Kumar, Margo A Sidell, Jason P Jones, and Edward F Vonesh. Survival of propensity matched incident peritoneal and hemodialysis patients in a united states health care system. *Kidney international*, 86(5):1016–1022, 2014.

Jialiang Li, Jason Fine, and Alan Brookhart. Instrumental variable additive hazards models. *Biometrics*, 71(1):122–130, 2015.

DY Lin, LJ 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.

Tom Loeys and Els Goetghebeur. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics*, 59(1):100–105, 2003.

Torben Martinussen, Stijn Vansteelandt, Eric J Tchetgen Tchetgen, and David M Zucker. Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. *Biometrics*, 2017.

Hui Nie, Jing Cheng, and Dylan S Small. Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics*, 67(4):1397–1405, 2011.

James 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.

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

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

James M Robins and Anastasios A Tsiatis. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in statistics-Theory and Methods*, 20(8):2609–2631, 1991.

Douglas E Schaubel and Guanghui Wei. Double inverse-weighted estimation of cumulative treatment effects under nonproportional hazards and dependent censoring. *Biometrics*, 67(1):29–38, 2011.

Man-Lai Tang and Sik-Yum Lee. Analysis of structural equation models with censored or truncated data via em algorithm. *Computational statistics & data analysis*, 27(1):33–46, 1998.

Lu Tian, Tianxi Cai, Els Goetghebeur, and LJ Wei. Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika*, 94(2):297–311, 2007.

Lu Tian, Lihui Zhao, and L J Wei. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics*, 15(2):222–233, 2014.

Hajime Uno, Tianxi Cai, Michael J Pencina, Ralph B D'Agostino, and LJ Wei. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*, 30(10):1105–1117, 2011.

Hans C Van Houwelingen. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, 34(1):70–85, 2007.

Hans C Van Houwelingen and Hein Putter. Comparison of stopped cox regression with direct methods such as pseudo-values and binomial regression. *Lifetime data analysis*, 21(2):180–196, 2015.

Xin Wang and Douglas E. Schaubel. Modeling restricted mean survival time under general censoring mechanisms. *Lifetime Data Analysis*, pages 1–24, 2017. ISSN 1572-9249. doi: 10.1007/s10985-017-9391-6.

Eric D Weinhandl, Robert N Foley, David T Gilbertson, Thomas J Arneson, Jon J Snyder, and Allan J Collins. Propensity-matched mortality comparison of incident hemodialysis and peritoneal dialysis patients. *Journal of the American Society of Nephrology*, 21(3):499–506, 2010.

Russell Wiesner, Erick Edwards, Richard Freeman, Ann Harper, Ray Kim, Patrick Kamath, Walter Kremers, John Lake, Todd Howard, Robert M Merion, et al. Model for end-stage liver disease (meld) and allocation of donor livers. *Gastroenterology*, 124(1):91–96, 2003.

J.M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. Econometric Analysis of Cross Section and Panel Data. MIT Press, 2010.

Fang Xiang and Susan Murray. Restricted mean models for transplant benefit and urgency. *Statistics in medicine*, 31(6):561–576, 2012.

Ronghui Xu and John O'Quigley. Estimating average regression effect under non-proportional hazards. *Biostatistics*, 1(4):423–439, 2000.

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

Lihui Zhao, Brian Claggett, Lu Tian, Hajime Uno, Marc A Pfeffer, Scott D Solomon, Lorenzo Trippa, and LJ Wei. On the restricted mean survival time curve in survival analysis. *Biometrics*, 2015.

Cheng Zheng, Ran Dai, Parameswaran N Hari, and Mei-Jie Zhang. Instrumental variable with competing risk model. *Statistics in Medicine*, 2017.

David M Zucker. Restricted mean life with covariates: modification and extension of a useful survival analysis method. *Journal of the American Statistical Association*, 93(442):702–709, 1998.