SEMIPARAMETRIC ESTIMATORS IN COMPETING RISKS REGRESSION

by

Jonathan G. Yabes

B.S. Statistics, University of the Philippines, Philippines, 2003

M.S. Statistics, University of the Philippines, Philippines, 2006

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

Jonathan G. Yabes

It was defended on

July 25, 2012

and approved by

Dissertation Advisor:
Chung-Chou H. Chang, PhD
Associate Professor
Departments of Medicine and Biostatistics
School of Medicine and Graduate School of Public Health
University of Pittsburgh

Committee Member:
Jong-Hyeon Jeong, PhD
Associate Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Gong Tang, PhD
Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Mark Unruh, MD
Associate Professor
Department of Medicine
School of Medicine
University of Pittsburgh

Committee Member:
Lisa A. Weissfeld, PhD
Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

SEMIPARAMETRIC ESTIMATORS IN COMPETING RISKS REGRESSION

Jonathan G. Yabes, PhD

University of Pittsburgh, 2012

Clinical trials and cohort studies that collect survival data frequently involve patients who may fail from one of multiple causes (failure types). These causes are called competing risks. The cumulative incidence function (CIF), or subdistribution, is a commonly reported quantity that describes the crude failure type-specific probability of the study population. The proportional subdistribution hazards model has been widely applied to study the effects of covariates on the CIF. In practice however, the time of failure may be recorded but the cause may be unknown or missing. To avoid bias, we developed two semiparametric estimators of covariate effects: the inverse probability weighted (IPW) estimator and the augmented inverse probability weighted (AIPW) estimator. We showed that these estimators are consistent and asymptotically normal. Their finite sample size properties and robustness were demonstrated through simulations. In many situations, investigators are interested in the marginal survival distribution of latent failure times, rather than the CIF. Because of the identifiability problem in competing risks, we derived an estimator of covariate effects in the Cox proportional hazards model by incorporating the random signs censoring (RSC) principle, which assumes that the main event failure time is independent of the indicator that the main event precedes the competing event. Unlike identifying assumptions that are typically imposed in practice, RSC is verifiable via stochastic ordering in the observed data. We further relaxed the RSC assumption by positing that independence is achieved conditional on some covariates. We showed that the resulting estimator is not only easy to implement but also has desirable asymptotic properties. We evaluated the estimator's finite sample

size performance through simulations. Medical datasets were used to illustrate the proposed methods. **Public Health Significance:** Biomedical and public health studies with time-to-event endpoint are abundant and often influence regulatory decisions. Trustworthiness of the research results not only relies on the design quality, but also on the soundness of the analytical approach used. The methodologies we propose account for two potential sources of bias in the conduct of such studies – competing risks and missing data.

Keywords: Competing risks; cumulative incidence function; doubly-robust; inverse probability weighting; missing cause of failure; marginal survival function; proportional subdistribution hazards; random signs censoring.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Clinical trials and cohort studies that collect survival data frequently involve subjects that are at risk of failing from one of multiple causes or types. Examples in the literature in which these competing risks exist are abundant and diverse, including behavioral studies looking at time to reconviction due to different offenses and transplant studies that examine various causes of mortality.

The analysis of competing risks data is complicated by the well known issue of non-identifiability, i.e, there exists both an independent risks model and dependent risks model that would lead to the same distribution of the observable data (Tsiatis, 1975). This implies that the observed data is insufficient to identify the marginal distribution of the event type of interest. One can either focus on crude quantities to avoid the identifiability issue altogether, or assume a dependence structure among the risks to make the estimation of the marginal distribution possible.

In contexts wherein the focus is on one of the several failure types, it is natural to consider models that relate the event of interest to some covariates. Traditional approaches focus on modeling the crude quantities such as the cause-specific hazards (CSH) or the cumulative incidence function (CIF). The CSH represents the instantaneous failure rate of the cause of interest given that no failure of any type was observed in the past. If however a probabilistic interpretation is desired, the cumulative incidence function (CIF), also called the subdistribution, is of interest. This quantity is defined as the cumulative probability of failing from the event of interest by time t while accounting for the presence of other types of failure.

The subdistribution often appeals to clinicians because of its interpretability. One of the earliest and most widely used modeling approaches is the proportional subdistribution

hazards model proposed by Fine and Gray (1999). It has gained popularity in the medical literature due to its resemblance with the ordinary Cox model.

In practice, however, a subject's failure time may be observed but the cause may be unavailable or not known exactly. This renders the standard estimation approach under the Fine and Gray model invalid. In the first part of this dissertation, we adopt some techniques used in the missing data literature to derive semiparametric estimators that can adequately handle this issue.

Often, the scientific question involves characterizing the "pure" or "net" effect of some covariates on survival, i.e., the covariate effect when the competing event is removed or is absent. In this situation, the crude quantities are, in general, not appropriate. Instead, identifying assumptions on the relationship between the competing risks time must be invoked to make a marginal analysis plausible. Independence is one such assumption that is commonly applied. This allows the use of standard survival analysis techniques to get consistent estimates of the marginal survival distribution. This is the only case where the CSH is equivalent to the marginal hazards. Other assumptions assume multivariate lifetime distributions (Moeschberger, 1974), random frailties (Clayton, 1978; Oakes et al., 1982), or copulas (Zheng and Klein, 1995; Escarela and Carriere, 2003; Lo and Wilke, 2010; Chen, 2010) to define the relationship between the competing events. These assumptions, however, are unverifiable from the observed data.

In the second part of this dissertation, we aim to develop an analysis approach that can accommodate positively dependent competing risks. We consider a semiparametric Cox model under the assumption of random signs censoring (RSC) (Cooke, 1993, 1996). RSC has been used in reliability studies focusing on a component failure in the presence of a competing event such as preventive maintenance. It posits that the main event failure time is independent of the event that it is observed. In other words, the probability that the main event is preceded by the occurrence of a competing event does not depend on the time at which the main event failure would occur. RSC can further be relaxed by assuming that the main event failure time and the its indicator are related, but this dependence can be explained by a set of common covariates. We refer to this as the conditional random signs censoring (CRSC) assumption.

To summarize, the goals of this dissertation are two-fold:

1. propose semiparametric estimators that are theoretically valid to handle missing cause of failure under the proportional subdistribution hazards model; and
2. propose a marginal analysis approach in the presence of positively related competing risks that satisfies the random signs censoring or conditional random signs censoring assumption.

2.0 PROPORTIONAL SUBDISTRIBUTION HAZARDS REGRESSION WITH MISSING CAUSE OF FAILURE

2.1 INTRODUCTION

In trials that involve assessing the treatment effect on the risk of failing from one of several competing risks, it is not uncommon to have the cause of failure be unknown or missing for some individuals. This occurs when the information about the cause responsible for the failure was poorly documented or handled, or when the diagnosis of the cause of death is difficult for some patients (Andersen et al., 1996). Heuristic and nonheuristic approaches have been used to deal with this problem. Heuristic approaches involve either a complete-case analysis (CC) or an analysis in which an extra category (EC) is formed for the missing cases. Albeit convenient, these methods could lead to substantial bias (Bakoyannis et al., 2010).

Many of the nonheuristic approaches have been developed under the framework of cause-specific hazards (CSH) regression. Under a proportional hazards assumption, Goetghebeur and Ryan (1995) proposed a model relating the CSH of the event of interest to that of the competing events, while Lu and Tsiatis (2001) suggested a multiple imputation (MI) approach. Using the inverse probability weighted (IPW) and augmented IPW (AIPW) estimators, Gao and Tsiatis (2005) focused on linear transformation models, whereas Lu and Liang (2008) worked on the additive hazards model. All of these strategies assume that the cause of failure is missing at random (MAR) (Rubin, 1976). Unfortunately, these methods are not directly applicable to modeling the subdistribution, which is also known as the cumulative incidence function.

The subdistribution is frequently used by clinical researchers because its interpretation is straightforward. However, if we wish to model it directly in settings where the cause of failure is not observed for all subjects, the available methods are relatively limited. Recently, Bakoyannis et al. (2010) adopted an MI approach under the proportional subdistribution hazards model of Fine and Gray (1999); their simulations showed that this technique outperforms the heuristic methods with respect to bias under different missingness assumptions.

In this chapter, our objective is to apply semiparametric theory to derive two additional estimators for proportional subdistribution hazards regression with missing cause of failure. By using inverse weighting and augmented inverse weighting, the proposed estimators yield theoretically valid estimates under the MAR assumption. We study their properties and conduct simulations to compare their performance with that of estimators obtained via the MI method of Bakoyannis et al. (2010), via a naïve complete-case analysis, and via a method in which an extra category is formed for the missing cases.

The motivation of this work ensues from data which consists of consecutive kidney transplant recipients, transplanted between January 2001 and January 2009 at the University of Pittsburgh Starzl Transplant Institute. Induction therapies are used to reduce the risk of acute rejection or infection in transplant recipients. Due to the immunosuppressive nature of these drugs, they may be linked to increased vulnerability to cancer. Thus, the aim is to examine the association between induction therapy and the risk of malignancy or malignancy-related death. Three induction therapy groups were compared: no induction, Alemtuzumab-, and Thymoglobulin-based protocols. Graft failure or death unrelated to malignancy constitutes a competing event. A complication that needs to be addressed, however, is that the cause of death cannot be ascertained for some subjects.

This article is organized as follows. In Section 2.2, we introduce some notations and revisit the proportional subdistribution hazards model. In Sections 2.3 and 2.4, we present the proposed IPW estimator and AIPW estimator, respectively, and discuss their asymptotic properties. In Section 2.5, we assess and compare through simulations the performance of the two estimators with that of the current approaches. In Section 2.6, we apply the proposed methods to analyze the induction therapy data, then conclude in the final section.

2.2 NOTATIONS AND MODEL

Suppose we have n independent individuals in a study. Let T be the failure time and ε be the failure type. Without loss of generality, consider two failure types where $\varepsilon = 1$ represents the cause of interest, and $\varepsilon = 2$ the competing event. When there are no missing cases, for each subject i we observe $\{X_i, \varepsilon_i, \boldsymbol{Z}_i\}$, where $X_i = T_i \wedge C_i$; C_i is the censoring time assumed to be independent of T_i; and $\boldsymbol{Z}_i$ is a p-dimensional vector covariates. For simplicity, we restrict $\boldsymbol{Z}$ to be fixed covariates although the results can be extended to external time-dependent covariate processes (Kalbfleisch and Prentice, 2002). Here we define $\varepsilon_i = 0$ whenever $X_i = C_i$. The proportional sudistribution hazards model of Fine and Gray (1999) has the form

$$\lambda_1(t|\boldsymbol{Z}) = \lambda_{10}(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}), \tag{2.1}$$

where the baseline subdistribution hazards $\lambda_{10}(\cdot)$ is left unspecified. This reminds us of the regular Cox model except that it models the event-1 subdistribution hazards $\lambda_1(\cdot)$ rather than the marginal hazards. The corresponding subdistribution can be calculated as

$$F_1(t|\boldsymbol{Z}) = 1 - \exp\left\{ - \int_0^t \lambda_{10}(u) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \, du \right\}.$$

The regression coefficients are estimated through a partial likelihood approach with modified risk sets. In particular, when there is no censoring, the risk set at time t includes both individuals who have yet to fail from the event of interest and those that have already failed from the competing cause prior to time t. When only administrative censoring is present, that is, the potential censoring time is known for all individuals, the risk sets are redefined such that an individual who failed from the competing event remains at risk only up to his potential censoring time. In the presence of random right censoring, the inverse probability of censoring weighting (IPCW) technique (Robins and Rotnitzky, 1992) is utilized to reweight subjects who experienced the competing event. More formally, by

defining at time t the counting process $N_i(t) = I(T_i \leq t, \varepsilon_i = 1)$, and the at risk process $Y_i(t) = I(T_i < t, \varepsilon_i = 2) + I(T_i \geq t)$, the reweighted at risk process for subject i is given by

$$\begin{aligned}\omega_i(t) &= I(C_i \geq T_i \wedge t)\frac{\hat{G}(t)}{\hat{G}(T_i \wedge t)}Y_i(t) \\ &= \frac{\hat{G}(t)}{\hat{G}(X_i)}I(X_i < t, \varepsilon_i = 2) + I(X_i \geq t),\end{aligned} \tag{2.2}$$

where $\hat{G}(t)$ is usually taken to be the Kaplan-Meier estimator of the censoring survival distribution. The weighted partial likelihood score equation takes the form

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\int_0^{\tau}\left\{\boldsymbol{Z}_i - \frac{\sum_j \omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j}{\sum_j \omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)}\right\}\omega_i(t)dN_i(t), \tag{2.3}$$

where $\tau = sup\{t : \Pr(\omega(t) \geq \epsilon > 0) > 0\}$. The estimator $\hat{\boldsymbol{\beta}}$ can be obtained by finding the root of $\boldsymbol{U}(\boldsymbol{\beta})$. It has been shown that $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal for the true parameter value $\boldsymbol{\beta}_0$ (Fine and Gray, 1999; Geskus, 2011).

2.3 INVERSE PROBABILITY WEIGHTED ESTIMATOR

2.3.1 Estimating Equations

The estimation procedure discussed above becomes inadequate when the cause of failure is not observed for all subjects. Analogous to the Horvitz-Thompson inverse selection probability technique (Horvitz and Thompson, 1952), we propose an IPW score equations for such settings. The basic idea is to recreate a random sample of the population to correct the selection bias that might have been induced by the missingness process. This is accomplished by upweighting subjects that are less likely to be observed based on some background characteristics. To make this approach plausible, we assume that the failure type is MAR, that is, the missingness mechanism depends only on fully observed quantities and not on the unobserved ε. This assumption has been widely employed in the missing data literature and is also the basis of many other missing data methods such as MI.

Let R_i be the complete-data indicator, that is, $R_i = 1$ if ε_i is observed and $R_i = 0$ if it is missing. We implicitly assume that censoring status is always observed so that we take $R_i = 1$ whenever $\varepsilon_i = 0$. Auxiliary covariates, denoted by $\boldsymbol{A}_i$, may also have been collected for each subject. These variables are not used to model the subdistribution but may be needed to fully account for the missingness process. The MAR assumption gives

$$\Pr(R_i = 1|\varepsilon_i, \varepsilon_i > 0, \boldsymbol{W}_i) = \Pr(R_i = 1|\varepsilon_i > 0, \boldsymbol{W}_i) \stackrel{set}{=} \pi_0(\boldsymbol{W}_i),$$

where $\boldsymbol{W}_i = (X_i, \boldsymbol{Z}_i, \boldsymbol{A}_i)$ are the fully observed variables, that is, R and ε are independent conditional on $\boldsymbol{W}$. This coupled with (2.3) leads us to consider the following IPW partial likelihood equations:

$$\boldsymbol{U}_w(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i)} \int_0^\tau \left\{ \boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, t) \right\} \omega_i(t) dN_i(t), \tag{2.4}$$

where

$$\begin{aligned}
\pi(\boldsymbol{W}_i) &= \Pr(R_i = 1|\boldsymbol{W}_i) \\
&= \pi_0(\boldsymbol{W}_i) I(\varepsilon_i > 0) + I(\varepsilon_i = 0), \\
\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, t) &= S_w^{(1)}(\boldsymbol{\beta}, t)/S_w^{(0)}(\boldsymbol{\beta}, t), \\
S_w^{(k)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{j=1}^{n} \frac{R_j}{\pi(\boldsymbol{W}_j)} \omega_j(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j) \boldsymbol{Z}_j^{\otimes k} \quad \text{for } k = 0, 1, 2
\end{aligned}$$

with $\boldsymbol{a}^{\otimes 0} = 1, \boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. By finding the root of $\boldsymbol{U}_w(\boldsymbol{\beta})$, we obtain the IPW estimator $\hat{\boldsymbol{\beta}}^w$ under the true $\pi(\boldsymbol{W})$.

Before we present the asymptotic properties of $\hat{\boldsymbol{\beta}}^w$, a few more notations are needed. Let $N^c(t) = \omega(t)N(t)$ be the counting process that incorporates the censoring information, and let $M^c(t) = N^c(t) - \int_0^t \omega(u) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) d\Lambda_{10}(u)$. In addition, let

$$\begin{aligned}
\boldsymbol{s}^{(k)}(\boldsymbol{\beta}, t) &= E\{\omega(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \boldsymbol{Z}^{\otimes k}\}, \qquad k = 0, 1, 2, \\
\boldsymbol{e}(\boldsymbol{\beta}, t) &= \boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)/s^{(0)}(\boldsymbol{\beta}, t), \\
\boldsymbol{v}(\boldsymbol{\beta}, t) &= \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} - \left\{ \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} \right\}^{\otimes 2}, \quad \text{and} \\
\mathbf{M} &= \int_0^\tau \{\boldsymbol{Z} - \boldsymbol{e}(\boldsymbol{\beta}, t)\} dM^c(t).
\end{aligned}$$

Geskus (2011) showed that $M^c(t)$ is a martingale. Note that $dN_i^c(t)$ can replace $\omega_i(t)dN_i(t)$ in (2.4) because $\omega_i(t) = 1$ whenever $N_i(t)$ has a jump.

THEOREM 1. *Under the regularity conditions given in the Appendix, $\hat{\boldsymbol{\beta}}^w$ is consistent for the true parameter $\boldsymbol{\beta}_0$ and*

$$n^{1/2}(\hat{\boldsymbol{\beta}}^w - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1}\boldsymbol{\Xi}_w\boldsymbol{\mathcal{I}}_\beta^{-1}),$$

where $\boldsymbol{\Xi}_w = E\{\pi^{-1}(\boldsymbol{W})\mathbf{M}^{\otimes 2}\}$ and $\boldsymbol{\mathcal{I}}_\beta = \int_0^\infty \boldsymbol{v}(\boldsymbol{\beta}, t)s^{(0)}(\boldsymbol{\beta}, t)d\Lambda_{10}(t)$.

$\boldsymbol{\mathcal{I}}_\beta$ and $\boldsymbol{\Xi}_w$ can be consistently estimated by, respectively, $\hat{\boldsymbol{\mathcal{I}}}_\beta = -\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_w(\hat{\boldsymbol{\beta}}^w)$ and

$$\hat{\boldsymbol{\Xi}}_w = \frac{1}{n}\sum_{i=1}^n \left\{\frac{R_i}{\pi(\boldsymbol{W}_i)}\hat{\mathbf{M}}_{w,i}(\hat{\boldsymbol{\beta}}^w)\right\}^{\otimes 2},$$

where $\hat{\mathbf{M}}_{w,i}(\boldsymbol{\beta}) = \int_0^\tau \{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, t)\}d\hat{M}_i^c(t)$, $d\hat{M}_i^c(t) = dN_i^c(t) - \omega_i(t)\exp(\hat{\boldsymbol{\beta}}^{wT}\boldsymbol{Z}_i)d\hat{\Lambda}_{10}(t)$, and $d\hat{\Lambda}_{10}(t) = [nS_w^{(0)}(\hat{\boldsymbol{\beta}}^w, t)]^{-1}\sum_j R_j\pi^{-1}(\boldsymbol{W}_j)dN_j^c(t)$.

In most situations, $\pi_0(\boldsymbol{W})$ is unknown and must be estimated from the data that are observed for everyone. To avoid the curse of dimensionality, we posit a parametric model $\pi_0(\boldsymbol{W};\boldsymbol{\gamma})$ for $\pi_0(\boldsymbol{W})$, where $\boldsymbol{\gamma}$ is a finite dimensional parameter. Although other parametric models can be employed, a logistic regression model of the form $\pi_0(\boldsymbol{W};\boldsymbol{\gamma}) = 1/(1+\exp\{-\boldsymbol{\gamma}^T\widetilde{\boldsymbol{W}}\})$ with $\widetilde{\boldsymbol{W}} = (1, \boldsymbol{W}^T)^T$ is often adopted since R_i is binary. Under correct model specification for $\pi_0(\boldsymbol{W};\boldsymbol{\gamma})$, a consistent estimator $\hat{\boldsymbol{\gamma}}$ of the true value $\boldsymbol{\gamma}_0$ of $\boldsymbol{\gamma}$ may be obtained via maximum likelihood estimation (MLE). Equation (2.4) then becomes

$$\boldsymbol{U}_w(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^n \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\int_0^\tau \left\{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t)\right\}dN_i^c(t), \tag{2.5}$$

where $\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t) = S_w^{(1)}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t)/S_w^{(0)}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t)$ and for $k = 0, 1, 2$, $S_w^{(k)}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t)$ has the same form as $S_w^{(k)}(\boldsymbol{\beta}, t)$ but with $\pi(\boldsymbol{W}_j)$ replaced by $\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\gamma}})$. The solution to $\boldsymbol{U}_w(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) = 0$ gives the IPW estimator $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$ when π is estimated.

THEOREM 2. *Under the regularity conditions (see Appendix) and if $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ is correctly specified, $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$ is consistent for $\boldsymbol{\beta}_0$ and has the same asymptotic distribution as $\hat{\boldsymbol{\beta}}^w$.*

Theorem 2 implies that, unlike other settings in which IPW is used, we gain no efficiency by using the estimated rather than the true (if known) complete-data probabilities as inverse weights. The reason behind this is that the MAR assumption renders $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ predictable, leaving the covariance between the terms in the score equations for estimating $\boldsymbol{\beta}$ and that of $\boldsymbol{\gamma}$ equal to 0. Thus, to obtain a consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$, we simply replace $\pi(\boldsymbol{W}_i)$ with its estimate $\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})$ and $\hat{\boldsymbol{\beta}}^w$ with $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$ in computing $\hat{\boldsymbol{\mathcal{I}}}_\beta$ and $\hat{\boldsymbol{\Xi}}_w$. This result conveniently enables us to carry out the IPW method in standard software, which normally accepts fixed weights as input, and still get valid standard errors even when the weights were estimated. Details on the implementation are described next.

2.3.2 Software Implementation

Model (2.1) for data with complete information on the cause of failure can be fitted several ways using standard software. In R, `crr` in the `cmprsk` library (Gray, 2010) can conveniently be used. However, it does not accommodate weighted analysis. More flexible approaches that involve some data preprocessing to permit the use of standard Cox regression commands (e.g., `coxph` in `R` or `proc phreg` in `SAS`) have been suggested. One method involves multiply-imputing censoring times for competing events using the Kaplan-Meier (Ruan and Gray, 2008). Another suggests explicitly setting up the data to include time-varying IPCW weights for those that experienced any of the competing events (Geskus, 2011). The advantage of the latter is that it is amenable to left-censored data.

We adopt the second approach in the discussion below. The basic idea is to run a Cox regression on the event of interest with case weights equal to the product of the IPCW and IPW, but with the standard errors computed using the robust sandwich estimator. The following steps can be followed:

1. Set the IPW weights for the censored subjects to `w.ipw` $= 1$. Among the uncensored subjects, fit a logistic model to get $\pi_0(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})$ and set `w.ipw` $= 1/\pi_0(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})$.
2. Obtain the Kaplan-Meier estimate $\hat{G}(t)$ of the censoring distribution.
3. Restrict the data to the complete cases. Table 1(a) shows an example of an artificial dataset for a subset of 5 subjects after performing the previous steps. Suppose the unique

event-1 failure times are $t_1 < t_2 < \ldots < t_D$. We then need to restructure the data into the counting process style of input. That is, create the variable `Tstart` to represent the start of the time interval and the variable `Tstop` to mark the end of the interval. A subject i who is censored or who experienced event-1 will have one record in this dataset, with `Tstart` $= 0$ and `Tstop` $= X_i$. For a subject i who experienced event-2, the record will be expanded to D rows, with the first row representing `Tstart` $= 0$ and `Tstop` $= t_1$, the second representing `Tstart` $= t_1$ and `Tstop` $= t_2$, and so forth. Next, the IPCW weights of censored or event-1 subjects will be set to 1, i.e., `w.cens` $= 1$. For event-2 subjects, set `w.cens` $= 1$ if $X_i >$ `Tstop` and `w.cens` $= \hat{G}($`Tstop`$)/\hat{G}(X_i)$ if $X_i <$ `Tstop`. Table 1(b) illustrates how the data would look like after this step.

4. Fit a Cox model for the main event using case weights equal to `w.ipw` $*$ `w.cens` and use a robust variance estimator for the standard errors. For instance in `R`, we would fit model (2.1) for the data in Table 1(b) using the command

```
coxph(Surv(Tstart,Tstop,etype==1) ~ Z + cluster(subject),
   data=wData, weight=w.ipw*w.cens).
```

   In `SAS`, this is equivalent to

```
proc phreg data=wData covsandwich(aggregate);
    model (Tstart,Tstop)*etype(0 2) = Z;
    id subject;
    weight = w.ipw*w.cens;
run;
```

2.4 AUGMENTED INVERSE PROBABILITY WEIGHTED ESTIMATOR

Although the IPW estimator is consistent when π is correctly specified, it may perform poorly otherwise. Moreover, it may lose efficiency by only using the complete cases. We can improve these limitations by adopting the augmented inverse probability weighting (AIPW) technique developed by Robins et al. (1994). This method augments the IPW score equations with a term that uses information from both the complete cases and missing cases. We continue

Table 1: Preparing data for inverse probability weighted analysis

(a) Sample dataset

subect	X	etype	w.ipw	Ghat	Z
1	2	0	1.00	0.9	3
2	4	1	1.25	0.8	6
3	6	2	1.50	0.7	9
4	8	1	1.75	0.6	12
5	10	1	2.00	0.5	15

(b) Sample dataset in counting process style of input

subject	X	Tstart	Tstop	etype	w.ipw	w.cens	Z
1	2	0	2	0	1.00	1.000	3
2	4	0	4	1	1.25	1.000	6
3	6	0	4	2	1.50	1.000	9
3	6	4	8	2	1.50	0.750	9
3	6	8	10	2	1.50	0.625	9
4	8	0	8	1	1.75	1.000	12
5	10	0	10	1	2.00	1.000	15

to assume MAR and begin by considering the probability of observing an event-1 failure conditional on the observed data:

$$\rho(\boldsymbol{W}_i) = \Pr\{\varepsilon_i = 1 | \boldsymbol{W}_i, \varepsilon_i > 0\}.$$

We can think of this as an imputation model for $\varepsilon = 1$. Now define $N_i^{\varepsilon}(t) = I(T_i \leq t) I(\varepsilon_i > 0)$. Notice that we can write $N_i(t) = I(\varepsilon_i = 1) N_i^{\varepsilon}(t)$ and $Y_i(t) = I(\varepsilon_i = 2) N_i^{\varepsilon}(t-) + I(T_i \geq t)$. It follows that

$$\begin{aligned} dN_i^c(t) &= \omega_i(t) I(\varepsilon_i = 1) dN_i^{\varepsilon}(t) \quad \text{and} \\ \omega_i(t) &= I(\varepsilon_i = 2) N_i^{\varepsilon}(t-) \hat{G}(t) / \hat{G}(X_i) + I(X_i \geq t). \end{aligned}$$

Also define $d\widetilde{N}_i^c(t) = E\{dN_i^c(t) | \boldsymbol{W}_i, \varepsilon_i > 0\}$ and $\widetilde{\omega}_i(t) = E\{\omega_i(t) | \boldsymbol{W}_i, \varepsilon_i > 0\}$. Direct calculation gives

$$\begin{aligned} d\widetilde{N}_i^c(t) &= \rho(\boldsymbol{W}_i) dN_i^{\varepsilon}(t), \quad \text{and} \\ \widetilde{\omega}_i(t) &= \{1 - \rho(\boldsymbol{W}_i)\} N_i^{\varepsilon}(t-) \hat{G}(t) / \hat{G}(X_i) + I(X_i \geq t). \end{aligned}$$

For the complete cases, we are able to observe $N_i^c(t)$ (and $\omega_i(t)$ also), but for subjects with unknown failure type, we only observe $N_i^{\varepsilon}(t)$. We can however compute $d\widetilde{N}_i^c(t)$ and $\widetilde{\omega}_i(t)$ for everyone. Thus the proposed AIPW estimating equations can be constructed as

$$\begin{aligned} \boldsymbol{U}_{aw}(\boldsymbol{\beta}) = &\sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i)} \int_0^{\tau} \left\{ \boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) \right\} dN_i^c(t) \\ &- \sum_{i=1}^{n} \frac{R_i - \pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)} \int_0^{\tau} \left\{ \boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) \right\} d\widetilde{N}_i^c(t), \end{aligned} \tag{2.6}$$

where $\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) = S_{aw}^{(1)}(\boldsymbol{\beta}, t) / S_{aw}^{(0)}(\boldsymbol{\beta}, t)$ and

$$\begin{aligned} S_{aw}^{(k)}(\boldsymbol{\beta}, t) = &\frac{1}{n} \sum_{j=1}^{n} \frac{R_j}{\pi(\boldsymbol{W}_j)} \omega_j(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j) \boldsymbol{Z}_j^{\otimes k} \\ &- \frac{1}{n} \sum_{j=1}^{n} \frac{R_j - \pi(\boldsymbol{W}_j)}{\pi(\boldsymbol{W}_j)} \widetilde{\omega}_j(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j) \boldsymbol{Z}_j^{\otimes k}, \qquad k = 0, 1, 2. \end{aligned}$$

Note that unlike $\boldsymbol{U}_w(\boldsymbol{\beta})$, $\boldsymbol{U}_{aw}(\boldsymbol{\beta})$ is augmented with a second term and in addition, uses augmented averages $S^{(0)}_{aw}(\boldsymbol{\beta}, t)$ and $\boldsymbol{S}^{(1)}_{aw}(\boldsymbol{\beta}, t)$, each taking contributions from both the complete and incomplete cases. For completeness of presentation, first we assume in Theorem 3 that π and ρ are known and provide the asymptotic properties of the resulting estimator, $\hat{\boldsymbol{\beta}}^{aw}$. In Theorem 4, we then examine the estimators obtained by using auxiliary models for π and ρ and show that they share the same asymptotic properties as $\hat{\boldsymbol{\beta}}^{aw}$ and posses the property of double-robustness.

THEOREM 3. *Under the regularity conditions (see Appendix),* $\hat{\boldsymbol{\beta}}^{aw}$ *is consistent for* $\boldsymbol{\beta}_0$ *and*

$$n^{1/2}(\hat{\boldsymbol{\beta}}^{aw} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1} \boldsymbol{\Xi}_{aw} \boldsymbol{\mathcal{I}}_\beta^{-1}),$$

where $\boldsymbol{\Xi}_{aw} = E\{\mathbf{M}^{\otimes 2}\} + E\{\frac{1-\pi(\boldsymbol{W})}{\pi(\boldsymbol{W})} Var(\mathbf{M}|\boldsymbol{W}, \varepsilon > 0)\}$.

In practice, $\rho(\boldsymbol{W})$ is estimated from the observed data. This probability can be determined from the relationship of the cause-specific hazards of the latent failure times of the main event T_1^* and competing event T_2^*, i.e.,

$$\frac{\lambda_1^{csh}(t|\boldsymbol{Z}, \boldsymbol{A})}{\lambda_1^{csh}(t|\boldsymbol{Z}, \boldsymbol{A}) + \lambda_2^{csh}(t|\boldsymbol{Z}, \boldsymbol{A})},$$

where $\lambda_j^{csh}(\cdot|\boldsymbol{Z}, \boldsymbol{A})$ is the conditional cause-specific hazards of T_j^*, $j = 1, 2$. Rather than modeling two separate cause-specific hazards, we posit a parametric model $\rho(\boldsymbol{W}; \boldsymbol{\eta})$ where $\boldsymbol{\eta}$ is of finite dimension. Again, we employ a standard logistic formulation in which $\rho(\boldsymbol{W}_i; \boldsymbol{\eta}) = 1/\{1 + \exp(-\widetilde{\boldsymbol{W}}_i^T \boldsymbol{\eta})\}$. In the presence of missingness, obtaining an estimate of $\boldsymbol{\eta}$ is problematic. MAR however implies

$$\begin{aligned}\Pr(\varepsilon_i = 1|\boldsymbol{W}_i, \varepsilon_i > 0) &= \Pr(\varepsilon_i = 1|\boldsymbol{W}_i, \varepsilon_i > 0, R_i = 0) \\ &= \Pr(\varepsilon_i = 1|\boldsymbol{W}_i, \varepsilon_i > 0, R_i = 1),\end{aligned}$$

signifying that we can estimate $\boldsymbol{\eta}$ from the uncensored complete cases. If $\rho(\boldsymbol{W}_i; \boldsymbol{\eta})$ is correctly specified, a consistent estimator for the true value $\boldsymbol{\eta}_0$ is the MLE $\hat{\boldsymbol{\eta}}$.

THEOREM 4. *Under the regularity conditions (see Appendix),* $\hat{\boldsymbol{\beta}}^{aw}(\boldsymbol{\gamma}, \hat{\boldsymbol{\eta}}), \hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \boldsymbol{\eta})$ *and* $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}})$ *are consistent for* $\boldsymbol{\beta}_0$ *and have the same asymptotic distribution as* $\hat{\boldsymbol{\beta}}^{aw}$ *as long as either* $\pi(\boldsymbol{W}; \boldsymbol{\gamma})$ *or* $\rho(\boldsymbol{W}; \boldsymbol{\eta})$ *is specified correctly.*

Consistent estimation of the variances in Theorems 3 and 4 is very similar to that of the IPW estimators. Thus, we only demonstrate this for $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}})$. First we set an augmented Breslow-type estimator for $d\Lambda_{10}(t)$:

$$d\hat{\Lambda}_{10}(t) = \frac{1}{nS_{aw}^{(0)}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}), \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, t)} \sum_{j=1}^{n} \left\{ \frac{R_j}{\pi(\boldsymbol{W}_j; \hat{\boldsymbol{\gamma}})} dN_j^c(t) - \frac{R_j - \pi(\boldsymbol{W}_j; \hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_j; \hat{\boldsymbol{\gamma}})} d\widetilde{N}_j^c(t) \right\}. \quad (2.7)$$

Also define

$$\begin{aligned}
d\hat{M}_i^c(t) &= dN_i^c(t) - \omega_i(t) \exp\{[\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}})]^T \boldsymbol{Z}_i\} d\hat{\Lambda}_{10}(t), \\
d\widehat{\widetilde{M}}_i^c(t) &= d\widetilde{N}_i^c(t; \hat{\boldsymbol{\eta}}) - \widetilde{\omega}_i(t; \hat{\boldsymbol{\eta}}) \exp\{[\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}})]^T \boldsymbol{Z}_i\} d\hat{\Lambda}_{10}(t), \\
\hat{\mathbf{M}}_{aw,i} &= \int_0^\tau \{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{aw}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}), \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, t)\} d\hat{M}_i^c(t), \quad \text{and} \\
\widehat{\widetilde{\mathbf{M}}}_{aw,i} &= \int_0^\tau \{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{aw}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}), \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, t)\} d\widehat{\widetilde{M}}_i^c(t),
\end{aligned}$$

with $d\widetilde{N}_i^c(t; \hat{\boldsymbol{\eta}}) = \rho(\boldsymbol{W}_i; \hat{\boldsymbol{\eta}}) dN_i^\varepsilon(t)$ and $\widetilde{\omega}_i(t; \hat{\boldsymbol{\eta}}) = \{1 - \rho(\boldsymbol{W}_i; \hat{\boldsymbol{\eta}})\} N_i^\varepsilon(t-) \hat{G}(t)/\hat{G}(X_i) + I(X_i \geq t)$. Then a consistent variance estimator is $\hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\beta}}^{-1} \hat{\boldsymbol{\Xi}}_{aw} \hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\beta}}^{-1}$ where $\hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\beta}} = -\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{U}_{aw}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}))$ and

$$\hat{\boldsymbol{\Xi}}_{aw} = \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i; \hat{\boldsymbol{\gamma}})} \hat{\mathbf{M}}_{aw,i}^{\otimes 2} + \frac{1}{n} \sum_{i=1}^{n} \frac{R_i \left(1 - \pi(\boldsymbol{W}_i; \hat{\boldsymbol{\gamma}})\right)}{\pi^2(\boldsymbol{W}_i; \hat{\boldsymbol{\gamma}})} (\hat{\mathbf{M}}_{aw,i} - \widehat{\widetilde{\mathbf{M}}}_{aw,i})^{\otimes 2}.$$

Although Equation (2.7) is doubly-robust, the estimated cumulative baseline subdistribution hazards function may not be monotonically increasing. The numerator cannot be guaranteed to be positive, particularly among subjects that were observed to be fail from the competing event. This suggests that an estimator in which the denominator is augmented, but the numerator is not. That is,

$$d\hat{\Lambda}_{10}(t) = \frac{1}{nS_{aw}^{(0)}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}), \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, t)} \sum_{j=1}^{n} \frac{R_j dN_j^c(t)}{\pi(\boldsymbol{W}_j; \hat{\boldsymbol{\gamma}})}. \quad (2.8)$$

Alternatively, we can also use the simple IPW estimator of the baseline subdistribution hazards:

$$d\hat{\Lambda}_{10}(t) = \frac{1}{nS_w^{(0)}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}}), \hat{\boldsymbol{\gamma}}, t)} \sum_{j=1}^{n} \frac{R_j dN_j^c(t)}{\pi(\boldsymbol{W}_j; \hat{\boldsymbol{\gamma}})}. \tag{2.9}$$

As part of the simulation studies in the next section, we will examine the finite sample properties and robustness of (2.7) – (2.9).

2.5 SIMULATION STUDIES

In this section we conduct simulations to assess the properties of the IPW and AIPW estimators for small to moderate sample sizes. We also compare their performance to the full cohort (no missing cases), the compete-case analysis, extra-category analysis, and multiple imputation approach that uses the `R` package `mitools` (Lumley, 2010). Estimators of the cumulative baseline subdistribution hazards proposed in the previous section are also examined.

Sample sizes of $n = 200$ or 500 were chosen. Two failure types were considered, both of which depend on a single covariate Z. Z_i was generated from a balanced Bernoulli distribution, then the failure time for the event of interest $(T_i, \varepsilon_i = 1)$ was drawn from the subdistribution given by $F_1(t|Z_i) = 1 - [1 - p\{1 - \exp(-t)\}]^{\exp(\beta_1 Z_i)}$. We set $p = 0.3$ to give about 30% type-1 failures when $Z = 0$ in the absence of censoring. Taking $\Pr(\varepsilon_i = 2|Z_i) = 1 - \Pr(\varepsilon_i = 2|Z_i)$, type-2 failure times were then generated from the subdistribution obtained from $\Pr(T_i \leq t|\varepsilon_i = 2, Z_i) = 1 - \exp\{-\exp(\beta_2 Z_i)t\}$. In all the simulations, we chose $\beta_1 = 0.5$ and $\beta_2 = -0.5$. The censoring times were independently simulated from Uniform(1,2) which produced 28% censoring on the average.

We assumed that the missingness mechanism follows a logistic model with complete-data probability that depends on the observed time X_i, the covariate Z_i, an auxiliary variable A_i drawn from a standard normal distribution, and the two-way of the covariates interactions with X_i. That is, $\pi_0(\boldsymbol{W}_i; \boldsymbol{\gamma}) = \{1+\exp(-\boldsymbol{\gamma}^T \widetilde{\boldsymbol{W}}_i)\}^{-1}$ where $\widetilde{\boldsymbol{W}}_i = (1, X_i, Z_i, A_i, X_i Z_i, X_i A_i)^T$.

We set $\boldsymbol{\gamma} = (\gamma_0, 1, -0.5, 0.25, -0.5, -0.25)^T$, producing about 46% and 24% missing among the uncensored cases when the intercept are set to $\gamma_0 = 0$ and $\gamma_0 = 1$ respectively.

To derive the IPW and AIPW estimators, we fitted three models for π with varying levels of misspecifiaction with respect to the set of covariates included: (1) a correct model, (2) a main effects model, and (3) an intercept model. For comparison, we also included an IPW analysis that uses the true complete-data probability. We mimicked the set-up from the IPW fitted models to build models for ρ. That is, we also assumed that $\rho(\widetilde{\boldsymbol{W}}_i; \boldsymbol{\eta}) = \{1 + \exp(-\boldsymbol{\eta}^T \widetilde{\boldsymbol{W}}_i)\}^{-1}$ where (1) $\widetilde{\boldsymbol{W}}_i = (1, X_i, Z_i, A_i, X_i Z_i, X_i A_i)$, (2) $\widetilde{\boldsymbol{W}}_i = (1, X_i, Z_i, A_i)$, or (3) $\widetilde{\boldsymbol{W}}_i = 1$. These were also the models used to derive the MI estimates. Since the true ρ is a function of the two conditional CSH, note that the first fitted imputation model, though uses the correct set of covariates, is misspecified. This is a scenario that is more likely encountered in practice. There were a total of 9 AIPW fitted models comprising the different combinations of the models for π and ρ. We used (2.8) in the calculation of the standard errors of the AIPW estimators. Table 2 provides a summarized description of the methods being compared in this simulation study.

We calculated the bias, standard deviation of the estimates (SD), average of the standard error estimates (SE), and empirical coverage probability (CP) of the sample 95% confidence intervals from the 1000 simulated datasets. The results under 46% missing cases are shown in Table 3.

As expected, the full cohort analysis produced virtually unbiased estimates and coverage probability close to the nominal level. The naïve methods CC and EC both exhibited substantial bias and severe under coverage. Despite an incorrect imputation model, the MI with correct set of covariates (MI_0) performed reasonably well in terms of bias. The estimated standard errors were too conservative though. A low level of misspecification (MI_1) with respect to the set of covariates used appears to be tolerated by the MI approach in which the bias that was still very low. However the bias becomes substantial under an intercept only model (MI_2).

The simulation results confirm the equivalence of IPW using true probabilities (IPW_{true}) and that using probabilities of complete-data estimated from a correct model (IPW_0), both of which showed unbiasedness and correct coverage level. Although the IPW that uses

estimated weights from incorrect models (IPW_1 and IPW_2) showed some bias and slight undercoverage, they still performed better than either CC or EC. Overall, the AIPW estimators performed the best. Interestingly, even the AIPW estimator (AIPW_{22}) that uses constant (and highly misspecified) model for both π and ρ still exhibited low bias and good coverage.

If the percent of missing cases was reduced to 24% (see Table 4), we can draw the same conclusions as above. The noticeable difference is the reduction in the standard errors in the missing data methods due to the increase in effective sample size being used, and the reduction in bias in the naïve methods.

We now compare the three estimators of $\Lambda_{10}(t)$ as given in Equations (2.7) – (2.9). We use the same set-up as above focusing on the case in which there were about 46% missing cases. We calculated the relative bias at three timepoints, $t = 0.5, 1.0$, and 1.5, at which the true values of the cumulative subdistribution hazards are $\Lambda_{10}(0.5) = 0.126, \Lambda_{10}(1.0) = 0.210$, and $\Lambda_{10}(1.5) = 0.265$. The results are shown in Table 5. In general, the relative bias of the three estimators are comparable and were low under correct specification or low misspecification of the the complete-data probability model. Estimator (2.7) displayed some advantage in smaller sample sizes, in earlier timepoints, and when the model for π is severely misspecified. The performance of (2.8) and (2.9) were very similar.

Table 6 compares how these three estimators affect the AIPW standard error estimates (SE) of the regression parameters. The SE's and empirical coverage levels (CP) produced by (2.7) – (2.9) are nearly identical. This suggests that any of them can be used when the focus is on making inference on the regression parameters. If one wishes to estimate the subdistribution, (2.7) may provide more robust estimates pointwise. When preservation of monotonicity is desired, either (2.8) or (2.9) can be used making every effort to specify the complete-data probability model correctly.

Table 2: Fitted Models

Fitted Model	Description	Terms for logit(π)	Terms for logit(ρ)
FC	Full Cohort	–	–
CC	Complete Case	–	–
EC	Extra Cause	–	–
MI_0	Misspecified	–	$1, X, Z, A, XZ, XA$
MI_1	Misspecified: low	–	$1, X, Z, A$
MI_2	Misspecified: high	–	1
IPW_{true}	True probabilities	$1, X, Z, A, XZ, XA$	–
IPW_0	Correctly specified	$1, X, Z, A, XZ, XA$	–
IPW_1	Misspecified: low	$1, X, Z, A$	–
IPW_2	Misspecified: high	1	–
AIPW_{00}	$\text{IPW}_0, \text{MI}_0$	$1, X, Z, A, XZ, XA$	$1, X, Z, A, XZ, XA$
AIPW_{01}	$\text{IPW}_0, \text{MI}_1$	$1, X, Z, A, XZ, XA$	$1, X, Z, A$
AIPW_{02}	$\text{IPW}_0, \text{MI}_2$	$1, X, Z, A, XZ, XA$	1
AIPW_{10}	$\text{IPW}_1, \text{MI}_0$	$1, X, Z, A$	$1, X, Z, A, XZ, XA$
AIPW_{11}	$\text{IPW}_1, \text{MI}_1$	$1, X, Z, A$	$1, X, Z, A$
AIPW_{12}	$\text{IPW}_1, \text{MI}_2$	$1, X, Z, A$	1
AIPW_{20}	$\text{IPW}_2, \text{MI}_0$	1	$1, X, Z, A, XZ, XA$
AIPW_{21}	$\text{IPW}_2, \text{MI}_1$	1	$1, X, Z, A$
AIPW_{22}	$\text{IPW}_2, \text{MI}_2$	1	1

Table 3: Simulation results under 46% missing comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes.

	$n = 200$				$n = 500$			
	Bias	SE	SD	CP	Bias	SE	SD	CP
FC	0.024	0.271	0.284	0.936	0.012	0.171	0.177	0.942
CC	-0.247	0.376	0.395	0.891	-0.235	0.233	0.245	0.814
EC	-0.427	0.376	0.393	0.780	-0.412	0.232	0.243	0.559
MI_0	-0.043	0.439	0.378	0.985	-0.009	0.281	0.244	0.970
MI_1	-0.035	0.368	0.376	0.953	-0.017	0.229	0.240	0.934
MI_2	-0.308	0.333	0.228	0.926	-0.301	0.206	0.144	0.749
IPW_{true}	0.005	0.372	0.390	0.948	0.015	0.231	0.244	0.939
IPW_0	0.014	0.377	0.391	0.953	0.018	0.232	0.241	0.947
IPW_1	0.033	0.373	0.385	0.958	0.036	0.230	0.241	0.939
IPW_2	-0.113	0.366	0.384	0.932	-0.104	0.227	0.240	0.909
AIPW_{00}	0.004	0.403	0.388	0.963	0.007	0.235	0.239	0.953
AIPW_{01}	0.006	0.395	0.389	0.960	0.007	0.234	0.239	0.951
AIPW_{02}	0.006	0.391	0.389	0.965	0.007	0.234	0.239	0.953
AIPW_{10}	0.002	0.388	0.386	0.961	0.007	0.234	0.239	0.947
AIPW_{11}	0.005	0.382	0.385	0.959	0.008	0.232	0.239	0.947
AIPW_{12}	0.013	0.380	0.383	0.963	0.015	0.232	0.239	0.948
AIPW_{20}	0.004	0.366	0.383	0.954	0.007	0.226	0.239	0.942
AIPW_{21}	-0.015	0.364	0.376	0.954	-0.011	0.226	0.235	0.940
AIPW_{22}	-0.020	0.391	0.386	0.960	-0.022	0.239	0.241	0.950

Table 4: Simulation results under 24% missing comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes.

	$n = 200$				$n = 500$			
	Bias	SE	SD	CP	Bias	SE	SD	CP
FC	0.024	0.271	0.284	0.936	0.012	0.171	0.177	0.942
CC	-0.083	0.312	0.326	0.933	-0.101	0.196	0.200	0.921
EC	-0.199	0.312	0.325	0.887	-0.222	0.196	0.198	0.786
MI_0	0.008	0.329	0.321	0.954	0.001	0.211	0.200	0.964
MI_1	0.006	0.312	0.318	0.946	-0.006	0.196	0.197	0.952
MI_2	-0.148	0.303	0.255	0.951	-0.161	0.190	0.158	0.914
IPW_{true}	0.028	0.310	0.326	0.939	0.010	0.195	0.199	0.942
IPW_0	0.026	0.312	0.324	0.948	0.012	0.195	0.199	0.947
IPW_1	0.033	0.310	0.323	0.946	0.019	0.195	0.198	0.946
IPW_2	-0.035	0.308	0.322	0.944	-0.053	0.194	0.197	0.940
$AIPW_{00}$	0.016	0.311	0.324	0.945	0.000	0.194	0.197	0.951
$AIPW_{01}$	0.015	0.311	0.323	0.948	0.000	0.194	0.197	0.950
$AIPW_{02}$	0.015	0.313	0.323	0.950	0.000	0.195	0.198	0.950
$AIPW_{10}$	0.015	0.309	0.323	0.942	0.000	0.194	0.197	0.947
$AIPW_{11}$	0.015	0.309	0.322	0.943	0.000	0.193	0.197	0.946
$AIPW_{12}$	0.017	0.311	0.322	0.950	0.003	0.194	0.197	0.948
$AIPW_{20}$	0.016	0.307	0.322	0.944	0.000	0.193	0.196	0.953
$AIPW_{21}$	0.006	0.307	0.318	0.944	-0.010	0.193	0.195	0.947
$AIPW_{22}$	0.006	0.318	0.325	0.953	-0.012	0.199	0.199	0.947

Table 5: Simulation results under 46% missing comparing the relative bias of the estimators of the cumulative baseline subdistribution hazards for different cohort sizes.

			$n = 200$			$n = 500$		
		t	0.5	1.0	1.5	0.5	1.0	1.5
$d\hat{\Lambda}_{10}(t)$		$\Lambda_{10}(t)$	0.126	0.210	0.265	0.126	0.210	0.265
(2.7)	AIPW_{00}		0.004	0.006	0.008	0.003	0.002	0.004
	AIPW_{01}		0.004	0.006	0.008	0.003	0.002	0.004
	AIPW_{02}		0.005	0.007	0.008	0.003	0.002	0.004
	AIPW_{10}		0.005	0.006	0.008	0.003	0.002	0.004
	AIPW_{11}		0.006	0.007	0.010	0.004	0.003	0.005
	AIPW_{12}		0.010	0.008	0.009	0.007	0.003	0.003
	AIPW_{20}		0.005	0.006	0.008	0.003	0.002	0.004
	AIPW_{21}		0.005	0.009	0.013	0.003	0.005	0.009
	AIPW_{22}		-0.020	-0.018	-0.017	-0.024	-0.024	-0.023
(2.8)	AIPW_{00}		0.008	0.006	0.008	0.003	0.001	0.004
	AIPW_{01}		0.008	0.006	0.008	0.003	0.001	0.004
	AIPW_{02}		0.008	0.006	0.008	0.003	0.001	0.004
	AIPW_{10}		0.013	0.011	0.010	0.008	0.005	0.005
	AIPW_{11}		0.013	0.011	0.010	0.008	0.005	0.005
	AIPW_{12}		0.012	0.010	0.009	0.007	0.004	0.004
	AIPW_{20}		-0.071	-0.035	-0.009	-0.079	-0.043	-0.016
	AIPW_{21}		-0.067	-0.031	-0.004	-0.074	-0.038	-0.011
	AIPW_{22}		-0.069	-0.035	-0.009	-0.076	-0.042	-0.016
(2.9)	AIPW_{00}		0.008	0.005	0.007	0.003	0.001	0.004
	AIPW_{01}		0.008	0.005	0.007	0.003	0.001	0.004
	AIPW_{02}		0.008	0.005	0.007	0.003	0.001	0.003
	AIPW_{10}		0.013	0.011	0.010	0.009	0.006	0.006
	AIPW_{11}		0.013	0.011	0.010	0.008	0.006	0.006
	AIPW_{12}		0.011	0.009	0.009	0.006	0.004	0.004
	AIPW_{20}		-0.068	-0.035	-0.011	-0.074	-0.042	-0.016
	AIPW_{21}		-0.063	-0.030	-0.005	-0.069	-0.037	-0.010
	AIPW_{22}		-0.063	-0.030	-0.005	-0.069	-0.036	-0.010

Table 6: Simulation results under 46% missing comparing the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the AIPW estimators resulting from the use the different estimators of the cumulative baseline subdistribution hazards for different cohort sizes.

		$n = 200$			$n = 500$		
$d\hat{\Lambda}_{10}(t)$		SE	SD	CP	SE	SD	CP
(2.7)	AIPW_{00}	0.403	0.388	0.963	0.235	0.239	0.953
	AIPW_{01}	0.395	0.389	0.960	0.234	0.239	0.951
	AIPW_{02}	0.391	0.389	0.965	0.234	0.239	0.953
	AIPW_{10}	0.388	0.386	0.961	0.234	0.239	0.947
	AIPW_{11}	0.382	0.385	0.959	0.232	0.239	0.947
	AIPW_{12}	0.380	0.383	0.963	0.232	0.239	0.948
	AIPW_{20}	0.366	0.383	0.954	0.226	0.239	0.942
	AIPW_{21}	0.364	0.376	0.954	0.226	0.235	0.940
	AIPW_{22}	0.391	0.386	0.960	0.239	0.241	0.950
(2.8)	AIPW_{00}	0.404	0.388	0.963	0.235	0.239	0.953
	AIPW_{01}	0.394	0.389	0.960	0.234	0.239	0.952
	AIPW_{02}	0.390	0.389	0.966	0.234	0.239	0.953
	AIPW_{10}	0.387	0.386	0.962	0.234	0.239	0.947
	AIPW_{11}	0.382	0.385	0.959	0.232	0.239	0.947
	AIPW_{12}	0.380	0.383	0.962	0.231	0.239	0.947
	AIPW_{20}	0.368	0.383	0.954	0.228	0.239	0.943
	AIPW_{21}	0.366	0.376	0.956	0.227	0.235	0.942
	AIPW_{22}	0.395	0.386	0.963	0.242	0.241	0.955
(2.9)	AIPW_{00}	0.403	0.388	0.963	0.235	0.239	0.953
	AIPW_{01}	0.394	0.389	0.960	0.234	0.239	0.953
	AIPW_{02}	0.390	0.389	0.966	0.234	0.239	0.953
	AIPW_{10}	0.388	0.386	0.962	0.234	0.239	0.947
	AIPW_{11}	0.382	0.385	0.959	0.232	0.239	0.947
	AIPW_{12}	0.380	0.383	0.962	0.231	0.239	0.947
	AIPW_{20}	0.368	0.383	0.954	0.228	0.239	0.943
	AIPW_{21}	0.366	0.376	0.955	0.227	0.235	0.942
	AIPW_{22}	0.395	0.386	0.963	0.242	0.241	0.955

2.6 EXAMPLE

Depleting antibody induction therapy has been linked to increased vulnerability to cancer. Due to Alemtuzumab's increasing recent use, interest lies in examining its association with the cumulative incidence of malignancy or malignancy-related death in comparison to no induction and an alternate therapy, Thymoglobulin. In the induction therapy analysis dataset, a total of 1309 transplant subjects were included, 986 (75%) of which received Alemtuzumab-based therapy, 203 (16%) received Thymoglobulin-based therapy, and 120(9%) received no induction therapy. In this retrospective cohort study, the median follow-up time was 4 years. There were 498 (38%) subjects who either had malignancy, graft failure, or died. Of these, 42 (8%) had cancer or cancer-related death, 413 (83%) had graft or cancer-unrelated death, and 43 (9%) died due to unknown cause. Though the amount of missingness appears to be small, its frequency is similar to that of the event of interest and thus cannot just be ignored.

Model (2.1) was assumed in analyzing this data. Covariates that were used to adjust the analysis include age group (> 53 versus ≤ 53), and degree of Human Leukocyte Antigens (HLA) mismatching (> 1 versus 0). We applied our two estimators to account for the missing cases, and compared the results with MI method (with $m = 10$ imputations and fitted using `mitools` in `R`) and the naïve approaches. The fitted imputation model, in addition to the covariates in the subdistribution model, included sex, race (non-white versus white), presence of pre-transplant dialysis, donor source type (living versus cadaveric), and follow-up time in quartile groups. The fitted model for the complete-data probability included, in addition to these covariates, the interaction between age group and follow-up time.

Table 7 shows the results of the subdistribution analysis. In general, the conclusions that can be drawn are the same regardless of the method used. All analytical approaches agree that the risk of malignancy in the induction therapy groups is not statistically different from the no induction group. There is however a significant increase in risk of about $95\% - 114\%$ in the Thymoglobulin group compared to the Alemtuzumab group as indicated by all the methods except MI. Older age is also associated with a significantly higher likelihood of cancer.

For this dataset, employing more complicated methods like MI, IPW, or AIPW did not provide substantitial analytical benefit as opposed to just using naïve methods to deal with missing data. The low proportion of missing cases may have had little influence on the bias. Nonetheless, being able to see that the results were similar increases our confidence in the conclusions drawn.

2.7 DISCUSSION

In this article, we derived an inverse probability weighted estimator that is theoretically valid and computationally simple under a proportional subdistribution hazards model. We also demonstrated how it can be implemented in standard software. However, its consistency relies on a correctly specified model for the complete-data probability. Moreover, it is calculated only from the complete cases resulting to a potential loss in efficiency. This led us to develop a second estimator that augments the IPW score equations with a term that incorporates information from the missing cases and that uses an imputation model for the type-1 failure probability. Recognizing that our proposed AIPW estimator could only be used with specialized software, we developed a user-friendly `R`-implemented `C++` macro that will be available for download.

Our AIPW estimator is doubly-robust being valid when either the complete-data probability model or the imputation model is correct. On one hand, Kang and Schafer (2007) showed through simulation that the usual doubly-robust estimator can be severely biased when both models are misspecified. For this reason, Tan (2006) and Cao et al. (2009) proposed improvements in doubly-robust estimation. On the other hand, our simulations demonstrated that the AIPW estimator could potentially be robust to misspecification of both models under the proportional subdistribution hazards framework, a finding whose confirmation requires additional studies.

It is important to realize that both the IPW and AIPW estimators operate under the assumption of MAR. The MI approach as suggested by Bakoyannis et al. (2010) also works under the same assumption. A natural question to ask is which approach should be taken

Table 7: Analysis of Induction Therapy Data

		CC	EC	MI	IPW	AIPW
Thymoglobulin vs No Induction	Coefficient	0.557	0.553	0.593	0.617	0.774
	SHR	1.745	1.739	1.810	1.853	2.169
	SE	0.576	0.576	0.580	0.581	0.620
	p-value	0.334	0.337	0.306	0.288	0.212
Alemtuzumab vs No Induction	Coefficient	-0.138	-0.114	0.166	-0.144	0.040
	SHR	0.871	0.892	1.180	0.866	1.041
	SE	0.529	0.529	0.544	0.528	0.573
	p-value	0.794	0.829	0.761	0.785	0.944
Thymoglobulin vs Alemtuzumab	Coefficient	0.695	0.667	0.435	0.761	0.734
	SHR	2.003	1.949	1.545	2.140	2.083
	SE	0.343	0.342	0.358	0.345	0.347
	p-value	0.043	0.051	0.225	0.028	0.035
Age$>$ 53	Coefficient	1.089	1.075	1.142	1.196	1.167
	SHR	2.971	2.931	3.133	3.308	3.211
	SE	0.333	0.333	0.335	0.337	0.338
	p-value	0.001	0.001	0.001	$<$.001	0.001
HLA Mismatch$>$ 0	Coefficient	-0.136	-0.139	-0.154	-0.156	-0.133
	SHR	0.873	0.871	0.857	0.855	0.876
	SE	0.095	0.094	0.084	0.100	0.101
	p-value	0.155	0.141	0.066	0.117	0.188

CC for complete cases; EC for missing failure type as an extra category; MI for multiple imputation; IPW inverse weighting; and AIPW for augmented inverse weighting.
SHR is subdistribution hazards ratio.

in analyzing data with missing cause of failure. The advantages and disadvantages of these approaches have been the subject of some debate in the missing data literature (e.g., Carpenter et al. (2006); Vansteelandt et al. (2010)). If we can correctly specify the imputation model, then MI is generally favored for efficiency. However, correct specification is difficult in many settings, including the situation that we considered in this chapter. As noted earlier, a correctly specified imputation model depends on the CSH of all event types. In practice, we would tend to rely on simpler models such as logistic regression. Even if we were willing to model all the conditional CSH, the chance of specifying all of them correctly decreases as the number of failure types increases. On the other hand, models for the complete-data probability are generally simpler; we may thus have more confidence and success in correctly specifying them. This is a practical appeal that IPW has. If we are torn choosing between MI or IPW, the AIPW estimator can offer protection from misspecifying the model for either π or ρ, but not necessarily both. It may lose some efficiency relative to a correctly specified MI model, but its robustness remains a strong attraction for use in real-life data analysis.

3.0 COMPETING RISKS REGRESSION UNDER RANDOM SIGNS CENSORING

3.1 INTRODUCTION

Many biomedical and public health studies aim to estimate the survival distribution of the time until the occurrence an event of interest, T_1^*. In the conduct of these studies, however, some subjects may experience another type of event, termed competing risk, that prevents the observation of T_1^*. For instance, individuals may dropout, die from causes other than the one of interest, or fail from any other event that precludes the measurement of T_1^*. Instead of observing the pair (T_1^*, T_2^*), we observe (T, ε), where $T = T_1^* \wedge T_2^*$ and $\varepsilon = 1$ if $T = T_1^*$ and $\varepsilon = 2$ if $T = T_2^*$. Standard approaches in this setting focus on either the cause-specific hazards (CSH) or the cumulative incidence function (CIF), which is also called subdistribution. Both of these quantities describe the time to first failure T and its cause ε and are suitable for assessing risk in the presence of the other competing event.

The crude quantities CSH and CIF are related to the joint distribution of (T, ε) rather than that of the latent failure times (T_1^*, T_2^*). Hence, they are not well suited if one wish to disentangle the risks associated to each of the competing events in relation to some covariates or prognositic variables. One may be misled to conclude that a factor reduces the risk of the the event of interest when in fact this protective effect can be partly attributed to a high proportion of subjects failing from the competing event. Smoking, for instance, may be perceived as protective of dementia because smokers tend to die from other causes and not survive long enough to experience the disease.

The scientific problem may involve characterizing the "pure" or "net" effect of some covariates on survival, i.e., the covariate effect when the competing event is removed or is

absent. For example, the transplantation community has always been interested in developing a prioritization scheme for organ allocation in a waiting list for transplant. It involves identifying risk factors associated with the underlying mortality process even though datasets that have been used for this purpose included patients who drop out or who received transplant. Interest therefore lies in marginal quantities rather than crude quantities. A well known issue in this set up however is that the observed data provides insufficient information to characterize the joint distribution of the competing risks times, i.e, there exists both an independent model and dependent model that would lead to the same distribution of the observable data (Tsiatis, 1975). This problem is known as non-identifiability. Unless one is willing to impose identifying assumptions, obtaining bounds on the marginal survival distribution is the only plausible analysis (Peterson, 1976).

One widely used assumption consider the independence of T_1^* and T_2^*, amounting to a non-informative censoring situation. This allows for consistent estimation of the survival distribution of T_1^* (Moeschberger and Klein, 1995). Standard survival analysis techniques, such as the Kaplan-Meier estimator or the Nelson-Aalen estimator to estimate survival function, and Cox proportional hazards model in the regression setting, can be applied.

Despite its usefulness in some settings, independence of the risks may be dubious in many situations. The transplantation process, for example, may be related to the death process because sicker patients are prioritized to receive transplant. Lagakos (1979) described other similar scenarios in which dependent competing risks may arise. If one proceeds with a marginal event time analysis assuming independence, the results may be seriously biased and appreciably misleading (Zheng and Klein, 1995; Huang and Zhang, 2008). One approach suggests that a set of covariates is sufficient to explain the dependence between the competing risks, i.e., the risks are independent conditional on the covariates (Zeng, 2004, 2005). After accounting for these covariates, the analysis then proceeds as in the independent case. In another collection of related work, some type of dependence structure on the joint distribution of the latent failure times is assumed. Moeschberger (1974) provided one of the early methods in this framework by suggesting joint lifetime distributions that maybe bivariate Weibull or normal. Later, Clayton (1978) and Oakes et al. (1982) introduced the notion of a common random effect, known as frailty, to model bivariate survival data. By

using copulas, Zheng and Klein (1995) derived non-parametric estimators of the marginal survival distributions that reduces to the Kaplan-Meier estimator in the case of independent risks. This was later extended to more than two competing events and to the regression setting (Escarela and Carriere, 2003; Lo and Wilke, 2010; Chen, 2010).

In this chapter, our goal is to develop an analysis method that can accommodate positively associated competing risks when the research question calls for a marginal analysis of the event of interest. We consider the Cox proportional hazards model under the assumption of random signs censoring (RSC) (Cooke, 1993, 1996).

Random signs censoring has been used in reliability studies to model the degradation of a system component (Cooke, 1993, 1996; Lindqvist and Skogsrud, 2008). It posits that the potential failure time of the event of interest (T_1^*) is independent of the event that a competing risk precedes it ($I(T_2^* < T_1^*)$). The basic idea is that the component emits a signal prior to failure, and detection of the signal would lead to some preemptive actions to prevent failure. In a biomedical setting, the component may be a biological system in which an impending failure (e.g., death) is indicated when a clinical marker reaches some critical threshold (e.g., poor health status). In this system, appropriate actions take place (e.g. transplant) when the signal is recognized (e.g., poor health status is determined) and hence no failure would be observed. However, if the signal is not detected, then failure would proceed. Under RSC, the chance that the signal is detected does not depend on the time at which failure would occur.

Current latent failure analysis methods in competing risks assume a certain dependence structure between T_1^* and T_2^* that, unfortunately, cannot be verified empirically. In contrast, the RSC assumption can be checked from the observed data. Cooke (1993) showed that a joint distribution of T_1^* and T_2^* that satisfies the RSC assumption exists if and only if the normalized subdistribution function of the main event is stochastically lower than that of the competing event. We can imagine checking the RSC requirement by a visual inspection of the empirical normalized subdistributions. This is useful in increasing our confidence in the validity of the assumption being imposed.

We can envision situations in which the RSC assumption may not hold. For example, the signal (e.g., poor health status) may be less likely to manifest among long term survivors

than among short term survivors. We posit that this association can be accounted for by a set of observable covariates. In this case, T_1^* is independent of $I(T_2^* < T_1^*)$ conditional on these covariates. That is, $T_1^* \perp \varepsilon_1 | \boldsymbol{Z}$ for a set of measured covariates $\boldsymbol{Z}$. We refer to this as conditional random sign censoring (CRSC).

This work fills some of the existing gaps in the literature about RSC. In particular, this study

1. incorporates covariates in a regression setting,
2. uses a semiparametric approach rather than parametric,
3. extends RSC to CRSC,
4. accommodates independent censoring, and
5. considers applications in the biomedical field.

In the next section, we give an expanded discussion on RSC and conditional RSC and how they are related with missing data theory. In Section 3.3, we introduce our proposed estimators for the regression coefficients in the Cox model when the data, with or without the usual independent right censoring, can be assumed to arise from RSC or CRSC. Next, we study their finite sample size properties through simulations in Section 3.4. Section 3.5 illustrates the proposed methods through the liver transplant data. We close this chapter through a discussion in Section 3.6.

3.2 RANDOM SIGNS CENSORING

3.2.1 Notations

For a pair of continuous life variables (T_1^*, T_2^*) with $\Pr(T_1^* = T_2^*) = 0$ and where T_1^* is the event time of interest, let the respective distribution function and survival function be denoted by $F_j^*(t) = \Pr(T_j^* \leq t)$ and $S_j^*(t) = 1 - F_j^*(t)$, respectively, for $j = 1, 2$. The corresponding subdistribution functions are defined by $F_1(t) = \Pr(T \leq t, \varepsilon = 1) = \Pr(T_1^* \leq t, T_1^* < T_2^*)$ and $F_2(t) = \Pr(T \leq t, \varepsilon = 2) = \Pr(T_2^* \leq t, T_2^* < T_1^*)$. F_1 and F_2 are such that $F_1(0) = F_2(0) = 0$ and $F_1(\infty) + F_2(\infty) = 1$ and referred to as a subdistribution

pair. We will denote the conditional distribution functions $\tilde{F}_1(t) = \Pr(T_1^* \leq t | T_1^* < T_2^*)$ and $\tilde{F}_2(t) = \Pr(T_2^* \leq t | T_2^* < T_1^*)$. These are also referred to as normalized subdistribution functions since $\tilde{F}_j(t) = F_j(t)/F_j(\infty)$.

3.2.2 Random Signs Censoring Principle

Cooke (1993, 1996) introduced the notion of random signs censoring as an alternative framework to the independent competing risks assumption. For life variables (T_1^*, T_2^*), T_2^* is called a random signs censoring of T_1^* if $I(T_2^* < T_1^*)$ (or equivalently if $\varepsilon_1 = I(T_1^* \leq T_2^*)$) is stochastically independent of T_1^*. This means that the event that a competing event failure precedes the main event failure is not influenced by the potential time T_1^* at which the subject fails or would have failed without the competing event. In this situation, the marginal distribution of the time to main event is identifiable, but that of the competing event is not (Lindqvist and Skogsrud, 2008). It follows that

$$\tilde{F}_1(t) = \Pr(T_1^* \leq t | T_1^* < T_2^*) = F_1^*(t)$$

so that the marginal distribution of T_1^* is the same as its normalized subdistribution. Note that T_2^* does not exhibit this property since it may or may not depend on ε_1.

RSC has been applied in reliability studies where component failure is of interest, but preventive maintenance (PM) precludes the observation of the actual failure time. One model that suits the RSC requirement assumes that the competing event happens just around the true potential main event time, and hence defines

$$T_2^* = T_1^* - (1 - 2\varepsilon_1)V$$

where ε_1 is independent of T_1^* and $V \in (0, T_1^*)$ is a random variable (Cooke, 1993). An equivalent representation defines $T_2^* = T_1^* - V$ where $-\infty < V \leq T_1^*$ with its sign, $\text{sgn}(V)$, independent of T_1^*.

Recently, Lindqvist and Skogsrud (2008) proposed modeling dependent competing risks via the first passage time of a Wiener process. A Wiener process with a drift, viewed as the underlying degradation process and formulated to satisfy the RSC leads to an inverse

Gaussian distribution for the component failure time. The time to PM, given that it has been performed before the component fails, also follows an inverse Gaussian distribution.

3.2.3 Verifying Random Signs Censoring

We restate the theorem that characterizes the "blueprint" of RSC adopted from Cooke (1993) and Lindqvist and Skogsrud (2008):

THEOREM 5. *Let K_1 and K_2 be a subdistribution pair. Then the following are equivalent:*

1. *There exists a pair (T_1^*, T_2^*) of life variables such that T_2^* is a random signs censoring of T_1^*, and such that*

$$F_1(t) = K_1(t), F_2(t) = K_2(t) \quad \text{for all } t \geq 0.$$

2.

$$\frac{K_1(t)}{K_1(\infty)} < \frac{K_2(t)}{K_2(\infty)} \quad \text{for all } t \geq 0.$$

The theorem implies that T_2^* is RSC of T_1^* if and only if $\tilde{F}_1(t) < \tilde{F}_2(t)$ or, equivalently, $\tilde{S}_1(t) = 1 - \tilde{F}_1(t) > \tilde{S}_2(t) = 1 - \tilde{F}_2(t)$. Since both of these normalized subdistributions are estimable, this is useful in checking whether RSC is compatible with the data. A graphical approach would involve looking for stochastic ordering in a plot of the estimated normalized subsurvival functions. If one can see that the estimated normalized subsurvival function of the main event dominates that of the competing event, then RSC may be a reasonable assumption.

When RSC does not hold, we assume that T_1^* and ε_1 share a set of common covariates $\boldsymbol{Z}$ so that they are dependent only through $\boldsymbol{Z}$. This implies that $T_1^* \perp \varepsilon_1 | \boldsymbol{Z}$. We will refer to this as the conditional random signs censoring (CRSC) assumption.

3.2.4 Missing Data Connection

In a competing risks data, T_1^* is observed whenever $\varepsilon_1 = 1$ and is missing otherwise. One then can view a competing risks data as being a selective sample from T_1^*, where the selection probability is equal to $E(\varepsilon_1) = \Pr(\varepsilon_1 = 1)$.

RSC implies that T_1^* is independent of ε_1 so that the observed occurrences of T_1^* is a random sample from $F_1^*(t)$. This is equivalent to missing completely at random (MCAR) in the missing data context. On the other hand, CSRC implies that $T_1^* \perp \varepsilon_1 | \boldsymbol{Z}$ and is analogous to the missing at random (MAR) assumption.

3.3 MODEL AND ESTIMATION

Suppose T_1^* follows the Cox proportional hazards model, i.e., the (marginal) hazards $\lambda_1^*(t|\boldsymbol{Z})$ of T_1^* has the form

$$\lambda_1^*(t|\boldsymbol{Z}) = \lambda_{10}^*(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z})$$

where the baseline hazard $\lambda_{10}^*(t)$ is unspecified and $\boldsymbol{Z}$ is a $p \times 1$ vector of covariates. In the absence of censoring and competing events, one can consistently estimate $\boldsymbol{\beta}$ through the partial likelihood score equations given by

$$\boldsymbol{U}^*(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^n Y_{1j}^*(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j) \boldsymbol{Z}_j}{\sum_{j=1}^n Y_{1j}^*(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j)} \right\} dN_{1i}^*(t) \tag{3.1}$$

derived from a random sample $\{(T_{1i}^*, \boldsymbol{Z}_i); i = 1, \ldots, n\}$ of $\{T_1^*, \boldsymbol{Z}\}$, where $N_{1i}^*(t) = I(T_{1i}^* \leq t)$, $Y_{1i}^*(t) = I(T_{1i}^* \geq t)$, and τ is any time point larger than the maximum observed main event time.

When a competing event, T_2^*, exists, the observed data consists of $(T_i, \varepsilon_{1i}, \boldsymbol{Z}_i)$ with $T_i = T_{1i}^* \wedge T_{2i}^*$ and ε_{1i} indicating whether the main event time is observed for subject i. If the two events were independent, the situation reduces to the standard censored data model, and one can employ the score equations (3.1) by using $N_i(t) = N_{1i}^*(t)\varepsilon_{1i}$ and $Y_i(t) = I(T_i \geq t)$ in place of $N_{1i}^*(t)$ and $Y_{1i}^*(t)$ respectively.

3.3.1 Without Independent Censoring

As noted earlier, RSC is equivalent to MCAR in the missing data literature. In the absence of independent censoring, this suggests that a complete-case (CC) analysis would be reasonable. This leads us to the following estimating equations for $\boldsymbol{\beta}$:

$$\boldsymbol{U}_{cc}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} \varepsilon_{1j} Y_j(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j) \boldsymbol{Z}_j}{\sum_{j=1}^{n} \varepsilon_{1j} Y_j(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j)} \right\} dN_i(t). \tag{3.2}$$

If the data can be assumed to arise from CRSC, it follows that

$$\Pr\{T \in [t+dt), \varepsilon_1 = 1 | \boldsymbol{Z}\} = \Pr\{T_1^* \in [t+dt) | \boldsymbol{Z}\} \Pr\{\varepsilon_1 = 1 | \boldsymbol{Z}\}, \text{ and}$$
$$\Pr\{T \geq t, \varepsilon_1 = 1 | \boldsymbol{Z}\} = \Pr\{T_1^* \geq t | \boldsymbol{Z}\} \Pr\{\varepsilon_1 = 1 | \boldsymbol{Z}\}.$$

Hence,

$$\begin{aligned}
\lim_{dt \to 0} \frac{\Pr\{T \in [t+dt), \varepsilon_1 = 1 | \boldsymbol{Z}\}}{dt \Pr\{T \geq t, \varepsilon_1 = 1 | \boldsymbol{Z}\}} &= \lim_{dt \to 0} \frac{\Pr\{T_1^* \in [t+dt), \varepsilon_1 = 1 | \boldsymbol{Z}\}}{dt \Pr\{T_1^* \geq t, \varepsilon_1 = 1 | \boldsymbol{Z}\}} \\
&= \lim_{dt \to 0} \frac{\Pr\{T_1^* \in [t+dt) | \boldsymbol{Z}\} \Pr\{\varepsilon_1 = 1 | \boldsymbol{Z}\}}{dt \Pr\{T_1^* \geq t | \boldsymbol{Z}\} \Pr\{\varepsilon_1 = 1 | \boldsymbol{Z}\}} \\
&= \lim_{dt \to 0} \frac{\Pr\{T_1^* \in [t+dt) | \boldsymbol{Z}\}}{dt \Pr\{T_1^* \geq t | \boldsymbol{Z}\}} \\
&= \lambda_1^*(t | \boldsymbol{Z}).
\end{aligned}$$

This implies that when modeling the marginal hazards under CRSC, the score equations (3.2) can still be validly applied. Unlike most modeling situations under MAR, specifying a model for $E(\varepsilon_1 | \boldsymbol{Z})$ to serve as inverse probability weights is unnecessary.

The estimating function (3.2) can be written as

$$\boldsymbol{U}_{cc}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ \boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{cc}(\boldsymbol{\beta}, t) \right\} dN_i(t) \tag{3.3}$$

where

$$\begin{aligned}
\bar{\boldsymbol{Z}}_{cc}(\boldsymbol{\beta}, t) &= \frac{S_{cc}^{(1)}(\boldsymbol{\beta}, t)}{S_{cc}^{(0)}(\boldsymbol{\beta}, t)} \\
S_{cc}^{(k)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{j=1}^{n} \varepsilon_{1j} Y_j(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j) \boldsymbol{Z}_j^{\otimes k}, \qquad k = 0, 1, 2.
\end{aligned}$$

Let $\boldsymbol{\beta}^{cc}$ be the root of $\boldsymbol{U}_{cc}(\boldsymbol{\beta})$. The asymptotic properties of this estimator can be established using standard partial likelihood inference. To this end, define the following:

$$\begin{aligned}
\boldsymbol{s}_1^{(k)}(\boldsymbol{\beta}, t) &= E\{Y_1(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\boldsymbol{Z}^{\otimes k}\}, \qquad k = 0, 1, 2, \\
\boldsymbol{e}_1(\boldsymbol{\beta}, t) &= \boldsymbol{s}_1^{(1)}(\boldsymbol{\beta}, t)/s_1^{(0)}(\boldsymbol{\beta}, t), \\
\boldsymbol{v}_1(\boldsymbol{\beta}, t) &= \frac{\boldsymbol{s}_1^{(2)}(\boldsymbol{\beta}, t)}{s_1^{(0)}(\boldsymbol{\beta}, t)} - \left\{\frac{\boldsymbol{s}_1^{(1)}(\boldsymbol{\beta}, t)}{s_1^{(0)}(\boldsymbol{\beta}, t)}\right\}^{\otimes 2}, \quad \text{and} \\
\mathbf{M} &= \int_0^\tau \{\boldsymbol{Z} - \boldsymbol{e}_1(\boldsymbol{\beta}, t)\} dM(t),
\end{aligned}$$

where $M = \varepsilon_1 M_1^*(t)$ and $M_1^*(t) = N_1^*(t) - \int_0^t Y_1^*(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z})d\Lambda_{10}^*(t)$.

THEOREM 6. *Under the (conditional) random signs censoring and the regularity conditions given in the Appendix B, $\hat{\boldsymbol{\beta}}^{cc}$ is consistent for the true parameter $\boldsymbol{\beta}_0$ and*

$$n^{1/2}(\hat{\boldsymbol{\beta}}^{cc} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1}),$$

where $\boldsymbol{\mathcal{I}}_\beta = \int_0^\tau \boldsymbol{v}_1(\boldsymbol{\beta}, t)\pi_1(\boldsymbol{Z})s_1^{(0)}(\boldsymbol{\beta}, t)d\Lambda_{10}^*(t)$ and $\pi_1(\boldsymbol{Z}) = E(\varepsilon_1|\boldsymbol{Z})$.

That is, the resulting estimator, $\hat{\boldsymbol{\beta}}_{cc}$ is consistent and asymptotically normal with asymptotic variance equal to the inverse of the Fisher information matrix, $\boldsymbol{\mathcal{I}}_\beta$, which can be consistently estimated by $\hat{\boldsymbol{\mathcal{I}}}_\beta = -\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_{cc}(\hat{\boldsymbol{\beta}}^{cc})$.

3.3.2 With Independent Censoring

If present, the usual independent right censoring poses another type of missingness that has to be addressed. We can adopt the inverse probability of censoring weighting (IPCW) approach (Robins and Rotnitzky, 1992) to construct unbiased estimating function from the score equations (3.2). Denote by C the censoring time which we assume to be independent of T_1^* and T_2^* and does not depend on $\boldsymbol{Z}$. Let $\{T_{1i}^*, T_{2i}^*, C_i, \boldsymbol{Z}_i; i = 1, \ldots, n\}$ be independent replicates of $\{T_1^*, T_2^*, C, \boldsymbol{Z}\}$. The observed data consists of $\{X_i, \delta_i\varepsilon_{1i}, \boldsymbol{Z}_i\}$ where $X = C \wedge T$ and $\delta = I(T \leq C)$ is the indicator that an observation is uncensored (i.e., have failed from either the main event or competing event). Note that under this setting, the score equations (3.2) are no longer computable from the data. For an uncensored subject i at

time t, we have that $\delta_i = 1$ and both $\varepsilon_{1i}Y_i(t)$ and $N_i(t)$ can be derived from the observed data. Although $\varepsilon_{1i}Y_i(t)$ and $N_i(t)$ are not observable when $\delta_i = 0$, $\delta_i\varepsilon_{1i}Y_i(t)$ and $\delta_i N_i(t)$ are computable for $\delta_i = 0, 1$.

Let $G(t) = \Pr(C > t)$ be the underlying survival distribution of C. For $k = 0, 1, 2$, consider the quantity

$$\begin{aligned}\frac{\delta_i\varepsilon_{1i}Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\boldsymbol{Z}_i^{\otimes k}}{G(X_i)} &= \frac{I(C_i > T_{1i}^* \wedge T_{2i}^*)\varepsilon_{1i}Y_{1i}^*(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\boldsymbol{Z}_i^{\otimes k}}{G(X_i)} \\ &= \frac{I(C_i > T_{1i}^*)\varepsilon_{1i}Y_{1i}^*(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\boldsymbol{Z}_i^{\otimes k}}{G(T_{1i}^*)}.\end{aligned}$$

Taking expectations conditional on T_{1i}^*, T_{2i}^*, and $\boldsymbol{Z}_i$, we have that

$$\begin{aligned}E\left\{\frac{\delta_i\varepsilon_{1i}Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\boldsymbol{Z}_i^{\otimes k}}{G(X_i)}\middle| T_{1i}^*, T_{2i}^*, \boldsymbol{Z}_i\right\} &= \frac{E\{I(C_i > T_{1i}^*)|T_{1i}^*\}}{G(T_{1i}^*)}\varepsilon_{1i}Y_{1i}^*(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\boldsymbol{Z}_i^{\otimes k} \\ &= \varepsilon_{1i}Y_i(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\boldsymbol{Z}_i^{\otimes k}.\end{aligned}$$

Similarly,

$$E\left\{\frac{\delta_i N_i(t)}{G(X_i)}\middle| T_{1i}^*, T_{2i}^*\right\} = N_i(t).$$

This leads us to form the following IPCW score equations:

$$\boldsymbol{U}_{cw}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\int_0^\tau \left\{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{cw}(\boldsymbol{\beta}, t)\right\}\frac{\delta_i}{G(X_i)}dN_i(t) \tag{3.4}$$

where

$$\begin{aligned}\bar{\boldsymbol{Z}}_{cw}(\boldsymbol{\beta}, t) &= \frac{S_{cw}^{(1)}(\boldsymbol{\beta}, t)}{S_{cw}^{(0)}(\boldsymbol{\beta}, t)} \\ S_{cw}^{(k)}(\boldsymbol{\beta}, t) &= n^{-1}\sum_{j=1}^{n}\frac{\delta_j}{G(X_j)}\varepsilon_{1j}Y_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}, \qquad k = 0, 1, 2.\end{aligned}$$

Note that if there are no censored observations $G(t) = 1$ for all t and equation (3.4) reduces to equation (3.2). We will denote the IPCW estimator by $\hat{\boldsymbol{\beta}}^{cw}$.

THEOREM 7. *Under the (conditional) random signs censoring and the regularity conditions given in the Appendix B,* $\hat{\boldsymbol{\beta}}^{cw}$ *is consistent for the true parameter* $\boldsymbol{\beta}_0$ *and*

$$n^{1/2}(\hat{\boldsymbol{\beta}}^{cw} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1} \boldsymbol{\Xi}_{cw} \boldsymbol{\mathcal{I}}_\beta^{-1}),$$

where $\boldsymbol{\Xi}_{cw} = \mathrm{Var}\{\delta G^{-1}(X)\mathbf{M}\}$ *and* $\boldsymbol{\mathcal{I}}_\beta = \int_0^\tau \boldsymbol{v}_1(\boldsymbol{\beta}, t)\pi_1(\boldsymbol{Z}) s_1^{(0)}(\boldsymbol{\beta}, t) d\Lambda_{10}^*(t)$.

$\boldsymbol{\mathcal{I}}_\beta$ and $\boldsymbol{\Xi}_{cw}$ can be consistently estimated by, respectively, $\hat{\boldsymbol{\mathcal{I}}}_\beta = -\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{cw}(\hat{\boldsymbol{\beta}}^{cw})$ and

$$\hat{\boldsymbol{\Xi}}_{cw} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i}{G(X_i)}\hat{\mathbf{M}}_{cw,i}(\hat{\boldsymbol{\beta}}^{cw})\right\}^{\otimes 2},$$

where $\hat{\mathbf{M}}_{cw,i}(\boldsymbol{\beta}) = \int_0^\tau \{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{cw}(\boldsymbol{\beta}, t)\} d\hat{M}_i(t)$, $d\hat{M}_i(t) = dN_i(t) - \varepsilon_{1i} Y_i(t) \exp(\hat{\boldsymbol{\beta}}^{cwT}\boldsymbol{Z}_i) d\hat{\Lambda}_{10}^*(t)$, and $d\hat{\Lambda}_{10}^*(t) = [nS_{cw}^{(0)}(\hat{\boldsymbol{\beta}}^{cw}, t)]^{-1} \sum_j \delta_j G(X_j) dN_j(t)$.

In practice, $G(t)$ has to be estimated. It may be replaced in the score equation (3.4) by a consistent estimator, say $\hat{G}(t)$. A natural candidate is the Kaplan-Meier estimator based on $\{X_i, 1-\delta_i; i = 1, ..., n\}$. Alternatively, the Kaplan-Meier estimator calculated only from the main events and censored cases may also be used. Using simulations, we will explore how these two compare.

Noting that $G(t)$ goes to zero as t increases, extreme weights may be encountered in the tail. In addition, the support of C could be shorter than the support of T in many datasets. As a result, the IPCW estimator may be inefficient or may give numerically unstable values. A potential solution is to choose a known truncation point t_0 such that $\Pr(X > t_0) > 0$. τ is then replaced by t_0 in equation (3.4) and usual analysis proceeds. Fine et al. (1998) noted that t_0 could be chosen adaptively such that about $10\% - 20\%$ of the observed main event times fall beyond it.

3.3.3 Verifying Random Signs Censoring in the Presence of Independent Censoring

The normalized subsurvival functions cannot be directly estimated from the data in the presence of independent censoring. As in section 3.3.2, the principles of IPCW can be adopted.

Conditional on $\boldsymbol{Z}_i$, we first note that

$$E\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{G(X_i)}I(X_i\in[t,t+dt),\varepsilon_{1i}=k)\right\}=\Pr\{T\in[t,t+dt),\varepsilon_1=k\}$$

and

$$E\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{G(X_i)}I(X_i\geq t,\varepsilon_{1i}=k)\right\}=\Pr\{T\geq t,\varepsilon_1=k\}$$

for $k=0,1$. This implies that

$$\begin{aligned}\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{G(X_i)}I(X_i\in[t,t+dt),\varepsilon_{1i}=k)}{\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{G(X_i)}I(X_i\geq t,\varepsilon_{1i}=k)}&\xrightarrow{a.s.}\frac{\Pr\{T\in[t,t+dt),\varepsilon_1=k\}}{\Pr\{T\geq t,\varepsilon_1=k\}}\\&=\frac{\Pr\{T\in[t,t+dt),\varepsilon_1=k\}/\Pr(\varepsilon_1=k)}{\Pr\{T\geq t,\varepsilon_1=k\}/\Pr(\varepsilon_1=k)}\end{aligned}$$

which is the hazard function of the normalized subsurvival distribution as $dt\to 0$. This suggests an IPCW product-limit type estimator of the form

$$\hat{\tilde{S}}_k(t)=\prod_{s\leq t}\left(1-\frac{\sum_{i=1}^{n}\frac{\delta_i}{G(X_i)}d\{I(X_i\leq s,\varepsilon_{1i}=k)\}}{\sum_{i=1}^{n}\frac{\delta_i}{G(X_i)}I(X_i\geq s,\varepsilon_{1i}=k)}\right).$$

3.4 SIMULATION STUDIES

In this section we conduct simulations to assess the finite sample properties of the proposed estimators. We simulate scenarios under both RSC and CRSC, and under different competing event and censoring proportions.

Sample sizes of $n = 200, 500$, or 1000 were chosen. A single covariate of interest, Z, was simulated from a Bernoulli distribution with success probability of 0.6. The main event time T_1^* was generated from the Cox model with a standard exponential baseline distribution, i.e., $\lambda_1^*(t|Z) = \exp(\beta Z)$, and where $\beta = 1$. As a point of comparison, the fitted model for this full cohort (FC) are also shown in the simulation tables.

To simulate the competing event times, we assumed the CRSC model $T_2^* = T_1^* - (1 - 2\varepsilon_1)V$, where $V \sim \text{Uniform}(0, T_1^*)$. ε_1 was generated from a logistic model given by $E(\varepsilon_1|\boldsymbol{W}) = \{1 + \exp(-\boldsymbol{\gamma}^T\widetilde{\boldsymbol{W}})\}^{-1}$ where $\widetilde{\boldsymbol{W}} = (1, Z, A)^T$ and A is a standard normal random deviate. The censoring times were independently generated from the exponential distribution with rates $\lambda_c = 0.5, 1$, or 2 that produce varying amounts of censoring.

Choosing $\boldsymbol{\gamma} = (0.5, 0, 0)$ gives an RSC scenario. We calculated the bias, standard deviation of the estimates (SD), average of the standard error estimates (SE), and empirical coverage probability (CP) of the sample 95% confidence intervals from 1000 simulated datasets. The results are shown in Table 8. Not surprisingly, the full cohort (FC) analysis in which there were no competing or censoring events performed well even with just $n = 200$. In the presence of a competing event (but no independent censoring), the complete-case analysis (CC) is unbiased with good coverage level especially under larger sample sizes. In the presence of censoring, however, this method becomes severely biased. An analysis in which the competing event was treated as independent censoring (COX) or simply excluded (RSC) showed lower though still noticeable bias.

With low level of censoring (19%), the IPCW estimator in which the true censoring distribution was used to calculate the weights (RSC-W_0) is virtually unbiased even in small samples. Larger sample sizes however are needed to get nearly correct coverage level. IPCW with censoring weights estimated from the entire data (RSC-W_1) gave slightly lower bias in large samples than that estimated from data in which the competing events were excluded

(RSC-W_2), but the latter showed smaller bias when $n = 200$. There is no noticeable advantage in using their respective estimators with 10% truncation point (RSC-W_{1t} and RSC-W_{2t}) except some efficiency gain.

With moderate level of censoring (30%), RSC-W_0 still performed well in terms of bias, but the coverage level has suffered. RSC-W_2 meanwhile outperformed RSC-W_1 in terms of bias, and further using a 10% truncation point improved the coverage level. With high proportion of censoring (45%), none of the IPCW estimators gave satisfactory results. In fact, one may even do better by treating the competing events as independently censored or simply excluding them. This may be because the contribution of the independently censored observation increased in the estimation, while the influence of the competing event diminished.

Results when $\boldsymbol{\gamma} = (0.5, -0.25, 0.5)$, $\boldsymbol{\gamma} = (0.5, -1, 1)$, and $\boldsymbol{\gamma} = (0.5, 1, 1)$ are shown, respectively, in Tables 9, 10, and 11. Note that these are CRSC situations. In general, we can observe the same pattern as in the RSC scenario. The performance of the IPCW estimators were acceptable under moderate censoring ($< 35\%$), but can be really bad otherwise. We see that in the absence of independent censoring, CC still performs well. When usual censoring exists, using censoring weights estimated from the entire data was better under low censoring ($< 20\%$), but using censoring weights estimated only from subjects who experienced the main event or were censored showed better performance when censoring is moderate ($20\% - 45\%$).

We repeated the simulation scenarios above with a lower main event rate by setting $\beta = 0.5$. Readers can refer to the appendix tables (C) for the results.

3.5 EXAMPLE

Since 2002, the Pediatric End-stage Liver Disease (PELD) score has been used to estimate the 90-day mortality rate for pediatric patients who need a liver transplant and are on the United Network for Organ Sharing (UNOS) transplant waiting list. A higher PELD score corresponds to poorer health status. However, physicians believe that the PELD score underestimates the mortality rate (Shneider et al., 2006; Freeman Jr, 2006) that they routinely

Table 8: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes. Parameters set to $\beta = 1, \gamma = (0.5, 0, 0)$

			$n = 200$				$n = 500$				$n = 1000$			
λ_c	$(\%T_1^*, \%T_2^*, \%C)$	method	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
–	(100, 0, 0)	FC	0.006	0.163	0.163	0.949	-0.003	0.102	0.100	0.954	0.001	0.072	0.073	0.952
–	(62, 38, 0)	CC	0.009	0.207	0.210	0.949	-0.001	0.130	0.129	0.950	$< .001$	0.091	0.095	0.942
0.5	(48, 33, 19)	COX	0.086	0.235	0.238	0.938	0.077	0.146	0.143	0.934	0.075	0.103	0.103	0.899
		RSC	0.068	0.233	0.235	0.945	0.057	0.146	0.143	0.936	0.062	0.103	0.107	0.904
		CC	-0.234	0.236	0.236	0.819	-0.231	0.146	0.148	0.646	-0.238	0.103	0.102	0.367
		RSC-W_0	$< .001$	0.299	0.326	0.923	0.009	0.198	0.214	0.925	-0.003	0.142	0.147	0.951
		RSC-W_1	-0.026	0.295	0.315	0.928	-0.012	0.194	0.191	0.952	-0.015	0.141	0.133	0.955
		RSC-W_{1t}	-0.033	0.300	0.301	0.936	-0.016	0.194	0.185	0.959	-0.018	0.138	0.127	0.960
		RSC-W_2	0.010	0.309	0.334	0.929	0.026	0.205	0.201	0.952	0.025	0.150	0.140	0.958
		RSC-W_{2t}	0.003	0.313	0.313	0.940	0.023	0.203	0.190	0.954	0.022	0.145	0.130	0.975
1	(40, 30, 30)	COX	0.081	0.260	0.262	0.934	0.075	0.162	0.157	0.940	0.074	0.114	0.113	0.908
		RSC	0.085	0.257	0.259	0.946	0.081	0.161	0.154	0.932	0.082	0.113	0.117	0.886
		CC	-0.378	0.262	0.262	0.669	-0.376	0.162	0.161	0.362	-0.379	0.114	0.115	0.102
		RSC-W_0	-0.031	0.397	0.497	0.877	-0.008	0.285	0.342	0.902	-0.012	0.217	0.263	0.903
		RSC-W_1	-0.089	0.383	0.441	0.893	-0.068	0.268	0.291	0.911	-0.053	0.205	0.211	0.921
		RSC-W_{1t}	-0.102	0.387	0.392	0.922	-0.066	0.259	0.252	0.934	-0.053	0.190	0.179	0.945
		RSC-W_2	-0.042	0.405	0.480	0.890	-0.020	0.288	0.318	0.916	-0.001	0.225	0.234	0.929
		RSC-W_{2t}	-0.056	0.409	0.417	0.936	-0.016	0.277	0.267	0.950	0.003	0.206	0.190	0.967
2	(30, 25, 45)	COX	0.085	0.306	0.305	0.956	0.074	0.190	0.182	0.959	0.072	0.133	0.135	0.926
		RSC	0.109	0.301	0.306	0.948	0.097	0.189	0.178	0.937	0.096	0.133	0.139	0.889
		CC	-0.567	0.310	0.303	0.532	-0.548	0.191	0.189	0.183	-0.547	0.134	0.137	0.016
		RSC-W_0	-0.198	0.528	0.852	0.726	-0.109	0.420	0.617	0.790	-0.114	0.339	0.498	0.797
		RSC-W_1	-0.302	0.503	0.683	0.754	-0.190	0.392	0.493	0.808	-0.174	0.320	0.399	0.812
		RSC-W_{1t}	-0.303	0.517	0.573	0.850	-0.184	0.376	0.388	0.888	-0.154	0.288	0.300	0.874
		RSC-W_2	-0.267	0.532	0.751	0.754	-0.146	0.423	0.548	0.813	-0.127	0.351	0.449	0.823
		RSC-W_{2t}	-0.267	0.548	0.621	0.858	-0.134	0.405	0.421	0.903	-0.099	0.315	0.329	0.905

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_0, RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights from true censoring distribution, from censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively.
$(\%T_1^*, \%T_2^*, \%C)$ for (% main events, %competing events, %censored).

Table 9: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes. Parameters set to $\beta = 1, \gamma = (0.5, -0.25, 0.5)$

			$n = 200$				$n = 500$				$n = 1000$			
λ_c	$(\%T_1^*, \%T_2^*, \%C)$	method	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
–	$(100, 0, 0)$	FC	0.011	0.163	0.166	0.953	0.004	0.102	0.099	0.963	0.003	0.072	0.071	0.952
–	$(58, 42, 0)$	CC	0.013	0.214	0.224	0.947	0.006	0.133	0.130	0.957	0.004	0.094	0.095	0.956
0.5	$(45, 37, 18)$	COX	0.039	0.242	0.249	0.946	0.028	0.150	0.150	0.951	0.023	0.106	0.108	0.942
		RSC	0.062	0.239	0.250	0.954	0.051	0.149	0.146	0.953	0.052	0.105	0.104	0.934
		CC	-0.225	0.242	0.245	0.825	-0.231	0.150	0.151	0.651	-0.236	0.105	0.105	0.380
		RSC-W_0	0.013	0.306	0.336	0.925	0.008	0.200	0.215	0.927	0.002	0.144	0.149	0.939
		RSC-W_1	-0.011	0.302	0.316	0.938	-0.015	0.196	0.201	0.943	-0.007	0.143	0.137	0.960
		RSC-W_{1t}	-0.022	0.308	0.304	0.952	-0.020	0.196	0.192	0.946	-0.014	0.141	0.134	0.964
		RSC-W_2	0.028	0.316	0.332	0.932	0.025	0.208	0.212	0.945	0.036	0.152	0.145	0.947
		RSC-W_{2t}	0.017	0.321	0.315	0.954	0.020	0.206	0.198	0.948	0.030	0.149	0.140	0.964
1	$(37, 33, 30)$	COX	0.030	0.267	0.273	0.951	0.024	0.166	0.168	0.948	0.017	0.116	0.121	0.942
		RSC	0.072	0.263	0.273	0.948	0.065	0.165	0.165	0.946	0.062	0.116	0.117	0.925
		CC	-0.379	0.269	0.294	0.666	-0.372	0.166	0.170	0.380	-0.379	0.116	0.118	0.105
		RSC-W_0	-0.032	0.402	0.543	0.834	-0.002	0.283	0.349	0.896	-0.016	0.219	0.259	0.924
		RSC-W_1	-0.098	0.387	0.484	0.859	-0.049	0.272	0.300	0.914	-0.048	0.208	0.213	0.938
		RSC-W_{1t}	-0.108	0.394	0.440	0.900	-0.059	0.264	0.263	0.940	-0.052	0.194	0.183	0.943
		RSC-W_2	-0.054	0.411	0.527	0.852	0.006	0.294	0.332	0.912	0.009	0.230	0.238	0.940
		RSC-W_{2t}	-0.064	0.418	0.471	0.908	-0.002	0.285	0.281	0.946	0.009	0.212	0.196	0.965
2	$(28, 28, 44)$	COX	0.022	0.314	0.327	0.950	0.015	0.194	0.198	0.940	0.009	0.136	0.140	0.941
		RSC	0.081	0.309	0.320	0.947	0.065	0.193	0.196	0.943	0.062	0.135	0.136	0.927
		CC	-0.559	0.319	0.358	0.537	-0.547	0.195	0.203	0.211	-0.543	0.136	0.141	0.022
		RSC-W_0	-0.197	0.533	0.880	0.694	-0.109	0.421	0.652	0.769	-0.111	0.343	0.485	0.807
		RSC-W_1	-0.290	0.511	0.759	0.732	-0.184	0.398	0.523	0.790	-0.167	0.322	0.394	0.824
		RSC-W_{1t}	-0.280	0.529	0.649	0.830	-0.183	0.382	0.424	0.851	-0.155	0.291	0.306	0.872
		RSC-W_2	-0.256	0.543	0.840	0.725	-0.135	0.433	0.590	0.793	-0.115	0.357	0.447	0.841
		RSC-W_{2t}	-0.243	0.563	0.706	0.835	-0.130	0.415	0.469	0.870	-0.096	0.320	0.338	0.896

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_0, RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights from true censoring distribution, from censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively.
$(\%T_1^*, \%T_2^*, \%C)$ for (% main events, %competing events, %censored).

Table 10: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes. Parameters set to $\beta = 1, \gamma = (0.5, -1, 1)$

			$n = 200$				$n = 500$				$n = 1000$			
λ_c	$(\%T_1^*, \%T_2^*, \%C)$	method	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
–	$(100, 0, 0)$	FC	0.011	0.163	0.166	0.953	0.004	0.102	0.099	0.963	0.003	0.072	0.071	0.952
–	$(48, 52, 0)$	CC	0.016	0.232	0.238	0.945	0.003	0.144	0.143	0.965	0.006	0.101	0.101	0.946
0.5	$(36, 46, 18)$	COX	-0.159	0.262	0.266	0.896	-0.173	0.163	0.161	0.815	-0.177	0.115	0.113	0.657
		RSC	0.006	0.258	0.262	0.945	-0.013	0.161	0.158	0.959	-0.006	0.113	0.107	0.954
		CC	-0.219	0.262	0.265	0.844	-0.230	0.162	0.163	0.682	-0.237	0.113	0.113	0.460
		RSC-W_0	0.022	0.321	0.355	0.924	0.013	0.210	0.225	0.936	0.001	0.152	0.159	0.942
		RSC-W_1	-0.004	0.318	0.335	0.932	-0.013	0.206	0.214	0.943	-0.008	0.150	0.149	0.951
		RSC-W_{1t}	-0.020	0.325	0.328	0.939	-0.020	0.208	0.206	0.953	-0.014	0.149	0.143	0.956
		RSC-W_2	0.044	0.334	0.357	0.928	0.036	0.219	0.228	0.937	0.042	0.161	0.158	0.939
		RSC-W_{2t}	0.026	0.342	0.346	0.932	0.028	0.221	0.216	0.954	0.036	0.159	0.149	0.963
1	$(30, 42, 29)$	COX	-0.177	0.288	0.289	0.888	-0.188	0.179	0.178	0.800	-0.193	0.126	0.129	0.660
		RSC	-0.017	0.283	0.285	0.952	-0.035	0.177	0.175	0.938	-0.032	0.124	0.124	0.937
		CC	-0.374	0.290	0.303	0.715	-0.372	0.179	0.180	0.433	-0.378	0.125	0.122	0.152
		RSC-W_0	-0.021	0.420	0.559	0.857	0.002	0.294	0.351	0.899	-0.013	0.227	0.263	0.926
		RSC-W_1	-0.094	0.405	0.504	0.869	-0.043	0.283	0.311	0.921	-0.049	0.216	0.218	0.943
		RSC-W_{1t}	-0.105	0.414	0.460	0.902	-0.057	0.280	0.279	0.937	-0.052	0.205	0.190	0.953
		RSC-W_2	-0.039	0.434	0.559	0.867	0.023	0.310	0.352	0.916	0.018	0.241	0.246	0.939
		RSC-W_{2t}	-0.051	0.445	0.501	0.909	0.009	0.306	0.306	0.946	0.017	0.227	0.206	0.961
2	$(22, 36, 42)$	COX	-0.197	0.337	0.338	0.889	-0.212	0.209	0.207	0.819	-0.215	0.146	0.150	0.688
		RSC	-0.053	0.330	0.333	0.943	-0.079	0.206	0.207	0.924	-0.077	0.144	0.146	0.909
		CC	-0.546	0.344	0.367	0.608	-0.541	0.209	0.213	0.269	-0.539	0.146	0.152	0.059
		RSC-W_0	-0.162	0.556	0.904	0.740	-0.088	0.430	0.638	0.806	-0.087	0.352	0.509	0.837
		RSC-W_1	-0.241	0.534	0.780	0.771	-0.164	0.410	0.529	0.824	-0.140	0.334	0.425	0.843
		RSC-W_{1t}	-0.254	0.552	0.686	0.833	-0.178	0.401	0.434	0.881	-0.139	0.310	0.332	0.889
		RSC-W_2	-0.190	0.574	0.879	0.764	-0.100	0.452	0.608	0.822	-0.072	0.376	0.496	0.857
		RSC-W_{2t}	-0.204	0.596	0.759	0.830	-0.114	0.442	0.486	0.893	-0.065	0.348	0.375	0.912

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_0, RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights from true censoring distribution, from censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively.
$(\%T_1^*, \%T_2^*, \%C)$ for (% main events, %competing events, %censored).

Table 11: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes. Parameters set to $\beta = 1, \gamma = (0.5, 1, 1)$

			$n = 200$				$n = 500$				$n = 1000$			
λ_c	$(\%T_1^*, \%T_2^*, \%C)$	method	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
–	(100, 0, 0)	FC	0.011	0.163	0.166	0.953	0.004	0.102	0.099	0.963	0.003	0.072	0.071	0.952
–	(71, 29, 0)	CC	0.015	0.202	0.206	0.952	0.005	0.126	0.124	0.959	0.003	0.089	0.089	0.956
0.5	(56, 25, 19)	COX	0.248	0.229	0.228	0.835	0.228	0.142	0.143	0.650	0.222	0.100	0.100	0.405
		RSC	0.125	0.228	0.232	0.924	0.106	0.142	0.142	0.888	0.106	0.100	0.097	0.822
		CC	-0.223	0.230	0.232	0.814	-0.231	0.143	0.143	0.629	-0.238	0.100	0.099	0.335
		RSC-W_0	0.010	0.297	0.321	0.920	0.006	0.196	0.207	0.928	-0.001	0.141	0.144	0.947
		RSC-W_1	-0.017	0.292	0.300	0.928	-0.019	0.191	0.188	0.940	-0.012	0.139	0.131	0.965
		RSC-W_{1t}	-0.019	0.298	0.290	0.944	-0.021	0.189	0.180	0.959	-0.016	0.136	0.126	0.963
		RSC-W_2	0.015	0.304	0.313	0.931	0.016	0.201	0.196	0.946	0.027	0.148	0.137	0.963
		RSC-W_{2t}	0.014	0.309	0.297	0.946	0.015	0.198	0.184	0.963	0.023	0.143	0.129	0.970
1	(46, 22, 31)	COX	0.253	0.254	0.253	0.864	0.236	0.158	0.161	0.700	0.228	0.111	0.113	0.477
		RSC	0.171	0.252	0.257	0.908	0.154	0.158	0.160	0.858	0.151	0.111	0.111	0.742
		CC	-0.380	0.257	0.274	0.656	-0.371	0.159	0.158	0.341	-0.378	0.111	0.109	0.077
		RSC-W_0	-0.041	0.395	0.525	0.831	-0.006	0.281	0.338	0.883	-0.012	0.217	0.256	0.918
		RSC-W_1	-0.112	0.376	0.458	0.849	-0.053	0.268	0.292	0.900	-0.048	0.204	0.207	0.928
		RSC-W_{1t}	-0.107	0.380	0.413	0.901	-0.058	0.257	0.248	0.934	-0.048	0.188	0.173	0.947
		RSC-W_2	-0.079	0.395	0.490	0.845	-0.007	0.288	0.318	0.904	0.001	0.223	0.228	0.940
		RSC-W_{2t}	-0.071	0.399	0.432	0.906	-0.010	0.273	0.263	0.945	0.005	0.202	0.183	0.964
2	(35, 19, 46)	COX	0.263	0.301	0.307	0.887	0.248	0.187	0.189	0.763	0.237	0.131	0.131	0.572
		RSC	0.222	0.298	0.307	0.909	0.201	0.185	0.190	0.834	0.194	0.130	0.131	0.706
		CC	-0.562	0.306	0.325	0.518	-0.544	0.188	0.192	0.182	-0.544	0.131	0.134	0.021
		RSC-W_0	-0.235	0.516	0.852	0.687	-0.134	0.413	0.636	0.737	-0.109	0.341	0.486	0.788
		RSC-W_1	-0.315	0.492	0.719	0.705	-0.204	0.389	0.511	0.768	-0.167	0.320	0.391	0.813
		RSC-W_{1t}	-0.281	0.513	0.607	0.841	-0.184	0.370	0.406	0.854	-0.152	0.285	0.297	0.862
		RSC-W_2	-0.296	0.515	0.777	0.704	-0.169	0.416	0.563	0.766	-0.125	0.348	0.435	0.822
		RSC-W_{2t}	-0.255	0.538	0.647	0.845	-0.142	0.395	0.439	0.870	-0.101	0.308	0.324	0.890

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_0, RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights from true censoring distribution, from censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively.
$(\%T_1^*, \%T_2^*, \%C)$ for (% main events, %competing events, %censored).

overadjust the score by adding about 5 points to it. The underlying model associated with the PELD score was originally based on a small sample of 884 patients, a large number of whom had received a transplant (McDiarmid et al., 2002). The PELD model was based on a Cox proportional hazards model and hence treats the receipt of a transplant as a noninformative censoring event. However, since patients who receive transplant tend to be much sicker than those who do not, we can surmise that receiving transplant may be positively correlated with a higher risk in the underlying death process and thus should be considered as an informative censoring event. This makes it more appropriate to apply an RSC model to estimate the 90-day mortality in patients who are awaiting transplant.

To develop our RSC model, we used data for 3,231 pediatric patients who were on the UNOS waiting list during the PELD era (i.e., on or after February 27, 2002). This sample excluded patients who were status 7 (inactive), were status 1 (at risk of imminent death), or had a diagnosis of cancer. Most patients were infants under 1 year old (53%), female (51%), and white (51%); 28%, 25.4%, 24.4%, and 22.3% were from the northeastern, midwestern, southern and western regions of the United States, respectively. The median follow-up time was 88 days. Of the patients, 161(5.0%) died before receiving a transplant, 1,315 (40.7%) received transplant, and 1,755 (54.3%) were censored (still alive or were removed from the study by June 2011).

We used the IPCW product-limit type estimator to estimate the normalized subsurvival distributions of death and transplant. We found that the distribution of death dominates that of transplant until about 77 days (Figure 3.5), indicating that RSC is a reasonable assumption.

We assumed a Cox proportional hazards model for time to death. The model included the following covariates: demographic characteristics (age, sex, race/ethnicity, and region); disease group (acute liver disease [ALD], auto-immune [AI], metabolic disorder [MD], biliary atresia (BA), and others); presence/absence of encephalopathy, growth failure, or ascites; and laboratory values (log-transformed total bilirubin level, log-transformed albumin level, and log-transformed international normalized ratio [INR]).

Table 12 shows the results based on different analysis methods. When we fit a standard Cox regression analysis in which transplantation was treated as independent censoring

(COX), we found that younger age, the presence of growth failure, a higher total bilirubin level, a lower albumin level, and a higher INR corresponded to a significant increase in the risk of death. These covariates in fact constitute the PELD score. If we fit an RSC Cox model in which patients who received transplant were excluded from the analysis (RSC), we found these same covariates to be significant, and in addition, that patients from the southern region had a higher risk of mortality than the those from the northeastern region. When we used a complete-case (CC) analysis in which only the deaths were included, we found that only the disease group and INR were significant. If we incorporated IPCW into the analysis in which the censoring weights were estimated from the entire data (RSC-W_1), we found that being in the southern region, having AI disorder, or having BA significantly reduced the risk of death, whereas having encephalopathy or higher INR increased mortality risk. When we used a 10% truncation point in the IPCW analysis (RSC-W_{1t}), we found similar results except that having an AI disorder or BA was significant only at the 10% level. When we estimated the censoring weights only from patients who died or were censored (RSC-W_2,RSC-W_{2t}), the estimates and p-values were similar.

In the identified set of important covariates, the fact that there were differences between treating transplant as independent censoring event and treating it as RSC event suggests that the PELD score is inadequate in predicting the risk of mortality pediatric patients awaiting for a liver transplant. Because of the potential positive association between death and transplantation, RSC may be a more reasonable assumption than non-informative censoring. However in this dataset, the RSC analysis using IPCW should be interpreted with caution because a large portion of the observations were independently censored.

3.6 DISCUSSION

In this study, we developed estimators of the covariate effects on survival under the assumption of random signs censoring or conditional random signs censoring in a competing risks setting. In contrast to existing methods that operate under RSC, a semiparametric Cox proportional hazards model was chosen to avoid more restrictive parametric assumptions.

The type of analysis is marginal in a sense that these estimators are consistent with respect to the parameters of the assumed main event latent failure time distribution. The associated hazard ratios are interpreted as the covariate effects if the competing risk was removed. We discussed the estimation both with and without the usual independent right censoring. We demonstrated that a complete-case analysis is valid in the absence of independent censoring, while an IPCW version can be used if some observations are independently censored. Not only do these estimators possess desirable asymptotic properties, they are also computationally easy to implement.

Unlike current methods that permit marginal analysis in a competing risks setting, the RSC assumption we employed can be checked empirically. One simply look for stochastic ordering in the plots of the observed normalized subsurvival functions. This increases our confidence in the conclusions that we draw.

Though the proposed methods work well in general, there are several issues that still need to be addressed in future studies. First, formal methods are needed in checking the RSC assumption. Although visual inspection of the normalized subsurvival functions provide an important diagostic tool, a testing procedure will provide a valuable alternative. In addition, it is not clear how we can check CRSC in continuous covariates. Second, the properties of the IPCW estimators have to be established analytically when the censoring distribution is estimated. Our simulations showed that censoring weights estimated by excluding the competing events perform better in small samples, but the reason needs further investigation. Next, simulations also showed that the performance of the IPCW estimators in finite samples suffer when the proportion of censored data is large. The standard error estimators may also need tweaking in this situation. An augmented IPCW estimator could be developed to improve robustness and efficiency. Finally, we note that under RSC or CRSC, the main event and competing event cannot be switched roles, and that extension to multiple competing events may be limited to preemptive actions that have similar relationship with the main event.

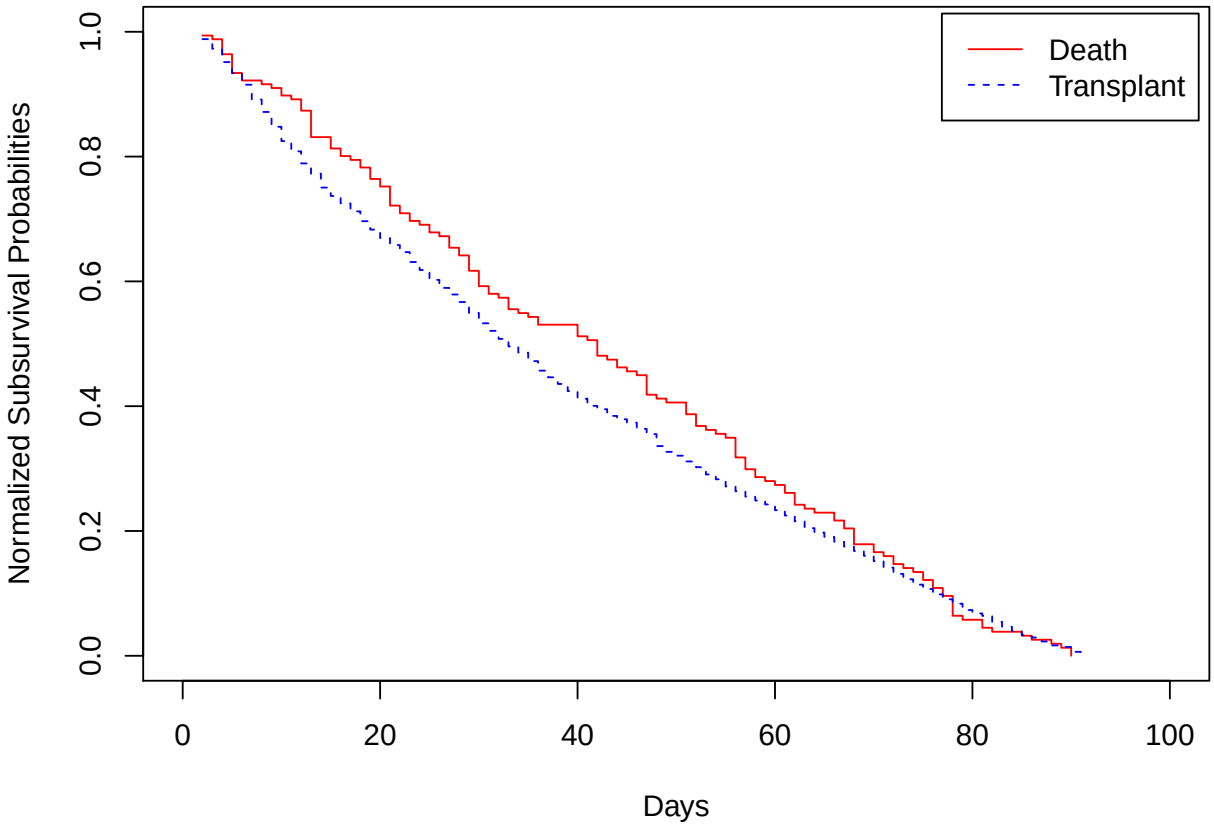

Figure 1: Checking the Random Signs Censoring Assumption in the Liver Transplant Data

Table 12: Analysis Results for the Liver Transplant Data

Variable (Reference)	Level		COX	RSC	CC	RSC-W_1	RSC-W_{1t}	RSC-W_2	RSC-W_{2t}
log Bilirubin		Coef±SE	1.19±0.14	1.26±0.14	0.08±0.13	0.08±0.13	0.12±0.15	0.08±0.13	0.12±0.15
		HR(p)	3.3(<.001)[a]	3.5(<.001)[a]	1.1(0.511)	1.1(0.538)	1.1(0.428)	1.1(0.540)	1.1(0.428)
log Albumin		Coef±SE	-1.61±0.32	-1.72±0.32	0.36±0.38	0.37±0.33	0.33±0.36	0.37±0.33	0.33±0.36
		HR(p)	0.2(<.001)[a]	0.2(<.001)[a]	1.4(0.335)	1.5(0.254)	1.4(0.359)	1.5(0.251)	1.4(0.356)
log INR		Coef±SE	0.89±0.16	0.96±0.14	1.10±0.29	1.10±0.27	1.14±0.29	1.10±0.27	1.14±0.29
		HR(p)	2.4(<.001)[a]	2.6(<.001)[a]	3.0(<.001)[a]	3.0(<.001)[a]	3.1(<.001)[a]	3.0(<.001)[a]	3.1(<.001)[a]
Age	[1, 2)	Coef±SE	0.73±0.35	0.72±0.35	0.12±0.37	0.13±0.36	0.17±0.40	0.13±0.37	0.17±0.40
(2+)		HR(p)	2.1(0.035)[b]	2.1(0.038)[b]	1.1(0.740)	1.1(0.726)	1.2(0.675)	1.1(0.724)	1.2(0.674)
	< 1	Coef±SE	1.01±0.29	1.09±0.29	0.40±0.30	0.40±0.31	0.44±0.35	0.40±0.31	0.44±0.35
		HR(p)	2.8(<.001)[a]	3.0(<.001)[a]	1.5(0.184)	1.5(0.194)	1.5(0.213)	1.5(0.197)	1.5(0.216)
Growth Failure	Yes	Coef±SE	0.62±0.17	0.47±0.17	0.07±0.19	0.07±0.17	0.04±0.17	0.07±0.17	0.04±0.17
(No)		HR(p)	1.9(<.001)[a]	1.6(0.005)[a]	1.1(0.733)	1.1(0.681)	1.0(0.832)	1.1(0.681)	1.0(0.833)
Sex	Male	Coef±SE	-0.01±0.16	-0.11±0.16	-0.04±0.18	-0.04±0.18	0.06±0.19	-0.04±0.18	0.06±0.19
(Female)		HR(p)	1.0(0.961)	0.9(0.492)	1.0(0.829)	1.0(0.832)	1.1(0.764)	1.0(0.837)	1.1(0.756)
Race	White	Coef±SE	-0.10±0.19	-0.07±0.19	0.06±0.22	0.05±0.20	0.07±0.22	0.05±0.20	0.07±0.22
(Others)		HR(p)	0.9(0.598)	0.9(0.721)	1.1(0.791)	1.1(0.801)	1.1(0.728)	1.0(0.809)	1.1(0.736)
	Black	Coef±SE	-0.36±0.24	-0.40±0.24	-0.53±0.29	-0.53±0.27	-0.54±0.29	-0.53±0.27	-0.54±0.29
		HR(p)	0.7(0.143)	0.7(0.102)	0.6(0.073)[c]	0.6(0.051)[c]	0.6(0.065)[c]	0.6(0.051)[c]	0.6(0.066)[c]
Region	S	Coef±SE	0.28±0.22	0.46±0.23	0.48±0.25	0.49±0.24	0.56±0.26	0.49±0.24	0.56±0.26
(NE)		HR(p)	1.3(0.207)	1.6(0.044)[b]	1.6(0.054)[c]	1.6(0.045)[b]	1.7(0.034)[b]	1.6(0.044)[b]	1.7(0.033)[b]
	MW	Coef±SE	0.22±0.22	0.33±0.23	0.04±0.26	0.04±0.29	0.16±0.31	0.04±0.29	0.17±0.31
		HR(p)	1.2(0.329)	1.4(0.155)	1.0(0.891)	1.0(0.891)	1.2(0.603)	1.0(0.885)	1.2(0.595)
	W	Coef±SE	-0.13±0.26	0.05±0.26	0.07±0.29	0.06±0.32	0.23±0.33	0.06±0.32	0.23±0.34
		HR(p)	0.9(0.605)	1.0(0.862)	1.1(0.816)	1.1(0.851)	1.3(0.484)	1.1(0.856)	1.3(0.485)
Disease Group	AI	Coef±SE	-0.36±0.50	-0.18±0.50	-1.08±0.54	-1.10±0.55	-0.94±0.57	-1.10±0.55	-0.94±0.57
(Acute)		HR(p)	0.7(0.474)	0.8(0.724)	0.3(0.046)[b]	0.3(0.044)[b]	0.4(0.098)[c]	0.3(0.043)[b]	0.4(0.098)[c]
	Metabolic	Coef±SE	-0.23±0.57	0.05±0.56	0.25±0.59	0.27±0.40	0.24±0.43	0.27±0.41	0.24±0.43
		HR(p)	0.8(0.679)	1.1(0.925)	1.3(0.667)	1.3(0.510)	1.3(0.576)	1.3(0.502)	1.3(0.569)
	BA	Coef±SE	-0.70±0.36	-0.60±0.35	-0.61±0.37	-0.63±0.31	-0.58±0.33	-0.63±0.31	-0.58±0.33
		HR(p)	0.5(0.051)[c]	0.5(0.087)[c]	0.5(0.096)[c]	0.5(0.046)[b]	0.6(0.084)[c]	0.5(0.045)[b]	0.6(0.082)[c]
	Other	Coef±SE	-0.30±0.38	-0.39±0.38	-0.41±0.41	-0.42±0.34	-0.40±0.36	-0.42±0.34	-0.40±0.36
		HR(p)	0.7(0.430)	0.7(0.309)	0.7(0.322)	0.7(0.225)	0.7(0.275)	0.7(0.222)	0.7(0.273)
Encephalophaty	Yes	Coef±SE	0.37±0.60	0.51±0.60	1.25±0.66	1.28±0.41	1.29±0.41	1.28±0.41	1.30±0.41
(No)		HR(p)	1.5(0.530)	1.7(0.400)	3.5(0.057)[c]	3.6(0.002)[a]	3.6(0.001)[a]	3.6(0.002)[a]	3.7(0.001)[a]
Ascites	Yes	Coef±SE	-0.01±0.07	0.02±0.07	0.01±0.09	0.01±0.10	-0.04±0.10	0.01±0.10	-0.04±0.10
(No)		HR(p)	1.0(0.889)	1.0(0.819)	1.0(0.899)	1.0(0.901)	1.0(0.712)	1.0(0.906)	1.0(0.706)

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights with censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively. [a]$p \leq .10$; [b]$p \leq .05$; [c]$p \leq .01$

4.0 CONCLUSIONS

In this dissertation, we have considered the analysis of time-to-event data in the presence of competing events. We considered a crude quantity approach in Chapter 2 and a marginal analysis approach in Chapter 3. Semiparametric models were employed to avoid imposing more restrictive parametric assumptions. Throughout this dissertation, weighting techniques have been useful in addressing the special features of the datasets.

In Chapter 2, we considered the popular proportional subdistribution hazards model when the failure type is not observed for everyone. By assuming that cause of failure is missing-at-random, we proposed two semiparametric estimators that corrects the selection bias arising from the missingness process. The first estimator uses inverse probability weights (IPW) to increase the contribution of the underrepresent subjects in the observed data. This approach is theoretically valid and yet easily implemented using standard packages. We showed analytically and through simulations that it possesses desirable properties such as consistency and asymptotic normality when the complete data probability model is correctly specified. The second estimator augments the IPW estimating equations with a term that involves an imputation model. Thus, one can think of this as a hybrid between IPW and multiple imputation (MI). The augmented inverse probability weighted (AIPW) estimator is doubly-robust, i.e., it remains consistent as long as either the complete-data probability model or imputation model (but not necessarily both) is correctly specified. We showed analytically that it possess desirable asymptotic properties. Simulations also showed that it is robust to certain model misspecifications.

Directions for future work can be based on several aspects related to this work. Our simulations demonstrated that the AIPW estimator could potentially be robust to misspecification of both constituent models under the proportional subdistribution hazards

framework, a finding whose confirmation requires additional studies. The assumption of proportionality in the Fine and Gray (1999) model may be violated in many cases. Thus, future work may extend the IPW and AIPW to other subdistribution modeling approaches such as the pseudo-observations approach (Klein and Andersen, 2005) or weighted-binomial model (Scheike et al., 2008). In addition to missingness in the cause of failure, some covariates could also be missing. The weighting methods considered here can be modified to accommodate this situation.

In Chapter 3, we adopted the random signs censoring (RSC) or conditional random signs censoring (CRSC) principle to allow a marginal time-to-event analysis in the Cox proportional hazards model when a positively related competing event exists. Unlike commonly used identifying assumptions, RSC and CRSC is verifiable from the observed data. In the absence of additional independent censoring, we showed that the complete case analysis in which the competing events were dropped is theoretically valid. When independent censoring exists, we employed the inverse probability of censoring weighting (IPCW) technique to develop consistent estimating functions for the Cox regression parameters. We derived the asymptotic properties of the resulting estimator, and simulations showed that it performs well under mild amounts of censoring.

There are several issues that can be addressed in future studies. First, developing a testing procedure to verify RSC will provide a valuable alternative to visual inspection of the normalized subsurvival functions. Hopefully, it will also be able to accommodate diagonsing CRSC for continuous covariates. Second, the properties of the IPCW estimators have to be established analytically when the censoring distribution is estimated. Lastly, an augmented IPCW estimator could be developed to improve robustness and efficiency of the estimators.

APPENDIX A

PROOFS OF THEOREMS IN CHAPTER 2

The following regularity conditions are needed in the proofs:

1. $\Lambda_{10}(\tau) < \infty$,
2. $\Pr\{\omega(t) \geq \epsilon > 0, \forall t \in [0, \tau]\} > 0$,
3. $\boldsymbol{\mathcal{I}}_\beta$ is positive definite,
4. $\pi(\boldsymbol{W}) \geq \epsilon > 0$, and
5. $\boldsymbol{Z}$ is time independent and bounded.

A Lemma from Qi et al. (2005) is useful in many portions of the proofs. We thus restate it here.

LEMMA 1. *Suppose that* $\sup_{t\in[0,\tau]} |h_n(t) - h(t)| \to 0$, $\sup_{t\in[0,\tau]} |g_n(t) - g(t)| \to 0$, *as* $n \to \infty$, *where* h *is continuous on* $[0, \tau]$ *and* $g_n(\cdot)$ *and* $g(\cdot)$ *are left-continuous on* $[0, \tau]$, *with their total variations bounded by a constant that is independent of* n. *Then as* $n \to \infty$,

$$\sup_{t\in[0,\tau]} \left| \int_0^t h_n(u) dg_n(u) - \int_0^t h(u) dg(u) \right| \to 0,$$
$$\sup_{t\in[0,\tau]} \left| \int_0^t h_n(u) dg_n(u) - \int_0^t h_n(u) dg(u) \right| \to 0.$$

Proof of Theorem 1. First we show that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \left\| \boldsymbol{S}_w^{(k)}(\boldsymbol{\beta}, t) - \boldsymbol{s}^{(k)}(\boldsymbol{\beta}, t) \right\| \xrightarrow{a.s} 0, \tag{A.1}$$

where $\mathcal{B}$ is a compact neighborhood of $\boldsymbol{\beta}_0$. This result can be deduced using similar arguments as in step A1 of the proof of theorem 1 of Qi et al. (2005) by noting that the functions $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}, t)$ are bounded and $s^{(0)}(\boldsymbol{\beta}, t)$ is bounded away from 0 on $[0, \tau] \times \mathcal{B}$, and that $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}, t)$ is an equicontinuous family at $\boldsymbol{\beta}$.

Next we show the asymptotic normality of $n^{-1/2}\boldsymbol{U}_w(\boldsymbol{\beta})$. Using straightforward calculations, one can everywhere replace $N_i^c(t)$ by $M_i^c(t)$ in (2.4). Thus we can write $\boldsymbol{U}_w(\boldsymbol{\beta}) = A_1 - A_2$ where

$$A_1 = \sum_{i=1}^{n} \int_0^{\infty} \{\boldsymbol{Z}_i - \boldsymbol{e}(\boldsymbol{\beta}, t)\} \frac{R_i}{\pi(\boldsymbol{W}_i)} dM_i^c(t), \quad \text{and}$$
$$A_2 = \int_0^{\infty} \{\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, t) - \boldsymbol{e}(\boldsymbol{\beta}, t)\} \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i)} dM_i^c(t).$$

Following the proof of theorem 1 of Qi et al. (2005) but with their $\boldsymbol{E}_{SW}(\boldsymbol{\beta}, \pi, t)$ replaced by $\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, t)$, V by R, π by $\pi(\boldsymbol{W})$, and $d\bar{M}_n(t)$ by $d\bar{M}_n^c(t) = n^{-1/2}\sum_{i=1}^n R_i \pi^{-1}(\boldsymbol{W}_i) dM_i^c(t)$, it can be shown by applying Lemma 1 and the strong embedding theorem (Shorack and Wellner, 1986, pp.47-48) that $n^{-1/2}A_2 \xrightarrow{\mathcal{P}} 0$. It follows that $n^{-1/2}\boldsymbol{U}_w(\boldsymbol{\beta})$ can be approximated by the sum of mean-0 iid random variables $n^{-1/2}\sum_i R_i \pi^{-1}(\boldsymbol{W}_i)\mathbf{M}_i$, with variance equal to $\boldsymbol{\Xi}_w = \mathrm{Var}\{R\pi^{-1}(\boldsymbol{W})\mathbf{M}\} = E\{\pi^{-1}(\boldsymbol{W})\mathbf{M}^{\otimes 2}\}$. The asymptotic normality follows from the central limit theorem.

We now establish the limit of $-\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_w(\boldsymbol{\beta})$. By applying (A.1) and Lemma 1, we see that $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left\| -\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_w(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}_\beta \right\| \xrightarrow{a.s} 0$.

Our next task is to establish the existence and consistency of $\hat{\boldsymbol{\beta}}^w$. Now, $n^{-1}\boldsymbol{U}_w(\boldsymbol{\beta}) \xrightarrow{\mathcal{P}} 0$ follows from the weak convergence of $n^{-1/2}\boldsymbol{U}_w(\boldsymbol{\beta})$ and $\boldsymbol{\mathcal{I}}_\beta$ is positive definite by condition (3). Similar arguments as in the proof of theorem 2 of Foutz (1977) can be made to show that $\hat{\boldsymbol{\beta}}^w$ exists and is unique in $\mathcal{B}$ with probability converging to 1 as $n \to \infty$, and $\hat{\boldsymbol{\beta}}^w \xrightarrow{\mathcal{P}} \boldsymbol{\beta}_0$.

Finally, we show the asymptotic normality of $n^{1/2}\hat{\boldsymbol{\beta}}^w$. By routine Taylor series expansion,

$$n^{1/2}(\hat{\boldsymbol{\beta}}^w - \boldsymbol{\beta}_0) = -\left[\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_w(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\right]^{-1} n^{-1/2}\boldsymbol{U}_w(\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}^*$ is in the line segment formed by $\hat{\boldsymbol{\beta}}^w$ and $\boldsymbol{\beta}_0$. From the previous results, we have $n^{1/2}(\hat{\boldsymbol{\beta}}^w - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1}\boldsymbol{\Xi}_w\boldsymbol{\mathcal{I}}_\beta^{-1})$.

Proof of Theorem 2. We begin by showing that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \left\| \boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},t) - \boldsymbol{s}^{(k)}(\boldsymbol{\beta},t) \right\| \xrightarrow{\mathcal{P}} 0. \tag{A.2}$$

It is easily seen that (A.1) can be extended to

$$\sup_{\substack{t\in[0,\tau],\\(\boldsymbol{\beta},\boldsymbol{\gamma})\in\mathcal{B}\times\mathcal{C}}} \left\| \boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},t) - \boldsymbol{s}^{(k)}(\boldsymbol{\beta},t) \right\| \xrightarrow{a.s} 0, \tag{A.3}$$

where $\mathcal{C}$ is a compact neighborhood of $\boldsymbol{\gamma}_0$. By a Taylor series expansion about $\boldsymbol{\gamma}$, we can write $\boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},t)$ as

$$\boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},t) - (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\frac{1}{n}\sum_{j=1}^{n}\frac{R_j}{\pi^2(\boldsymbol{W}_j;\boldsymbol{\gamma})}\frac{\partial\pi(\boldsymbol{W}_j;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T}\omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k} + o_p(1).$$

Then if $\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})$ is correctly specified, the second term converges to 0 in probability by the property of maximum likelihood estimators. This coupled with (A.3) shows (A.2).

To show the asymptotic normality $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}})$, write $\boldsymbol{U}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}}) = B_1 - B_2$ where

$$B_1 = \sum_{i=1}^{n}\frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\int_0^\tau \{\boldsymbol{Z}_i - \boldsymbol{e}(\boldsymbol{\beta}.t)\}\, dM_i^c(t) \quad \text{and}$$

$$B_2 = \int_0^\tau \left\{\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},t) - \boldsymbol{e}(\boldsymbol{\beta},t)\right\}\sum_{i=1}^{n}\frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}dM_i^c(t).$$

If we can show that $n^{-1/2}\sum_{i=1}^n R_i\pi^{-1}(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})M_i^c(t)$ converges to 0 almost surely, then by Lemma 1 and the strong embedding theorem, we can deduce that B_2 converges to 0. By a Taylor series expansion about $\boldsymbol{\gamma}$, we have that B_2 is approximately

$$n^{-1/2}\sum_{i=1}^{n}\frac{R_i}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}M_i^c(t) - (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})n^{-1/2}\sum_{i=1}^{n}\frac{R_i}{\pi^2(\boldsymbol{W}_i;\boldsymbol{\gamma})}\frac{\partial\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T}M_i^c(t)$$

which clearly converges to 0 by the boundedness of π away from 0. Next, a Taylor series expansion of B_1 about $\boldsymbol{\gamma}$ gives

$$\sum_{i=1}^{n}\frac{R_i}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}\mathbf{M}_i - (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})B_{\beta\gamma} + o_p(1),$$

where $B_{\beta\gamma} = \sum_{i=1}^n R_i\pi^{-2}(\boldsymbol{W}_i;\boldsymbol{\gamma})\{\partial\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^T\}\mathbf{M}_i$. $n^{-1/2}B_{\beta\gamma}$ (and hence the second term in the equation above) then converges to 0 by the martingale property. Thus, $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}})$ can be approximated by a sum of mean-0 iid random variables $n^{-1/2}\sum_i R_i\pi^{-1}(\boldsymbol{W}_i;\boldsymbol{\gamma})\mathbf{M}_i$

with variance $E\{\pi^{-1}(\boldsymbol{W};\boldsymbol{\gamma})\mathbf{M}^{\otimes 2}\}$. Normality immediately follows from the central limit theorem.

Now, (A.2) and Lemma 1 gives $\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left\|-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_w(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}})-\boldsymbol{\mathcal{I}}_\beta\right\| \xrightarrow{\mathcal{P}} 0$. The consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$ can then be established in the same fashion as in Theorem 1.

Proof of Theorem 3. We first show that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}}\left\|\boldsymbol{S}^{(k)}_{aw}(\boldsymbol{\beta},t)-\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)\right\| \xrightarrow{a.s} 0. \tag{A.4}$$

We can write $\boldsymbol{S}^{(k)}_{aw}(\boldsymbol{\beta},t)=S_1+S_2$ with

$$\begin{aligned}
S_1 &= n^{-1}\sum_{j=1}^{n}\omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}\\
S_2 &= n^{-1}\sum_{j=1}^{n}\frac{R_j-\pi(\boldsymbol{W}_j)}{\pi(\boldsymbol{W}_j)}\{\omega_j(t)-\widetilde{\omega}_j(t)\}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}.
\end{aligned}$$

Since the first term converges to $\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)$, it is left to show that second term converges to 0. This however immediately follows from MAR and recalling the definition of $\pi(\boldsymbol{W})$ and $\rho(\boldsymbol{W})$.

We now establish the asymptotic normality of $n^{-1/2}\boldsymbol{U}_{aw}(\boldsymbol{\beta},t)$. Note that we can replace $dN_i(t)$ and $d\widetilde{N}^c_i(t)$ in (2.6) by, respectively, $dM^c_i(t)$ and $d\widetilde{M}^c_i(t)$ where $d\widetilde{M}^c(t)=E\{dM^c(t)|\boldsymbol{W},\varepsilon>0\}=d\widetilde{N}^c(t)-\widetilde{\omega}(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z})d\Lambda_{10}(t)$. We can write

$$\begin{aligned}
\boldsymbol{U}_{aw}(\boldsymbol{\beta}) = &\sum_{i=1}^{n}\frac{R_i}{\pi(\boldsymbol{W}_i)}\int_0^\tau\{\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},t)\}dM^c_i(t)\\
&-\sum_{i=1}^{n}\frac{R_i-\pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)}\int_0^\tau\{\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},t)\}d\widetilde{M}^c_i(t),
\end{aligned}$$

which can be further decomposed as $C_1 - C_2 - C_3 + C_4$ where

$$
\begin{aligned}
C_1 &= \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i)} \mathbf{M}_i \\
C_2 &= \int_0^{\tau} \{\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) - \boldsymbol{e}(\boldsymbol{\beta}, t)\} \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i)} dM_i^c(t) \\
C_3 &= \sum_{i=1}^{n} \frac{R_i - \pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)} \widetilde{\mathbf{M}}_i \\
C_4 &= \int_0^{\tau} \{\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) - \boldsymbol{e}(\boldsymbol{\beta}, t)\} \sum_{i=1}^{n} \frac{R_i - \pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)} d\widetilde{M}_i^c(t),
\end{aligned}
$$

with $\widetilde{\mathbf{M}} = \int_0^{\tau} \{\boldsymbol{Z} - \boldsymbol{e}(\boldsymbol{\beta}, t)\} d\widetilde{M}^c(t)$. Applying (A.4), the strong embedding theorem, and Lemma 1, it can be shown that $n^{-1/2} C_2$ and $n^{-1/2} C_4$ goes to 0 in probability. It follows that

$$
n^{-1/2} \boldsymbol{U}_{aw}(\beta) = n^{-1/2} \sum_{i=1}^{n} \mathbf{M}_i + n^{-1/2} \sum_{i=1}^{n} \frac{R_i - \pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)} (\mathbf{M}_i - \widetilde{\mathbf{M}}_i) + o_p(1), \tag{A.5}
$$

with $\widetilde{\mathbf{M}} = \int_0^{\tau} \{\boldsymbol{Z} - \boldsymbol{e}(\boldsymbol{\beta}, t)\} d\widetilde{M}^c(t)$. Thus, $n^{-1/2} \boldsymbol{U}_{aw}(\boldsymbol{\beta})$ is approximately a sum of mean-0 iid random variables with variance

$$
\begin{aligned}
\boldsymbol{\Xi}_{aw} &= E\left\{ \mathbf{M} + \frac{R - \pi(\boldsymbol{W})}{\pi(\boldsymbol{W})} (\mathbf{M} - \widetilde{\mathbf{M}}) \right\}^{\otimes 2} \\
&= E\left\{ \mathbf{M}^{\otimes 2} \right\} + E\left\{ \frac{1 - \pi(\boldsymbol{W})}{\pi(\boldsymbol{W})} \mathrm{Var}(\mathbf{M} | \boldsymbol{W}, \varepsilon > 0) \right\},
\end{aligned}
$$

Applying the central limit theorem proves the asymptotic normality of $n^{-1/2} \boldsymbol{U}_{aw}(\boldsymbol{\beta})$.

We now have to show that

$$
\sup_{\boldsymbol{\beta} \in \mathcal{B}} \left\| -\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\beta}^T} \boldsymbol{U}_{aw}(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}_{\beta} \right\| \xrightarrow{\mathcal{P}} 0. \tag{A.6}
$$

Taking the derivative of (2.6) with respect to $\boldsymbol{\beta}$, we have

$$
\begin{aligned}
-\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\beta}^T} \boldsymbol{U}_{aw}(\boldsymbol{\beta}) = &\int_0^{\tau} \frac{\partial}{\partial \boldsymbol{\beta}^T} \bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{R_i}{\pi(\boldsymbol{W}_i)} dM_i^c(t) - \frac{R_i - \pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)} d\widetilde{M}_i^c(t) \right\} \\
&+ \int_0^{\tau} \frac{\partial}{\partial \boldsymbol{\beta}^T} \bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) S_{aw}^{(0)}(\boldsymbol{\beta}, t) d\Lambda_{10}(t),
\end{aligned}
$$

where

$$
\frac{\partial}{\partial \boldsymbol{\beta}^T} \bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}_{aw}^{(2)}(\boldsymbol{\beta}, t)}{S_{aw}^{(0)}(\boldsymbol{\beta}, t)} - \left(\frac{\boldsymbol{S}_{aw}^{(1)}(\boldsymbol{\beta}, t)}{S_{aw}^{(0)}(\boldsymbol{\beta}, t)} \right)^{\otimes 2}.
$$

Because $n^{-1}\sum_i R_i\pi^{-1}(\boldsymbol{W}_i)dM_i^c(t)$ and $n^{-1}\sum_i [R_i-\pi(\boldsymbol{W}_i)]\pi^{-1}(\boldsymbol{W}_i)d\widetilde{M}_i^c(t)$ converges to 0, applying (A.4) and Lemma 1 proves (A.6). It is then straightforward to show the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}^{aw}$ following the same steps used in Theorem 1.

Proof of Theorem 4. Here we just derive the asymptotic results for $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ (similar steps can be followed for $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\boldsymbol{\eta})$ and $\boldsymbol{\beta}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$). The estimating equation is given by

$$\begin{aligned}\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}}) = &\sum_{i=1}^n \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\int_0^\tau \left\{\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)\right\}dN_i^c(t)\\ &-\sum_{i=1}^n \frac{R_i-\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\int_0^\tau \left\{\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)\right\}d\widetilde{N}_i^c(t;\hat{\boldsymbol{\eta}}),\end{aligned}$$

where $\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},t)=S_{aw}^{(1)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)/S_{aw}^{(0)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)$, and

$$\begin{aligned}S_{aw}^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t) = &\frac{1}{n}\sum_{j=1}^n \frac{R_j}{\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\gamma}})}\omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}\\ &-\frac{1}{n}\sum_{j=1}^n \frac{R_j-\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\gamma}})}\widetilde{\omega}_j(t;\hat{\boldsymbol{\eta}})\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}, \qquad k=0,1,2.\end{aligned}$$

We begin by showing that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}}\left\|\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)-\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)\right\|\xrightarrow{\mathcal{P}}0. \tag{A.7}$$

We note that (A.4) can be extended to

$$\sup_{\substack{t\in[0,\tau],\\(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta})\in\mathcal{B}\times\mathcal{C}\times\mathcal{H}}}\left\|\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},t)-\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)\right\|\xrightarrow{a.s}0, \tag{A.8}$$

where $\mathcal{H}$ is a compact neighborhood of $\boldsymbol{\eta}_0$. An expansion on $\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)$ about $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$, results to $\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)=\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},t)-D_{11}-D_{12}+o_p(1)$ where

$$\begin{aligned}D_{11} &= (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi^2(\boldsymbol{W}_i;\boldsymbol{\gamma})}\frac{\partial\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\{\omega_i(t)-\widetilde{\omega}_i(t;\boldsymbol{\eta})\} \quad \text{and}\\ D_{12} &= (\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})\frac{1}{n}\sum_{i=1}^n \frac{R_i-\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\frac{\partial}{\partial\boldsymbol{\eta}^T}\widetilde{\omega}_i(t;\boldsymbol{\eta}).\end{aligned}$$

Clearly, D_{11} and D_{12} converge to 0 when either $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ or $\rho(\boldsymbol{W};\boldsymbol{\eta})$ is correct. Applying (A.8) gives the desired result.

To establish the normality of $n^{-1/2}\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$, write $\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}}) = D_{21} - D_{22}$ where

$$D_{21} = \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\mathbf{M}_i - \sum_{i=1}^{n} \frac{R_i - \pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\widetilde{\mathbf{M}}(\hat{\boldsymbol{\eta}}), \quad \text{and}$$

$$D_{22} = \int_0^{\tau} \left\{\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t) - \boldsymbol{e}(\boldsymbol{\beta},t)\right\} \sum_{i=1}^{n} \left\{ \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})} dM_i^c(t) - \frac{R_i - \pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})} d\widetilde{M}_i^c(t;\hat{\boldsymbol{\eta}}) \right\},$$

with

$$\widetilde{\mathbf{M}}(\boldsymbol{\eta}) = \int_0^{\tau} \left\{\boldsymbol{Z}_i - \boldsymbol{e}(\boldsymbol{\beta},t)\right\} d\widetilde{M}_i^c(t;\boldsymbol{\eta}) \quad \text{and}$$

$$d\widetilde{M}_i^c(t;\boldsymbol{\eta}) = d\widetilde{N}_i(t;\boldsymbol{\eta}) - \widetilde{\omega}_i(t;\boldsymbol{\eta}) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) d\Lambda_{10}(t).$$

Again by expansion about $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$, it can be shown that the summation in D_{22} converges to 0 when either $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ or $\rho(\boldsymbol{W};\boldsymbol{\eta})$ is correct. Applying (A.7), the strong embedding theorem, and Lemma 1 leads to $n^{-1/2}D_{22}$ converging to 0 in probability. Following a similar expansion, we can further decompose $n^{-1/2}D_{21}$ into $n^{-1/2}D_{31} - n^{-1/2}D_{32} - n^{-1/2}D_{33} + o_p(1)$ where

$$n^{-1/2}D_{31} = n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{R_i}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}\mathbf{M}_i - \frac{R_i - \pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}\widetilde{\mathbf{M}}_i(\boldsymbol{\eta}) \right\},$$

$$n^{-1/2}D_{32} = n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{\pi^2(\boldsymbol{W}_i;\boldsymbol{\gamma})} \frac{\partial \pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^T} \left\{\mathbf{M}_i - \widetilde{\mathbf{M}}_i(\boldsymbol{\eta})\right\} \quad \text{and}$$

$$n^{-1/2}D_{33} = n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \frac{1}{n} \sum_{i=1}^{n} \frac{R_i - \pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})} \frac{\partial}{\partial \boldsymbol{\eta}^T} \widetilde{\mathbf{M}}_i(\boldsymbol{\eta}).$$

If either $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ or $\rho(\boldsymbol{W};\boldsymbol{\eta})$ is correct, it can be immediately deduced that $n^{-1/2}D_{32}$ and $n^{-1/2}D_{33}$ converges to 0 in probability. Thus, $n^{-1/2}\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ is approximately a sum of mean-0 iid random variables that has the same form as (A.5). Asymptotic normality follows from the central limit theorem.

Next by (A.7) and Lemma 1, convergence of $-\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ to $\boldsymbol{\mathcal{I}}_{\beta}$ can be shown. The consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ can then be established.

APPENDIX B

PROOFS OF THEOREMS IN CHAPTER 3

The following regularity conditions are needed in the following proofs:

1. $\Lambda_{10}^*(\tau) < \infty$,
2. $\Pr\{Y_1(t) = 1\} > 0$,
3. $\Pr\{\varepsilon_1\} \geq \epsilon > 0$,
4. $G(\tau) > 0$
5. $\boldsymbol{\mathcal{I}}_\beta$ is positive definite, and
6. $\boldsymbol{Z}$ is time independent and bounded.

We only show proofs under CRSC since RSC is a special case.

Proof of Theorem 6. First we show that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \left\| \boldsymbol{S}_{cc}^{(k)}(\boldsymbol{\beta},t) - \pi_1(\boldsymbol{Z})\boldsymbol{s}_1^{(k)*}(\boldsymbol{\beta},t) \right\| \xrightarrow{a.s} 0, \tag{B.1}$$

where $\mathcal{B}$ is a compact neighborhood of $\boldsymbol{\beta}_0$. Under CRSC and noting that (1) the functions $\pi_1(\boldsymbol{Z})\boldsymbol{s}_1^{(k)}(\boldsymbol{\beta},t)$ are bounded, (2) $\pi_1(\boldsymbol{Z})s_1^{(0)}(\boldsymbol{\beta},t)$ is bounded away from 0 on $[0,\tau]\times\mathcal{B}$, and (3) $\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)$ is an equicontinuous family at $\boldsymbol{\beta}$, the result immediately follows from Theorem III.1 of Andersen and Gill (1982).

Next we show the asymptotic normality of $n^{-1/2}\boldsymbol{U}_{cc}(\boldsymbol{\beta})$. Using straightforward calculations, one can everywhere replace $N_i(t)$ by $M_i(t)$ in (3.2). Thus we can write $\boldsymbol{U}_{cc}(\boldsymbol{\beta}) = A_1 - A_2$

where

$$A_1 = \sum_{i=1}^{n} \int_0^\tau \{\boldsymbol{Z}_i - \boldsymbol{e}_1(\boldsymbol{\beta}, t)\}\, dM_i(t), \quad \text{and}$$
$$A_2 = \int_0^\tau \left\{\bar{\boldsymbol{Z}}_{cc}(\boldsymbol{\beta}, t) - \boldsymbol{e}_1(\boldsymbol{\beta}, t)\right\} \sum_{i=1}^{n} dM_i(t).$$

Under regularity condition (6) and invoking the CRSC assumption, it follows that $M_i(t)$ is also a martingale. The fact that $\bar{M}_n(t) = \sum_{i=1}^n dM_i(t)$ is the difference of two nondecreasing processes, $\bar{M}_n(t)$ converges weakly to a process $W_M(t)$ with continuous sample paths. Using similar argument as in the proof of theorem 1 of Qi et al. (2005), applying Lemma 1 and the strong embedding theorem (Shorack and Wellner, 1986, pp.47-48), we have that $n^{-1/2}A_2 \xrightarrow{\mathcal{P}} 0$. Hence, $n^{-1/2}\boldsymbol{U}_{cc}(\boldsymbol{\beta})$ can be approximated by the sum of mean-0 iid random variables $n^{-1/2}\sum_i \mathbf{M}_i$, with variance equal to $\boldsymbol{\Xi}_{cc} = \text{Var}\{\mathbf{M}\}$. It can be shown by straightforward algebra and calculating the predictable variation process of $\mathbf{M}$ that $\boldsymbol{\Xi}_{cc} = \boldsymbol{\mathcal{I}}_\beta$. The asymptotic normality follows from the central limit theorem.

We now establish the limit of $-\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{cc}(\boldsymbol{\beta})$. By applying (B.1) and Lemma 1, we see that $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left\| -\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{cc}(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}_\beta \right\| \xrightarrow{a.s} 0$.

The existence and consistency of $\hat{\boldsymbol{\beta}}^{cc}$ follows from the weak convergence of $n^{-1/2}\boldsymbol{U}_{cc}(\boldsymbol{\beta})$ and condition (5).

Finally, we show the asymptotic normality of $n^{1/2}\hat{\boldsymbol{\beta}}^{cc}$. By routine Taylor series expansion,

$$n^{1/2}(\hat{\boldsymbol{\beta}}^{cc} - \boldsymbol{\beta}_0) = -\left[\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{cc}(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\right]^{-1} n^{-1/2}\boldsymbol{U}_{cc}(\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}^*$ is in the line segment formed by $\hat{\boldsymbol{\beta}}^{cc}$ and $\boldsymbol{\beta}_0$. From the previous results, we have $n^{1/2}(\hat{\boldsymbol{\beta}}^{cc} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1})$.

Proof of Theorem 7. First we show that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \left\| \boldsymbol{S}_{cw}^{(k)}(\boldsymbol{\beta}, t) - \pi_1(\boldsymbol{Z})\boldsymbol{s}_1^{(k)*}(\boldsymbol{\beta}, t) \right\| \xrightarrow{a.s} 0, \tag{B.2}$$

where $\mathcal{B}$ is a compact neighborhood of $\boldsymbol{\beta}_0$. This can be proven by following the same logic in showing (B.1) under regularity condition (4).

Now to show the asymptotic normality of $n^{-1/2}\boldsymbol{U}_{cc}(\boldsymbol{\beta})$, we first everywhere replace $N_i(t)$ by $M_i(t)$ in (3.4) and write $\boldsymbol{U}_{cw}(\boldsymbol{\beta}) = B_1 - B_2$ where

$$B_1 = \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i - \boldsymbol{e}_1(\boldsymbol{\beta}, t)\} \frac{\delta_i}{G(X_i)} dM_i(t), \quad \text{and}$$
$$B_2 = \int_0^{\tau} \left\{\bar{\boldsymbol{Z}}_{cw}(\boldsymbol{\beta}, t) - \boldsymbol{e}_1(\boldsymbol{\beta}, t)\right\} \sum_{i=1}^{n} \frac{\delta_i}{G(X_i)} dM_i(t).$$

From regularity conditions (4) and (6), and noting that $M_i(t)$ is a martingale and that $\bar{M}_n(t) = \sum_{i=1}^{n} \frac{\delta_i}{G(X_i)} dM_i(t)$ is the difference of two nondecreasing processes, the same steps as in the proof of Theorem 6 can be followed to show that $n^{-1/2}A_2 \xrightarrow{\mathcal{P}} 0$. Thus, $n^{-1/2}\boldsymbol{U}_{cw}(\boldsymbol{\beta})$ is approximately the sum of mean-0 iid random variables $n^{-1/2}\sum_i \delta_i G^{-1}(X_i)\mathbf{M}_i$, with variance equal to $\boldsymbol{\Xi}_{cw} = \text{Var}\{\delta G^{-1}(X)\mathbf{M}\}$. The asymptotic normality follows from the central limit theorem.

We now establish the limit of $-\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{cw}(\boldsymbol{\beta})$. By applying (B.2) and Lemma 1, we see that $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left\| -\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{cw}(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}_{\beta} \right\| \xrightarrow{a.s} 0$.

The existence and consistency of $\hat{\boldsymbol{\beta}}^{cc}$ follows from the weak convergence of $n^{-1/2}\boldsymbol{U}_{cw}(\boldsymbol{\beta})$ and condition (5).

Finally, by routine Taylor series expansion,

$$n^{1/2}(\hat{\boldsymbol{\beta}}^{cw} - \boldsymbol{\beta}_0) = -\left[\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{cw}(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\right]^{-1} n^{-1/2}\boldsymbol{U}_{cw}(\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}^*$ is in the line segment formed by $\hat{\boldsymbol{\beta}}^{cw}$ and $\boldsymbol{\beta}_0$. From the previous results, we have $n^{1/2}(\hat{\boldsymbol{\beta}}^{cw} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_{\beta}^{-1}\boldsymbol{\Xi}_{cw}\boldsymbol{\mathcal{I}}_{\beta}^{-1})$, which shows the asymptotic normality of $\hat{\boldsymbol{\beta}}^{cw}$.

APPENDIX C

TABLES

Table 13: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes. Parameters set to $\beta = 0.5, \gamma = (0.5, 0, 0)$

			$n = 200$				$n = 500$				$n = 1000$			
λ_c	$(\%T_1^*, \%T_2^*, \%C)$	method	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
–	$(100, 0, 0)$	FC	0.001	0.151	0.149	0.955	-0.004	0.095	0.093	0.941	$< .001$	0.067	0.068	0.950
–	$(62, 38, 0)$	CC	0.001	0.192	0.192	0.944	-0.003	0.120	0.119	0.943	-0.002	0.085	0.088	0.940
0.5	$(45, 32, 23)$	COX	0.042	0.227	0.230	0.941	0.035	0.141	0.139	0.943	0.035	0.100	0.099	0.943
		RSC	0.037	0.225	0.229	0.946	0.028	0.141	0.137	0.944	0.033	0.099	0.102	0.940
		CC	-0.141	0.228	0.227	0.908	-0.139	0.142	0.145	0.822	-0.142	0.100	0.100	0.690
		RSC-W_0	-0.006	0.309	0.353	0.903	0.001	0.211	0.237	0.907	-0.002	0.154	0.163	0.933
		RSC-W_1	-0.031	0.304	0.343	0.908	-0.017	0.204	0.221	0.915	-0.014	0.152	0.156	0.934
		RSC-W_{1t}	-0.032	0.308	0.311	0.933	-0.016	0.199	0.202	0.948	-0.015	0.143	0.139	0.954
		RSC-W_2	-0.012	0.320	0.370	0.904	0.003	0.217	0.239	0.910	0.006	0.163	0.169	0.937
		RSC-W_{2t}	-0.013	0.324	0.329	0.939	0.005	0.210	0.213	0.947	0.007	0.152	0.147	0.958
1	$(36, 28, 36)$	COX	0.039	0.257	0.255	0.947	0.037	0.160	0.159	0.955	0.034	0.113	0.112	0.947
		RSC	0.046	0.254	0.255	0.946	0.043	0.159	0.154	0.949	0.043	0.112	0.114	0.937
		CC	-0.222	0.260	0.259	0.854	-0.220	0.161	0.159	0.731	-0.222	0.113	0.115	0.502
		RSC-W_0	-0.035	0.421	0.573	0.828	-0.013	0.313	0.409	0.861	-0.024	0.245	0.320	0.865
		RSC-W_1	-0.073	0.405	0.505	0.863	-0.053	0.291	0.348	0.868	-0.064	0.229	0.268	0.875
		RSC-W_{1t}	-0.085	0.410	0.439	0.917	-0.058	0.277	0.284	0.925	-0.052	0.206	0.211	0.921
		RSC-W_2	-0.052	0.431	0.557	0.847	-0.029	0.315	0.389	0.864	-0.041	0.254	0.304	0.875
		RSC-W_{2t}	-0.064	0.436	0.474	0.916	-0.032	0.299	0.310	0.930	-0.025	0.226	0.232	0.930
2	$(25, 23, 52)$	COX	0.045	0.311	0.305	0.968	0.035	0.193	0.184	0.968	0.034	0.136	0.136	0.948
		RSC	0.064	0.307	0.309	0.956	0.048	0.192	0.180	0.964	0.050	0.135	0.140	0.940
		CC	-0.329	0.317	0.321	0.810	-0.308	0.195	0.195	0.647	-0.306	0.137	0.137	0.400
		RSC-W_0	-0.244	0.554	1.721	0.701	-0.093	0.452	0.726	0.748	-0.125	0.380	0.601	0.756
		RSC-W_1	-0.252	0.528	0.789	0.735	-0.138	0.425	0.604	0.778	-0.148	0.357	0.490	0.772
		RSC-W_{1t}	-0.225	0.554	0.675	0.845	-0.126	0.413	0.461	0.887	-0.121	0.322	0.359	0.884
		RSC-W_2	-0.249	0.560	0.875	0.710	-0.124	0.460	0.678	0.769	-0.131	0.393	0.562	0.773
		RSC-W_{2t}	-0.215	0.590	0.739	0.842	-0.105	0.447	0.506	0.893	-0.097	0.354	0.404	0.879

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_0, RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights from true censoring distribution, from censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively.
$(\%T_1^*, \%T_2^*, \%C)$ for (% main events, %competing events, %censored).

Table 14: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes. Parameters set to $\beta = 0.5, \gamma = (0.5, -0.25, 0.5)$

			$n = 200$				$n = 500$				$n = 1000$			
λ_c	$(\%T_1^*, \%T_2^*, \%C)$	method	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
–	$(100, 0, 0)$	FC	0.007	0.151	0.154	0.946	0.001	0.095	0.092	0.964	0.003	0.067	0.066	0.959
–	$(58, 42, 0)$	CC	0.007	0.198	0.208	0.946	0.001	0.123	0.121	0.956	0.004	0.087	0.089	0.954
0.5	$(42, 35, 22)$	COX	-0.011	0.233	0.238	0.946	-0.019	0.145	0.146	0.945	-0.022	0.102	0.105	0.946
		RSC	0.024	0.231	0.239	0.953	0.013	0.144	0.140	0.957	0.017	0.102	0.100	0.953
		CC	-0.135	0.235	0.244	0.895	-0.138	0.145	0.148	0.831	-0.142	0.102	0.103	0.710
		RSC-W_0	0.003	0.318	0.371	0.882	0.005	0.212	0.237	0.911	$< .001$	0.157	0.167	0.925
		RSC-W_1	-0.018	0.311	0.354	0.895	-0.016	0.207	0.228	0.913	-0.011	0.154	0.159	0.938
		RSC-W_{1t}	-0.017	0.315	0.331	0.919	-0.020	0.203	0.205	0.944	-0.014	0.146	0.143	0.950
		RSC-W_2	0.002	0.328	0.381	0.890	0.004	0.221	0.248	0.911	0.012	0.167	0.174	0.927
		RSC-W_{2t}	0.003	0.332	0.350	0.921	0.002	0.215	0.217	0.949	0.010	0.156	0.152	0.956
1	$(33, 31, 36)$	COX	-0.021	0.264	0.272	0.947	-0.024	0.164	0.167	0.941	-0.028	0.116	0.118	0.934
		RSC	0.023	0.261	0.269	0.954	0.014	0.163	0.163	0.954	0.012	0.115	0.115	0.955
		CC	-0.226	0.268	0.297	0.834	-0.217	0.165	0.173	0.721	-0.221	0.116	0.117	0.510
		RSC-W_0	-0.051	0.428	0.623	0.814	-0.013	0.313	0.419	0.855	-0.023	0.249	0.315	0.891
		RSC-W_1	-0.095	0.408	0.564	0.826	-0.043	0.298	0.372	0.862	-0.051	0.235	0.267	0.894
		RSC-W_{1t}	-0.092	0.416	0.495	0.895	-0.048	0.284	0.310	0.924	-0.048	0.211	0.214	0.936
		RSC-W_2	-0.079	0.436	0.626	0.809	-0.017	0.327	0.420	0.852	-0.022	0.262	0.306	0.893
		RSC-W_{2t}	-0.075	0.445	0.539	0.893	-0.018	0.310	0.341	0.923	-0.017	0.233	0.236	0.944
2	$(23, 26, 51)$	COX	-0.022	0.319	0.330	0.935	-0.033	0.198	0.200	0.943	-0.035	0.139	0.142	0.930
		RSC	0.021	0.315	0.322	0.955	0.002	0.197	0.198	0.953	0.003	0.138	0.139	0.949
		CC	-0.324	0.326	0.352	0.824	-0.302	0.200	0.207	0.659	-0.302	0.139	0.147	0.417
		RSC-W_0	-0.174	0.565	0.970	0.716	-0.112	0.462	0.778	0.736	-0.103	0.381	0.587	0.763
		RSC-W_1	-0.209	0.540	0.826	0.749	-0.149	0.435	0.630	0.770	-0.134	0.359	0.497	0.791
		RSC-W_{1t}	-0.203	0.565	0.708	0.846	-0.132	0.423	0.497	0.881	-0.117	0.325	0.374	0.882
		RSC-W_2	-0.200	0.574	0.922	0.730	-0.133	0.474	0.718	0.751	-0.113	0.399	0.573	0.784
		RSC-W_{2t}	-0.191	0.604	0.779	0.839	-0.110	0.462	0.556	0.883	-0.090	0.361	0.421	0.888

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_0, RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights from true censoring distribution, from censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively.
$(\%T_1^*, \%T_2^*, \%C)$ for (% main events, %competing events, %censored).

Table 15: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes. Parameters set to $\beta = 0.5, \gamma = (0.5, -1, 1)$

			$n = 200$				$n = 500$				$n = 1000$			
λ_c	$(\%T_1^*, \%T_2^*, \%C)$	method	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
–	$(100, 0, 0)$	FC	0.007	0.151	0.154	0.946	0.001	0.095	0.092	0.964	0.003	0.067	0.066	0.959
–	$(48, 52, 0)$	CC	0.009	0.216	0.224	0.941	-0.002	0.134	0.133	0.958	0.005	0.094	0.095	0.946
0.5	$(34, 44, 21)$	COX	-0.224	0.253	0.257	0.851	-0.237	0.158	0.155	0.662	-0.237	0.111	0.111	0.428
		RSC	-0.059	0.250	0.252	0.939	-0.078	0.157	0.153	0.924	-0.069	0.110	0.106	0.917
		CC	-0.129	0.255	0.264	0.906	-0.139	0.158	0.161	0.842	-0.143	0.111	0.111	0.742
		RSC-W_0	0.013	0.337	0.390	0.904	0.007	0.225	0.253	0.911	-0.002	0.166	0.179	0.925
		RSC-W_1	-0.008	0.330	0.375	0.912	-0.014	0.219	0.242	0.919	-0.013	0.163	0.175	0.933
		RSC-W_{1t}	-0.016	0.337	0.355	0.933	-0.023	0.218	0.224	0.940	-0.016	0.157	0.154	0.952
		RSC-W_2	0.019	0.353	0.412	0.896	0.015	0.238	0.269	0.910	0.015	0.180	0.197	0.920
		RSC-W_{2t}	0.011	0.360	0.382	0.924	0.005	0.235	0.244	0.946	0.014	0.171	0.167	0.953
1	$(27, 39, 34)$	COX	-0.239	0.286	0.283	0.862	-0.254	0.178	0.178	0.695	-0.254	0.125	0.127	0.482
		RSC	-0.098	0.282	0.279	0.947	-0.121	0.177	0.176	0.893	-0.116	0.124	0.125	0.839
		CC	-0.219	0.290	0.313	0.853	-0.215	0.179	0.181	0.779	-0.217	0.125	0.123	0.577
		RSC-W_0	-0.034	0.453	0.660	0.812	0.001	0.326	0.422	0.867	-0.011	0.260	0.325	0.899
		RSC-W_1	-0.083	0.433	0.594	0.828	-0.032	0.312	0.385	0.871	-0.039	0.244	0.274	0.905
		RSC-W_{1t}	-0.084	0.443	0.526	0.877	-0.043	0.304	0.324	0.921	-0.040	0.225	0.223	0.941
		RSC-W_2	-0.057	0.471	0.681	0.805	0.010	0.350	0.451	0.862	0.001	0.280	0.326	0.902
		RSC-W_{2t}	-0.058	0.484	0.587	0.866	-0.002	0.339	0.368	0.915	0.002	0.255	0.252	0.945
2	$(19, 32, 49)$	COX	-0.263	0.346	0.347	0.871	-0.281	0.214	0.211	0.742	-0.276	0.150	0.151	0.560
		RSC	-0.154	0.340	0.341	0.923	-0.178	0.212	0.212	0.863	-0.169	0.149	0.149	0.792
		CC	-0.298	0.355	0.380	0.844	-0.291	0.217	0.225	0.709	-0.299	0.150	0.155	0.482
		RSC-W_0	-0.102	0.591	1.021	0.736	-0.067	0.474	0.789	0.775	-0.076	0.400	0.618	0.809
		RSC-W_1	-0.137	0.570	0.893	0.762	-0.108	0.452	0.662	0.804	-0.104	0.378	0.512	0.827
		RSC-W_{1t}	-0.143	0.596	0.774	0.841	-0.110	0.446	0.526	0.892	-0.102	0.348	0.387	0.903
		RSC-W_2	-0.117	0.616	1.026	0.726	-0.078	0.505	0.785	0.763	-0.070	0.433	0.621	0.813
		RSC-W_{2t}	-0.122	0.650	0.875	0.822	-0.078	0.501	0.608	0.885	-0.064	0.399	0.456	0.905

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_0, RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights from true censoring distribution, from censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively.
$(\%T_1^*, \%T_2^*, \%C)$ for (% main events, %competing events, %censored).

Table 16: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for different cohort sizes. Parameters set to $\beta = 0.5, \gamma = (0.5, 1, 1)$

			$n = 200$				$n = 500$				$n = 1000$			
λ_c	$(\%T_1^*, \%T_2^*, \%C)$	method	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
–	(100, 0, 0)	FC	0.007	0.151	0.154	0.946	0.001	0.095	0.092	0.964	0.003	0.067	0.066	0.959
–	(71, 29, 0)	CC	0.010	0.187	0.192	0.943	0.001	0.117	0.115	0.965	0.002	0.082	0.084	0.956
0.5	(52, 24, 24)	COX	0.212	0.220	0.220	0.865	0.196	0.137	0.139	0.712	0.191	0.097	0.097	0.516
		RSC	0.107	0.219	0.223	0.934	0.089	0.137	0.137	0.914	0.092	0.097	0.095	0.849
		CC	-0.132	0.222	0.229	0.892	-0.138	0.138	0.137	0.814	-0.143	0.097	0.095	0.691
		RSC-W_0	$< .001$	0.305	0.348	0.893	0.002	0.206	0.228	0.909	-0.002	0.152	0.160	0.935
		RSC-W_1	-0.019	0.297	0.332	0.897	-0.019	0.200	0.215	0.919	-0.014	0.149	0.152	0.937
		RSC-W_{1t}	-0.014	0.302	0.309	0.928	-0.019	0.195	0.193	0.941	-0.014	0.140	0.135	0.954
		RSC-W_2	-0.006	0.309	0.349	0.892	-0.004	0.210	0.228	0.921	0.002	0.158	0.164	0.943
		RSC-W_{2t}	$< .001$	0.313	0.320	0.932	-0.003	0.203	0.200	0.950	0.003	0.147	0.141	0.953
1	(41, 21, 37)	COX	0.220	0.251	0.253	0.865	0.208	0.156	0.159	0.745	0.199	0.110	0.110	0.559
		RSC	0.151	0.249	0.253	0.915	0.133	0.156	0.158	0.882	0.130	0.110	0.110	0.776
		CC	-0.226	0.254	0.275	0.821	-0.213	0.157	0.158	0.709	-0.220	0.110	0.109	0.477
		RSC-W_0	-0.062	0.412	0.599	0.790	-0.017	0.305	0.403	0.845	-0.019	0.242	0.306	0.876
		RSC-W_1	-0.109	0.392	0.534	0.803	-0.047	0.289	0.357	0.842	-0.052	0.226	0.255	0.878
		RSC-W_{1t}	-0.094	0.400	0.467	0.889	-0.045	0.274	0.290	0.921	-0.044	0.201	0.203	0.942
		RSC-W_2	-0.102	0.411	0.574	0.798	-0.029	0.309	0.390	0.838	-0.030	0.246	0.285	0.876
		RSC-W_{2t}	-0.082	0.420	0.494	0.892	-0.025	0.291	0.311	0.919	-0.020	0.217	0.219	0.946
2	(29, 17, 53)	COX	0.241	0.305	0.310	0.906	0.223	0.189	0.190	0.816	0.212	0.132	0.131	0.670
		RSC	0.195	0.301	0.306	0.921	0.174	0.188	0.192	0.868	0.168	0.132	0.130	0.786
		CC	-0.324	0.311	0.322	0.797	-0.299	0.191	0.195	0.643	-0.305	0.133	0.137	0.381
		RSC-W_0	-0.218	0.539	0.950	0.693	-0.134	0.441	0.744	0.706	-0.114	0.374	0.574	0.743
		RSC-W_1	-0.241	0.515	0.800	0.719	-0.162	0.415	0.617	0.740	-0.138	0.351	0.478	0.766
		RSC-W_{1t}	-0.200	0.546	0.673	0.837	-0.134	0.405	0.476	0.873	-0.117	0.315	0.350	0.889
		RSC-W_2	-0.244	0.537	0.861	0.708	-0.155	0.441	0.676	0.729	-0.124	0.378	0.529	0.763
		RSC-W_{2t}	-0.196	0.572	0.718	0.836	-0.120	0.430	0.515	0.879	-0.098	0.338	0.382	0.898

FC for full cohort; CC for complete cases only; COX for competing event as independent censoring; RSC for competing event excluded analysis; RSC-W_0, RSC-W_1, RSC-W_{1t}, RSC-W_2, and RSC-W_{2t} for inverse probability censoring weights from true censoring distribution, from censoring distribution estimated from entire data, from entire data with 10% trunctation point, from main events and censored cases, and from main event and censored cases with 10% truncation point, respectively.
$(\%T_1^*, \%T_2^*, \%C)$ for (% main events, %competing events, %censored).

BIBLIOGRAPHY

J. Andersen, E. Goetghebeur, and L. Ryan. Missing cause of death information in the analysis of survival data. *Statistics in Medicine*, 15(20):2191–2201, 1996.

P.K. Andersen and R.D. Gill. Cox's regression model for counting processes: a large sample study. *The annals of statistics*, 10(4):1100–1120, 1982.

G. Bakoyannis, F. Siannis, and G. Touloumi. Modelling competing risks data with missing cause of failure. *Statistics in Medicine*, 29(30):3172–3185, 2010.

W. Cao, A.A. Tsiatis, and M. Davidian. Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96(3):723–734, 2009.

J.R. Carpenter, M.G. Kenward, and S. Vansteelandt. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3):571–584, 2006.

Y.H. Chen. Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(2):235–251, 2010.

D.G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1): 141–151, 1978.

R.M. Cooke. The total time on test statistic and age-dependent censoring. *Statistics & probability letters*, 18(4):307–312, 1993.

R.M. Cooke. The design of reliability data bases, part i and ii: review of standard design concepts. *Reliability Engineering & System Safety*, 51(2):137–146,209–223, 1996.

G. Escarela and J.F. Carriere. Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research*, 12(4):333, 2003.

J.P. Fine and R.J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, pages 496–509, 1999.

J.P. Fine, Z. Ying, and L.G. Wei. On the linear transformation model for censored data. *Biometrika*, 85(4):980–986, 1998.

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

R.B. Freeman Jr. Mortality risk, behavior, and pediatric liver allocation. *Liver Transplantation*, 12(1):12–15, 2006.

G. Gao and A.A. Tsiatis. Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika*, 92 (4):875–891, 2005.

R.B. Geskus. Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics*, 67(1):39–49, 2011.

E. Goetghebeur and L. Ryan. Analysis of competing risks survival data when some failure types are missing. *Biometrika*, 82(4):821–833, 1995.

R.J. Gray. *cmprsk: Subdistribution Analysis of Competing Risks*, 2010. URL http://CRAN.R-project.org/package=cmprsk. R package version 2.2-1.

D.G. Horvitz and D.J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

X. Huang and N. Zhang. Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. *Biometrics*, 64(4):1090–1099, 2008.

J.D. Kalbfleisch and R.L. Prentice. *The Statistical Analysis of Failure Time Data*, volume 360. Wiley-Interscience, 2002.

J.D.Y. Kang and J.L. Schafer. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22 (4):523–539, 2007.

J.P. Klein and P.K. Andersen. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61(1):223–229, 2005.

S.W. Lagakos. General right censoring and its impact on the analysis of survival data. *Biometrics*, pages 139–156, 1979.

B.H. Lindqvist and G. Skogsrud. Modeling of dependent competing risks by first passage times of wiener processes. *IIE Transactions*, 41(1):72–80, 2008.

S. Lo and R.A. Wilke. A copula model for dependent competing risks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(2):359–376, 2010.

K. Lu and A.A. Tsiatis. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, 57(4):1191–1197, 2001.

W. Lu and Y. Liang. Analysis of competing risks data with missing cause of failure under additive hazards model. *Statistica Sinica*, 18:219–234, 2008.

T. Lumley. *mitools: Tools for multiple imputation of missing data*, 2010. URL http://CRAN.R-project.org/package=mitools. R package version 2.0.1.

S.V. McDiarmid, R. Anand, A.S. Lindblad, et al. Development of a pediatric end-stage liver disease score to predict poor outcome in children awaiting liver transplantation1. *Transplantation*, 74(2):173, 2002.

M.L. Moeschberger. Life tests under dependent competing causes of failure. *Technometrics*, pages 39–47, 1974.

M.L. Moeschberger and J.P. Klein. Statistical methods for dependent competing risks. *Lifetime Data Analysis*, 1(2):195–204, 1995.

D. Oakes et al. A concordance test for independence in the presence of censoring. *Biometrics*, 38(2):451, 1982.

A.V. Peterson. Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. *Proceedings of the National Academy of Sciences*, 73(1): 11, 1976.

L. Qi, CY Wang, and R.L. Prentice. Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, 100(472):1250–1263, 2005.

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

J.M. Robins, A. Rotnitzky, and L.P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, pages 846–866, 1994.

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

D.B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

T.H. Scheike, M.J. Zhang, and T.A. Gerds. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95(1):205, 2008.

B.L. Shneider, F.J. Suchy, and S. Emre. National and regional analysis of exceptions to the pediatric end-stage liver disease scoring system (2003–2004). *Liver transplantation*, 12(1): 40–45, 2006.

G.R. Shorack and J.A. Wellner. *Empirical Processes with Applications to Statistics*. Wiley (New York), 1986.

Z. Tan. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006.

A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20, 1975.

S. Vansteelandt, J. Carpenter, and M.G. Kenward. Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(1):37–48, 2010.

D. Zeng. Estimating marginal survival function by adjusting for dependent censoring using many covariates. *The Annals of Statistics*, 32(4):1533–1555, 2004.

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

M. Zheng and J.P. Klein. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82(1):127–138, 1995.