Methods for Clustered Competing Risks Data and Causal Inference using Instrumental Variables for Censored Time-to-event Data

by

Sai Hurrish Dharmarajan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2018

Doctoral Committee:

Professor Douglas E. Schaubel, Chair
Research Assistant Professor Kevin He
Professor Jack D. Kalbfleisch
Professor Rajiv Saran

Sai H. Dharmarajan

shdharma@umich.edu

ORCID iD: 0000-0002-9287-5810

To my parents and my sister

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, Prof. Douglas Schaubel, for being an incredible mentor and guide. I am extremely grateful for his unflinching support and patience over the last four years.

I would like to thank Prof. Kevin He, Prof. Jack Kalbfleisch and Prof. Rajiv Saran for taking the time to serve on my dissertation committee as well as reading this dissertation thoroughly and providing valuable feedback to improve its quality.

I would like to thank the many friends I've met over the past five years, for making my graduate school experience memorable and ensuring that I'll look back fondly on the time I spent in the city of Ann Arbor. I would like to thank Tingting for being supportive and bringing cheer always. Finally, I would like to thank my family for their encouragement, inspiration and constant support. I dedicate this dissertation to them.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Competing risks data are commonly encountered in biomedical studies when subjects are subject to failure from many distinct causes. In this dissertation, we propose new methods for analyzing clustered competing risks data (Chapters 1 and 2) and for instrumental variable analysis of censored time-to-event data (Chapters 3 and 4).

In Chapter 1, we consider the problem of evaluation of center performance on multiple competing events. We propose estimating center effects through cause-specific proportional hazards frailty models that allow correlation among a centers cause-specific effects. Estimation of our model proceeds via penalized partial likelihood and is implemented in R. To evaluate center performance, we also propose a directly standardized excess cumulative incidence (ECI) measure. Therefore, based on our proposed methods, practitioners can evaluate centers either through the cause-specific hazards or the cumulative incidence functions. We demonstrate, through simulations, the advantages of the proposed methods to detect outlying centers, by comparing the proposed methods and existing methods which assume uncorrelated random center effects. In addition, we develop a Correlation Score Test to test the null hypothesis that the competing event processes within a center are uncorrelated. Using data from the Scientific Registry of Transplant Recipients, we apply our method to evaluate the performance of Organ Procurement Organizations on two competing risks: (i) receipt of a kidney transplant and (ii) death on the wait-list.

In Chapter 2, we propose to model the effects of cluster and individual-level covariates directly on the cumulative incidence functions of each risk through semiparametric additive regression models containing cluster-specific random effects. A unique

feature of our approach is that we model the dependency of failure times both within and across causes among individuals within a cluster by allowing for the correlation of cluster-specific random effects across causes. By decomposing the cause-specific cumulative incidence functions using a mixture model representation, we are able to estimate model parameters associated with all competing risks under consideration, satisfying the constraint that the sum of cumulative incidence functions does not exceed one. We develop estimating equations for parameter estimation and test our estimation procedure via simulations. We apply our method to multicenter competing risks data from the Scientific Registry of Transplant Recipients.

In Chapter 3, we turn our focus to causal inference in the censored time-to-event setting in the presence of unmeasured confounders. Unmeasured confounding of the relationship between a treatment and outcome of interest is a major concern in any observational study. Instrumental variable (IV) analysis methods are able to control for unmeasured confounding. However, IV analysis methods developed for censored time-to-event data tend to rely on assumptions that may not be reasonable in many practical applications, making them unsuitable for use in observational studies. In this chapter, we develop weighted estimators of the complier average causal effect on the restricted mean survival time. Our method is able to accommodate instrument-outcome confounding and adjust for covariate dependent censoring, making it particularly suited for causal inference from observational studies. We establish the asymptotic properties and derive easily implementable asymptotic variance estimators for the proposed estimators. Through simulation studies, we show that the proposed estimators tend to be more efficient than propensity score matching based estimators or inverse probability of treatment weighted estimators in certain situations, and tend to perform as well in other situations. We apply our method to compare HD and PD modalities for end stage renal disease (ESRD) patients using data from the United States Renal Data System (USRDS).

In Chapter 4, we develop an instrumental variable analysis method for competing risks data. Very few methods have been developed to address the problem of unmeasured confounding in the competing risks setting. Further, existing methods focus on estimating causal effects on a single, primary cause of interest. In doing so, these methods tend to overlook important features of the exposure-outcome relationship and ignore the interplay between causes. We develop a method that permits simultaneous inference of causal effects on the absolute risk or cumulative incidence of all causes. By using a semiparametric mixture component model, we ensure that the additivity constraint for the cause-specific cumulative incidence functions is satisfied. Our method makes no restriction about the type of exposure or IV and is able to accommodate exposure dependent censoring. We demonstrate finite sample properties through simulation studies. Using data from the United States Renal Data System (USRDS), we apply our method to compare HD and PD modalities for end stage renal disease (ESRD) patients with respect to two competing outcomes: (i) risk of death from cardiovascular diseases and (ii) risk of death from other causes.

CHAPTER I

Evaluating center performance in the competing risks setting: Application to outcomes of wait-listed end-stage renal disease patients

1.1 Introduction

The availability of electronic health records and the demand for value-driven healthcare have led to greatly increased interest in the methods for evaluation of center performance (Ash et al., 2012). For continuous or binary outcomes, center effects are usually estimated as either fixed or random effects models. Evaluation of center performance is then generally carried out by comparing these estimated risk-adjusted center effects to some fixed quantity, or the average center effect, or by using graphical checks (Spiegelhalter et al., 2012).

The proposed methods are motivated by the end-stage renal disease (ESRD) setting. There are thousands more patients in need of transplantation than there are donor kidneys. As a result medically suitable ESRD patients are placed on a waiting list. For example, in 2015, there were 98,956 patients on the kidney waiting list at year-end, but only 11,594 deceased-donor kidney transplants (Hart et al., 2016). In the United States, there are 58 wait-lists, each administered by an Organ Procurement Organization (OPO). Our objective here is to evaluate OPOs with respect

to (i) kidney transplantation and (ii) pre-transplant death (competing risks) among wait-listed patients.

While there has been extensive research conducted into establishing methods for institutional comparisons with respect to binary and continuous outcomes, apart from a few recent studies, time-to-event outcomes have received considerably less attention. He and Schaubel (2014a) assessed the standardized mortality ratio (SMR) measure based on the Cox model and developed an alternative based on stratification. In another study, He and Schaubel (2014b) developed a direct standardized measure of center performance.

Oftentimes in clinical and epidemiological settings, there is more than one competing outcome of interest. In such cases, there are two approaches to conceptualize the event times for the competing risks. The first approach assumes that, for every patient, a latent event time (Gail, 1975; Crowder, 2001) exists for each outcome and only the minimum of these (Cox, 1959) is observed. Under this conceptualization, latent event times must act independently in order for marginal quantities (e.g., cause- or event-specific survival function) to be identifiable. A second approach, adopted by us, assumes that only one event time, pertaining to the cause of failure, exists for each subject (Kalbfleisch and Prentice, 2002). Data from such settings can now be analyzed through the analysis of cause-specific hazards (Kalbfleisch and Prentice, 2002; **?**).

With competing risks data, a comparison of centers with respect to all-cause mortality has the potential to obscure important findings by averaging of dissimilar results (Van Rompaye et al., 2010). An analysis by cause has the potential to yield more interpretable and insightful conclusions (Putter et al., 2007). Fan and Schaubel (2016) proposed, as a center performance measure, the difference between the estimated cumulative incidence of transplant for patients at a given center and the average of the estimated cumulative incidences. Based on similar techniques, Van Rompaye, Erik-

son and Goetghebeur (2015) developed an 'excess cause-specific cumulative incidence' (ECI). For indirectly standardized measures, center performance is evaluated at the patient mix or covariate distribution of each center. Although useful for internal benchmarking, directly standardized measures are preferred for comparisons across centers (Varewyck et al., 2014). Note that random center effects may be preferable to fixed effects in the presence of small center sizes (Ash et al., 2012; Ohlssen et al., 2006; Kalbfleisch and Wolfe, 2013).

Most existing methods for clustered competing risks model the within-cluster dependence through a random effect, and concentrate on a single risk (or separate models for each risk) (Katsahian and Boudreau, 2011; Do Ha et al., 2014). In contrast, we propose a class of frailty models which allow a centers cause-specific random effects be correlated. This approach utilizes the additional information available in the form of correlation between cause-specific random effects within a center.

In this chapter, we develop a directly standardized ECI measure to contrast center performance on competing outcomes. We utilize an easily implementable penalized partial likelihood method (Ripatti and Palmgreen, 2000). Note that Gorfine and Hsu (2011) and Gorfine et al. (2014) also developed frailty models for correlated event times within-cluster. However, an Expectation-Maximization (EM) algorithm was used which requires numerical integration at each E-step. In comparison, our estimation procedure does not require any numerical integration and is implemented through a single call to **coxme** function of the **coxme** package (Therneau, 2009).

If competing events are indeed uncorrelated, fitting separate models is appropriate and easier than the proposed methods. Therefore, we also develop a convenient score test for the presence of correlation between competing risks within-center. The score test does not require fitting the joint model and, thus, provides an *a priori* checks the appropriateness of using separate cause-specific models, in lieu of the proposed methods.

1.2 Proposed Methods

1.2.1 Model and Likelihood

There are J centers or clusters, with each center j having n_j members ($j = 1, \ldots, J$) so that there are $\sum_{j=1}^{J} n_j = n$ individuals in the entire sample. For each subject $i (i = 1, ..., n_j)$ in center j, let T_{ij}^0 and C_{ij} denote the failure time and the censoring time, respectively, and let $\boldsymbol{X}_{ij}$ be a vector of time-independent covariates. The observed event time is then defined as $T_{ij} = \min(T_{ij}^0, C_{ij})$. Each subject fails due to one of K causes, we use Δ_{ij} ($\Delta_{ij} \in \{0, ..., K\}$) to indicate the cause of the observed failure for subject i in center j, with $\Delta_{ij} = 0$ if $T_{ij}^0 > C_{ij}$. The observed data consist of $\{T_{ij}, \Delta_{ij}, \boldsymbol{X}_{ij}, A_{ij}\}$ for $i = 1, \ldots, n_j$ and $(j = 1, \ldots, J)$, where $A_{ij} = 1$ if subject i belongs to center j and 0 otherwise.

Additionally, we define a vector of center-specific random effects or frailties, for the jth center, $\boldsymbol{\gamma}_j = (\gamma_{j1}, ..., \gamma_{jK})^T$, given which the event times for all subjects within that center are assumed to be conditionally independent. Thus, the cause-specific hazard function for cause k, for the subject i in the center j, is given by:

$$\lambda_{ijk}(t|\boldsymbol{X}_{ij}, \gamma_{jk}) = \lim_{h \downarrow 0} \frac{1}{h} Pr(t \leq T_{ij}^0 < t + h, \Delta_{ij} = k | T_{ij}^0 \geq t, \boldsymbol{X}_{ij}, \gamma_{jk})$$

and is assumed to be following the proportional hazards model:

$$\lambda_{ijk}(t|\boldsymbol{X}_{ij}, \gamma_{jk}) = \lambda_{0k}(t) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \gamma_{jk}\} \tag{1.1}$$

for $k = 1, .., K$ where $\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_k$ and $\lambda_{01}, ...\lambda_{0k}$ are cause-specific regression coefficients and cause-specific baseline hazards respectively. Here, we assume that the vector of covariates $\boldsymbol{X}_{ij}$ is the same for all causes, but it can be replaced by cause-specific vectors of covariates $\boldsymbol{X}_{ijk}$. The center-specific random effects imply a correlation between the cause-specific hazards across subjects within a center. Further, by as-

suming that the center-specific random effect vectors arise from a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{V}_j$, i.e., $\boldsymbol{\gamma}_j \sim MVN(\mathbf{0}, \boldsymbol{V}_j)$, our model allows for the association of different cause-specific hazards across individuals within a center. It is important to note that our model implies that the cause-specific hazards for different causes may be correlated across individuals within a center and not that the cause-specific event times within each individual are correlated. Indeed, as we do not adopt the latent failure time paradigm, our model is agnostic about the existence of different cause-specific event times within each individual.

We focus on the case of $K = 2$ competing causes, and allow for center-specific random effects for the two different causes to be negatively associated, i.e., $\text{Corr}(\gamma_{j1}, \gamma_{j2}) \leq 0$. To this end we reformulate the cause-specific hazards in equation (1.1) as

$$\lambda_{ij1}(t|\boldsymbol{X}_{ij}, b_j^0, b_j^1) = \lambda_{01}(t) \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + b_j^1 + b_j^0\} \tag{1.2}$$

$$\lambda_{ij2}(t|\boldsymbol{X}_{ij}, b_j^0, b_j^2) = \lambda_{02}(t) \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + b_j^2 - b_j^0\} \tag{1.3}$$

where $b_j^1 + b_j^0 = \gamma_{j1}$ and $b_j^2 - b_j^0 = \gamma_{j2}$. We have decomposed a center's cause-specific random-effect into two independent components: a shared random-effect, b_j^0, acting in opposite directions on the hazards of the two different risks, and a cause-specific random effect component b_j^k. This implies that $\text{Cov}(\gamma_{j1}, \gamma_{j2}) = -\text{Var}(b_j^0)$. We further assume that jointly $\boldsymbol{b}_j = (b_j^0, b_j^1, b_j^2) \sim p(\boldsymbol{b}_j; \boldsymbol{D}_j) = MVN(\mathbf{0}, \boldsymbol{D}_j(\boldsymbol{\theta}_j))$, where $\boldsymbol{D}_j(\boldsymbol{\theta}_j)$ is a diagonal covariance matrix with unknown parameters denoted by the vector $\boldsymbol{\theta}_j$.

We now construct the likelihood function for the model implied in equation (1.1) in terms of the parameters $(\lambda_0(t), \boldsymbol{\beta}_k^T, \boldsymbol{\theta}_j)$. Note that, for any given subject, $\lambda_{ij}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j) = \sum_{k=1}^K \lambda_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j)$. Thus, the cause-specific densities can be represented as $f_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j) = \lambda_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j) S_{ij}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j)$ for $k = \{1, .., K\}$, where $S_{ij}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j) = \exp\{-\sum_{k=1}^K \lambda_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{b}_j)\}$. Hence, the likelihood function can be written in terms of cause-specific hazard functions. Let the at-risk indicator for subject i in center j be given by $Y_{ij}(t) = I(T_{ij} \geq t)$.

Using the notation given in Section 1.2.1, we write the likelihood for subjects in center j as:

$$L_j = \int \prod_{i=1}^{n_j} \prod_{k=1}^{K} \left\{ \lambda_{0k}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\} \right\}^{I(\Delta_{ij}=k)} \times \left[\exp(-\int_0^t Y_{ij}(u)\lambda_{0k}(u) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\} du) \right] p(\boldsymbol{b}_j; \boldsymbol{D}_j(\boldsymbol{\theta}_j)) d\boldsymbol{b}_j \quad (1.4)$$

where the integral sign represents the unobserved frailties given by $\boldsymbol{b}_j$ being integrated out and $\boldsymbol{Z}_{ijk}$ are design vectors setup to obtain the cause-specific hazard models in equations (1.2) and (1.3). Specifically, if subject i is in center j then $\boldsymbol{Z}_{ij1} = (1, 1, 0)$ and $\boldsymbol{Z}_{ij2} = (-1, 0, 1)$, and if subject i does not belong to center j then $\boldsymbol{Z}_{ij1} = \boldsymbol{Z}_{ij2} = (0, 0, 0)$. It is important to note that for the construction of the above likelihood, we assumed the following: (1) Conditional on $\{\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ijk}, \boldsymbol{b}_j\}$, the event times and censoring times are independent and the censoring times are non-informative for $\{\boldsymbol{\beta}_k, \lambda_{0k}, k = 1, 2\}$, (2) $\boldsymbol{X}_{ij}$ and $\boldsymbol{b}_j$ are independent.

1.2.2 Estimation

It follows from equation (1.4) above that the overall likelihood of the data is given by:

$$L = \int \prod_{j=1}^{J} \prod_{i=1}^{n_j} \prod_{k=1}^{K} \{\lambda_{0k}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\}\}^{I(\Delta_{ij}=k)} \times [\exp(-\Lambda_{0k}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\})] \times p(\boldsymbol{b}_j; \boldsymbol{D}_j(\boldsymbol{\theta}_j)) d\boldsymbol{b}_j, \quad (1.5)$$

where $\Lambda_{0k}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\} = \int_0^{T_{ij}} Y_{ij}(u)\lambda_{0k}(u) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j^T \boldsymbol{Z}_{ijk}\} du$.

Let $\boldsymbol{b} = \{\boldsymbol{b}_1{}^T, ..., \boldsymbol{b}_J{}^T\}^T$ be a vector of all random-effects, obtained by stacking the center-specific vectors of random effects $\boldsymbol{b}_j, j = 1, ..., J$. Correspondingly, we define $p(\boldsymbol{b}; \boldsymbol{D}(\boldsymbol{\theta})) = MVN(\boldsymbol{0}, \boldsymbol{D}(\boldsymbol{\theta}))$ such that $\boldsymbol{D}(\boldsymbol{\theta})$ is a block-diagonal covariance

matrix composed of blocks formed by $\boldsymbol{D}_j(\boldsymbol{\theta_j})$. We further assume that $\boldsymbol{\theta}_j = \boldsymbol{\theta}_{j'} = (\theta_0, \theta_1, \theta_2)$; i.e., the center-specific random effect vectors, $\boldsymbol{b}_j$ are i.i.d with $\text{Var}(b_j^l) = \theta_l, l = \{0, 1, 2\}$.

The integrand in equation (1.5) above can be viewed as the full likelihood of the data under our model, composed of the conditional likelihood of the data given random effects $\boldsymbol{b}$, multiplied by the likelihood of the random effects. Taking the log, we define:

$$l_{full} = l_{cond} + l_{\boldsymbol{b}} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{k=1}^{K} I(\Delta_{ij} = k)\{\log(\lambda_{0k}(T_{ij})) + \boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j{}^T \boldsymbol{Z}_{ijk}\} - \sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{k=1}^{K} \Lambda_{0k}(T_{ij}) + \boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j{}^T \boldsymbol{Z}_{ijk} + \log|\boldsymbol{D}|^{-\frac{1}{2}} - \frac{1}{2}\boldsymbol{b}^T \boldsymbol{D}^{-1}\boldsymbol{b} \quad (1.6)$$

The above equation is a penalized log-likelihood for the observed data. As in Ripatti and Palmgren (2000), treating $\boldsymbol{b}$ as a fixed effect and using profile likelihood to estimate $\Lambda_{0k}(t)$ parameters, then plugging back the resulting Breslow (1974) estimator $\hat{\Lambda}_{0k}(t)$ into equation (1.6) yields the following penalized partial log-likelihood (PPLL):

$$l_{ppll} = l_1 + l_{\boldsymbol{b}} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{k=1}^{K} I(\Delta_{ij} = k)\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} + \boldsymbol{b}_j{}^T \boldsymbol{Z}_{ijk} - \log\sum_{r=1}^{J}\sum_{q=1}^{n_j} Y_{qr}(T_{ij}) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_{qr} + \boldsymbol{b}_r{}^T \boldsymbol{Z}_{qrk}\}\} + \log|\boldsymbol{D}|^{-\frac{1}{2}} - \frac{1}{2}\boldsymbol{b}^T \boldsymbol{D}^{-1}\boldsymbol{b}. \quad (1.7)$$

As recommended in Ripatti and Palmgren (2000), we suggest obtaining the estimates of $((\boldsymbol{\beta}_k, \boldsymbol{b}), k = \{1, 2\})$ as solutions to the PPLL. To estimate $\boldsymbol{\theta}$ we need to integrate out $\boldsymbol{b}$. As in Breslow and Clayton (1993), we use a Laplace saddle point approximation to the integration of penalized partial likelihood $L_{PPLL} = \exp(l_{ppll})$, with respect to $d\boldsymbol{b}$. Doing so, we obtain an expression for the log of the integrated likelihood as:

$$l_{INT} = -\frac{1}{2}\log|\boldsymbol{D}| - \frac{1}{2}\log|K"(\hat{\boldsymbol{b}})| - K(\hat{\boldsymbol{b}})$$

$$K(\hat{\boldsymbol{b}}) = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{k=1}^{K} I(\Delta_{ij}=k)\{\boldsymbol{\beta}_k^T\boldsymbol{X}_{ij}+\hat{\boldsymbol{b}}_j^T\boldsymbol{Z}_{ijk}-\log\sum_{r=1}^{J}\sum_{q=1}^{n_j} Y_{qr}(T_{ij})\exp\{\boldsymbol{\beta}_k^T\boldsymbol{X}_{qr}+\hat{\boldsymbol{b}}_r^T\boldsymbol{Z}_{qrk}\}\} + \log|\boldsymbol{D}|^{-\frac{1}{2}} - \frac{1}{2}\hat{\boldsymbol{b}}^T\boldsymbol{D}^{-1}\hat{\boldsymbol{b}}$$

and $\hat{\boldsymbol{b}}$ denotes the solution to the partial derivatives of $K(\boldsymbol{b})$ with respect to $\boldsymbol{b}$, i.e., $\hat{\boldsymbol{b}}$ solves:

$$K'(\boldsymbol{b}) = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{k=1}^{K} I(\Delta_{ij}=k)\left[\boldsymbol{Z}_{ijk}-\frac{\sum_{r=1}^{J}\sum_{q=1}^{n_j} Y_{qr}(T_{ij})\boldsymbol{Z}_{qrk}\exp\{\boldsymbol{\beta}_k\boldsymbol{X}_{qr}+\boldsymbol{b}_r{}^T\boldsymbol{Z}_{qrk}\}}{\sum_{r=1}^{J}\sum_{q=1}^{n_j} Y_{qr}(T_{ij})exp\{\boldsymbol{\beta}_k\boldsymbol{X}_{qr}+\boldsymbol{b}_r{}^T\boldsymbol{Z}_{qrk}\}}\right]-\boldsymbol{D}^{-1}\boldsymbol{b}=0. \tag{1.8}$$

The quantity $K"(\hat{\boldsymbol{b}})$ is the set of second partial derivatives of $K(\boldsymbol{b})$ at $\hat{\boldsymbol{b}}$. $K"(\hat{\boldsymbol{b}})$ is also the second partial derivative of l_{PPLL}, evaluated at $\hat{\boldsymbol{b}}$. If we define $\boldsymbol{H}$ as the matrix of second derivatives or Hessian of the PPLL with respect to $(\boldsymbol{\beta},\boldsymbol{b})$, such that:

$$\boldsymbol{H} = \begin{bmatrix}\boldsymbol{H}_{11} & \boldsymbol{H}_{12}\\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22}\end{bmatrix} = -\mathcal{I}(\boldsymbol{\beta},\boldsymbol{b}) + \begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}^{-1}\end{bmatrix}$$

where $\mathcal{I}(\boldsymbol{\beta},\boldsymbol{b}) = -\partial^2 l_1/\partial(\boldsymbol{\beta},\boldsymbol{b})\partial(\boldsymbol{\beta},\boldsymbol{b})'$, then $\boldsymbol{H}(\boldsymbol{\beta},\hat{\boldsymbol{b}})_{22} = K"(\hat{\boldsymbol{b}})$. We then have:

$$l_{INT} \approx l_1(\boldsymbol{\beta},\hat{\boldsymbol{b}}) + l_{\boldsymbol{b}}(\boldsymbol{\theta},\hat{\boldsymbol{b}}) - \frac{1}{2}\log|\boldsymbol{H}(\boldsymbol{\beta},\hat{\boldsymbol{b}})_{22}| \tag{1.9}$$

As demonstrated by Ripatti and Palmgren (2000), ignoring the last term on the right hand side of equation (1.9) while estimating $(\boldsymbol{\beta},\boldsymbol{b})$ leads to very little loss of information. This corresponds to using the PPLL to estimate $(\boldsymbol{\beta},\boldsymbol{b})$ via a Newton-Raphson algorithm. We have the following estimating equation for $\boldsymbol{\beta}$:

$$\partial l_{PPLL}/\partial\boldsymbol{\beta} =$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{k=1}^{K} I(\Delta_{ij}=k)\left[\boldsymbol{X}_{ij}-\frac{\sum_{j=1}^{J}\sum_{q=1}^{n_j} Y_{qr}(T_{ij})\boldsymbol{X}_{qr}\exp\{\boldsymbol{\beta}_k\boldsymbol{X}_{qr}+\boldsymbol{b}_r{}^T\boldsymbol{Z}_{qrk}\}}{\sum_{j=1}^{J}\sum_{q=1}^{n_j} Y_{qr}(T_{ij})\exp\{\boldsymbol{\beta}_k\boldsymbol{X}_{qr}+\boldsymbol{b}_r{}^T\boldsymbol{Z}_{qrk}\}}\right]=0 \tag{1.10}$$

The estimating equation for $\boldsymbol{b}$ is similarly obtained by setting $\partial l_{PPLL}/\partial \boldsymbol{b}$ to zero, and is identical to equation (1.8). Thus, equation (1.8), required for the saddle point Laplace approximation, is automatically satisfied when PPLL is used to estimate $\boldsymbol{b}$. To estimate $\boldsymbol{D}(\boldsymbol{\theta})$ we plug the estimated values $(\hat{\boldsymbol{\beta}})$ into equation (1.9) and solve for $\boldsymbol{\theta}$ that maximizes l_{INT}. This gives us the following estimating equation:

$$-\frac{1}{2}\left[tr(\boldsymbol{D}^{-1}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\theta}})+tr(\boldsymbol{H}_{22}^{-1}\frac{\partial \boldsymbol{D}^{-1}}{\partial \boldsymbol{\theta}})-\hat{\boldsymbol{b}}^T\boldsymbol{D}^{-1}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\theta}}\boldsymbol{D}^{-1}\hat{\boldsymbol{b}}\right]=0 \tag{1.11}$$

For a diagonal covariance matrix, as in our case, we obtain the following solution:

$$\hat{\theta}_l=\frac{(\hat{\boldsymbol{b}^l})^T(\hat{\boldsymbol{b}^l})+tr(\boldsymbol{H}_{22}^{l}(\hat{\boldsymbol{b}^l})^{-1})}{J}, l=\{0,1,2\} \tag{1.12}$$

where $\hat{\boldsymbol{b}^l}=\{\hat{\boldsymbol{b}_1^l},....,\hat{\boldsymbol{b}_J^l}\}$ and $\boldsymbol{H}_{22}^{l}(\hat{\boldsymbol{b}^l})$ is the sub-matrix corresponding to $\hat{\boldsymbol{b}^l}$ terms. The proposed estimation algorithm begins with an initial guess of $\boldsymbol{\theta}$, then alternates between using the PPLL to estimate $(\boldsymbol{\beta},\boldsymbol{b})$ as listed above and using equation (1.12) to update $\boldsymbol{\theta}$ until convergence. As suggested by Gray (1992), the variance of $(\hat{\boldsymbol{\beta}}^T,\hat{\boldsymbol{b}}^T)^T$ is obtained as:

$$\hat{V}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{b}})=\boldsymbol{H}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{b}})^{-1}\mathcal{I}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{b}})\boldsymbol{H}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{b}})^{-1} \tag{1.13}$$

To obtain the asymptotic distribution for $(\hat{\boldsymbol{\beta}},\hat{\mathbf{b}},\hat{\lambda}_{0k}(s))$, we assumed that the increments $\hat{\lambda}_{0k}(s)$ are independent of $(\hat{\boldsymbol{\beta}},\hat{\mathbf{b}})$. Under this assumption we estimated the variance of $\hat{\lambda}_{0k}(s)$ via a non-parametric bootstrap approach where the values of $(\hat{\boldsymbol{\beta}},\hat{\mathbf{b}})$ were treated as fixed by setting $\mathbf{X}\hat{\boldsymbol{\beta}}+\hat{\mathbf{b}}$ as an offset in the linear predictor of the instantaneous hazard. Thus, our desired asymptotic variance-covariance matrix for $(\hat{\boldsymbol{\beta}},\hat{\mathbf{b}},\hat{\lambda}_{0k}(s))$ was obtained using equation (1.13) to estimate the variance of $(\hat{\boldsymbol{\beta}},\hat{\mathbf{b}})$

and a non-parametric bootstrap approach to estimate the variance of $\hat{\lambda}_{0k}(s)$. In doing so we assume independence between $\hat{\lambda}_{0k}(s)$ and $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$. Our simulation studies suggest this to be a safe assumption. In reality, the increments of $\hat{\lambda}_{0k}(s)$ and $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$ may be weakly correlated. However, with increasing sample size one would expect this correlation to get weaker and have a negligible impact on the standard errors of estimates. Then, ignoring this correlation in return for substantial gains in computational efficiency seems appropriate. It should also be noted that, while using the Laplace approximation to the marginal log-likelihood leads to little loss of information, it might result in a slight underestimation of standard errors of fixed and random effect parameters if the cluster sizes are very small, as demonstrated in Ripatti and Palmgren (2000).

1.2.3 Center Effect Measures: Cumulative Incidence

We define the cumulative incidence function (CIF) of cause k for subject i at center j as:

$$F_{ijk}(t) = P(T_i^0 \leq t, \Delta_i = k | A_i = j, \boldsymbol{X_{ij}}), \tag{1.14}$$

the probability that an individual i in center j experiences a cause k event by time t. To evaluate the performance of center j with respect to type k events, we first define the average risk of events of type k at that center as $F_{jk}(t) = E_{\boldsymbol{X}}[F_{ijk}(t)]$, which is estimated as:

$$\hat{F}_{jk}(t) = \hat{E}_{\boldsymbol{X}}[F_{ijk}(t)] = \frac{\sum_{i=1}^{n_j} F_{ijk}(t)}{n_j} \tag{1.15}$$

Note that the above equation can be interpreted as potential risk for event k, at time t, that would be observed if the entire study population was treated at center j, assuming there are no unmeasured confounders. To compare the performance of center j to that of other centers we difference this potential risk with the average of such potential risks across all the centers. We call this measure the excess cumulative

incidence. This is denoted as $\delta_{jk}(t) = F_{jk}(t) - E_A[F_{jk}(t)]$ and estimated as:

$$\hat{\delta}_{jk}(t) = \hat{F}_{jk}(t) - \frac{\sum_{q=1}^{J} \hat{F}_{qk}(t)}{J} \tag{1.16}$$

1.2.4 Estimating Center Effects

We estimate cumulative incidence functions, defined in equation (1.14) using the cause-specific hazards estimated from section 1.2. We note that the cause-specific CIF for cause k, individual i at center j can be written as:

$$F_{ijk}(t) = \int_0^t S_{ij}(s)\lambda_{ijk}(s)ds, \tag{1.17}$$

for which an estimate $\hat{F}_{ijk}(t)$ is then obtained by plugging into equation (1.17) the following estimated quantities:

$$\hat{\lambda}_{ijk}(s) = \hat{\lambda}_{0k}(s)\exp(\hat{\boldsymbol{\beta}}_k \boldsymbol{X}_{ij} + \hat{b}_j \boldsymbol{Z}_{ijk}) \quad ; \quad \hat{S}_{ij}(s) = \exp\{-\sum_{k=1}^{2} \hat{\Lambda}_{0k}(s)\exp(\hat{\boldsymbol{\beta}}_k \boldsymbol{X}_{ij} + \hat{\boldsymbol{b}}_j \boldsymbol{Z}_{ijk})\}$$

where $\hat{\boldsymbol{\beta}}_k$, $\hat{\boldsymbol{b}}_j$ are estimates obtained as detailed in Section 1.2.2, and $\hat{\Lambda}_{0k}(t) = \int_0^t \hat{\lambda}_{0k}(s)ds$ is the cumulative cause-specific baseline hazard function obtained by integrating the Breslow-Aalen (Breslow 1974) estimate of the cause-specific baseline hazard function. Estimates of $F_{jk}(t)$ and the excess cumulative incidence at center j, $\hat{\delta}_{jk}(t)$, are subsequently obtained by plugging $\hat{F}_{ijk}(t)$ into equations (1.14) and (1.16) respectively.

To obtain the variance of the cause-specific cumulative incidence and excess cumulative incidence functions, we apply a parametric bootstrap approach. Specifically, we re-sample the estimated parameters $\hat{\boldsymbol{\beta}}_k$, $\hat{\boldsymbol{b}}_j$ and $\hat{\lambda}_{0k}(s)$ from their estimated asymptotic distributions to obtain bootstrapped estimates of the cumulative incidence functions. The variance of $\hat{F}_{jk}(t)$ and $\hat{\delta}_{jk}(t)$ are estimated as variance of the corresponding boot-

strapped estimates.

1.3 Score test of Correlation of Cause-specific Hazards

As mentioned in Section 1.2.1, equation (1.1), the cause-specific hazard function for cause k, for the ith subject in center j, is assumed to follow:

$$\lambda_{ijk}(t|\boldsymbol{X}_i, \gamma_{jk}) = \lambda_{0k}(t) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_i + \gamma_{jk}\}$$

Thus, the likelihood for the observed data in center j is:

$$L_j = \int \prod_{i=1}^{n_j} \prod_{k=1}^{K} \{\lambda_{0k}(t_i) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_i + \gamma_{jk}\}\}^{\Delta_{ik}(t)} \times \left[\exp(-\int_0^\tau Y_i(u)\lambda_{0k}(u) \exp\{\boldsymbol{\beta}_k^T \boldsymbol{X}_i + \gamma_{jk}\}dt)\right] p(\boldsymbol{\gamma}_j; \boldsymbol{V}(\boldsymbol{\theta}))d\boldsymbol{\gamma}_j \quad (1.18)$$

To develop a score test of the correlation of cause-specific hazards within centers, we consider a special case of the model in equation (1.1) when only $K = 2$ causes are present. Assume that the center-specific random effects or frailty for cause 2 and cause 1 differ by a multiplicative constant, i.e., $\gamma_{j2} = \omega\gamma_{j1}$, implying the following specification for the cause-specific hazards:

$$\lambda_{ij1}(t|\boldsymbol{X}_i) = \lambda_{01}(t) \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_i + \gamma_{j1}\}; \quad \lambda_{ij2}(t|\boldsymbol{X}_i) = \lambda_{02}(t) \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_i + \omega\gamma_{j1}\} \quad (1.19)$$

The presence of a correlation between the cause-specific hazards within centers is then assessed by testing $H_0 : \omega = 0$. When $\omega = 0$, there is little evidence for a linear relationship between center-specific random effects for causes 1 and 2. Conversely, even if the center-specific random effects are not perfectly correlated as implied by the specification in (19) but have a dependence of the form specified in model (1) we

would expect to reject the test of $H_0 : \omega = 0$ in favor of $H_a : \omega \neq 0$. This is because, in case of any non-zero correlation between the center-specific random effects, the specification in (19) with some $\omega \neq 0$ should provide a better fit to the observed data than that with $\omega = 0$. Thus, we propose to test for the presence of correlation between cause-sepcific hazards in model (1), i.e., $H_0 : Cov(\gamma_{j1}, \gamma_{j2}) = 0$, using the specification in (19) and testing $H_0 : \omega = 0$.

Under the joint model for the cause-specific hazards in (19), likelihood for observed data in center j is given by:

$$L_j = \int \prod_{i=1}^{n_j} \{\lambda_{01}(t_i) \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_i + \gamma_{j1}\}\}^{\Delta_{i1}(t)} \{\lambda_{02}(t_i) \exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i + \omega\gamma_{j1}\}\}^{\Delta_{i2}(t)}$$
$$\times [\exp(-\int_0^\tau Y_i(u)\lambda_{01}(u) \exp\{\boldsymbol{\beta}_1^T \boldsymbol{X}_i + \gamma_{j1}\} du)][\exp(-\int_0^\tau Y_i(u)\lambda_{02}(u) \exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i + \omega\gamma_{j1}\} du)]$$
$$\times p(\gamma_{j1}; \theta) d\gamma_{j1}. \quad (1.20)$$

The marginal log-likelihood for the observed data at all centers is then given by:

$$\log l(\omega, \boldsymbol{\beta}_k, \lambda_{0k}) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (\sum_{k=1}^{2} \Delta_{ik}(t) \{\log \lambda_{0k}(t_i) + \{\boldsymbol{\beta}_k^T \boldsymbol{X}_i\}\}) + \log \int K_j(z_j, t) p(z_j; \theta) dz_j$$

where $z_j = \log \gamma_{j1}$, and

$$K_j(z, t) = z_j^{\sum_{i=1}^{n_j} N_{i1}(t-) + \omega N_{i2}(t-)}$$
$$\times \exp\left\{ -z_j \left(\int_0^t Y_i(u)\lambda_{02}(u) \exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i\} du \right) - z_j^{\omega} \left(\int_0^t Y_i(u)\lambda_{02}(u) \exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i\} du \right) \right\}.$$

1.3.1 Correlation Score Test

Using the above formulation, the score test for correlation of the two cause-specific hazards tests $H_0 : \omega = 0$. The score function is:

$$U_\omega(\omega, \boldsymbol{\beta}_k, \lambda_{0k}) = \sum_j \frac{\int \{\sum_{i=1}^{n_j} N_{i2}(t_i) - (\int_0^t Y_i(u)\lambda_{02}(u)\exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i\}du)\} z_j^\omega \log z_j K_j(z_j) p(z_j;\theta) dz_j}{\int K_j(z_j) p(z_j;\theta) dz_j}$$

Setting $\omega = 0$ and replacing $\boldsymbol{\beta}_k, \lambda_{0k}$ and θ with their estimates when $\omega = 0$, we have:

$$\begin{aligned} U_\omega(\omega, \boldsymbol{\beta}_k, \lambda_{0k}) = \sum_j \frac{\int \{\sum_{i=1}^{n_j} N_{i2}(t_i) - (\int_0^t Y_i(u)\hat{\lambda}_{02}(u)\exp\{\hat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_i\}du)\} \log z_j \hat{K}_j(z_j) p(z_j;\hat{\theta}) dz_j}{\int \hat{K}_j(z_j) p(z_j;\theta) dz_j} \\ = \sum_j \widehat{M}_{2j}.\widehat{\log z_j} \end{aligned}$$

$\widehat{M}_{2j}$ is an estimate of the $\{\sum_{i=1}^{n_j} N_{i2}(t_i) - (\int_0^t Y_i(u)\lambda_{02}(u)\exp\{\boldsymbol{\beta}_2^T \boldsymbol{X}_i\}du)\}$, the sum of the martingale residuals for cause 2 at center j; and $\widehat{\log z_j} = E[\log z_j | O_j]$, i.e., the posterior expectation of the log frailties given the observed data in center j, O_j. If the frailties z_j are assumed to follow a log normal distribution, there is no closed form expression for $\widehat{\log z_j}$, however we can use the estimates $\hat{\gamma}_{j1}$ obtained by maximizing the penalized partial log-likelihood for cause 1. Balan et al. (2016) note that the test of $H_0 : \omega = 0$ can be carried out by testing if $\widehat{M}_{2j}$ and $\widehat{\log z_j}$ are correlated. Thus, the correlation score test (CST) tests if there is a linear dependency between $\widehat{M}_{2j}$ and $\widehat{\log z_j}$ and uses the regular t statistic from linear regression as the test statistic, $t = r\sqrt{(J-2)/(1-r^2)}$. Under $H_0 : \omega = 0$, asymptotically, t follows a t distribution with $J - 2$ degrees of freedom.

1.4 Simulation Studies

In the first (of two) set of simulations, we evaluated the fixed effect parameter estimators, variance components of the random effects, and Correlation Score

Test. There were $K = 2$ competing risks, and $J = 50$ or $J = 100$ centers (configurations 1 and 2, respectively). The center-specific random effects γ_{j1}, γ_{j2} followed a mean zero multivariate normal (MVN) distribution with variance components $\boldsymbol{\sigma}_j = (\sigma_1^2, \sigma_2^2, \rho_{12}) = (0.25, 0.25, -0.5)$. Using the re-parameterization described in Sections 1.2.1 and 1.2.2, this corresponds to the center-specific random effects vector $\boldsymbol{b}_j = (b_j^0, b_j^1, b_j^2)$ being generated from a MVN with mean zero and diagonal covariance matrix $\boldsymbol{D}$ with elements $\boldsymbol{\theta}_j = (\theta_0, \theta_1, \theta_2) = (0.125, 0.125, 0.125)$. The sample size within each center was fixed at $n_j = 20$ or $n_j = 50$ for different sub-configurations. In addition, we considered a single $N(0, 1)$ covariate X_i with regression coefficients $\beta_1 = 0.5$ and $\beta_2 = 1.25$ for causes $k = 1$ and $k = 2$ respectively. Given $\beta_k, \boldsymbol{\gamma}_j$ and the covariate X_i we generated a failure time T_i^0 for each subject within center j from an exponential distribution with rate parameter $\mu = \sum_{k=1}^{2} \mu_k = \sum_{k=1}^{2} \exp(\beta_k X_i + \gamma_{jk})$. We assigned a cause of failure for subject i in center j given a failure at time t using $Pr(\Delta_i = k | T_i^0 = t) = \mu_k / \mu$. Finally, all censoring occurred at time $\tau = 0.4$ in all configurations.

As shown in Table 1.1, the proposed method performs very well in estimating the parameters of interest. Also in Table 1.1, we present results of simulations where the center-specific random effects γ_{j1}, γ_{j2} were generated from a mean zero MVN with $\boldsymbol{\sigma}_j = (\sigma_1^2, \sigma_2^2, \rho_{12}) = (0.25, 0.25, 0)$, in order to assess the loss in efficiency due to unnecessarily estimating a correlation parameter when the true random effects are not correlated.

In Table 1.2, we evaluate the proposed CST and a likelihood ratio test (LRT) of the correlation between cause-specific hazards, via $H_0 : \rho = 0$. For each (J, n_j) configuration, the Type 1 error rate was calculated as the mean number of times H_0 when the random effects were generated from a mean zero MVN with $\boldsymbol{\sigma}_j = (0.25, 0.25, 0)$. Similarly, the Power was the mean number of rejections when the random effects were generated from a mean zero MVN with $\boldsymbol{\sigma}_j = (0.25, 0.25, -0.5)$.

Table 1.1: Estimating Regression Coefficients and Variance Components: Results from 500 Simulated Datasets

J	n_j		True Value	Bias	ESD	CP	True Value	Bias	ESD	CP
50	20	β_1	0.5	0.007	0.075	0.946	0.5	0.000	0.075	0.954
		β_2	1.25	0.002	0.072	0.950	1.25	0.002	0.074	0.942
		θ_1	0.125	-0.003	0.068	–	0	0.022	0.036	–
		θ_2	0.125	-0.001	0.088	–	0.125	-0.027	0.095	–
		θ_3	0.125	0.005	0.087	–	0.125	-0.021	0.089	–
50	50	β_1	0.5	-0.001	0.043	0.962	0.5	-0.001	0.043	0.962
		β_2	1.25	0.000	0.044	0.954	1.25	-0.004	0.046	0.944
		θ_1	0.125	-0.002	0.051	–	0	0.020	0.027	–
		θ_2	0.125	0.003	0.066	–	0.125	-0.020	0.073	–
		θ_3	0.125	-0.004	0.057	–	0.125	-0.026	0.069	–
100	20	β_1	0.5	0.003	0.050	0.960	0.5	0.000	0.051	0.960
		β_2	1.25	0.001	0.051	0.946	1.25	0.001	0.053	0.942
		θ_1	0.125	-0.005	0.053	–	0	0.017	0.029	–
		θ_2	0.125	0.003	0.066	–	0.125	-0.019	0.074	–
		θ_3	0.125	-0.001	0.064	–	0.125	-0.021	0.065	–
100	50	β_1	0.5	0.001	0.032	0.942	0.5	-0.001	0.033	0.944
		β_2	1.25	0.000	0.031	0.952	1.25	0.002	0.030	0.964
		θ_1	0.125	0.002	0.037	–	0	0.015	0.023	–
		θ_2	0.125	-0.001	0.043	–	0.125	-0.017	0.053	–
		θ_3	0.125	0.000	0.041	–	0.125	-0.012	0.049	–

Table 1.2: Power and Type I error of proposed Correlation Score Test (CST), and Likelihood Ratio (LR) tests. The null hypothesis is no correlation between cause-specific hazards within center: Results from 500 Simulated Datasets

Number of Centers	Subjects per Center	Type I Error		Power	
(J)	(n_j)	LRT	CST	LRT	CST
50	20	0.006	0.032	0.416	0.358
50	50	0.028	0.026	0.782	0.692
100	20	0.022	0.048	0.710	0.654
100	20	0.034	0.036	0.982	0.960

The CST seems to do almost as well as the LRT, attaining a type I error rate closer to the nominal 0.05 and achieving nearly as much power. More importantly, the CST is carried out in much less computation time, since it does not require fitting the full model.

In the second simulation study, we evaluated our estimators of the center-specific random effects $\{\gamma_{j1}, \gamma_{j2}\}$. Again, $K = 2$, $J = 50$, and $X_i \sim N(0, 1)$ with regression coefficients $\beta_1 = 0.5$ and $\beta_2 = 1.25$ for $k = 1$ and $k = 2$ respectively. Of the 50 centers, we fixed the value of the random effects for center j' and allowed the random effects for the remaining 49 centers to come from a mean 0 MVN with $\boldsymbol{\sigma}_j = (\sigma_1^2, \sigma_2^2, \rho_{12}) = (0.25, 0.25, -0.5)$. The sample size for each of these 49 centers, $n_j, j \neq j'$ was set equal to the random draw from a $N(100, 40^2)$ variate bounded at 20. Given $\beta_k, \boldsymbol{\gamma}_j$ and X_i, we generated T_i^0 from an exponential distribution with rate parameter $\mu_i = \sum_{k=1}^{2} \mu_{ik}$, where $\mu_{ik} = \exp(\beta_k X_i + \gamma_{jk})$, and assigned a cause of failure using $Pr(\Delta_i = k | T_i^0 = t) = \mu_{ik}/\mu$. Censoring again occurred at time $\tau = 0.4$.

We studied the performance of our estimators at different values of the random effects $\{\gamma_{j'1}, \gamma_{j'2}\}$ and at different $n_{j'}$ values. We compared the proposed method to an approach that fits separate frailty models for each k and therefore ignores the correlation between the center-specific random effects. As shown in Table 1.3, the

proposed method produces center effect estimates with smaller mean square error, regardless of the center size and effect.

An expanded version of Table 1.3 is made available in Appendix (see Table A.1). While both methods produce shrinkage, leveraging information on the correlation structure of the center-specific random effects leads to estimates with reduced shrinkage and higher rates of coverage. These gains in bias and coverage become more pronounced with decreasing sample sizes, and as the true values of the center effects deviate from the mean of the random effect distribution.

To examine our proposed excess cumulative incidence (ECI) center effect measure, we conducted simulations where the center-specific effects $\{\gamma_{j1}, \gamma_{j2}\}$ were known for all centers. We set $J = 50$, with n_j set equal to the maximum of 20 and a $N(100, 40^2)$ variate. Center-specific effects $\{\gamma_{j1}, \gamma_{j2}\}$ were each fixed at one realization from a MVN with mean 0 and $\boldsymbol{\sigma}_j = (\sigma_1^2, \sigma_2^2, \rho_{12}) = (0.25, 0.25, -0.5)$; theses were then treated as true center effects. We set $X_i \sim N(0, 1)$, with $\beta_1 = 0.5$ and $\beta_2 = 1.25$ for causes 1 and 2 respectively. Failure times and causes were then generated as presented earlier. Censoring was again at $\tau = 0.4$. The true ECI for each center was calculated at $t = 0.3$. In Table 1.4, we compare the proposed method with fitting separate cause-specific Cox frailty models. In terms of mean squared error of the ECI estimates, the proposed method generally out-performs the separate-models approach. A striking example, from Table 1.4, is the ECI estimates for Center $j = 23$, whose true ECI values for cause 1 and cause 2 are at opposite extremes.

1.5 Application

We applied the proposed methods to evaluate Organ Procurement Organizations (OPOs) with respect to two competing risks: (i) deceased-donor kidney transplantation (ii) death (prior to transplantation). We use data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donor,

Table 1.3: Estimating Center-Specific Effects: Results from 500 Simulations

			Proposed Method				Ignornig Correaltion of Random Effects
$n_{j'}$		True Value	Bias	ESD	ASE	CP	Relative MSE
20	$\gamma_{j'1}$	0.0	-0.019	0.231	0.322	0.988	1.113
	$\gamma_{j'2}$	0.0	-0.015	0.239	0.305	0.980	1.084
	$\gamma_{j'1}$	0.5	-0.175	0.241	0.297	0.970	1.232
	$\gamma_{j'2}$	-0.5	0.168	0.249	0.327	0.972	1.248
	$\gamma_{j'1}$	1.0	-0.276	0.244	0.276	0.870	1.252
	$\gamma_{j'2}$	-1.0	0.397	0.244	0.354	0.838	1.700
40	$\gamma_{j'1}$	0.0	-0.007	0.222	0.263	0.988	1.093
	$\gamma_{j'2}$	0.0	-0.015	0.209	0.243	0.986	1.041
	$\gamma_{j'1}$	0.5	-0.097	0.208	0.232	0.946	1.203
	$\gamma_{j'2}$	-0.5	0.108	0.214	0.271	0.974	1.274
	$\gamma_{j'1}$	1.0	-0.141	0.203	0.209	0.916	1.242
	$\gamma_{j'2}$	-1.0	0.268	0.221	0.306	0.914	1.799
60	$\gamma_{j'1}$	0.0	-0.010	0.202	0.231	0.968	1.098
	$\gamma_{j'2}$	0.0	-0.005	0.187	0.210	0.976	1.074
	$\gamma_{j'1}$	0.5	-0.066	0.195	0.200	0.962	1.142
	$\gamma_{j'2}$	-0.5	0.070	0.209	0.240	0.958	1.148
	$\gamma_{j'1}$	1.0	-0.097	0.168	0.178	0.948	1.227
	$\gamma_{j'2}$	-1.0	0.219	0.219	0.276	0.890	1.666
80	$\gamma_{j'1}$	0.0	-0.018	0.181	0.210	0.988	1.095
	$\gamma_{j'2}$	0.0	-0.020	0.179	0.191	0.964	1.080
	$\gamma_{j'1}$	0.5	-0.071	0.180	0.180	0.954	1.170
	$\gamma_{j'2}$	-0.5	0.066	0.180	0.218	0.970	1.153
	$\gamma_{j'1}$	1.0	-0.105	0.161	0.161	0.894	1.206
	$\gamma_{j'2}$	-1.0	0.193	0.201	0.256	0.918	1.554
100	$\gamma_{j'1}$	0.0	-0.020	0.170	0.194	0.976	1.095
	$\gamma_{j'2}$	0.0	-0.015	0.157	0.176	0.978	1.065
	$\gamma_{j'1}$	0.5	-0.076	0.15	0.167	0.964	1.171
	$\gamma_{j'2}$	-0.5	0.059	0.197	0.203	0.932	1.123
	$\gamma_{j'1}$	1.0	-0.095	0.141	0.149	0.928	1.218
	$\gamma_{j'2}$	-1.0	0.155	0.196	0.242	0.944	1.494

Table 1.4: Estimating Excess Cumulative Incidence: Results from 500 Simulation

			Proposed Method				Ignornig Correlation of Random Effects
Cause	Center	True Value	Bias	ESD	ASE	CP	Relative MSE
1	14	-0.170	0.029	0.022	0.027	0.850	1.018
	16	-0.096	0.024	0.029	0.034	0.926	0.750
	17	-0.181	0.003	0.017	0.023	0.990	3.432
	38	-0.138	0.017	0.024	0.029	0.958	1.226
	1	-0.179	0.023	0.020	0.026	0.918	1.462
	36	0.006	-0.009	0.038	0.038	0.932	0.950
	4	-0.033	0.001	0.034	0.037	0.948	1.010
	49	-0.070	0.018	0.029	0.034	0.944	0.777
	32	-0.047	0.010	0.031	0.035	0.960	0.969
	34	0.005	0.001	0.035	0.039	0.948	1.028
	23	0.344	-0.022	0.045	0.047	0.934	1.279
	19	0.118	-0.013	0.043	0.045	0.904	0.939
	13	0.142	-0.015	0.044	0.046	0.942	1.036
	15	0.127	-0.007	0.043	0.044	0.938	1.082
	18	0.210	-0.022	0.047	0.048	0.904	0.988
2	26	-0.222	0.020	0.019	0.025	0.932	2.265
	25	-0.140	0.014	0.027	0.029	0.936	1.196
	20	-0.137	0.013	0.025	0.030	0.952	1.209
	23	-0.199	0.014	0.021	0.025	0.950	2.122
	5	-0.078	0.006	0.030	0.032	0.950	1.056
	29	-0.017	0.004	0.034	0.033	0.922	0.954
	11	-0.020	0.002	0.031	0.034	0.954	0.987
	45	-0.043	0.006	0.032	0.033	0.934	0.965
	34	-0.058	0.007	0.029	0.033	0.952	0.993
	9	-0.009	0.001	0.031	0.033	0.946	1.016
	41	0.203	-0.016	0.037	0.037	0.928	1.065
	40	0.158	-0.012	0.035	0.036	0.934	1.076
	31	0.157	-0.016	0.038	0.038	0.900	0.969
	17	0.371	-0.020	0.037	0.034	0.902	1.327
	14	0.111	-0.009	0.033	0.035	0.926	1.245

wait-listed candidates, and transplant recipients in the U.S., submitted by the members of the Organ Procurement and Transplantation Network (OPTN), and has been described elsewhere. The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors.

The study cohort included patients wait-listed between 1/1/2010 and 4/30/2010. Patients were followed from the date of listing until the earliest of receipt of a kidney transplant, death, removal from wait-list, or the end of the observation period, 12/31/2012. Using the proposed methods, we compared OPOs across the U.S. with respect to the cumulative incidence of receiving a deceased-donor transplant and the cumulative incidence of death prior to transplantation. The time point we chose was two years post wait-listing, an appropriate time horizon based on previous related analyses (e.g., Fan and Schaubel, 2016). Patients receiving a living donor transplant were treated as independently censored, which is appropriate from the perspective that living-donor transplantation depends on many factors related to a patient's specific circumstances and largely independent of OPO. Note that living-donor transplantation was not a cause of our interest, rendering unappealing its inclusion as a separate cause.

Our study population included $n = 11,759$ patients across $J = 58$ OPOs across the U.S. A total of 2,408 patients (20.5%) received a deceased-donor kidney transplant, while 1,114 (9.5%) died first. We adjusted for the following patient-level covariates: age at listing, race, sex, body mass index, primary renal diagnosis, panel reactive antibody level and blood type. Owing to the large dimension of the covariate vector, we used a two-stage approach, as done in Kalbfleisch and Wolfe (2013), to obtain the risk-adjusted center effects (see also He and Schaubel, 2014b). Specifically, we estimated the patient-level covariates at the first stage by fitting a Cox model stratified by OPO. At the second stage, we estimated the cause-specific OPO effects by fitting

the proposed model, using the patient-level linear predictor from the first stage as an offset. The estimated variance components are given by $\hat{\boldsymbol{\sigma}}_j = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho}_{12}) = (0.619, 0.031, 0.210)$. The estimated correlation was determined to be statistically significant, with the CST yielding a p-value of 0.021.

Figure 1 displays the estimated OPO-specific ECI's at 2 years post-listing, along with 95% confidence intervals. The ECIs of transplantation ranged from -0.120 to 0.404, and the ECIs of death ranged from -0.126 to 0.115. For a given OPO, a high ECI for transplantation and a low ECI for death represent good performance. We classified OPOs as low- or high-outliers based on the 95% confidence intervals.

We compared the proposed method to a method that ignores the correlation between the cause-specific center effects with respect to outlier classification (Table A.2). While the two methods produced nearly identical classifications of OPOs based on the incidence of transplant, the proposed method classified 6 more OPOs as outliers than fitting separate frailty models by cause. This is a consequence of the reduction in shrinkage in the ECI estimates by the proposed method, due to leveraging the information on the correlation structure.

1.6 Discussion

In this chapter, we develop methods for evaluating center performance in the competing risks setting. We propose estimating center effects through cause-specific proportional hazards frailty models that allow correlation among a centers cause-specific hazards. We also propose a score test to test for the presence of correlation between a center's cause-specific hazards.

In our application, the cause-specific center effects do not seem to be strongly correlated. In scenarios where the correlation between cause-specific center effects is on the higher side, as maybe the case, for example, if there exists an unmeasured covariate influencing both outcomes, using the proposed method instead of currently

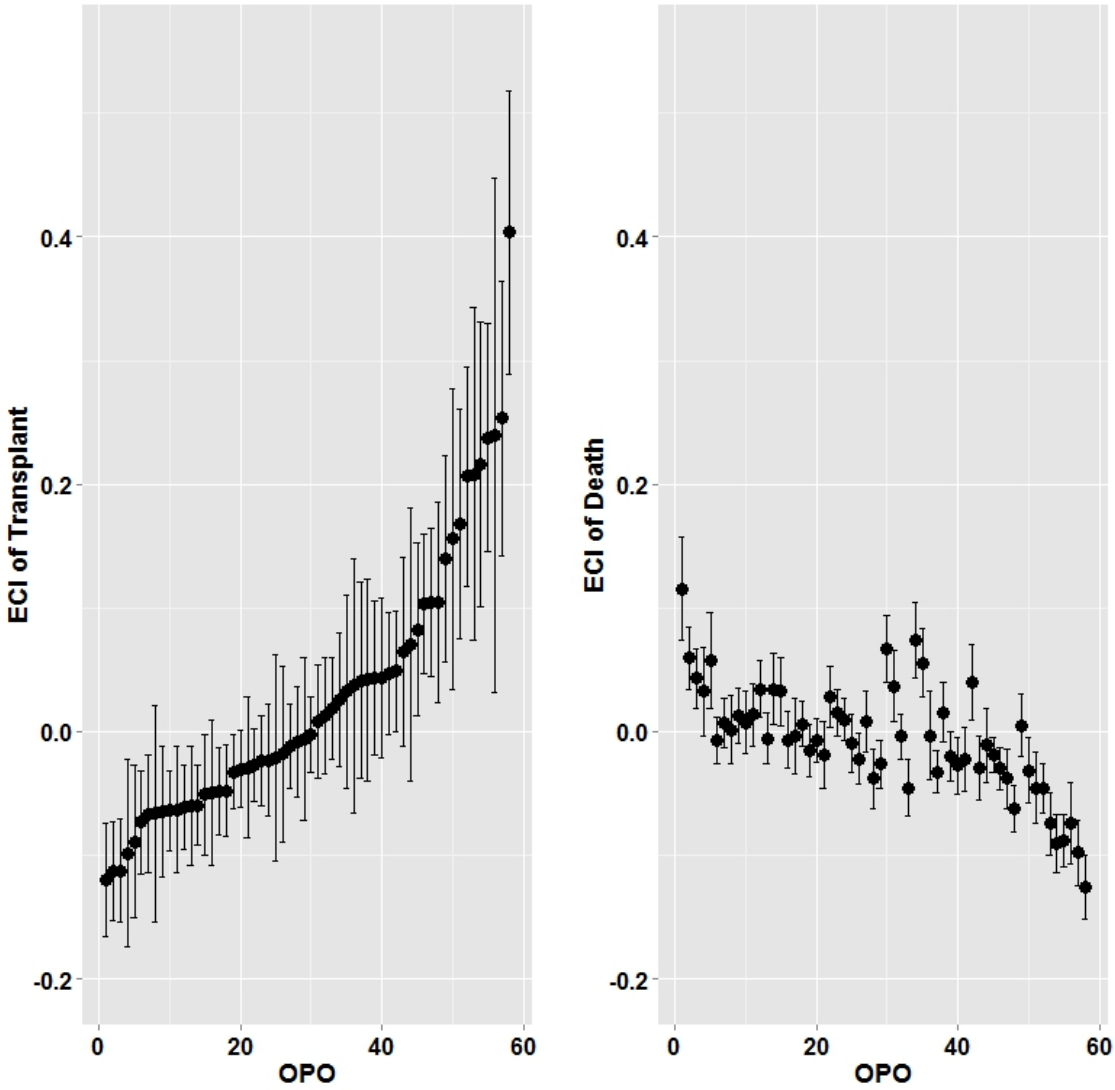

Figure 1.1: Analysis of Scientific Registry of Transplant Recipients (SRTR) Data: Caterpillar Plots of Excess Cause-specific Cumuluative Incidence of Death and Kidney Transplantation for 58 Organ Procurement Organizations

available methods may produce a larger change in classification of centers than seen here. Since fitting the proposed model may be computationally cumbersome, we recommend first using the proposed CST, to determine if the proposed model is warranted (the alternative being cause-specific frailty models).

As mentioned in Section 1.2.1, we assume that the patient-level covariates and center-level random-effects are uncorrelated. In practice, covariates and random-effects may be correlated, for example, sicker patients may prefer a center whose case-mix adjusted outcomes are better. This violation of our model assumption will lead to biased estimates of fixed effect parameters and consequently the center-level random effects in our model. In our application, to avoid problems due to confounding between the patient-level covariates and the OPO-specifc random-effects, we use a two-stage approach to estimate the center random effects. In the first stage, we fit a model stratified by OPO to estimate the regression parameters associated with a large number of patient characteristics. In the second stage, we use the estimated regression parameters as an offset in the linear predictor of the instantaneous hazard in a random-effects model, that is, we estimate the random effects given $\mathbf{X}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is estimated from the stratified model. This ensures that an unbiased estimate of $\hat{\boldsymbol{\beta}}$ is used while estimating the random effects. Apart from accommodating for confounding by patient-level covariates, the two-stage approach has an added benefit of easing computational burden. Given the potential for confounding by patient-level covariates, especially, in studies of center performance, we strongly recommend the use of the two-stage approach in lieu of a joint estimation approach which ignores the problem of confounding and is thereby bound to produce biased estimates of center effects.

The random effects estimated using our proposed two-stage approach represent an estimate of variation between centers after all the within center variation has been accounted for accurately. It is possible that the random-effects may still be correlated

with center-level averages of the covariates $\mathbf{X}$, and that this variation could further be partitioned into variation due to differences in center-level averages of the covariates $\mathbf{X}$ and other remaining variation between centers. The question of adjusting further for between-center differences while using a random-effects model may be a policy decision rather than a methodological decision. Further adjustment for between-center differences can be done using the between-method decomposition of covariates as suggested by Sjölander et al. (2013), where center-level averages of the covariates $\mathbf{X}$ are included as predictors.

CHAPTER II

A semiparametric mixture component model with random effects for the analysis of clustered competing risks data

2.1 Introduction

Competing risks data are encountered in biomedical studies when subjects are subject to failure from many distinct causes or events. Our work here is motivated by data arising from the end-stage renal disease (ESRD) setting where medically suitable patients in need of a kidney transplant are placed on a waiting list. For our purposes, this is when follow-up begins. While on the wait list, these patients may die before receiving a transplant. In this case, the competing events are: (i) receipt of a kidney transplant and (ii) pre-transplantation death while on the wait list.

In many situations, competing risks data cannot be considered independent and appropriate methods are needed to account for the correlation across subjects. This is also the case in our motivating application, since patients are clustered within transplant centers, i.e., the center where the patient's wait list registration was initiated. In this chapter, we develop a method for the simultaneous analysis of the absolute risk of different competing events in the ESRD setting, taking into consideration the correlation of failure times across patients within the same transplant center.

Early work on analyzing competing risks data was mostly focused on the estimation and modeling of cause-specific hazards or the instantaneous risk of an event (Prentice and Kalbfleisch, 1978). Let T denote the time to first event and $\epsilon \in \{1, ..., K\}$ denote the cause or type of failure, with K being the number of distinct causes. Then, the cause-specific hazard for an event $\epsilon = k$ at time t is $\lambda_k(t) = \lim_{\Delta t \to 0} Pr(t \leq T < T + \Delta t, \epsilon = k)/(\Delta t)$. Another identifiable quantity from competing risks data (T, ϵ) is the subdistribution or cumulative incidence function for cause k: $F_k(t) = Pr(T \leq t, \epsilon = k)$. The cumulative incidence function measures the absolute event-specific risk and presents a scientifically relevant alternative to the cause-specific hazard function. Given its scientific relevance, direct modeling of covariate effects on the cumulative incidence function has received a lot of focus recently. Fine and Gray (1999), building on earlier work by Gray (1988) proposed a Cox proportional hazards model for the subdistribution hazards; while, Sun et al. (2006) explored an additive hazards model. Andersen et al. (2003) and Klein and Anderson (2005) proposed a regression technique that utilizes pseudovalues from a jackknife statistic constructed from the cumulative incidence curve in a generalized estimating equation for estimation of covariate effects. Jeong and Fine (2007) proposed parametric regression of the cumulative incidence functions using a simple form of the Gompertz distribution for the log of the baseline cumulative subdistribution hazard function. Scheike et al. (2008) proposed direct binomial regression of the cumulative incidence function through a flexible semiparametric model where some covariates have time-varying effects and others have constant effects.

The aforementioned methods focus on estimating covariate effects on the cumulative incidence function of a primary cause of interest only, allowing an arbitrary structure for other causes. Oftentimes, investigators may be interested in simultaneous analysis of the different causes and their interrelationship. In such cases, fitting the aforementioned models separately by cause may not be desirable as information

on between cause association is generally not available. Recently, Mao and Lin (2016) developed a method based on semi-parametric transformation models that permits joint inference on all competing events. Another joint modeling approach, studied by Huang and Zhang (2008) and Chen (2010) takes into account the relationships among failure times through an assumed copula. Yet, none of the methods mentioned so far explicitly address the additivity constraint for the cumulative incidence functions, $\sum_k Pr(\epsilon = k) = 1$, that is, the constraint that a subject must eventually fail from one and only one of the distinct causes.

The additivity constraint can be explicitly incorporated by employing a mixture regression modeling framework. For example, Larson and Dinse (1985), Maller and Zhou (2002), Lu and Peng (2008) and Choi and Huang (2014) use the mixture model approach to decompose the model for cumulative incidence functions into a model for failure time conditional on the cause of failure, $Pr(T \leq t|\epsilon = k)$, and a model for the marginal probability of the cause of failure, $Pr(\epsilon = k)$. Here, we use the mixture regression modeling framework to develop semiparametric models for the cause-specific cumulative incidence functions in the clustered competing risks data setting.

In many applications, competing risks data cannot be considered independent. For instance, data from multicenter clinical trials and familial studies consist of clustered subjects whose failure time distributions may be correlated. The analysis of clustered competing risks data present two major challenges. First, methodologies developed for clustered single failure time data are not directly applicable to the competing risks setting. Second, methods need to appropriately account for correlation of event times among subjects within the same cluster. Methods developed for analyzing clustered competing risks data can broadly be categorized into: (1) methods designed for estimation and inference of appropriately defined measures of associations among failure times within clusters; and (2) regression methods for assessing covariate effects while accounting for within-cluster correlation of event times.

For measuring the association of the cause-specific failure times within a cluster, Bandeen-Roche and Liang (2002) introduced a nonparametric cause-specific cross hazard ratio for bivariate competing risks data; Cheng and Fine (2008) proposed an alternative representation using bivariate hazard functions. Cheng et al. (2007) derived nonparametric estimators of bivariate cause-specific hazard and cumulative incidence functions, and proposed two association measures in terms of these bivariate functions. These and a few other methods (Cheng 2010), do not accommodate covariate effects and require joint cause-specific intensities to be specified for all cause combinations. Scheike and Sun (2012) proposed a parametric regression model to estimate covariate effects on the cross-odds ratio for multivariate competing risks data. The cross odds ratio is a measure of association between cause-specific failure times within a cluster that can be represented in terms of the bivariate and univariate cumulative incidence functions.

Methods for regression analysis of clustered competing risks data can be divided into methods based on marginal and conditional approaches. Marginal approaches focus on estimating marginal effects of covariates from a population average regression model while accounting for the dependence across individuals within a cluster. For example, Zhou et al. (2012) proposed an extension of the proportional subdistribution hazards model with sandwich-type variance estimators to account for correlation within clusters. Conditional approaches, on the other hand, seek to estimate both the covariate effects and the within-cluster associations. Katsahian et al. (2006) and Katsahian and Bodreau (2011) extended the Fine-Gray proportional subdistribution hazards model for the clustered data setting by including a cluster-specific frailty or random-effect term. Scheike et al. (2010) proposed an alternative semiparametric random effects model under which the marginal model for the cumulative incidence function does not depend on the distribution of the random effects. These studies propose modeling strategies to estimate covariate and clustering effects on the abso-

lute risk of a single, primary cause of interest. If interest lies in assessing multiple causes, their models should be fitted repeatedly, focusing on each event separately. However, given that the additivity constraint is not explicitly addressed, these models may not hold simultaneously, making them difficult to interpret and unreliable for prediction purposes.

In this chapter we develop a semiparametric random effects model for the analysis of clustered competing risks data. Our method permits simultaneous inference of covariate effects on all competing risks and allows for correlation of failure times across subjects within a cluster. The dependence parameters in our model has an interpretation as a measure of association of failure times across subjects within a cluster. In the subsequent sections, we first introduce some notations and describe our modeling approach. Then, we describe an inference procedure for (i) regression parameters of the marginal model measuring covariate effects and (ii) the dependence parameters measuring the effect of clustering.

2.2 Setup and Model

Assume that there are n study subjects in total and that each comes from one of J centers or clusters (hereafter, we use the two terms interchangeably). Each center j , has n_j members, such that $\sum_{j=1}^{J} n_j = n$. For each subject $i(i = 1, ..., n_j)$ in cluster j, let T_{ij} and C_{ij} denote the failure time and the censoring time, respectively. Let $\boldsymbol{X}_{ij}$ be a vector of subject-specific time-independent covariates and let $\boldsymbol{W}_{ij}$ be another covariate vector that includes 1 and may share some common components with $\boldsymbol{X}_{ij}$. Assuming each subject fails due to one and only one of K causes, we let $\epsilon_{ij} \in \{1, ..., K\}$ denote the cause of failure, $N_{ijk}(t) = I(T_{ij} \leq t, \epsilon_{ij} = k)$ be the counting process for cause k, and let $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ be the observed-event indicator. Observed data for each subject i in cluster j then consist of $\{\boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \tilde{T}_{ij}, \tilde{\epsilon}_{ij}\}$, where $\tilde{T}_{ij} = \min(T_{ij}, C_{ij})$ is the observed event time, and $\tilde{\epsilon}_{ij} = \Delta_{ij}\epsilon_{ij}$ is the observed event indicator.

Further, to incorporate the additive constraint, we propose using the mixture model representation for competing risks. In particular, the model for cause-specific cumulative incidence functions is decomposed into a model for the distribution of ϵ_{ij} given the covariate vector $\boldsymbol{W}_{ij}$, $Pr(\epsilon_{ij} = k|\boldsymbol{W}_{ij})$, and a model for conditional cumulative incidence function $Pr(T_{ij} \leq t|\epsilon_{ij} = k, \boldsymbol{X}_{ij})$. We adapt this approach to the clustered data setting by including cluster-specific random effects in the models for conditional cumulative incidence functions. Specifically,for the jth center, we define a center-specific random effect a_j affecting the model for the distribution of ϵ_{ij} and a K-variate vector of center-specific random effects or frailties, $\boldsymbol{b}_j = (b_{j1}, ..., b_{jK})^T$ affecting model for conditional cumulative incidence function. Given these center-specific random effects, the event times for all subjects within that center are assumed to be conditionally independent.

For ease of representation and without loss of generality, we write down our models and describe our estimation procedure for $K = 2$, such that $\epsilon_{ij} \in \{1, 2\}$. We propose the following probit regression model for the marginal distribution of ϵ_{ij} given the covariate vector $\boldsymbol{W}_{ij}$,

$$\pi_k(\boldsymbol{W}_{ij}, a_j) = Pr(\epsilon_{ij} = k|\boldsymbol{W}_{ij}, a_j) = \Phi((-1)^{k-1}(\boldsymbol{\gamma}^T\boldsymbol{W}_{ij} + a_j)), k \in \{1, 2\}, \quad (2.1)$$

where Φ is the cumulative distribution function for a standard Normal variable. We assume the following model for conditional cumulative incidence functions:

$$Pr(T_{ij} \leq t|\epsilon_{ij} = k, \boldsymbol{X}_{ij}, b_{jk}) = 1 - \exp\{-b_{jk}t - \eta_k(t) - \boldsymbol{\beta}_k^T\boldsymbol{X}_{ij}t\}, k \in \{1, 2\}. \quad (2.2)$$

This leads us to the following random effects model for the cause-specific cumulative incidence function in the presence of clustering:

$$F_{ijk}(t|b_{jk}, \boldsymbol{X}_{ij}, a_j, \boldsymbol{W}_{ij}) = \pi_k(\boldsymbol{W}_{ij}, a_j)(1 - \exp\{-b_{jk}t - \eta_k(t) - \boldsymbol{\beta}_k^T \boldsymbol{X}_{ij}t\}), k \in \{1, 2\}. \tag{2.3}$$

Note that the additivity constraint is satisfied as $\sum_{k=1}^{2} \pi_k(\boldsymbol{W}_{ij}) = 1$ by design.

We assume that the center-specific random effect terms, a_j, in the probit regression model for the marginal distribution of ϵ_{ij} are uncorrelated with the random effect terms $\boldsymbol{b}_j = (b_{j1}, ..., b_{jK})^T$ in the additive hazards model for conditional cumulative incidence functions. Further, we assume that all random effects are normally distributed, such that, $a_j \sim N(0, \theta_a)$ and

$$\boldsymbol{b}_j = \begin{bmatrix} b_{j1} \\ b_{j2} \end{bmatrix} \sim MVN\Bigg(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \theta_{b11} & \theta_{b12} \\ \theta_{b12} & \theta_{b22} \end{bmatrix} \Bigg).$$

Then, under model (3), the marginal cumulative incidence function can be written as:

$$F_{ijk}(t|\boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}) = \tilde{\pi}_k(\boldsymbol{W}_{ij})(1 - \exp\{-H_k(t) - \boldsymbol{\beta}_k^T \boldsymbol{X}_{ij}t\})$$

where $\tilde{\pi}_k(\boldsymbol{W}_{ij}) = E_{\theta_a}[\Phi((-1)^{k-1}(\boldsymbol{\gamma}^T \boldsymbol{W}_{ij} + a_j))] = \Phi((-1)^{k-1}\tilde{\boldsymbol{\gamma}}^T \boldsymbol{W}_{ij})$, and $H_k(t) = \eta_k(t) - \log(E_{\theta_{bkk}}[\exp(-b_{jk}t)])$. Note that a conditional probit model with Normal random effects when marginalized yields a probit model with fixed-effect parameters equal to the corresponding parameters in the conditional model scaled by $\sqrt{(1 + \theta_a)}$, that is, $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}/\sqrt{(1 + \theta_a)}$.

We employ a two-stage estimation procedure for the parameters in model (3). In the first stage, we estimate parameters of marginal model (4), namely $\boldsymbol{\gamma}, \theta_a, \boldsymbol{\beta}_k$ and $H_k(t)$ and, in the second stage, we estimate the dependence parameters, $\theta_{b11}, \theta_{b22}, \theta_{b12}$, for model (3), i.e., the distributional parameters of the cluster-specific random effects

for both causes. We describe the proposed estimation procedure in the section that follows.

2.3 Inference Procedures

2.3.1 Estimation of Marginal Model Parameters

Let $G_{ij}(t) = Pr(C_{ij} > t|\boldsymbol{X}_{ij})$, then observe that, under conditional independence between C_{ij} and (T_{ij}, ϵ_{ij}), we have:

$$E\left[\frac{\Delta_{ij}N_{ijk}(t)}{Pr(C_{ij} > t)}\right] = E\left[E\left\{\frac{\Delta_{ij}N_{ijk}(t)}{G_{ij}(t)}\middle| T_{ij}, \epsilon_{ij}, \boldsymbol{X}_{ij}\right\}\right] = E(N_{ijk}(t)) = F_{ijk}(t). \quad (2.4)$$

Scheike et al. (2008) proposed to estimate parameters for a semiparametric regression model for the cumulative incidence functions by solving estimating equations based on the weighted response $\Delta_{ij}N_{ijk}(t)/G_{ij}(t)$. As $G_{ij}(t)$ is usually not known, it could be substituted with an estimate $\hat{G}_{ij}(t)$ obtained from either a Kaplan-Meier (1958) estimator of the censoring distribution, or an estimator based on a regression model, like the Cox (1972) proportional hazards model, relating the censoring distribution to covariates. Let $D_{H_k(t)}(t) = \partial F_{ijk}^{H_k(t),\boldsymbol{\beta}_k}(t)/\partial H_k(t)$ and $D_{\boldsymbol{\beta}_k}(t) = \partial F_{ijk}^{H_k(t),\boldsymbol{\beta}_k}(t)/\partial \boldsymbol{\beta}_k(t)$. If $\tilde{\pi}_k(\boldsymbol{W}_{ij})$ were known, the parameters $\{H_k(t), \boldsymbol{\beta}_k\}$ of the marginal model (3) can be estimated using the following estimating equations:

$$U_{H_k}(t, H_k(t), \boldsymbol{\beta}_k) = \sum_{j=1}^{J}\sum_{i=1}^{n_j} D_{H_k}(t)\left\{\frac{\Delta_{ij}N_{ik}(t)}{\hat{G}_{ij}(t)} - F_{ijk}(t)\right\} = 0 \quad (2.5)$$

$$U_{\boldsymbol{\beta}_k}(\tau, H_k(t), \boldsymbol{\beta}_k) = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\int_0^{\tau} D_{\boldsymbol{\beta}_k}(t)\left\{\frac{\Delta_{ij}N_{ijk}(t)}{\hat{G}_{ij}(t)} - F_{ijk}(t)\right\} = 0 \quad (2.6)$$

However, $\tilde{\pi}_k(\boldsymbol{W}_{ij})$ is unknown and depends on the unknown parameters $\{\boldsymbol{\gamma}, \theta_a\}$. To estimate $\{\boldsymbol{\gamma}, \theta_a\}$ simultaneously, we propose and approach similar to the generalized estimating equations of order 2 (GEE2) approach proposed by Feddag (2014).

Our estimating equations for $\{\boldsymbol{\gamma}, \theta_a\}$ are motivated by examining the conditional mean and empirical pairwise covariance of $\epsilon_{ij} = k$. We first note that ϵ_{ij} is not observed for all subjects. In the case that a subject is censored, i.e., when $\tilde{\epsilon}_{ij}$, and censoring occurs at time $\tilde{T}_{ij} = t$, the probability of $\epsilon_{ij} = 1$ conditional on $T \geq t$ is denoted by $g_{ijk}(t; \boldsymbol{\gamma}, \boldsymbol{\beta}_k, H_k(t), \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij})$, where

$$\begin{aligned} g_{ijk}(t; \boldsymbol{\gamma}, \boldsymbol{\beta}_k, H_k(t), \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij}) &= Pr(\epsilon = k | T_{ij} > t, \boldsymbol{W}_{ij}, \boldsymbol{\beta}_k, H_k(t)) \\ &= \frac{\pi_k(\boldsymbol{W}_{ij}) \exp\{-H_k(t) - \boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} t\}}{\sum_{k=1}^{2} \pi_k(\boldsymbol{W}_{ij}) \exp\{-H_k(t) - \boldsymbol{\beta}_k^T \boldsymbol{X}_{ij} t\}}. \end{aligned}$$

It follows that:

$$Pr(\epsilon_{ij} = k | \tilde{\epsilon}_{ij}, \tilde{T}_{ij} = t, \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij}) = E[I(\tilde{\epsilon}_{ij} = k) + I(\tilde{\epsilon}_{ij} = 0) g_{ijk}(t; \boldsymbol{\gamma}, \boldsymbol{\beta}_k, H_k(t), \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij})].$$

This leads us to the following first order estimating equation $\boldsymbol{\gamma}$ based on the conditional mean:

$$\boldsymbol{\delta}_{j1} = \sum_{i=1}^{n_j} \{I(\tilde{\epsilon}_{ij} = k) + I(\tilde{\epsilon}_{ij} = 0) g_{ijk}(\tilde{T}_{ij}) - \tilde{\pi}_k(\boldsymbol{W}_{ij})\}$$

The above equation is supplemented with an equation based on the empirical pairwise covariances within a cluster. Specifically, we define

$$\begin{aligned} \boldsymbol{\delta}_{j2} = \sum_{i=1}^{n_j} \sum_{i'=i}^{n_j} \frac{\Delta_{ij}}{\hat{G}_{ij}(\tilde{T}_{ij})} \frac{\Delta_{i'j}}{\hat{G}_{ij}(\tilde{T}_{i'j})} \Big\{ & I(\epsilon_{ij} = 1) I(\epsilon_{i'j} = 1) - \\ & E_{\theta_a}[\Phi((-1)^{k-1}(\boldsymbol{\gamma}^T \boldsymbol{W}_{ij} + a_j)) \Phi((-1)^{k-1}(\boldsymbol{\gamma}^T \boldsymbol{W}_{i'j} + a_j))] \Big\} \end{aligned}$$

Estimates of $\boldsymbol{\gamma}, \theta_a$ are then obtained by solving:

$$U_{\boldsymbol{\gamma}, \theta_a} = \boldsymbol{D}^T \sum_{j=1}^{J} [\boldsymbol{\delta}_{j1}, \boldsymbol{\delta}_{j2}]^T = 0 \tag{2.7}$$

where $\boldsymbol{D}$ is a matrix of derivatives, such that:

$$\boldsymbol{D} = \begin{bmatrix} \frac{\partial \boldsymbol{\delta}_{j1}}{\partial \boldsymbol{\gamma}} & \frac{\partial \boldsymbol{\delta}_{j1}}{\partial \theta_a} \\ \frac{\partial \boldsymbol{\delta}_{j2}}{\partial \boldsymbol{\gamma}} & \frac{\partial \boldsymbol{\delta}_{j2}}{\partial \theta_a} \end{bmatrix}.$$

The regression parameters of the marginal model can be estimated by starting with arbitrary initial values and solving the above estimating equations (2.5), (2.6) and (2.7) using an iterative algorithm until a pre-specified convergence criterion is satisfied. Upon convergence, we obtain estimators of $\boldsymbol{\gamma}, \theta_a, \boldsymbol{\beta}_k, H_k$, denoted by $\hat{\boldsymbol{\gamma}}, \hat{\theta}_a, \hat{\boldsymbol{\beta}}_k, \hat{H}_k$ respectively.

2.3.2 Estimation of Dependence Parameters

To estimate dependence parameters $\{\theta_{b11}, \theta_{b22}, \theta_{b22}\}$, we consider cross-moments between a pair of subjects i, i' within the same cluster, say j, experiencing a pair of causes k, k', where $k \in 1, 2, k' \in 1, 2$. Let:

$$V_{ii',j,kk'}(t) = (\hat{G}_{ij}(t))^{-1}\Delta_{ij}N_{ijk}(t)(\hat{G}_{i'j}(t))^{-1}\Delta_{i'j}N_{i'jk'}(t) \tag{2.8}$$

Let $v_{ii',j,kk'}(t) = E\{V_{ii',j,kk'}(t)\} = E\{(\hat{G}_{ij}(t))^{-1}\Delta_{ij}N_{ijk}(t)(\hat{G}_{i'j}(t))^{-1}\Delta_{i'j}N_{i'jk'}(t)\}$. As noted in Section 2, in our model, the cause-specific random effects within center j follows a multivariate normal distribution with $Var(b_{j1}) = \theta_{b11}$, $Var(b_{j2}) = \theta_{b22}$ and $\text{Cov}(b_{j1}, b_{j2}) = \theta_{b12}$. It follows that:

$$\begin{aligned} v_{ii',j,12}(t) = E_{\theta_a}[\pi_{1j}(\boldsymbol{W}_{ij}, a_j)\pi_{2j}(\boldsymbol{W}_{i'j}, a_j)](1 - \exp\{-H_1(t) - \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}t\} - \exp\{-H_2(t) \\ - \boldsymbol{\beta}_2^T\boldsymbol{X}_{i'j}t\} + \exp\{\theta_{12}t^2 - H_1(t) - \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}t - H_2(t) - \boldsymbol{\beta}_2^T\boldsymbol{X}_{i'j}t\}) \end{aligned}$$

$$v_{ii',j,11}(t) = E_{\theta_a}[\pi_{1j}(\boldsymbol{W}_{ij}, a_j)\pi_{1j}(\boldsymbol{W}_{i'j}, a_j)](1 - \exp\{-H_1(t) - \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}t\} - \exp\{-H_1(t)$$

$$- \boldsymbol{\beta}_1^T \boldsymbol{X}_{i'j} t\} + \exp\{\theta_1 t^2 - H_1(t) - \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} t - H_1(t) - \boldsymbol{\beta}_1^T \boldsymbol{X}_{i'j} t\})$$

$$v_{ii',j,22}(t) = E_{\theta_a}[\pi_{2j}(\boldsymbol{W}_{ij}, a_j)\pi_{2j}(\boldsymbol{W}_{i'j}, a_j)](1 - \exp\{-H_2(t) - \boldsymbol{\beta}_2^T \boldsymbol{X}_{ij} t\} - \exp\{-H_2(t) - \boldsymbol{\beta}_2^T \boldsymbol{X}_{i'j} t\} + \exp\{\theta_2 t^2 - H_2(t) - \boldsymbol{\beta}_2^T \boldsymbol{X}_{ij} t - H_2(t) - \boldsymbol{\beta}_2^T \boldsymbol{X}_{i'j} t\})$$

Let I_j denote the set index for jth cluster, the above results leads us to consider the following estimating equations for the dependence parameters:

$$U_{\theta_{12}}(\tau, \hat{\theta}_a, \hat{\boldsymbol{\gamma}}, \hat{H}_1, \hat{\boldsymbol{\beta}}_1, \hat{H}_2, \hat{\boldsymbol{\beta}}_2, \hat{G}) = \int_0^\tau \sum_{j=1}^{J} \sum_{i,i' \in I_j, i \neq i'} \{\hat{V}_{ii',j,12}(t) - v_{ii',j,12}(t, \theta_{12}, \hat{\theta}_a, \hat{\boldsymbol{\gamma}}, \hat{H}_1, \hat{\boldsymbol{\beta}}_1, \hat{H}_2, \hat{\boldsymbol{\beta}}_2)\} dt = 0 \quad (2.9)$$

$$U_{\theta_1}(\tau, \hat{\theta}_a, \hat{\boldsymbol{\gamma}}, \hat{H}_1, \hat{\boldsymbol{\beta}}_1, \hat{G}) = \int_0^\tau \sum_{j=1}^{J} \sum_{i,i' \in I_j, i < i'} \{\hat{V}_{ii',j,11}(t) - v_{ii',j,11}(t, \theta_1, \hat{\theta}_a, \hat{\boldsymbol{\gamma}}, \hat{H}_1, \hat{\boldsymbol{\beta}}_1)\} dt = 0 \quad (2.10)$$

$$U_{\theta_2}(\tau, \hat{\theta}_a, \hat{\boldsymbol{\gamma}}, \hat{H}_2, \hat{\boldsymbol{\beta}}_2, \hat{G}) = \int_0^\tau \sum_{j=1}^{J} \sum_{i,i' \in I_j, i < i'} \{\hat{V}_{ii',j,22}(t) - v_{ii',j,22}(t, \theta_2, \hat{\theta}_a, \hat{\boldsymbol{\gamma}}, \hat{H}_2, \hat{\boldsymbol{\beta}}_2)\} dt = 0 \quad (2.11)$$

We denote the estimators obtained by solving equations (2.9), (2.10) and (2.11) as $\hat{\theta}_{12}$, $\hat{\theta}_1$ and $\hat{\theta}_2$ respectively.

2.3.3 Variance Estimation using perturbation resampling

To estimate the variance and to construct confidence intervals for our proposed estimators, we use a perturbation-based resampling method. We apply perturbing ran-

dom variables directly to the estimating functions at the cluster-level to approximate the distribution of the aforementioned estimators. Specifically, let $\{\xi_j, j = 1, ..., J\}$ be J independent copies of a positive random variable ξ from a known distribution with unit mean and unit variance. Fixing the data at their observed values, perturbed estimators are obtained as the solution to the following perturbed estimating functions:

$$U^*_{H_k}(t, H_k(t), \boldsymbol{\beta}_k) = \sum_{j=1}^{J} \xi_j \sum_{i=1}^{n_j} D_{H_k}(t)\left\{ \frac{\Delta_{ij} N_{ik}(t)}{G^*_{ij}(t)} - F_{ijk}(t) \right\} = 0 \tag{2.12}$$

$$U^*_{\boldsymbol{\beta}_k}(\tau, H_k(t), \boldsymbol{\beta}_k) = \sum_{j=1}^{J} \xi_j \sum_{i=1}^{n_j} \int_0^{\tau} D_{\boldsymbol{\beta}_k}(t)\left\{ \frac{\Delta_{ij} N_{ijk}(t)}{G^*_{ij}(t)} - F_{ijk}(t) \right\} = 0 \tag{2.13}$$

$$U^*_{\boldsymbol{\gamma},\theta_a} = \boldsymbol{D}^T \sum_{j=1}^{J} \xi_j [\boldsymbol{\delta}_{j1}, \boldsymbol{\delta}_{j2}]^T = 0 \tag{2.14}$$

$$U^*_{\theta_{12}}(\tau, \boldsymbol{\gamma}^*, H^*_1, \boldsymbol{\beta}^*_1, H^*_2, \boldsymbol{\beta}^*_2, G^*) = \int_0^{\tau} \sum_{j=1}^{J} \xi_j \sum_{i,i' \in I_j, i \neq i'} \{V^*_{ii',j,12}(t) - v_{ii',j,12}(t, \theta_{12}, \theta^*_a, \boldsymbol{\gamma}^*, H^*_1, \boldsymbol{\beta}^*_1, H^*_2, \boldsymbol{\beta}^*_2)\} dt = 0 \tag{2.15}$$

$$U^*_{\theta_1}(\tau, \boldsymbol{\gamma}^*, H^*_1, \boldsymbol{\beta}^*_1, G^*) = \int_0^{\tau} \sum_{j=1}^{J} \xi_j \sum_{i,i' \in I_j, i < i'} \{V^*_{ii',j,11}(t) - v_{ii',j,11}(t, \theta_1, \theta^*_a, \boldsymbol{\gamma}^*, H^*_1, \boldsymbol{\beta}^*_1)\} dt = 0 \tag{2.16}$$

$$U^*_{\theta_2}(\tau, \boldsymbol{\gamma}^*, H^*_2, \boldsymbol{\beta}^*_2, G^*) = \int_0^{\tau} \sum_{j=1}^{J} \xi_j \sum_{i,i' \in I_j, i < i'} \{V^*_{ii',j,22}(t) - v_{ii',j,22}(t, \theta_2, \theta^*_a, \boldsymbol{\gamma}^*, H^*_2, \boldsymbol{\beta}^*_2)\} dt = 0 \tag{2.17}$$

where G^* is the perturbed version of $\hat{G}$, an estimator of the censoring distribution with weights $\{\xi_j, j = 1, ..., J\}$, and $V^*_{ii',j,kk'}(t) = (G^*_{ij}(t)^{-1} \Delta_{ij} N_{ijk}(t) (G^*_{ij}(t))^{-1} \Delta_{i'j} N_{i'jk'}(t)$. The above perturbed estimating equations (2.12)-(2.17) are solved using the same procedures used to solve their unperturbed counterparts, estimating equations (2.5)-(2.7)

and (2.9)-(2.11), to obtain estimators $\{\boldsymbol{\gamma}^*, \theta_a^*, \boldsymbol{\beta}_k^*, H_k^*, \theta_{12}^*, \theta_1^*, \theta_2^*\}$. By repeatedly generating $\{\xi_j, j = 1, ..., J\}$, say M times, we can obtain a large number of realizations of the perturbed estimators, say - $\{\boldsymbol{\gamma}^{*(m)}, \theta_a^{*(m)}, \boldsymbol{\beta}_k^{*(m)}, H_k^{*(m)}, \theta_{12}^{*(m)}, \theta_1^{*(m)}, \theta_2^{*(m)}, m = 1, ..., M\}$.

It can be shown that the unconditional distribution of estimates $\{\hat{\boldsymbol{\gamma}}, \hat{\theta}_a, \hat{\boldsymbol{\beta}}_k, \hat{H}_k, \hat{\theta}_{12}, \hat{\theta}_1, \hat{\theta}_2\}$ can be approximated by the conditional distribution of the perturbed estimates given the observed data (van der Vaart and Wellner, 1996). Thus, the variance and confidence intervals of estimates $\{\hat{\boldsymbol{\gamma}}, \hat{\theta}_a, \hat{\boldsymbol{\beta}}_k, \hat{H}_k, \hat{\theta}_{12}, \hat{\theta}_1, \hat{\theta}_2\}$ are estimated based on the empirical distribution of $\{\boldsymbol{\gamma}^{*(m)}, \theta_a^{*(m)}, \boldsymbol{\beta}_k^{*(m)}, H_k^{*(m)}, \theta_{12}^{*(m)}, \theta_1^{*(m)}, \theta_2^{*(m)}, m = 1, ..., M\}$.

2.4 Simulation Studies

In this section we describe simulation studies carried out to evaluate our proposed method. For the results presented in Tables 2.1 and 2.2, data were generated as follows. The failure time T_{ij} for each individual i in center j ($j \in \{1, ..., J\}$) was generated from model (2.3). First, a cause ϵ_{ij} was generated from model (2.1), where $W = \{1, Z_{ij}\}$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1) = (0.5, -0.5)$; and conditional on cause of failure a time was generated from model (2.2), where $X_{ij} = Z_{ij}$, $\beta_1 = 0.8, \beta_2 = 0.5, \eta_1(t) = t + 0.5t^2, \eta_2(t) = t + 0.5t^2$. The single covariate Z_{ij} was generated from a Bernoulli distribution with a success probability of 0.5. The cluster-specific random effects in the probit model (2.1), a_j, were generated from a $N(0, 0.49)$ distribution and the random effects in the model (2.2) for the conditional cumulative incidence, $\{(b_{j1}, b_{j2}), j = 1, ...J\}$, were generated from a multivariate normal distribution with mean $(0, 0)$ and variance components $(\theta_1, \theta_2, \theta_{12}) = (0.4, 0.4, -0.2)$. In each case, a censoring time for each individual was generated dependent on the covariate as follows $C_{ij} \sim Unif(1, 2) \times Z_{ij1} + Unif(0.25, 2) \times (1 - Z_{ij1})$. This setup lead to a censoring rate of approximately 18%, and an occurrence rate of approximately 47% for cause 1 failures and 35% for cause 2 failures in all sample size configurations.

Table 2.1: Simulation results for one covariate setting with negatively correlated cluster-specific random effects

Parameter	True Value	BIAS	SD	SE	CP
γ_0	0.5	0.026	0.116	0.111	0.936
γ_1	-0.5	-0.011	0.130	0.123	0.928
β_1	0.8	-0.002	0.232	0.220	0.940
β_2	0.5	0.053	0.251	0.258	0.928
θ_a	0.49	0.040	0.196	0.189	0.934
θ_{b11}	0.5	-0.069	0.321	0.276	0.882
θ_{b22}	0.5	-0.075	0.383	0.322	0.870
θ_{b12}	-0.25	-0.028	0.408	0.440	0.940

Bias, empirical bias of estimates; SD, empirical standard deviation; SE, mean of estimated standard error via the proposed resampling method; CP, empirical coverage probability

Tables 2.1 displays the results of our numerical investigations for a setting with $J = 100$ clusters and with varying sample size per cluster, generated as $n_j \sim Unif\{2, 3, 4, 5, 6, 7, 8, 9, 10\}$. To calculate the asymptotic standard errors of our estimators, we used the resampling procedure detailed in Section 2.3. Specifically, for each simulated dataset we generated $M = 100$ realizations of $\{\xi_j, j = 1, ..., J\}$; where ξ_j were independently distributed unit exponential random variables. Results shown in Tables 2.1 indicates that our method produces nearly ubiased estimators for both the marginal model parameters and the dependence parameters. The proposed resampling method produces standard error estimates that are close to the empirical standard errors and 95% confidence intervals that provide close to nominal coverage.

2.5 Application

We applied our proposed methods to data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data consist of information on all donors, transplant recipients and candidate patients on wait-lists across the U.S., submitted by the members of the Organ Procurement and Transplantation Network (OPTN), and

has been described elsewhere. We specifically studied the outcomes of candidates wait-listed for a kidney transplant. To this end, the competing events in our analysis were (i) receipt of a deceased donor kidney transplant (ii) death while on the wait-list.

As mentioned earlier, our data consisted of a natural clustering of patients within transplant centers. Specifically, our study cohort consisted of 3,114 patients wait listed for a kidney transplant between January 1st, 2002 and June 30th, 2002, across 52 transplant centers in UNOS Regions 1, 2 and 9. This included all transplant centers in the states of Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, West Virginia and Vermont. With these 52 transplant centers as the clustering units, we estimated the effect of patient-level covariates and the effects of clustering on cumulative incidence functions of the two competing events of interest. We followed patients from the date of listing until the earliest of receipt of a kidney transplant, death, removal from wait-list, or the end of the observation period, December 31st, 2008. As living-donor transplantation was not a cause of our interest, its inclusion as a separate cause was not considered and patients receiving a living donor transplant were treated as independently censored.

Of 3,114 patients, 51.2% received a deceased-donor kidney transplant and 19.2% died before end of observation period. Our model included following patient-level covariates for both causes: age at listing, race, sex, body mass index, primary renal diagnosis, peak renal reactive antibody level and blood type. In Tables 2.2 and 2.3, we present the results of our analysis. Firstly, we examine the various covariate effects on the marginal probability of transplantation, denoted by $\hat{\boldsymbol{\gamma}}$, and the covariate effects on conditional cumulative incidence of transplantation, $\hat{\boldsymbol{\beta}}_1$, and pre-transplant death, $\hat{\boldsymbol{\beta}}_2$. The reference categories are Caucasian, blood type O, female, glomerulonephritis for the primary renal disease diagnosis and BMI > 30.

Table 2.2: Results of application to SRTR data: Covariate Effects

	$\hat{\gamma}$		$\hat{\beta}_1$		$\hat{\beta}_2$	
Covariate	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	2.257	0.268	0	0	0	0
Race						
Black	-0.283	0.078	-0.117	0.02	-0.023	0.032
Asian	0.001	0.275	-0.184	0.034	-0.095	0.067
Hispanic	0.013	0.162	-0.085	0.033	-0.127	0.04
Other	-0.243	0.352	0.099	0.173	0.125	0.179
Blood Type						
A	0.216	0.107	0.109	0.028	0.016	0.025
AB	0.456	0.159	0.343	0.077	0.019	0.073
B	-0.103	0.115	0.042	0.023	-0.031	0.028
Gender						
Male	-0.028	0.069	0.006	0.023	0.006	0.02
Primary Renal Diagnosis						
Polycystic kidney disease	0.382	0.175	0.011	0.042	-0.067	0.072
Diabetes	-0.196	0.175	0.03	0.046	0.004	0.073
Hypertension	-0.018	0.174	0.009	0.044	-0.05	0.068
Other diagnosis	0.173	0.17	0.028	0.034	0.027	0.073
Peak Reactive Antibody	-0.007	0.002	-0.002	0.0005	-0.001	0.0005
BMI						
Low	-0.092	0.397	0.069	0.083	0.122	0.19
Normal	-0.036	0.082	0.019	0.022	0.063	0.035
Overweight	0.059	0.086	0.049	0.021	0.045	0.024
Age (in decades)	-0.295	0.041	0.015	0.01	0.019	0.009

Table 2.3: Results of application to SRTR data: Cluster Effects

Parameter	Estimate	SE
θ_a	0.040	0.017
θ_{b11}	0.010	0.008
θ_{b22}	0.018	0.009
θ_{b12}	0.010	0.007

From Table 2.2, age at listing, BMI > 30, panel reactive antibody and being African American demonstrated significant negative effects on the marginal probability of receiving a deceased donor kidney transplant. Being African American also seemed to have a significant effect on the timing of receipt of transplant for patients who eventually receive a transplant. In Table 2.3, we examine the effects of clustering on the data. The results seem to indicate that, for our data, the clustering of patients within transplant centers can be sufficiently addressed by including random effect terms in the probit regression model for the marginal probability of transplantation.

2.6 Discussion

In this chapter we propose a semiparametric random effects model for clustered competing risks data based on the mixture regression modeling framework. Our method presents an approach to estimate covariate effects on cumulative incidence functions of all competing causes simultaneously, while accounting for correlation of event times across subjects within cluster. To the best of our knowledge, there are no existing methods to deal with clustered or correlated competing risks data that allow for joint inference on all cause-specific cumulative incidence functions. Prominent methodologies focus on a single cause of interest and in doing so are not able to capture the interplay between competing events. By explicitly addressing the additivity constraint for cumulative incidence functions our method is able to provide meaningful predictions of absolute risk of all causes at a specified time and inference about marginal probabilities. Further, our method provides a quantification of the effects of clustering both within and across causes.

A potential drawback of our method is that it might require data with longer follow-up times. Essentially, our method requires sufficient information in the tail of the event time distribution to be able to distinguish between the marginal model for the type of event and the conditional model for the timing of an event given the

type.

CHAPTER III

Weighted estimators of the complier average causal effect on restricted mean survival time with observed instrument-outcome confounders

3.1 Introduction

A major concern in any study lacking randomized treatment assignment is the potential for confounding of the relationship between the treatment and outcome of interest. In the absence of randomization, estimation of the causal effect of treatment generally requires an untestable and often unrealistic assumption regarding the treatment selection mechanism. Specifically, it needs to be assumed that the treatment is randomly assigned conditional on the observed covariates, i.e., there are no unmeasured confounders of the treatment-outcome association. Unmeasured confounding can be overcome by conducting an instrumental variable (IV) analysis. This requires the availability of an IV, which is a variable (a) that is associated with the treatment of interest, (b) that has no direct effect on the outcome except through the treatment of interest and (c) whose association with the treatment and the outcome is not confounded by any unmeasured confounders. Such a variable, when available, can be used to identify treatment effects without knowledge of the treatment selection mechanism (Imbens and Angrist (1994), Angrist et al. (1996)). Some common examples

of IVs in the binary treatment setting include physician preferences for treatment prescription, randomized encouragement to treatment, and treatment assignment in randomized clinical trials with noncompliance.

With the use of instrumental variable methods in biomedical studies gaining popularity only recently, there has been very little research into developing methods for IV analysis of right censored time-to-event data. For the randomized study setting, Robins and Tsiatsis (1991) developed semiparametric estimators of the treatment effect under a semiparametric structural accelerated failure time model for the outcome. Loeys and Goetghebeur (2003) extended the approach proposed by Robins and Tsiatsis (1991) to a proportional hazards model of treatment effect working with the restriction that subjects randomized to control have no access to treatment. Baker (1998) worked with discrete-time survival data and developed an estimator of the difference in hazards at a specific time between compliers in the treatment and control groups of a randomized trial. This estimator is analagous to a standard IV estimator applied to a survival outcome at a specific time but can result in negative estimates of hazards and can be inefficient in certain situations. Seeking to gain efficiency, Nie et al. (2011) utilized the mixture structure implied by the latent compliance model to develop a plug-in non-parametric empirical maximum likelihood estimation approach for the difference between compliers in the treatment and control groups of a randomized trial, with respect to survival probability at a specific time point. However, like the method proposed by Loeys and Goetghebeur (2003), Nie et al. (2011) requires that subjects randomized to the control group have no access to treatment. More recently, Richardson et al. (2016) considered competing risks data and proposed nonparametric estimators, decomposing the overall causal effect of treatment on survival probability at a fixed time point into the sum of causal effects on cause-specific cumulative incidence functions. Like the other authors mentioned above, these authors assume independent censoring.

A major drawback of all the aforementioned methods and a few others (eg. Elashoff et al. (2012)), is that they do not take covariates into consideration and thus do not allow for IV-outcome confounding. As such, these methods are not suited for causal inference from observational data. Among the methods that do permit inclusion of covariates, Mark and Robins (1993) considered an accelerated failure time model for the outcome, Cuzick et al. (2007) considered a proportional hazards model and Gong (2008) considered parametric survival models. More recently, Tchetgen Tchetgen et al. (2015) proposed a control function approach to estimate the difference in hazards at a specific time under an additive hazards model for the outcome. This approach shares some similarity with approaches developed for continuous instrumental variables in the censored time-to-event data setting which assume an additive hazards model for the outcome (e.g., Li et al. (2015)). Yu et al. (2016) extended the work of Cuzick et al. (2007) to the class of semiparametric linear transformation models which include the proportional hazards model and proportional odds model as special cases.

In this chapter, we develop an estimator for the treatment effect on the restricted mean survival time (RMST) under unmeasured confounding of the treatment-outcome association using a binary IV analysis. Being a cumulative treatment effect, the effect of treatment on RMST may be of greater interest than the effect on survival or hazard at a specific time point, especially in the presence of a treatment effect that changes over time (Schaubel and Wei (2011)). To the best of our knowledge, only one other author has studied IV analysis of RMST. Kjaersgaard and Parner (2016) proposed a pseudo-outcome approach to determine treatment effects on RMST in a setting with a continuous IV.

Our motivating example inovolves the comparison between peritoneal dialysis (PD) and haemodialysis (HD), the two most frequently used dialysis modalities, with respect to 5-year RMST among end stage renal disease (ESRD) patients. While

kidney transplantation remains is preferred treatment for patients with ESRD, most patients are placed on dialysis until a transplant is available or as their only therapy. While many studies have compared the two modalities currently in use, results have been conflicting with some studies have showing PD to be associated with a survival advantage initially but no significant difference afterward (Fenton et al. (1997), Jaar et al. (2005), Heaf et al. (2002), Kumar et al. (2014)) and others showing that mortality rate is higher in patients receiving PD, especially older patients, than those receiving HD (Kim et al. (2014), Weinhandl et al. (2010)). A key concerns in most of the studies of this comparison is strong selection bias, as PD patients tend to be younger and healthier, and as studies mostly only adjust for observed confounders. This leads to the question, if unmeasured confounders were accounted for, which dialysis modality would be preferred in terms of patient survival?

In an observational setting, such as ours, the assumption that the instrument of choice is completely randomly assigned might not be valid. However, the random assignment requirement may be met after adjusting for a set of observed instrument-outcome confounders, making the instrument conditionally distributed "as good as random". These measured instrument-outcome confounders can be adjusted for by including them in two stage regression models or through matching. Two stage regression modeling in the survival setting often requires additional modeling assumptions (Li et al. (2015), Tchetgen Tchetgen et al. (2015)). Matching, on the other hand, may be infeasible in the presence of even a moderate number of covariates, as in our setting. For such situations, some authors (e.g., Frolich (2007)) have proposed matching using the instrument propensity score, i.e., the conditional probability of assignment to the instrument group encouraging treatment given covariates. An alternative to matching and regression based estimators are inverse weighting based estimators with weights based on the estimated instrument propensity score. Previously, Tan (2007) proposed an inverse probability of instrument weighted (IPIW) IV

estimator, where subjects in each instrument group are weighted by the inverse of the conditional probability of assignment to that instrument group. The weights proposed by Tan (2006) are analogous to the traditional inverse probability of treatment weight (IPTW) (Robins et al. (2000)) used for treatment comparisons in the unconfounded setting. In this chapter, we propose to use weights that, as will be shown later, tend to produce more efficient estimators than matching or IPIW proposed by Tan (2006). Further, unlike matching based estimators, we are able to derive easily implementable asymptotic variance estimators for our proposed treatment effect estimators and thus do not have to rely on resampling based methods.

In Section 3.2, we begin with a description of the notation and assumptions required for our method. We then derive the asymptotic properties of our proposed estimators; proofs of which are provided in the Appendix. In Section 3.3, we evaluate the performance of our estimators in finite samples. In Section 3.4, we apply our methods to compare HD and PD modalities using data from the United States Renal Data System (USRDS). Finally, in Section 3.5, we provide a discussion.

3.2 Methods

3.2.1 Notation

Our data consist of subjects randomized to one of two levels of a binary instrumental variable. Henceforth, we refer to these two levels of the IV as encouragement toward the treatment and encouragement towards control. For each subject $i(i = 1, ..., n)$, the binary IV is denoted by Z_i, with $Z_i = 1$ for subjects randomized to receive encouragement toward treatment and $Z_i = 0$ for subjects randomized to receive encouragement toward control. While the effect of the randomly assigned encouragement status may sometimes be of interest, the goal of an IV analysis is to estimate the causal effect of the treatment actually received.

To this end, we use Rubin's potential outcome framework to define quantities of interest. We first define a vector of potential treatment outcomes for each subject as $A_i{}^* = (A_i(0), A_i(1))$, where $A_i(0)$ and $A_i(1)$ denote the treatment that subject i would have received had they been randomized to receive $Z_i = 0$ and $Z_i = 1$, respectively. Then, under the so-called consistency assumption (Rubin, 2005), the observed treatment $A_i = A_i(0)I(Z_i = 0) + A_i(1)I(Z_i = 1)$. Based on the vector of potential treatment outcomes, subjects can be grouped into four complier classes: subjects with $A_i{}^* = (0, 1)$ are *compliers* (i.e., they receive treatment only if encouraged toward treatment); subjects with $A_i{}^* = (1, 1)$ are *always takers* (i.e., they always receive treatment); subjects with $A_i{}^* = (0, 0)$ are *never takers* (i.e., they never receive treatment); and subjects with $A_i{}^* = (1, 0)$ are *defiers* (i.e., they receive treatment only if encouraged toward control and vice-versa). We further define $T_i(z, a_i(z))$, the potential time-to-event that would be observed if subject i is randomized to $Z = z$ and actually receives treatment $a_i(z)$, for all combinations of $(z, a_i(z))$. In the absence of censoring, an IV analysis estimates the causal effect of the treatment received on subjects in the complier class, i.e., an IV analysis estimates $\delta = E[T_i(1, 1) - T_i(0, 0)|A_i{}^* = (0, 1)]$. This is commonly referred to as the local average treatment effect (LATE) or the complier average causal effect (CACE).

Let T_i denote the time-to-event, which is subject to right censoring by C_i. We let $\boldsymbol{X}_i$ be a vector of observed time-independent covariates that, in the absence of adjustment, could potentially confound the instrument-outcome relationship. We let $\Delta_i = I(T_i \leq C_i)$ be the observed event indicator. The observed data for each subject i, then consists of $\{\tilde{T}_i, \Delta_i, \boldsymbol{X}_i, A_i\}$, where $\tilde{T}_i = \min(T_i, C_i)$ represents observation time. The presence of censoring generally implies that the mean survival time is not identifiable. As such, we propose to measure the causal effect in terms of the RMST. The RMST at a given time L is defined as the $E(min(T, L))$, where it is required that $L \leq \tau$ with τ denoting the maximum observation time. Note that RMST is also

equal to the area under the survival curve in the interval $[0, L]$.The causal effect of interest then becomes, $\delta(L) = E[min(T_i(1,1), L) - min(T_i(0,0), L)|A_i^* = (0,1)]$.

3.2.2 Assumptions

To estimate the CACE, we make the following six assumptions:

A1. Stable unit treatment value assumption (SUTVA).

Also called the no interference assumption, SUTVA states that each subject's potential outcomes are not affected by the randomly assigned encouragement status of other subjects in the population. This allows us to consider each subject's potential outcomes as a function of only their encouragement status and treatment.

A2. Independence of instrument. $Z_i \perp A_i(0), A_i(1), T_i(0,0), T_i(0,1), T_i(1,0), T_i(1,1)|\boldsymbol{X}_i$.

This assumption states that the potential outcomes are independent of the randomly assigned encouragement status, conditional on the observed covariates. Essentially, we are assuming that conditioning on the observed covariate vector, the IV is independent of unmeasured confounders.

A3. Exclusion Restriction. $T_i(0,1) = T_i(1,1)$, $T_i(1,0) = T_i(0,0)|\boldsymbol{X}_i$.

This assumption states that the IV can affect the outcome only by affecting the treatment received.

A4. Non-zero causal effect of Z on A: $E(A_i(1) - A_i(0)) > 0$.

The IV, encouragement status, is assumed to have a positive effect on the actual treatment received.

A5. Monotonicity: $A_i(1) \geq A_i(0)$.

This asumption says that the potential treatment received under encouragement toward treatment is greater than or equal to that under encouragement toward control. This rules out the existence of the *defiers* class.

A6. Independent Censoring. $C_i \perp T_i | Z_i, \boldsymbol{X}_i$.

Censoring is assumed to be independent of time-to-event given the observed covariates and encouragement status.

Under assumptions *A1-A5* and in the absence of censoring, the CACE can be recovered from observed quantities as:

$$\delta(L) = \frac{E[min(T_i, L)|Z_i = 1] - E[min(T_i, L)|Z_i = 0]}{E[A_i|Z_i = 1] - E[A_i|Z_i = 0]} \tag{3.1}$$

However, in the presence of censoring, assumption A6 permits the construction of an inverse probability of censoring weighted estimator of the quantity in the numerator of equation (3.1).

3.2.3 Weighting

To account for measured confounders of the instrument-outcome relationship, we re-weight the data using weights based on the instrument propensity score, defined as,

$$e(\boldsymbol{X}) = Pr(Z = 1|\boldsymbol{X}). \tag{3.2}$$

Owing to the balancing property of the propensity score, conditioning on $e(\boldsymbol{X})$ retains independence of the IV and unmeasured confounders, and using weights based on $e(\boldsymbol{X})$ sufficiently adjusts for confounding due to $\boldsymbol{X}$. Furthermore, assumptions A2, A3 and A6 which condition on $\boldsymbol{X}$, can also be written by conditioning on $e(\boldsymbol{X})$ instead.

We propose using two different weights, namely, a matching weight intially studied by Li et al. (2013), and an overlap weight initially developed by Li et al. (2016). Both of these weights are based on the instrument propensity score. The matching weight (MW) was developed as a weighting analogue to paired matching on the propensity

score and is expressed as,

$$MW = \frac{min\{e(\boldsymbol{X}), (1 - e(\boldsymbol{X}))\}}{Ze(\boldsymbol{X}) + (1 - Z)(1 - e(\boldsymbol{X}))}. \tag{3.3}$$

Li et al. (2013) show that, in the unconfounded setting, the estimator of treatment effect obtained using the MW is asymptotically equivalent to the estimator obtained from one-to-one paired matching on the propensity score. Essentially, using the MW then provides a method to make treatment comparisons using all the subjects in the data and thus provides a more efficient alternative to matching where unmatched subjects are discarded. Further, unlike matching, the weighting approach leads to more accurate variance calculation and simpler asymptotic analysis. These appealing attributes of the MW motivate us to investigate its use for the setting with unmeasured confounders. Thus, we develop a MW estimator of the CACE that can be viewed as a more efficient alternative to the instrument propensity score matching based estimator.

The overlap weights (OW) were proposed by Li et al. (2016) for the comparison of binary treatments in the unconfounded setting. The OW are target to estimate the treatment effect in a sub-population with the most overlap, that is, a sub-population of subjects who could appear with substantial probability in either treatment group. The overlap weights belong to a broader class of covariate balancing weights in the binary treatment setting which includes MW and IPIW. As Li et al. (2016) show, among this class, the overlap weights based estimator has the minimum asymptotic variance. Further, as the authors note, the overlap weights are in a sense asymptotically equivalent to matching. Specifically, a matching analysis based on exact matching of subjects on the same discrete design points (or small neighborhood, for continuous variables) would use weights equivalent to the overlap weights estimated from a saturated propensity score model with an indicator for each design point.

Here, we propose using the OW to estimate the CACE, with the weights expressed as,

$$OW = \frac{e(\boldsymbol{X})(1 - e(\boldsymbol{X}))}{Ze(\boldsymbol{X}) + (1 - Z)(1 - e(\boldsymbol{X}))}. \tag{3.4}$$

We compare the proposed weights to the inverse instrument probability weights (IPIW) proposed by Tan (2006), and expressed as,

$$IPIW = \frac{1}{Ze(\boldsymbol{X}) + (1 - Z)(1 - e(\boldsymbol{X}))} \tag{3.5}$$

As mentioned earlier, the IPIW weights seek to achieve covariate balance across instrument groups by re-weighting subjects in each group by the probability of being assigned to the observed instrument group. The IPIW weights may become very large when the propensity score approaches 0 or 1, leading to biased and highly inefficient estimates of treatment effect. As will be seen in our simulation studies, the weights we propose are able to guard against this as they are bounded between 0 and 1.

3.2.4 Estimation

In this section we develop an estimator, $\hat{\delta}(L)$, of the CACE on the RMST in presence of measured time-independent confounders of the instrument-outcome and instrument-treatment relationship, $\boldsymbol{X}_i$. We first concentrate on estimating the quantities in the numerator of (3.1).

Let $N_i(t) = \Delta_i I(T_i \leq t)$ be the observed event counting process indicator and $Y_i(t) = \Delta_i I(T_i \geq t)$ be the observed at-risk process indicator. Further, define IV level-specific versions $dN_{iz}(t) = I(Z_i = z)dN_i(t)$ and $Y_{iz}(t) = I(Z_i = z)Y_i(t)$, where $z \in \{0, 1\}$. Then, in the absence of censoring, an estimate of the cumulative hazard $\Lambda_z(t)$ when randomly assigned to IV level $Z = z$ can be obtained using the weighted

Nelson-Aalen estimator,

$$\hat{\Lambda}_z(t) = \sum_{i=1}^{n} \int_0^t \frac{w_i^e(\hat{\boldsymbol{\beta}}) dN_{iz}(s)}{\sum_{i=1}^n w_i^e(\hat{\boldsymbol{\beta}}) Y_{iz}(s)}, \tag{3.6}$$

where $w_i^e(\hat{\boldsymbol{\beta}})$ are weights based on a model-based estimate of the IV propensity score, $e(\boldsymbol{X}_i)$. We assume that the IV assignment given $\boldsymbol{X}_i$ follows a logistic regression model with parameters $\boldsymbol{\beta}$, such that:

$$Pr(Z_i = 1|\boldsymbol{X}_i) = logit^{-1}(\boldsymbol{X}_i^T \boldsymbol{\beta}).$$

Here, we do not distinguish between the type of weight as the expression for the estimator is the same regardless of the type of weight, that is, $w_i^e(\hat{\boldsymbol{\beta}})$ could be equal to the MW, OW or IPIW.

To adjust for covariate dependent censoring, we weight each subject's contribution at time t by the inverse of the probability of being uncensored at time t, i.e., each subject's contribution in (3.3) is multiplied by $w_i^c(t) = P(C_i \geq t|\boldsymbol{X}_i)^{-1}$. As the true censoring distribution is unknown in most cases, an estimate $\hat{w}_i^c(t)$ can be obtained non-parametrically or by fitting separate Cox proportional hazards models to the censoring distribution at each level of the IV. For example, if the censoring time for subjects assigned to IV level $Z = z$ is modeled using the following proportional hazards model $\lambda_{iz}^c(t) = \lambda_{0z}(t) \exp(\boldsymbol{X}_i^T \boldsymbol{\theta}_{0z})$, then, an estimate for $w_i^c(t)$ when randomly assigned to IV level $Z = z$ is given by $\hat{w}_i^c(t; \hat{\theta}_z) = \exp(\hat{\Lambda}_{0z}(t; \hat{\boldsymbol{\theta}}_z) \exp(\boldsymbol{X}_i^T \hat{\boldsymbol{\theta}}_z))$. Thus, in the presence of covariate-dependent censoring, the weighted Nelson-Aalen estimator of the cumulative hazard $\Lambda_z(t)$ is given by

$$\hat{\Lambda}_z(t) = \sum_{i=1}^{n} \int_0^t \frac{\hat{w}_i^c(s; \hat{\boldsymbol{\theta}}_z) w_i^e(\hat{\boldsymbol{\beta}}) dN_{iz}(s)}{\sum_{i=1}^n \hat{w}_i^c(s; \hat{\boldsymbol{\theta}}_z) w_i^e(\hat{\boldsymbol{\beta}}) Y_{iz}(s)}. \tag{3.7}$$

Let $\mu_{T,z}(L) = E[min(T_i, L)|Z_i = z]$ denote the average RMST at IV level $Z = z$. An estimator $\hat{\mu}_{T,z}$ is obtained as:

$$\hat{\mu}_{T,z}(L) = \int_0^L \hat{S}_z(t)dt, \tag{3.8}$$

where $\hat{S}_z(t) = \exp(-\hat{\Lambda}_z(t))$. Thus, an estimate of the numerator in equation (3.1) is given by $\hat{\mu}_{T,1}(L) - \hat{\mu}_{T,0}(L)$.

The denominator in (3.1) is estimated as the difference in the weighted average of actual treatment received between the two IV levels. Let $\mu_{A,z} = E[A_i|Z_i = z]$, then an estimate of the denominator in (3.1) is given by:

$$\hat{\mu}_{A,1} - \hat{\mu}_{A,0} = \frac{\sum_{i=1}^n w_i^e(\hat{\boldsymbol{\beta}})A_i I(Z_i = 1)}{\sum_{i=1}^n w_i^e(\hat{\boldsymbol{\beta}})I(Z_i = 1)} - \frac{\sum_{i=1}^n w_i^e(\hat{\boldsymbol{\beta}})A_i I(Z_i = 0)}{\sum_{i=1}^n w_i^e(\hat{\boldsymbol{\beta}})I(Z_i = 0)}. \tag{3.9}$$

Thus, an estimate of the complier average causal effect in the presence of covariate-dependent censoring and observed instrument-outcome confounders is given by:

$$\hat{\delta}(L) = \frac{\hat{\mu}_{T,1}(L) - \hat{\mu}_{T,0}(L)}{\hat{\mu}_{A,1} - \hat{\mu}_{A,0}}. \tag{3.10}$$

3.2.5 Asymptotic Properties

The following two theorems summarize the asymptotic behavior of our proposed estimator.

THEOREM 1: Under assumed regularity conditions (a.) to (g.) in Appendix, $\hat{\Lambda}_z(t) = \hat{\Lambda}_z(t; \hat{\boldsymbol{\beta}})$ *converges almost surely and uniformly to* $\Lambda_z(t)$ *for* $t \in [0, \tau]$, *and* $n^{1/2}\{\hat{\Lambda}_z(t) - \Lambda_z(t)\}$ *converges asymptotically to a zero mean Gaussian process with covariance func-*

tion $\sigma_z(s,t) = E\{\boldsymbol{\Phi}_{iz}(s)\boldsymbol{\Phi}_{iz}(t)\}$, *where*

$$\boldsymbol{\Phi}_{iz}(t) = \boldsymbol{\Phi}_{iz1}(t) + \boldsymbol{\Phi}_{iz2}(t) + \boldsymbol{\Phi}_{iz3}(t) + \boldsymbol{\Phi}_{iz4}(t);$$

$$\boldsymbol{\Phi}_{iz1}(t) = \{\boldsymbol{h}_z(t) + \boldsymbol{d}_z(t)\}^T \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0})$$

$$\boldsymbol{\Phi}_{iz2}(t) = \{\boldsymbol{g}_z(t) + \boldsymbol{f}_z(t)\}^T \boldsymbol{\Omega_C^{-1}}(\boldsymbol{\theta_{0z}}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i^C}(\boldsymbol{\theta_{0z}})$$

$$\boldsymbol{\Phi}_{iz3}(t) = \sum_{i=1}^{n} \int_0^t q_z(s,t) b(s; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z})^{-1} dM_i^C(s)$$

$$\boldsymbol{\Phi}_{iz4}(t) = \sum_{i=1}^{n} \int_0^t w_i^e(\boldsymbol{\beta}_0), w_i^C(s; \boldsymbol{\theta}_{0z}) b(s; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z})^{-1} dM_{iz}(s)$$

$$dM_{iz}(s) = dN_{iz}(s) - Y_{iz}(s) d\Lambda_z(s)$$

$$dM_i^C(s) = dN_i^C(s) - Y_i(s) d\Lambda_z^C(s)$$

The expressions $\boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0})$ and $\boldsymbol{\Omega_C^{-1}}(\boldsymbol{\theta_{0z}}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i^C}(\boldsymbol{\theta_{0z}})$ represent the influence functions for the logistic model (for the IV assignment) and Cox model (for censoring) respectively. Explicit expressions for these influence functions and for the other parameters above are given in the Appendix. The consistency of $\hat{\Lambda}_z(t; \hat{\boldsymbol{\beta}})$ is proved through the consistency of $\hat{\boldsymbol{\beta}}$ and application of the continuous mapping theorem and the Strong Law of Large Numbers. The proof of asymptotic normality follows from expressing $n^{1/2}\{\hat{\Lambda}_z(t) - \Lambda_z(t)\}$ as a sum of independent and identically distributed mean zero variates. Multivariate Central Limit theorem can then be used to show asymptotic normality at a fixed time point, t, while various empirical process results can be used to establish convergence to a Gaussian process. We refer to the Appendix for a detailed sketch of the proof which utilizes some of the ideas in Schaubel and Wei (2011), where, in Theorem 1, the authors establish the asymptotic properties for the doubly inverse weighted estimator of treatment group-specific cumulative hazard functions.

THEOREM 2: Under regularity conditions (a.) - (g.) in Appendix, $\hat{\delta}(t) = \frac{\hat{\mu}_{T,1}(t;\hat{\boldsymbol{\beta}}) - \hat{\mu}_{T,0}(t;\hat{\boldsymbol{\beta}})}{\hat{\mu}_{A,1}(\hat{\boldsymbol{\beta}}) - \hat{\mu}_{A,0}(\hat{\boldsymbol{\beta}})}$ *converges almost surely to* $\delta(t)$ *for* $t \in [0, \tau]$, *and* $n^{1/2}\{\hat{\delta}(t) - \delta(t)\}$ *converges asymptotically to a mean zero Gaussian process with covariance function* $\sigma^{\delta}(s,t) = E\{\boldsymbol{\xi}_i(s)\boldsymbol{\xi}_i(t)\}$, *where*

$$\boldsymbol{\xi}_i(t) = \boldsymbol{\xi}_{i1}(t) + \boldsymbol{\xi}_{i2}(t);$$

$$\boldsymbol{\xi}_{i1}(t) = \boldsymbol{Q}^T \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0})$$

$$\boldsymbol{\xi}_{i2}(t) = \{\mu_{A,1}(t) - \mu_{A,0}(t)\}^{-1} \sum_{i=1}^{n} \boldsymbol{\Sigma}_i(t)$$

A detailed sketch of the proof with explicit expressions and definitions for the quantities $\boldsymbol{Q}^T$ and $\boldsymbol{\Sigma}_i(t)$ is given in the Appendix. The covariance function is estimated by replacing the limiting values with their empirical counterparts.

3.3 Simulation Studies

Simulation studies were conducted to assess the finite sample properties of the proposed CACE estimator and the associated asymptotic standard error estimator. We also demonstrate the benefits of using an IV analysis method by comparing the proposed IV analysis methods to a 'naive' analysis. In a 'naive' analysis, it is assumed that there is no unmeasured confounding. Thus, one proceeds to estimate the average treatment effect by adjusting only for the observed covariates, as done in the unconfounded setting. In the unconfounded setting, treatment comparisons can still be made by using propensity score matching or inverse weighting based methods but rather than the instrument propensity score, the treatment propensity score, that is, the probability of receiving the actual treatment is used. For example, to obtain a matching based estimator of the average treatment effect in the 'naive' analysis,

subjects are matched across treatment groups on the treatment propensity score and the difference estimates of RMST between treated and untreated subjects is averaged across matched sets. Similarly, the inverse-weighting based estimator of the treatment effect is simply equal to the difference in inverse weighted estimates of RMST in the treatment and untreated groups, using weights based on the treatment propensity score.

For our simulations, under assumptions $A1 - A6$, data were generated for a setting with two observed instrument-outcome confounders, $\{X_1, X_2\}$ and one unmeasured confounder, X_u. For each subject, X_2 was generated from a $Bern(0.6)$ distribution and X_1 and X_u were generated from two separate univariate $N(0, 0.5)$ distributions. Given covariate values, the level of IV that each subject was randomized to was generated from the logistic model, $Pr(Z = 1) = logit^{-1}(-0.5 + 3X_1 + X_2)$. The actual treatment receipt status was then generated from the logistic model: $Pr(A = 1) = logit^{-1}(-0.5+Z+0.25X_1+0.25X_2+0.25X_u)$. Event times T were then generated from an exponential model with rate $\lambda_T = 0.01(-2-A-0.5X_1+0.5X_2-0.25X_u)$. Censoring times C were generated from an exponential model with rate $\lambda_c = 0.01(-\lambda_0-1.5X_2)$, where λ_0 was set to $\{1, 2.5\}$ to correspond to a high ($\sim 47\%$) and moderate ($\sim 25\%$) level of censoring. For each censoring scenario, the performance of the estimators was evaluated at sample sizes $n = 500, 1000, 2000$. In all scenarios, we were interested in estimating the complier average causal effect on the RMST at $L = 1825$ (i.e., 5 years, if the time scale were in days). The results discussed in this section are based on 1000 Monte Carlo simulations.

Tables 3.1 and 3.2 display a comparison of the three different weighting estimators and a propensity score matching approach in an IV analysis and a naive analysis at sample sizes $n = 500, 1000$ and 2000 in a setting with $\sim 25\%$ censoring. As expected, a naive analysis that ignored confounding seemed to produce systematically biased

Table 3.1: Simulation results: Proposed IV estimators and propensity score matching with $\approx 25\%$ censored before $L = 1825$ and $\delta(L) = 499$

	Naive Analysis		IV Analysis				
Method	Percent Bias	ESD	Percent Bias	ESD	Relative MSE	ASE	CP
n = 500							
IPTW / IPIW	-53	54	6	436	1.52	476	0.99
MW	-53	54	5	301	0.72	334	0.97
OW	-53	55	4	309	0.77	340	0.97
Matching	-53	60	1	354	1	1	1
n = 1000							
IPTW / IPIW	-55	40	2	313	1.75	336	0.99
MW	-55	40	2	207	0.77	230	0.98
OW	-55	40	2	211	0.79	234	0.98
Matching	-55	44	-5	236	1	1	1
n = 2000							
IPTW / IPIW	-55	40	-4	170	1.13	224	0.99
MW	-55	40	-2	139	0.75	158	0.98
OW	-55	40	-1	141	0.78	161	0.98
Matching	-55	44	-10	155	1	1	1

For Naive Analysis
IPTW $= \{Ae(\boldsymbol{X}) + (1-A)(1-e(\boldsymbol{X}))\}^{-1}$
MW - Matching Weight $= min\{e(\boldsymbol{X}), (1-e(\boldsymbol{X}))\}\{Ae(\boldsymbol{X})+(1-A)(1-e(\boldsymbol{X}))\}^{-1}$
OW - Overlap Weight $= e(\boldsymbol{X})(1-e(\boldsymbol{X}))\{Ae(\boldsymbol{X}) + (1-A)(1-e(\boldsymbol{X}))\}^{-1}$
For IV Analysis
IPIW $= \{Ze(\boldsymbol{X}) + (1-Z)(1-e(\boldsymbol{X}))\}^{-1}$
MW - Matching Weight $= min\{e(\boldsymbol{X}), (1-e(\boldsymbol{X}))\}\{Ze(\boldsymbol{X})+(1-Z)(1-e(\boldsymbol{X}))\}^{-1}$
OW - Overlap Weight $= e(\boldsymbol{X})(1-e(\boldsymbol{X}))\{Ze(\boldsymbol{X}) + (1-Z)(1-e(\boldsymbol{X}))\}^{-1}$

Table 3.2: Simulation results: Proposed IV estimators and propensity score matching with $\approx 25\%$ censored before $L = 1825$ and $\delta(L) = 499$

	Naive Analysis		IV Analysis				
Method	Percent Bias	ESD	Percent Bias	ESD	Relative MSE	ASE	CP
n = 500							
IPTW / IPIW	-45	71	-5	512	1.14	474	0.97
MW	-45	72	-3	411	0.73	376	0.96
OW	-44	72	-4	421	0.77	384	0.96
Matching	-42	82	11	478	1	1	1
n = 1000							
IPTW / IPIW	-49	51	-4	395	1.59	343	0.96
MW	-49	52	-3	284	0.82	269	0.95
OW	-48	52	-3	284	0.82	275	0.95
Matching	-47	58	-3	314	1	1	1
n = 2000							
IPTW / IPIW	-49	51	-3	230	1.1	238	0.97
MW	-49	52	0	193	0.77	192	0.96
OW	-48	52	0	197	0.8	195	0.95
Matching	-47	58	-6	218	1	1	1

For Naive Analysis
IPTW $= \{Ae(\boldsymbol{X}) + (1-A)(1-e(\boldsymbol{X}))\}^{-1}$
MW - Matching Weight $= min\{e(\boldsymbol{X}), (1-e(\boldsymbol{X}))\}\{Ae(\boldsymbol{X})+(1-A)(1-e(\boldsymbol{X}))\}^{-1}$
OW - Overlap Weight $= e(\boldsymbol{X})(1-e(\boldsymbol{X}))\{Ae(\boldsymbol{X}) + (1-A)(1-e(\boldsymbol{X}))\}^{-1}$
For IV Analysis
IPIW $= \{Ze(\boldsymbol{X}) + (1-Z)(1-e(\boldsymbol{X}))\}^{-1}$
MW - Matching Weight $= min\{e(\boldsymbol{X}), (1-e(\boldsymbol{X}))\}\{Ze(\boldsymbol{X})+(1-Z)(1-e(\boldsymbol{X}))\}^{-1}$
OW - Overlap Weight $= e(\boldsymbol{X})(1-e(\boldsymbol{X}))\{Ze(\boldsymbol{X}) + (1-Z)(1-e(\boldsymbol{X}))\}^{-1}$

estimates at all sample sizes. With $\sim 25\%$ of the data being censored, the naive analysis estimates produced estimates that had a relative bias of $53 - 55\%$. At a higher censoring rate of $\sim 47\%$, the relative bias for the naive analysis estimates ranged from $42 - 50\%$.

For the IV analysis, in terms of bias, the performance of the inverse weighted estimators seemed comparable to that of the propensity score based matching estimator. The proposed MW and OW estimators, however, outperformed the matching based estimator with respect to efficiency, measured in terms of mean squared error (MSE), at all sample sizes and all levels of censoring. The OW and MW based estimators had a MSE ranging from 0.7 - 0.8 times of that of the matching based estimators. The IPIW weights, on the other hand, were less efficient than matching with relative MSE ranging from 1.1 - 1.6. The asymptotic standard error estimators approximated the true standard deviation well and, correspondingly provided confidence intervals with appropriate coverage probabilities. The proposed asymptotic variance estimator provided 95% confidence intervals that covered the true parameter with probability 97% - 99% when $\sim 25\%$ of the data were censored and with probability 95% - 97% when $\sim 47\%$ of the data were censored.

3.4 Application

We applied our methods compare HD and PD modalities for end stage renal disease (ESRD) patients using data from the United States Renal Data System (USRDS). Previous studies of this question have yielded conflicting results providing no conclusive evidence for or against the use of PD. This suggests that an IV analysis might be in order to adequately address unmeasured treatment-outcome confounding and shed some valuable insight on the problem.

We conducted an IV analysis to estimate the effect of dialysis modality on the restricted mean survival at 5 years ($L = 1825$ days) since ESRD incidence. Our

study population consisted of incident dialysis patients initiating dialysis between 01/01/2009 and 12/31/2014. A potential instrument is the dialysis facility-level variation in PD usage, defined as the facility-specific proportion of patients initiating dialysis with PD. We used a dichotomized version of this instrument, with patients in facilities with PD usage above the national average of PD usage defined as being randomized to receiving the instrument of encouragement toward PD. Owing to the nature of our analysis we excluded small dialysis facilities defined as having < 10 PD patients and < 50 patients in total. After this step, our study cohort had 164,837 patients distributed across 929 dialysis facilities in the United States. To avoid introducing patient-level confounding between the instrument and unmeasured confounders, historical data from 2006-08 was used to determine PD usage. The mean PD usage rate within facility varied from 1.8% to 54.6% with a mean of 14.5% and median of 12.5%. The correlation between facility-level mean PD usage in 2006-08 and 2009-14 was 0.57, and the dichotomized PD encouragement status was significantly associated ($\beta = 0.1, p < 0.0001$) with individual PD uasge in a model adjusting for available patient-level covariates, suggesting potential for a good instrument.

Table 3.3 presents a comparison of patients initiating dialysis on PD and HD with respect to age, comorbidities and primary renal diagnosis. On average, PD patients were indeed 5 years younger and healthier in terms of having fewer comorbidities than HD patients. While these patient-level factors are observed and can be adjusted for in a regression analysis, it is plausible that other unmeasured patient-level confounders might influence both the choice of dialysis modality and survival, thus, necessitating an IV analysis. The likelihood of unmeasured confounding seems greater knowing that every available risk factor in Table 3.3 is more prevalent for HD than PD patients.

Based on historical evidence as important predictors, we included the following patient-level covariates in the logistic regression model for estimating the instru-

Table 3.3: Analysis of USRDS Data: Description of Study Cohort by Dialysis Modality

Covariate	Haemodialysis	Peritoneal Dialysis	Standardized Difference
Percent Died	53	36	-36.3
Age (Years)	63.6	58.1	-36.5
Primary Renal Diagnosis			
Diabetes	46	43	-5.9
Hypertension	28	26	-4.6
Glomerulonephritis	8	15	22.3
Other	17	15	-5.5
Comorbidities			
Alcohol Use	2	1	-11.6
ASHD	21	13	-21.3
Cancer	8	5	-11.9
CHF	33	16	-39.5
COPD	10	4	-22.4
CVA	10	6	-13.7
Diabetes	11	7	-12.1
Drug Use	1	0	-10.7
PVD	14	9	-17.4
Tobacco Use	7	6	-2.9

Table 3.4: Analysis of USRDS Data: Results from IV and Naive analysis

Method	IV Analysis		Naive Analysis	
Method	Estimate	95% Interval	Estimate	95% Interval
IPIW	87.5	(17.8 , 157.1)	118.4	(107.5 , 129.1)
MW	88.5	(18.1 , 158.8)	102.0	(88.5 , 115.5)
OW	87.5	(17.9 , 157.0)	118.2	(108.9 , 127.6)

ment propensity score: year of ESRD incidence, age at dialysis initiation, gender, race, ethnicity (Hispanic or not), primary renal diagnosis (glomerulonephritis (GN) diabetes, hypertension, and others), and binary comorbidity indicators for the presence of cancer, diabetes, athlero-sclerotic heart disease (ASHD), congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), cerebrovascular accident (CVA), peripheral vascular disease (PVD), elicit drug use, tobacco use, and alcohol consumption. The censoring distribution in each instrument-level was estimated using separately fitted Cox hazards regression models including all the aforementioned patient-level covariates. The estimated propensity score and censoring probabilities was then used to construct the weighted estimator in equation (3.7).

In Table 3.4 we present the results from the IV analysis and the corresponding 'naive' analysis for each of the proposed inverse-weighting based estimators. For the 'naive' analysis, we ignored the presence of unmeasured confounders of the treatment-outcome relationship, i.e., we estimated the causal effect as the difference in inverse weighted estimates of RMST in the treatment and control groups, using weights based on the treatment propensity score. Both the IV and naive analysis showed the use of PD to be beneficial. The IV analysis indicated that initiating dialysis using PD may lead to a gain in 5-year RMST of nearly 3 months.

3.5 Discussion

In this chapter, we develop a weighted estimator of the complier average causal effect on the restricted mean survival time. The proposed method addresses unmeasured confounding of the treatment-outcome relationship in the censored time-to-event setting. A unique feature of our approach it that it accomodates observed instrument-outcome confounding, such that one only needs to assume that the instrument is randomly assigned conditional on some observed covariates. This makes the proposed estimator particularly suited for causal inference from observational studies.

The weights we propose to use, namely the matching weight and the overlap weight tend to outperform the IPIW proposed by Tan (2006) and propensity score matching in terms of MSE even in the presence of moderate variability in instrument propensity score. This was seen to be the case even when we explored using a variance stabilized version of IPIW weights (results not shown here), with simulation results being nearly identical to that presented here. This is mainly because, unlike the IPIW, both the MW and OW are bounded between 0 and 1 and are thus less sensitive to extreme weights. Further, an advantage of using the proposed weighted estimators over matching is the availability of easily implementable asymptotic variance estimators which are derived in this chapter. Future research could concentrate on improving these estimators further by developing a doubly robust version of the MW and OW based estimators of the CACE.

To the best of our knowledge, this is the first study of a binary IV analysis of RMST. The only other study of an IV analysis method for estimating causal effects on the RMST (Kjaersgaard and Parner (2016)) was in the context of a contionuous IV and used a pseudo-observation appraoch. As Kjaersgaard and Parner (2016) point out, a major limitation of the pseudo-observation approach is that censoring is required to be independent of covariates and instrument. In comparison, the proposed method only requires the assumption that censoring and survival times are independent conditional on covariates and instrument. Future research could concentrate on developing methods that further relax this assumption by allowing censoring to depend on treatment.

CHAPTER IV

Instrumental variable estimators of exposure effects for competing risks data

4.1 Introduction

With the issue of unmeasured confounding arising in many clinical investigations based on observational data, instrumental variable (IV) analysis methods have been gaining in popularity in biomedical research. As the name suggests, these methods exploit the availability of an IV, a variable that is (a) associated with the exposure of interest, (b) has no direct effect on the outcome except through the exposure of interest and (c) is not subject to any unmeasured confounding itself. Such methods seek to overcome unmeasured confounding of the exposure-outcome association. Essentially, the availability of an IV permits the identification of exposure or treatmet effects without knowledge of the treatment selection mechanism (Imbens and Angrist (1994), Angrist et al. (1996)).

With an origination in econometrics, IV analysis methods have been widely used in empirical economic research, but have only recently made their way to biomedical studies. Consequently, while IV estimation of exposure effects is well established for continuous and binary outcomes, there has been little research into developing methods for IV analysis of censored time-to-event data. In one of the first methodological

investigations in this area, Robins and Tsiatis (1991) developed semiparametric estimators of the treatment effect under a semiparametric structural accelerated failure time model for the outcome in the randomized study setting, that is, a setting with a binary exposure and no exogenous covariates. The authors required that all censoring be administrative or fixed at a certain known time point. Loeys and Goethburger (2003) extended the approach proposed by Robins and Tsiatsis to a proportional hazards model of the treatment effect, under independent censoring. Other notable studies of IV methods for censored time-to-event data in the randomized setting, include studies by Baker (1998), Nie et al. (2011), Elashoff et al. (2012). More recently, Richardson et al. (2016) proposed non-parametric treatment effect estimators in the competing risks setting, decomposing the overall causal effect of treatment on survival probability at a fixed time point into the sum of causal effects on the cause-specific cumulative incidence functions. Other studies (e.g., Mark and Robins (1993), Cuzick et al. (2007), Frolich (2007)) have proposed methods to control for exogenous covariates in the binary exposure setting.

For a continuous exposure, and in the presence of exogenous covariates, Li et al. (2015) proposed a two stage least squares (2SLS) approach to estimate the effect of exposure assuming an additive hazards model for the outcome. Chan (2015) subsequently showed that for the additive hazards model, exposure-dependent censoring could be accommodated using a control function or two stage residual inclusion (2SRI) approach for estimation. This was further elaborated on by Tchetgen Tchetgen (2015), who also considered binary exposures. Kjaersgaard and Parner (2016) proposed a pseudo-observation approach which requires censoring to be independent of covariates and instrument. Zheng et al. (2017) extended the method proposed by Li et al. (2015) to the competing risk setting by assuming an additive subdistribution hazards model for the cause of interest. Like Li et al. (2015), the authors also proposed to use a 2SLS approach to estimate the exposure effect, and required that

censoring be independent of exposure.

In this chapter, we develop an IV analysis method to estimate the effect of an exposure of interest on the cumulative incidence functions (CIFs) in the competing risks setting. Competing risks data are encountered in biomedical studies when subjects are subject to failure from many distinct causes or events. Unlike the method proposed by Zheng et al. (2017), which also address competing risks but focus on a single primary cause of interest, the method developed in this chapter permits simultaneous inference of the exposure effect on all competing causes. Our work here is specifically motivated by the comparison between peritoneal dialysis (PD) and haemodialysis (HD), the two most frequently used dialysis modalities, with respect to cardiovascular (CVD) and non-cardiovascular (non-CVD) mortality among end stage renal disease (ESRD) patients in the 5 years following initiation of dialysis. Thus, in our case, the two different competing events are: cardiovascular death and non-cardiovascular death. While many studies have compared the two modalities currently in use with respect to overall survival, very few studies have examined the difference in cause-specific mortality rates across modalities. Some registry based studies seem to suggest that PD use may be associated with an increased risk in myocardial infarction (Johnson et al. (2009), Kim et al. (2015)), contributing to a higher rate of CVD death in PD patients. However, as PD patients tend to be younger and healthier, and as studies mostly only adjust for observed confounders, selection bias is a concern for most of the observational studies of comparison between PD and HD patients. Thus, here, we seek to answer the question, if unmeasured confounders were accounted for, which dialysis modality is more successful in terms of CVD and non-CVD mortality?

Competing risks data are often analyzed through modeling of cause-specific hazards or the instantaneous risk of an event (Prentice and Kalbfleisch, 1978). Let T denote the time to first event and $\epsilon \in \{1, ..., K\}$ denote the cause or type of failure,

with K being the number of distinct causes. Then, the cause-specific hazard for an event $\epsilon = k$ at time t is $\lambda_k(t) = \lim_{\Delta t \to 0} Pr(t \leq T < T + \Delta t, \epsilon = k)/(\Delta t)$. More recent wrk has focused on modeling another identifiable quantity from competing risks data, the subdistribution or cumulative incidence function for the kth cause-specific event: $F_k(t) = Pr(T \leq t, \epsilon = k)$. The cumulative incidence function measures the absolute event-specific risk and presents a scientifically relevant alternative to the cause-specific hazard function. Many techniques have been proposed for direct modeling of covariate effects on the subdistributuion function. For example, Fine and Gray (1999) proposed a Cox proportional hazards model for the subdistribution hazards, Sun et al. (2006) explored an additive hazards model, Andersen et al. (2003) and Klein and Anderson (2005) proposed a pseudo-value based approach, Jeong and Fine (2007) proposed parametric regression of the cumulative incidence functions, and recently, Scheike et al. (2008) proposed direct binomial regression of the cumulative incidence function through a flexible semiparametric model where some covariates have time-varying effects and others have constant effects.

The aforementioned methods focus on estimating covariate effects on the cumulative incidence function of a primary cause of interest only, allowing an arbitrary structure for other causes. Investigators may often be interested in a simultaneous analysis of the different cause-specific events. In such cases, while using any of the aforementioned methods, the model for each cause has to be fitted separately. However, fitting the models repeatedly, for each cause, may not be appropriate as these models may not hold simultaneously. Recently, Mao and Lin (2016) developed a method based on semi-parametric transformation models that permits joint inference on all competing events. However, none of the methods mentioned above explicitly address the additivity constraint for the cumulative incidence functions, $\sum_k Pr(\epsilon = k) = 1$, that is, the constraint that a subject must eventually fail from one and only one of the many distinct causes.

The additivity constraint can be explicitly incorporated using a mixture regression modeling framework. Notable studies of the mixture regression modeling framework include Larson and Dinse (1985), Maller and Zhou (2002), Lu and Peng (2008) and Choi and Huang (2014). In this approach, the model for cumulative incidence functions is decomposed into a model for failure time conditional on the cause of failure, $Pr(T \leq t|\epsilon = k)$, and a model for the marginal probability of the cause of failure, $Pr(\epsilon = k)$. In this chapter, we develop IV-estimators for competing risks data under a semiparametric mixture component model on the cumulative incidence functions. In essence, we develop a method to identify exposure effects on the cumulative incidence functions of all causes simultaneously in the presence of unmeasured confounding. Our method is able to accommodate exposure dependent censoring and control for observed instrument outcome covariates.

In Section 4.2, we begin with a description of the semiparametric mixture regression model and lay out the assumptions required for our method. In Section 4.3, we describe the estimation procedure and the resampling method used to obtain asymptotic variance estimators. In Section 4.4, we evaluate the performance of our estimators in finite samples.

4.2 Setup and Model

Consider a setting where subjects can fail from one and only one of K causes with $\epsilon \in \{1, ..., K\}$ denoting the cause of failure. Let T and C denote the failure time and the censoring time, respectively, with $\tilde{T} = \min(T, C)$ denoting the observed event time. Further, let $\Delta = I(T \leq C)$ be the observation indicator, so that $\tilde{\epsilon} = \Delta\epsilon$ is the observed event indicator. We are interested in estimating the effect of an exposure X_e on the cumulative incidence functions for each competing risk. However, we assume that this effect is subject to confounding by an unobserved covariate X_u which is correlated with X_e. We also allow for the presence of some observed subject-specific

time-independent covariates $\boldsymbol{X}_o$.

As noted earlier, we seek to develop a method that permits simultaneous inference on all cause-specific cumulative incidence functions. To this end, using the mixture model representation for competing risks we decompose the model for cause-specific cumulative incidence functions into a model for the cause of failure, ϵ, given $(X_e, X_u, \boldsymbol{X}_o)$, and a model for conditional cumulative incidence function $Pr(T \leq t|\epsilon = k, X_e, X_u, \boldsymbol{X}_o)$. By doing so, we ensure that the additivity constraint for cumulative incidence functions is met.

For ease of representation and without loss of generality, we write down our models and describe our estimation procedure for $K = 2$, such that $\epsilon \in \{1, 2\}$. We assume the following probit regression model for the marginal distribution of ϵ given $(X_e, X_u, \boldsymbol{X}_o)$:

$$I\{\epsilon = 1|X_e, X_u, \boldsymbol{X}_o\} = I\{\gamma_0 + \boldsymbol{\gamma}_o^T \boldsymbol{X}_o + \gamma_e X_e + \xi(X_u) > 0\}, \tag{4.1}$$

where $\xi(X_u)$ is a mean-zero residual error dependent on the unobserved confounder X_u but independent of $\boldsymbol{X}_o$. We further assume the following additive hazards model for conditional cumulative incidence functions:

$$Pr(T \leq t|\epsilon = k, X_e, X_u, \boldsymbol{X}_o) = 1-\exp\{-\eta_k(t)-\boldsymbol{\beta}_{ko}^T \boldsymbol{X}_o t+\beta_{ke} X_e t+b_k(X_u)t\}, k \in \{1, 2\}, \tag{4.2}$$

where $b_k(X_u)$ is dependent on X_u. This leads us to the following model for the cumulative incidence function of cause 1:

$$F_1(t|X_e, X_u, \boldsymbol{X}_o) =$$
$$E[I\{\gamma_0+\boldsymbol{\gamma}_o^T \boldsymbol{X}_o+\gamma_e X_e+\xi(X_u) > 0\}]\{1-\exp\{-\eta_1(t)-\boldsymbol{\beta}_{1o}^T \boldsymbol{X}_o t+\beta_{1e} X_e t+b_1(X_u)t\}\}, \tag{4.3}$$

and the following model for the cumulative incidence function of cause 2:

$$F_2(t|X_e, X_u, \boldsymbol{X}_o) =$$
$$E[1-I\{\gamma_0+\boldsymbol{\gamma}_o^T\boldsymbol{X}_o+\gamma_e X_e+\xi(X_u)>0\}]\{1-\exp\{-\eta_2(t)-\boldsymbol{\beta}_{2o}^T\boldsymbol{X}_o t+\beta_{2e}X_e t+b_2(X_u)t\}\}. \tag{4.4}$$

As X_u is not observed and as a confounder correlated with X_e, naively fitting models (4.3) and (4.4) using observed data will lead to inconsistent estimates of the effect of the endogenous variable on the cause-specific cumulative incidence. To this end, we propose an instrumental variable approach that enables the consistent estimation of $\gamma_e, \beta_{1e}, \beta_{2e}$. For this, we assume the availability of an instrumental variable X_I, such that:

$$X_e = \alpha_c + \boldsymbol{\alpha}_o^T\boldsymbol{X}_o + \alpha_I X_I + \nu, \tag{4.5}$$

where $\alpha_c, \boldsymbol{\alpha}_o^T$ and α_I are coefficients for a constant intercept term, the observed or exogenous covariates and the instrumental variable, respectively in a linear model generating the endogenous variable X_e. The above model also contains a residual error term, ν, that depends on the unobserved confounder X_u, so that $Cov(X_u, \nu|X_I, \boldsymbol{X}_o) \neq 0$. This induces confounding of the exposure-outcome relationship.

Given model (4.5), consistent estimation of $\gamma_e, \beta_{1e}, \beta_{2e}$ is possible using the control function approach assuming the following standard IV conditions:

A1. ν is a mean zero normal random variable with $E(\nu|X_I, \boldsymbol{X}_o) = 0$, that is, ν is uncorrelated with $X_I, \boldsymbol{X}_o$.

A2. $\alpha_I \neq 0$ in model (4.5)

A3. X_I is uncorrelated with (X_u, T, ϵ) given X_e

A4. $\boldsymbol{X}_o$ is uncorrelated with X_u

A5. $\xi(X_u)$, $b_1(X_u)$, $b_2(X_u)$ are mean zero normal random variables such that:

(a) $\xi(X_u) = \gamma_\nu \nu + e_\xi$, where $e_\xi \sim N(0, \sigma^2_\xi)$

(b) $b_1(X_u) = \beta_{1\nu}\nu + e_{b_1}$, where $e_{b_1} \sim N(0, \sigma^2_{b_1})$

(c) $b_2(X_u) = \beta_{2\nu}\nu + e_{b_2}$, where $e_{b_2} \sim N(0, \sigma^2_{b_2})$,

where e_ξ, e_{b_1} and e_{b_2} are independent and normally distributed residual errors that are uncorrelated with ν and $X_I, \boldsymbol{X}_o$.

A6. $C \perp T | X_e, X_I \boldsymbol{X}_o$.

Assumption *A1* states that the residual error ν in model (4.5) is uncorrelated with the observed covariates and the instrumental variable implying that there is no unobserved confounding of the effect of X_I and X_e. Further, *A1* implies that model (4.5) can be fitted using ordinary least squares regression using $X_I, \boldsymbol{X}_o$ as covariates. Assumption *A2* and *A3* are standard IV conditions that specify that an instrumental variable must have a non-zero effect on the exposure and must only affect the outcome through the exposure variable. Assumption *A4* emphasizes that $\boldsymbol{X}_o$ are exogenous variables independent of X_u. Assumption *A5* relates the residual terms in models (4.2), (4.3) and (4.4) to the residual term in model (4.5). This relation allows the use of a control function or two-stage residual inclusion approach to consistently estimate the parameters in these models. To see this, consider the model for the CIF of cause 1 given $(X_e, \boldsymbol{X}_o, X_u)$. Taking expectation over the distribution of observed variables, i.e., with respect to $(X_u | X_e, X_I, \boldsymbol{X}_o)$, we have:

$$
\begin{aligned}
F_1(t|X_e, X_I, \boldsymbol{X}_o) &= E\{F_1(t|X_e, X_u, \boldsymbol{X}_o)|X_e, X_I, \boldsymbol{X}_o\} \\
&= E\{I\{\gamma_0 + \boldsymbol{\gamma}_o^T \boldsymbol{X}_o + \gamma_e X_e + \xi(X_u) > 0\} \\
&\quad \{1 - \exp[-\eta_1(t) - \boldsymbol{\beta}_{1o}^T \boldsymbol{X}_o t + \beta_{1e} X_e t + b_1(X_u)t]\}|X_e, X_I, \boldsymbol{X}_o\} \\
&= E\{I\{\gamma_0 + \boldsymbol{\gamma}_o^T \boldsymbol{X}_o + \gamma_e X_e + \gamma_\nu \nu + e_\xi > 0\}
\end{aligned}
$$

$$
\begin{aligned}
&\{1-\exp[-\eta_1(t)-\boldsymbol{\beta}_{1o}^T\boldsymbol{X}_o t+\beta_{1e}X_e t+\beta_{1\nu}\nu t+e_{b_1}t]\}|X_e,X_I,\boldsymbol{X}_o\}\\
&=\left\{\int_{-\infty}^{\infty} I\{\gamma_0+\boldsymbol{\gamma}_o^T\boldsymbol{X}_o+\gamma_e X_e+\gamma_\nu\nu+e_\xi>0\}dF(e_\xi)\right\}\\
&\left\{\int_{-\infty}^{\infty}\{1-\exp[-\eta_1(t)-\boldsymbol{\beta}_{1o}^T\boldsymbol{X}_o t+\beta_{1e}X_e t+\beta_{1\nu}\nu t+e_{b_1}t]\}dF(e_{b_1})\right\}\\
&=\Phi\{\gamma_0^*+\boldsymbol{\gamma}_o^{*T}\boldsymbol{X}_o+\gamma_e^* X_e+\gamma_\nu^*\nu\}\\
&\{1-\exp[-H_1(t)-\boldsymbol{\beta}_{1o}^T\boldsymbol{X}_o t+\beta_{1e}X_e t+\beta_{1\nu}\nu t]\}
\end{aligned}
$$

The second and third equality above follow from assumption *A5*. In the last equality, we write $H_1(t)=\eta_1(t)+log(E_{\sigma_{b_1}^2}[\exp(-e_{b_1}t)]$ and $\gamma^*=\gamma/\sqrt{1-\rho_{\xi,\nu}^2}$, where $\rho_{\xi,\nu}$ is the correlation between $\xi(X_u)$ and ν. More generally, we can write the model for cause k CIF given (X_e,X_I,X_o) as:

$$
\begin{aligned}
&F_1(t|X_e,X_I,\boldsymbol{X}_o)=\\
&\Phi\{\gamma_0^*+\boldsymbol{\gamma}_o^{*T}\boldsymbol{X}_o+\gamma_e^* X_e+\gamma_\nu^*\nu\}\{1-\exp[-H_k(t)-\boldsymbol{\beta}_{ko}^T\boldsymbol{X}_o t+\beta_{ke}X_e t+\beta_{k\nu}\nu t]\} \quad (4.6)
\end{aligned}
$$

Intuitively, the residual ν can be thought of as capturing any variation in the distribution of (T,ϵ) that is attributable to unobserved correlates of X_e. As these unobserved correlates must include all confounders of the relationship between (T,ϵ) and X_e, ν can be used as a proxy for the unobserved confounders. For this reason, in the above parametrization of the cause-specific CIFs, the terms $\gamma_\nu^*\nu$ and $\beta_{k\nu}\nu$ are referred to as control functions, similar to the control functions used in IV estimation of linear and non-linear models. As ν is not observed, for estimation, we use a consistent estimate obtained from an OLS fit of model (4.5) instead, that is, we use $\hat{\nu}=X_e-\hat{\alpha}_0+\hat{\boldsymbol{\alpha}}_o\boldsymbol{X}_o-\hat{\alpha}_I X_I$ while estimating parameters in model (4.6). Thus, parameters in model (4.6) is estimated through a 'two-stage residual inclusion' (2SRI) procedure.

Note that by estimating parameters in model (4.6) we are able to recover $(\gamma_0, \boldsymbol{\gamma}_o, \gamma_e)$ up to a multiplicative constant; that is, we identify $(\gamma_0^*, \boldsymbol{\gamma}_o^*, \gamma_e^*)$. However, if $\xi(X_u)$ is assumed to be a realization of standard normal random variable, i.e., $Var(\xi(X_u)) = 1$, then the parameter $\rho_{\xi,\nu}$ is identified using an estimate of $Var(\nu)$ from the first stage OLS regression. Subsequently parameters $(\gamma_0, \boldsymbol{\gamma}_o, \gamma_e)$ are identified exactly.

The independent censoring assumption *A6* required to accommodate right censored data, is weaker than that used by Li et al. (2015), Zheng et al. (2017) and other papers which use a two-stage least squares (2SLS) approach to IV analysis of censored time-to-event data. As demonstrated by Chan (2015), the 2SRI or control function approach can accommodate censoring that depends on the endogenous variable X_e whereas 2SLS approaches require that censoring be independent of X_e.

4.3 Inference

4.3.1 Estimation

Assuming that the study contains n subjects, we note that the observed data for each subject $i, i \in \{1, ..., n\}$, consists of $\{X_{ei}, X_{Ii}, \boldsymbol{X}_{oi}, \tilde{T}_i, \tilde{\epsilon}_i\}$. In the first stage of our estimation procedure we regress X_{ei} on $(X_{Ii}, \boldsymbol{X}_{oi})$. We use OLS to obtain parameter estimates $(\hat{\alpha}_0, \hat{\alpha}_o, \hat{\alpha}_I)$ for model (4.5). Using these estimates and the observed X_{ei}, we estimate the subject-specific residuals in model (4.5) as $\hat{\nu}_i = X_{ei} - \hat{\alpha}_0 + \hat{\boldsymbol{\alpha}}_o \boldsymbol{X}_{oi} - \hat{\alpha}_I X_{Ii}$. These $\hat{\nu}_i$ are used instead of ν as predictors in model (4.6). Thus, the second stage of our estimation procedure involves estimating parameters of model (4.6) with predictors $(\boldsymbol{X}_{oi}, X_{ei}, \hat{\nu}_i)$.

To describe the estimation of model (4.6), we introduce some additional notation. For ease of presentation, we use the following vector notation for the predictors: $\boldsymbol{X}_i = \{\boldsymbol{X}_{oi}, X_{ei}, \hat{\nu}_i\}$ and $\boldsymbol{W}_i = \{1, \boldsymbol{X}_{oi}, X_{ei}, \hat{\nu}_i\}$. The corresponding parameters are denoted by $\boldsymbol{\beta} = \{\beta_0, \boldsymbol{\beta}_o, \beta_e, \beta_\nu\}$ and $\boldsymbol{\gamma}^* = \{\gamma_0^*, \boldsymbol{\gamma}_o^*, \gamma_e^*, \gamma_\nu^*\}$. Using this notation, we

estimate the following models for cause-specific CIFs:

$$F_{ik}(t|X_e, X_I, \boldsymbol{X}_o) = \Phi\{(-1)^k \boldsymbol{\gamma}^{*T} \boldsymbol{W}_i\}\{1 - \exp[-H_k(t) - \boldsymbol{\beta}_k^T \boldsymbol{X}_i t]\}. \tag{4.7}$$

Let $N_{ik} = I(T_i \leq t, \epsilon_i = k)$ be the counting process indicator for cause k and let $G_i(t) = Pr(C_i > t|\boldsymbol{X}_i)$, then observe that, under conditional independence between C_i and (T_i, ϵ_i), we have:

$$E\left\{\frac{\Delta_i N_{ik}(t)}{G_i(t)}\right\} = E\left[E\left\{\frac{\Delta_i N_{ik}(t)}{G_i(t)}|T_i, \epsilon_i\right\}\right] = E(N_{ik}(t)) = F_{ik}(t),$$

where we condition on the predictors $\{\boldsymbol{X}_i, \boldsymbol{W}_i\}$. Scheike et al. (2008) proposed to estimate parameters for a semiparametric regression model for the cumulative incidence functions by solving estimating equations based on the weighted response $\Delta_i N_{ik}(t)/G_i(t)$. As $G_i(t)$ is usually not known, it could be substituted with an estimate $\hat{G}_i(t)$ obtained from either a non-parametric Kaplan-Meier estimator of the censoring distribution or an estimator based on a regression model, like the Cox proportional hazards model, relating the censoring distribution to covariates. Let $D_{H_k(t)}(t) = \partial F_{ik}^{H_k(t),\boldsymbol{\beta}_k}(t)/\partial H_k(t)$ and $D_{\boldsymbol{\beta}_k}(t) = \partial F_{ik}^{H_k(t),\boldsymbol{\beta}_k}(t)/\partial \boldsymbol{\beta}_k(t)$. If $\Phi\{(-1)^k \boldsymbol{\gamma}^{*T} \boldsymbol{W}_i\}$ were known, the parameters $\{H_k(t), \boldsymbol{\beta}_k\}$ in (4.7) can be estimated using the following estimating functions:

$$U_{H_k(t)}(t, H_k(t), \boldsymbol{\beta}_k) = \sum_{i=1}^{n} D_{H_k}(t)\left\{\frac{\Delta_i N_{ik}(t)}{\hat{G}_i(t)} - F_{ik}(t)\right\} \tag{4.8}$$

$$U_{\boldsymbol{\beta}_k}(\tau, H_k(t), \boldsymbol{\beta}_k) = \sum_{i=1}^{n} \int_0^{\tau} D_{\boldsymbol{\beta}_k}(t)\left\{\frac{\Delta_i N_{ik}(t)}{\hat{G}_i(t)} - F_{ik}(t)\right\}. \tag{4.9}$$

However, $\Phi\{(-1)^k \boldsymbol{\gamma}^{*T} \boldsymbol{W}_i\}$ is unknown and depends on the unknown parameter vector $\boldsymbol{\gamma}^*$. To estimate $\boldsymbol{\gamma}^*$, we propose another set of estimating equations motivated by examining the conditional probability that $\epsilon_i = k$. Note that ϵ_i is not observed

for all subjects. In the case that a subject is censored, i.e., when $\tilde{\epsilon}_i$, and censoring occurs at time $\tilde{T}_i = t$, the probability of $\epsilon_i = 1$ conditional on $T \geq t$ is denoted by $g_{ik}(t; \boldsymbol{\gamma}^*, \boldsymbol{\beta}_k, H_k(t), \boldsymbol{W}_i, \boldsymbol{X}_i)$, where

$$g_{ik}(t; \boldsymbol{\gamma}^*, \boldsymbol{\beta}_k, H_k(t), \boldsymbol{W}_i, \boldsymbol{X}_i) = Pr(\epsilon = k | T_i > t, \boldsymbol{W}_i, \boldsymbol{\beta}_k, H_k(t)) \\ = \frac{\Phi\{(-1)^k \boldsymbol{\gamma}^{*T} \boldsymbol{W}_i\} \exp\{-H_k(t) - \boldsymbol{\beta}_k^T \boldsymbol{X}_i t\}}{\sum_{k=1}^{2} \Phi\{(-1)^k \boldsymbol{\gamma}^{*T} \boldsymbol{W}_i\} \exp\{-H_k(t) - \boldsymbol{\beta}_k^T \boldsymbol{X}_i t\}}.$$

It follows that,

$$Pr(\epsilon_i = k | \tilde{\epsilon}_i, \tilde{T}_i = t, \boldsymbol{W}_i, \boldsymbol{X}_i) = E[I(\tilde{\epsilon}_i = k) + I(\tilde{\epsilon}_i = 0) g_{ik}(t; \boldsymbol{\gamma}, \boldsymbol{\beta}_k, H_k(t), \boldsymbol{W}_i, \boldsymbol{X}_i)].$$

This leads us to the following estimating function for $\boldsymbol{\gamma}$:

$$U_{\boldsymbol{\gamma}^*} = \sum_{i=1}^{n} \boldsymbol{W}_i \{I(\tilde{\epsilon}_i = k) + I(\tilde{\epsilon}_i = 0) g_{ik}(t; \boldsymbol{\gamma}^*, \boldsymbol{\beta}_k, H_k(t), \boldsymbol{W}_i, \boldsymbol{X}_i) - \Phi\{(-1)^k \boldsymbol{\gamma}^{*T} \boldsymbol{W}_i\}\}. \tag{4.10}$$

Thus, the regression parameters of the marginal model can be estimated by starting with arbitrary initial values and solving the above estimating equations (4.8), (4.9) and (4.10) using an iterative algorithm until a pre-specified convergence criterion is satisfied. Upon convergence, we obtain estimators of $\boldsymbol{\gamma}^*, \boldsymbol{\beta}_k, H_k$, denoted by $\hat{\boldsymbol{\gamma}}^*, \hat{\boldsymbol{\beta}}_k, \hat{H}_k$ respectively.

As noted earlier, in model (4.7), $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}/\sqrt{1 - \rho_{\xi,\nu}}$. For a probit outcome model, under the assumption that $Var(\xi(X_u)) = 1$, we have $\gamma_\nu = \rho_{\xi,\nu}/\sigma_\nu$, where σ_ν is the standard deviation of ν. An estimate of the standard deviation of ν, $\hat{\sigma}_\nu$ can be obtained from the first stage OLS fit of model (4.5). The estimated quantities $\hat{\sigma}_\nu$ and $\hat{\gamma}_\nu^*$ can then be used to obtain an estimate of $\rho_{\xi,\nu}$ using the relation $\gamma_\nu^* = (\rho_{\xi,\nu}/\sigma_\nu)/\sqrt{1 - \rho_{\xi,\nu}}$. Thus, the parameters $\boldsymbol{\gamma} = \{\gamma_0, \boldsymbol{\gamma}_o, \gamma_e, \gamma_\nu\}$ are identified as both $\hat{\gamma}_\nu^*$ and the residual variance of the so-called selection model (4.5), σ_ν^2, are identified.

4.3.2 Variance Estimation using perturbation resampling

To estimate the variance and to construct confidence intervals for our proposed estimators, we use a perturbation-based resampling method. We apply perturbing random variables directly to the contributions of each subject. In the first stage, this translates to fitting a weighted least squares regression model to the data with subject-specific weights set equal to the value of perturbing random variables. In the second stage, the perturbing random variables are applied directly to the estimating functions at the subject-level to approximate the distribution of the estimators. Specifically, let $\{\omega_i, i = 1, ..., n\}$ be n independent realizations of a positive random variable ω from a known distribution with unit mean and unit variance. Fixing the data at their observed values, perturbed estimators $\{\tilde{\alpha}_0, \tilde{\boldsymbol{\alpha}}_o, \tilde{\alpha}_e\}$ are obtained by fitting a WLS regression with subject-sepcific weights ω_i. These parameter estimates are then used to estimate $\tilde{\nu}$, a perturbed version of $\hat{\nu}$. Using $\tilde{\nu}$ and with the remaining data fixed at their observed values, perturbed estimators of parameters in model (4.7) are obtained as the solution to the following perturbed estimating functions:

$$\tilde{U}_{H_k(t)}(t, H_k(t), \boldsymbol{\beta}_k) = \sum_{i=1}^{n} \omega_i D_{H_k}(t) \left\{ \frac{\Delta_i N_{ik}(t)}{\tilde{G}_i(t)} - F_{ik}(t) \right\} \tag{4.11}$$

$$\tilde{U}_{\boldsymbol{\beta}_k}(\tau, H_k(t), \boldsymbol{\beta}_k) = \sum_{i=1}^{n} \omega_i \int_0^{\tau} D_{\boldsymbol{\beta}_k}(t) \left\{ \frac{\Delta_i N_{ik}(t)}{\tilde{G}_i(t)} - F_{ik}(t) \right\} \tag{4.12}$$

$$\tilde{U}_{\boldsymbol{\gamma}^*} = \sum_{i=1}^{n} \omega_i \boldsymbol{W}_i \{ I(\tilde{\epsilon}_i = k) + I(\tilde{\epsilon}_i = 0) g_{ik}(t; \boldsymbol{\gamma}, \boldsymbol{\beta}_k, H_k(t), \boldsymbol{W}_i, \boldsymbol{X}_i) - \Phi\{(-1)^k \boldsymbol{\gamma}^{*T} \boldsymbol{W}_i\} \}, \tag{4.13}$$

where $\tilde{G}$ is the perturbed version of $\hat{G}$, an estimator of the censoring distribution with weights $\{\omega_i, i = 1, ..., n\}$. The above perturbed estimating equations (4.11)-(4.13) are solved using the same procedures used to solve their unperturbed counterparts, estimating equations (4.8)-(4.10), to obtain estimators $\{\tilde{\nu}, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\beta}}_k, \tilde{H}_k\}$. By repeatedly

generating $\{\omega_i, i = 1, ..., n\}$, say M times, we can obtain a large number of realizations of the perturbed estimators, say - $\{\tilde{\nu}^{(m)}, \tilde{\boldsymbol{\gamma}}^{(m)}, \tilde{\boldsymbol{\beta}}_k^{(m)}, \tilde{H}_k^{(m)}, m = 1, ..., M\}$.

It can be shown that the unconditional distribution of estimates $\{\hat{\nu}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_k, \hat{H}_k\}$ can be approximated by the conditional distribution of the perturbed estimates given the observed data (van der Vaart and Wellner, 1996). Thus, the variance and confidence intervals of estimates $\{\hat{\nu}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_k, \hat{H}_k\}$ are estimated based on the empirical distribution of $\{\tilde{\nu}, \tilde{\boldsymbol{\gamma}}^{(m)}, \tilde{\boldsymbol{\beta}}_k^{(m)}, \tilde{H}_k^{(m)}, m = 1, ..., M\}$.

4.4 Simulation Studies

Simulation studies were conducted to assess the finite sample properties of the proposed method. In the first simulation study, the performance of the proposed method was examined in a setting with a continuous instrumental variable and indpendent censoring. To do so, under assumptions $A1 - A6$, data for $K = 2$ competing risks were generated for a setting with one observed exogenous covariate, X_o, generated from a $Bern(0.5)$ distribution, and one instrumental variable X_I, generated from a $N(0.25, 1)$ distribution. Subsequently, the exposure or endogenous variable was generated from the following linear model: $X_e = 0.5X_I + \nu$, where $\nu \sim N(0, 0.5)$, represented the influence of an unmeasured confounder, say X_u, on the exposure. The residual error terms correlated with the unmeasured confounder in models (4.3) and (4.4) were then generated from the following three linear models: (i) $\xi(X_u) = \nu + e_\xi$, where $e_\xi \sim N(0, 0.1)$, (ii) $b_1(X_u) = \nu + e_{b_1}$, where $e_{b_1} \sim N(0, 0.1)$ and (iii) $b_2(X_u) = \nu + e_{b_2}$, where $e_{b_2} \sim N(0, 0.75)$. Given the generated residual error terms, the exogenous covariate and exposure values, a cause and time of failure were generated from models (4.3) and (4.4) for each subject with the parameter values set to $\{\gamma_0, \gamma_o, \gamma_e\} = \{-0.5, 0.5, 0.5\}$, $\{\beta_{1o}, \beta_{1e}\} = \{0.5, 0.5\}$, $\{\beta_{2o}, \beta_{2e}\} = \{0.5, 0.5\}$ and $\eta_1(t) = \eta_2(t) = t + 1.5t^2$. Finally, a censoring time for each subject was generated from a $Unif(0.5, 1.25)$ distribution.

Table 4.1: Simulation results comparing proposed IV method to a benchmark method and naive regression method at different sample sizes under independent censoring ($\sim 16\%$)

	Benchmark		Naive Method		Proposed IV Method			
Parameter	Bias	ESD	Bias	ESD	Bias	ESD	ASE	CP
				n = 250				
γ_e	0.018	0.223	0.583	0.182	0.016	0.222	0.282	0.990
β_{1e}	-0.038	0.726	0.408	0.546	-0.017	0.667	0.710	0.947
β_{2e}	-0.037	0.297	0.412	0.209	-0.043	0.288	0.295	0.938
				n = 500				
γ_e	0.018	0.143	0.507	0.138	0.019	0.161	0.181	0.980
β_{1e}	-0.023	0.479	0.430	0.376	-0.023	0.483	0.479	0.940
β_{2e}	-0.041	0.197	0.403	0.140	-0.041	0.202	0.199	0.944
				n = 1000				
γ_e	0.007	0.098	0.506	0.091	0.009	0.113	0.123	0.970
β_{1e}	-0.023	0.319	0.435	0.258	-0.022	0.322	0.332	0.940
β_{2e}	-0.029	0.135	0.403	0.100	-0.028	0.140	0.136	0.930

Bias, Empirical Bias of Estimates; ESD, sample standard deviation; ASE, mean of estimated standard error via resampling method; CP, empirical coverage probability of 95% interval

Simulation results at sample sizes $n = 250, 500$ and 1000 are presented in Table 4.1. The proposed method was compared to a benchmark method, wherein the true value of ν was used while estimating the parameters and a naive method wherein a model using only the observed covariates X_o and X_e was fit to the data. As shown in Table 4.1, ignoring unmeasured confounding, as done in the naive method, seemed to lead to highly biased estimates of the exposure effect. The proposed IV method, on the other hand, did almost as well as the benchmark method with respect to bias and empirical stnadard deviation. The proposed resampling-based standard errors provided a reasonable approximation of the empirical standard deviation and coverage rates around the nominal rate of 95%.

In a second simulation study, the performance of the proposed method in a setting with a binary instrumental variable and exposure-dependent censoring was examined.

Table 4.2: Simulation results comparing proposed IV method to a benchmark method and naive regression method at different sample sizes under censoring ($\sim 16\%$)

	Benchmark		Naive Method		Proposed IV Method			
Parameter	Bias	ESD	Bias	ESD	Bias	ESD	ASE	CP
				n = 250				
γ_e	0.039	0.208	0.522	0.173	0.028	0.233	0.286	0.982
β_{1e}	-0.036	0.563	0.549	0.388	-0.044	0.573	0.603	0.976
β_{2e}	0.027	0.259	0.483	0.183	0.019	0.268	0.264	0.942
				n = 500				
γ_e	0.016	0.149	0.497	0.126	0.015	0.168	0.179	0.966
β_{1e}	-0.044	0.384	0.552	0.26	-0.044	0.389	0.399	0.956
β_{2e}	0.003	0.173	0.467	0.121	0.002	0.179	0.178	0.946
				n = 1000				
γ_e	0.017	0.104	0.497	0.087	0.014	0.115	0.122	0.962
β_{1e}	-0.033	0.26	0.575	0.174	-0.035	0.261	0.272	0.964
β_{2e}	-0.003	0.126	0.459	0.085	-0.006	0.129	0.124	0.944

Bias, Empirical Bias of Estimates; ESD, sample standard deviation; ASE, mean of estimated standard error via resampling method; CP, empirical coverage probability of 95% interval

The exogenous covariate X_o and the residual error terms reflecting the influence of an unmeasured confounder on the exposure-outcome relationship were generated in the same manner as in the first simulation study. However, the instrumental variable, X_I was generated from a $Bern(0.6)$ distribution and the exposure was generated from the linear model: $X_e = X_I + \nu$. The parameter values of models (4.3) and (4.4) were fixed at $\{\gamma_0, \gamma_o, \gamma_e\} = \{-0.5, 0.5, 0.5\}$, $\{\beta_{1o}, \beta_{1e}\} = \{0.5, 0.5\}$, $\{\beta_{2o}, \beta_{2e}\} = \{0.5, 0.5\}$ and $\eta_1(t) = \eta_2(t) = 0.8t + 1t^2$. The censoring time was dependent on the value of the exposure and was generated as: $Unif(0.4, 1.2) \times I(X_e > 0) + Unif(0.8, 1.2) \times I(X_e < 0)$.

Table 4.2 displays results of the second simulation study at sample sizes $n = 250, 500$ and 1000. As was the case with independent censoring, the naive method seemed to produce highly biased estimates, while the proposed IV method seemed to do nearly as well as the method using the true value of the unmeasured confounder. The per-

turbation resampling method seemed to estimate the standard error of the parameter estimates well, with the corresponding 95% confidence intervals producing a coverage rate ranging from 94% - 98% in all sample sizes. All results discussed in this section are based on 500 Monte Carlo simulations.

4.5 Application

We applied our method to compare HD and PD modalities for end stage renal disease (ESRD) patients using data from the United States Renal Data System (USRDS). Specifically, we sought to differentiate the two modalities with respect to two competing risks: (i) death from cardiovascular diseases (CVD) and (ii) death from other causes. While very few studies have studied the association between cardiovascular mortality and dialysis modality, many previous studies have compared HD and PD modalities with respect to all cause mortality. However, these studies have yielded conflicting results providing no conclusive evidence for or against the use of PD. This suggests that an IV analysis might be in order to adequately address unmeasured treatment-outcome confounding and shed some valuable, new insight on the comparison between modalities, especially for a cause-specific comparison.

We conducted an IV analysis to examine the association between dialysis modality and the risk of death from cardiovascular diseases and other causes at 5 years since ESRD incidence. Our study population consisted of incident dialysis patients initiating dialysis between 01/01/2009 and 12/31/2009 and belonging to ESRD Network 11. This network serves dialysis patients in the midwestern states of Michigan, Minnesota, North Dakota, South Dakota and Wisconsin. The instrument in our analysis was the dialysis facility-level variation in PD usage, defined as the facility-specific proportion of patients initiating dialysis with PD. Given the nature of our analysis we excluded small dialysis facilities defined as having < 10 PD patients and < 50 patients in total. After this step, our study cohort had 2,001 patients distributed

across 59 dialysis facilities. To avoid introducing patient-level confounding between the instrument and unmeasured confounders, historical data from 2006-08 was used to determine PD usage. The mean PD usage rate within facility varied from 3.5% to 33.9% with a mean of 12.1% and median of 10.2%. The correlation between facility-level mean PD usage in 2006-08 and 2009-14 was 0.57, and the mean PD usage within facility was significantly associated ($\beta = 0.531, p < 0.0001$) with individual PD usage in a model adjusting for available patient-level covariates, suggesting potential for a good instrument.

Table 4.3 presents a comparison of patients initiating dialysis on PD and HD with respect to age, comorbidities and primary renal diagnosis. On average, PD patients were younger by about 5 years and healthier in terms of having a lower prevalence of comorbidities than HD patients. While these observed patient-level factors can be adjusted for in a regression analysis, it is plausible that other unmeasured patient-level confounders might influence both the choice of dialysis modality and survival, thus, necessitating an IV analysis. Further, given that almost every available risk factor in Table 4.3 is more prevalent for HD than PD patient, unmeasured confounding seems likely.

Based on historical evidence as important predictors, we adjusted for the following patient-level risk-factors: age at dialysis initiation, gender, race, ethnicity (Hispanic or not), primary renal diagnosis (glomerulonephritis (GN) diabetes, hypertension, and others), and binary comorbidity indicators for the presence of cancer, diabetes, athlero-sclerotic heart disease (ASHD), congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), cerebrovascular accident (CVA), peripheral vascular disease (PVD), elicit drug use, tobacco use, and alcohol consumption. The censoring distribution was estimated using a Cox hazards regression model including all the aforementioned patient-level covariates and the treatment.

In Table 4.4 we present the results from the IV analysis and the corresponding

Table 4.3: Analysis of USRDS Data: Description of Study Cohort by Dialysis Modality

Covariate	Haemodialysis	Peritoneal Dialysis	Standardized Difference
Died	69 %	50 %	-29 %
CVD death	22.4 %	21.2 %	-2.0 %
Non-CVD death	69.1 %	49.5 %	-28.7 %
Age (Years)	64.2	59.3	-0.229
Primary Renal Diagnosis			
Diabetes	41.2	42.4	1.8
Hypertension	26.8	27.8	1.5
Glomerulonephritis	8.4	14.1	12.8
Other	23.6	15.6	-14.1
Comorbidities			
Alcohol Use	2.8	2.5	-1.1
ASHD	26.1	19.7	-10.7
Cancer	11.6	9.1	-5.8
CHF	37.4	19.7	-28.2
COPD	11.8	7.1	-11.4
CVA	9.5	9.1	-1.1
Diabetes	11.9	9.1	-6.5
Drug Use	1.4	1.0	-2.4
PVD	15.6	9.1	-14.0
Tobacco Use	7.9	10.1	5.4

Table 4.4: Analysis of USRDS Data: Results from IV and Naive analysis

Parameter	IV Analysis		Naive Analysis	
Parameter	Estimate	95% Interval	Estimate	95% Interval
γ_e	1.507	(0.035 , 2.979)	0.458	(0.131 , 0.785)
β_{1e}	0.102	(-0.473 , 0.678)	-0.055	(-0.097 , -0.014)
β_{2e}	0.190	(-0.409 , 0.788)	-0.046	(-0.088, -0.005)

'naive' analysis for each of the proposed inverse-weighting based estimators. For the 'naive' analysis, we ignored the presence of unmeasured confounders of the treatment-outcome relationship and proceeded as we did for the naive analysis in the Simulation studies. The naive analysis showed the use of PD to be associated with a statistically significant increase in the marginal risk of death from CVD. This result was confirmed in the IV analysis. The cumulative incidence curves for CVD mortality and non-CVD mortality for a subject with average covariate values, from the naive and IV analysis, are presented in Figure 4.1 and Figure 4.2 respectively. Also plotted in these figures are the non-parametric estimate of the CIF obtained using the Aalen-Johansen estimator. The non-parametric estimate is not adjusted for unobserved and observed confounders. In terms of non-CVD mortality, both the IV and naive analysis indicated a significant survival benefit for PD patients at 4.5 years.

Our findings seem to support findings of Johnson et al. (2009) and Kim et al. (2015), who noted a significantly increased risk of death from CVD and CVD events for PD patients. Both these studies were registry analyses of a patient population from a country other than the United States. Further, while they adjusted for a rich set of observed covariates, neither study addressed unmeasured confounding.

4.6 Discussion

In this chapter we develop an instrumental variable analysis method for addressing unmeasured confounding in the competing risks setting. Our method can be used for

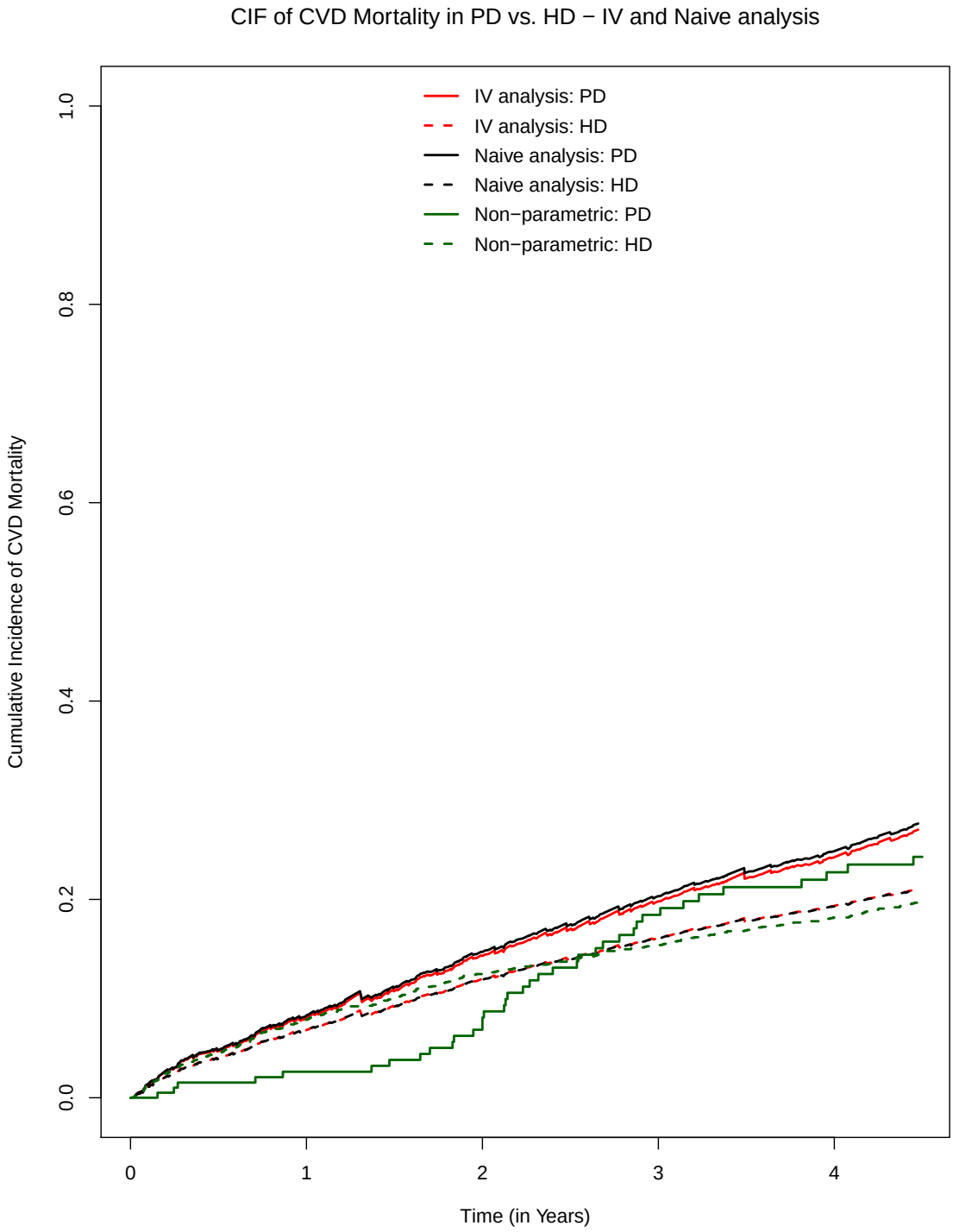

Figure 4.1: Cumulative Incidence of death from CVD by dialysis modality, in the first 4.5 years after initiation of dialysis

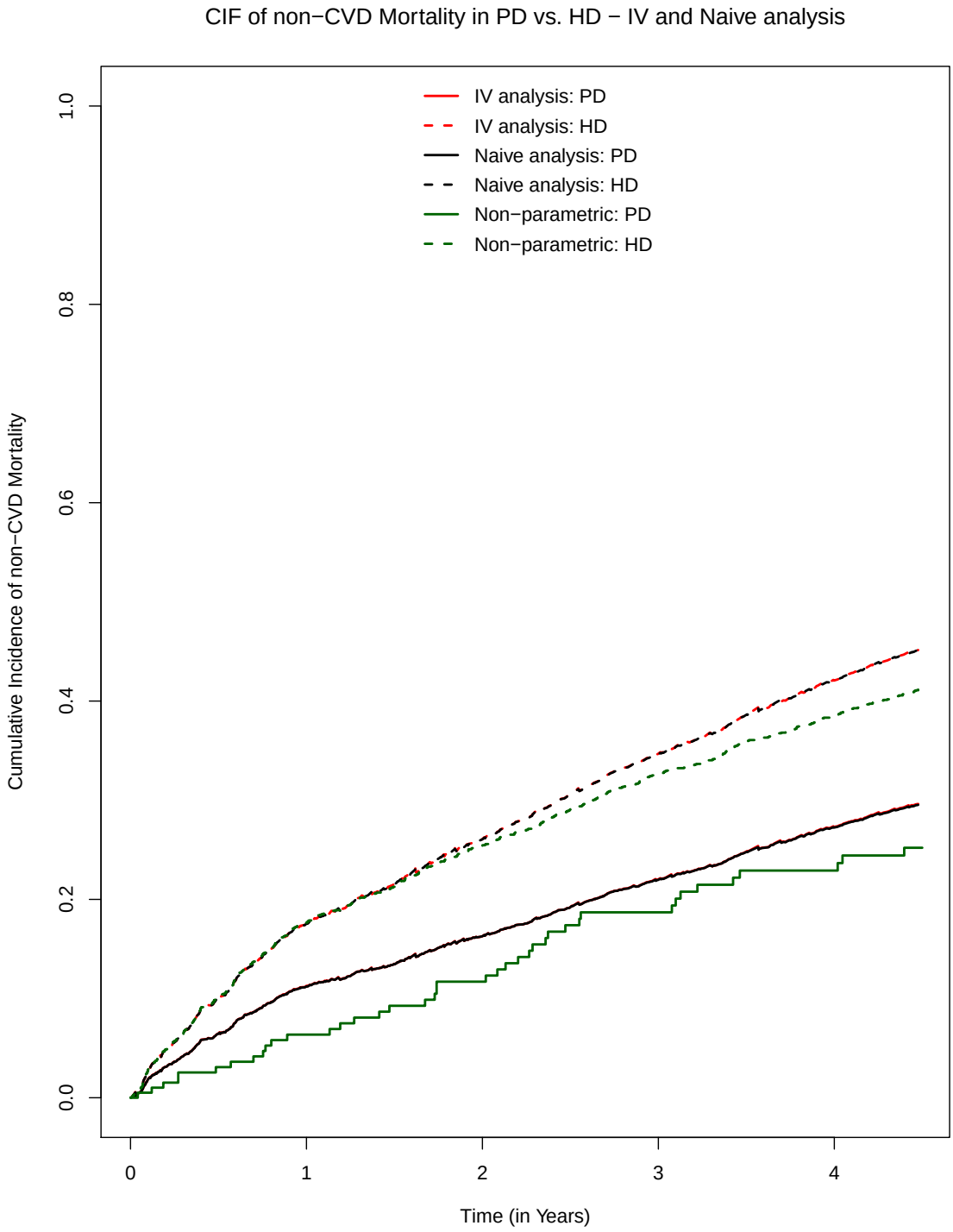

Figure 4.2: Cumulative Incidence of non-CVD death by dialysis modality, in the first 4.5 years after initiation of dialysis

binary or continuous exposures, and unlike previously developed methods for competing risks data (Zheng et al. (2017), Richardson et al. (2016)), can accommodate exposure-dependent censoring. The main strength of the developed method, however, lies in its ability to estimate the effect of an exposure of interest on the absolute risk of all causes, simultaneously. We use a semiparametric mixture component model that guarantees that $\sum_{k=1}^{K} F_k(\infty) = 1$. In this regard, our method permits an investigator to take into account all causes or event types. Prominent methodologies centered on direct modeling of the CIFs force investigators to focus on a single cause of interest. When these methodologies are extended to perform IV analysis in the competing risks setting, as done in Zheng et al. (2017), important features of the exposure-outcome relationship and the interplay between causes tend to get overlooked. Thus, the method developed in this chapter provides a unique approach to adjusting for unmeasured confounding in the competing risks setting and by simultaneously modeling exposure effects on all causes, provides a deeper understanding of the exposure-outcome relationship.

APPENDICES

APPENDIX A

Appendix for Chapter I

In this Appendix we present an expanded version of simulation results presented in Table 3 of Chapter II, a tabulation of the results mentioned in Section 2.5 and a scatter plot of the estimated center-specific random effects and ECIs of the two outcomes - transplantation and death - for the data analysis detailed in Section 2.5.

Table 1 in this document compares the proposed method to a method ignoring the correlation between cause-specific random effects within a center with respect to Bias, Empirical Standard Deviation, Asymptotic Standard Error and Coverage Probability of center effect estimates.

Table 2 in this document compares the proposed method to a method that ignores the correlation between the cause-specific center effects with respect to outlier classification.

Figure 1 contains two scatterplots. Above is a scatterplot of the center-specific random effects for cause 1 (Transplantation) and cause 2 (Death). Below is a scatterplot of the ECI for cause 1 (Transplantation) and ECI for cause 2 (Death).

Table A.1: Estimating Center-Specific Effects: Results from 500 Simulated Datasets

			Proposed Method				Ignornig Random Effects Correlation			
$n_{j'}$		True Value	Bias	ESD	ASE	CP	Bias	ESD	ASE	CP
20	$\gamma_{j'1}$	0.0	-0.022	0.234	0.324	0.985	-0.027	0.246	0.337	0.995
	$\gamma_{j'2}$	0.0	-0.013	0.243	0.306	0.975	-0.018	0.253	0.315	0.975
	$\gamma_{j'1}$	0.5	-0.175	0.241	0.298	0.965	-0.217	0.249	0.310	0.965
	$\gamma_{j'2}$	-0.5	0.172	0.249	0.328	0.975	0.227	0.250	0.337	0.940
	$\gamma_{j'1}$	1.0	-0.276	0.243	0.277	0.880	-0.326	0.249	0.285	0.825
	$\gamma_{j'2}$	-1.0	0.403	0.243	0.355	0.835	0.566	0.233	0.363	0.680
40	$\gamma_{j'1}$	0.0	-0.011	0.226	0.264	0.985	-0.016	0.237	0.273	0.990
	$\gamma_{j'2}$	0.0	-0.008	0.208	0.243	0.985	-0.012	0.213	0.250	0.980
	$\gamma_{j'1}$	0.5	-0.098	0.214	0.233	0.940	-0.129	0.222	0.240	0.925
	$\gamma_{j'2}$	-0.5	0.112	0.217	0.271	0.970	0.155	0.227	0.278	0.935
	$\gamma_{j'1}$	1.0	-0.139	0.204	0.208	0.910	-0.177	0.210	0.214	0.890
	$\gamma_{j'2}$	-1.0	0.265	0.220	0.307	0.920	0.406	0.226	0.313	0.780
60	$\gamma_{j'1}$	0.0	-0.003	0.197	0.231	0.970	-0.008	0.205	0.237	0.965
	$\gamma_{j'2}$	0.0	-0.006	0.188	0.210	0.975	-0.009	0.193	0.214	0.980
	$\gamma_{j'1}$	0.5	-0.057	0.196	0.199	0.965	-0.084	0.199	0.204	0.965
	$\gamma_{j'2}$	-0.5	0.065	0.213	0.241	0.960	0.100	0.214	0.245	0.935
	$\gamma_{j'1}$	1.0	-0.091	0.169	0.178	0.955	-0.122	0.172	0.181	0.920
	$\gamma_{j'2}$	-1.0	0.218	0.217	0.276	0.880	0.333	0.219	0.281	0.800
80	$\gamma_{j'1}$	0.0	-0.015	0.175	0.209	0.995	-0.024	0.183	0.215	0.995
	$\gamma_{j'2}$	0.0	-0.022	0.178	0.191	0.970	-0.028	0.185	0.194	0.965
	$\gamma_{j'1}$	0.5	-0.071	0.177	0.180	0.960	-0.095	0.183	0.184	0.950
	$\gamma_{j'2}$	-0.5	0.065	0.181	0.218	0.970	0.089	0.186	0.222	0.965
	$\gamma_{j'1}$	1.0	-0.107	0.160	0.161	0.895	-0.134	0.164	0.163	0.855
	$\gamma_{j'2}$	-1.0	0.195	0.198	0.255	0.920	0.283	0.202	0.259	0.810
100	$\gamma_{j'1}$	0.0	-0.026	0.164	0.195	0.990	-0.036	0.171	0.199	0.985
	$\gamma_{j'2}$	0.0	-0.018	0.154	0.176	0.980	-0.025	0.159	0.179	0.970
	$\gamma_{j'1}$	0.5	-0.081	0.147	0.167	0.965	-0.103	0.150	0.170	0.945
	$\gamma_{j'2}$	-0.5	0.058	0.200	0.203	0.925	0.077	0.207	0.206	0.915
	$\gamma_{j'1}$	1.0	-0.099	0.137	0.149	0.930	-0.124	0.141	0.151	0.885
	$\gamma_{j'2}$	-1.0	0.157	0.191	0.241	0.945	0.232	0.196	0.245	0.890

Table A.2: Analysis of Scientific Registry of Transplant Recipients (SRTR) Data: Comparing Classification of Organ Procurement Organizations (OPOs) based on Excess Cumulative Incidence (ECI) of Death and Kidney Transplantation

Classification of OPOs Ignornig Correlation of Random Effects	Classification of OPOs Using Proposed Method					
	Based on ECI of Transplant			Based on ECI of Death		
	Low Outlier	Not an Outlier	High Outlier	Low Outlier	Not an Outlier	High Outlier
Low Outlier	17	1	0	16	1	0
Not an Outlier	0	24	2	5	23	1
High Outlier	0	1	13	0	0	12

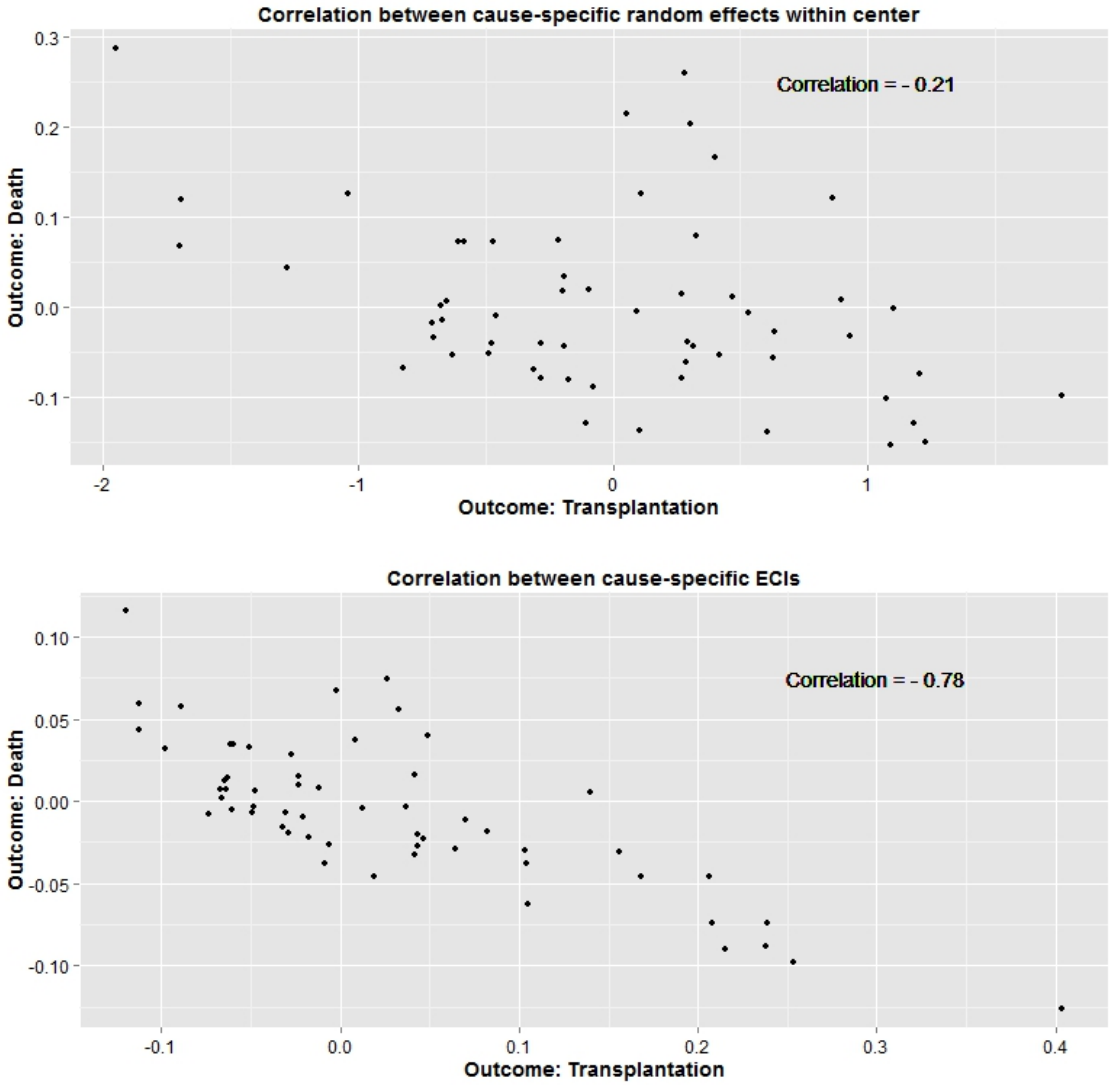

Figure A.1: Scatter Plots of center-specific random effects (above) and Excess Cause-specific Cumuluative Incidence (below) for the outcomes of Transplant and Death for 58 OPOs

APPENDIX B

Appendix for Chapter III

In this Appendix we present a sketch of the proof of THEOREM 1 and THEOREM 2 in Chapter 4.

Proof of THEOREM 1

To establish the asymptotic properties of the estimator proposed in the section 2.4, we first assume the following regularity conditions for $i = 1, ..., n$:

a. $\{\tilde{T}_i, \Delta_i, \boldsymbol{X}_i, A_i\}$are independent and identically distributed

b. $\boldsymbol{X}_i$ is bounded almost surely

c. $\Lambda_z(\tau) < \infty$, where τ is some pre-specified time-point

d. $Pr(Z_i = z|\boldsymbol{X}_i) > 0$ for $z = \{0, 1\}$

e. Positive definiteness of the matrix $\boldsymbol{\Omega(\beta_0)} = E[\boldsymbol{X}_i^{\otimes 2} p_i(\boldsymbol{\beta_0})(1 - p_i(\boldsymbol{\beta_0}))]$, where $p_i(\boldsymbol{\beta_0}) = Pr(Z_i = 1|\boldsymbol{X_i})$

With respect to censoring, we assume that the censoring time for subjects assigned to IV level $Z = z$ follows a proportional hazards model $\lambda_{iz}^c = \lambda_{0z}(t)\exp(\boldsymbol{X}_i^T\boldsymbol{\theta}_{0z})$. Then, an estimate for $w_i^c(t)$ when randomly assigned to IV level $Z = z$ is given by $\hat{w}_i^c(t;\hat{\theta}_z) = \exp(\hat{\Lambda}_{0z}(t;\hat{\boldsymbol{\theta}}_z)\exp(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}}_z))$. Here, we assume that $\hat{\Lambda}_{0z}(t;\hat{\boldsymbol{\theta}}_z)$ is the Breslow-Aalen estimator for $\Lambda_{0z}(t)$, so that:

$$\hat{\Lambda}_{0z}(t;\hat{\boldsymbol{\theta}}_z) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{dN_i^c(s)}{B_c^{(0)}(s;\hat{\theta})}, \tag{B.1}$$

where $B_c^{(d)}(s;\hat{\theta}) = n^{-1}\sum_{i=1}^n Y_i(s)X_i^{\otimes d}\exp\{(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}}_z)\}$ for $d = 0, 1, 2$ with $a^{\otimes 0} = 1, a^{\otimes 1} = a, a^{\otimes 2} = aa^T$ for a vector a. Further, $\boldsymbol{\theta}_{0z}$ is assumed to be estimated through partial likelihood by $\hat{\theta}_z$, the solution of the score equation, $U_c(\boldsymbol{\theta}) = 0$, where:

$$U_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\int_0^\tau \boldsymbol{X}_i - \bar{\boldsymbol{X}}_c(t,\boldsymbol{\theta})dN_i^c(t), \tag{B.2}$$

$$\bar{\boldsymbol{X}}_c(t,\boldsymbol{\theta}) = \frac{B_c^{(1)}(s;\theta)}{B_c^{(0)}(s;\theta)}. \tag{B.3}$$

With these assumptions on the censoring model, to accomodate covariate-dependent censoring, we assume the following additional regularity conditions:

f. Continuity of the following functions:
$b_c^{(1)}(s;\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}b_c^{(0)}(s;\boldsymbol{\theta}), b_c^{(2)}(s;\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}b_c^{(0)}(s;\boldsymbol{\theta})$,
where $b_c^{(d)}(s;\boldsymbol{\theta})$ is the limiting value of $B_c^{(d)}(s;\boldsymbol{\theta})$ for$d = 0, 1, 2$ with $b_c^{(0)}(s;\boldsymbol{\theta})$ and $b_c^{(1)}(s;\boldsymbol{\theta})$ bounded and $b_c^{(2)}(s;\boldsymbol{\theta})$ bounded away from 0 for $t \in [0,\tau]$ and $\boldsymbol{\theta}$ in an open set

g. Positive definiteness of the matrix $\boldsymbol{\Omega}_c(\boldsymbol{\theta}) = \int_0^\tau \boldsymbol{v}_c(t,\boldsymbol{\theta})b_c(t,\boldsymbol{\theta})d\Lambda_0^c(t)$, where $\boldsymbol{v}_c(t,\boldsymbol{\theta}) = b_c^{(2)}(t,\boldsymbol{\theta})/b_c^{(0)}(t,\boldsymbol{\theta}) - \bar{\boldsymbol{x}}_c(t,\boldsymbol{\theta})$ and $\bar{\boldsymbol{x}}_c(t,\boldsymbol{\theta}) = b_c^{(1)}(s;\theta)/b_c^{(0)}(s;\theta)$ is the limiting value of $\bar{\boldsymbol{X}}_c(t,\boldsymbol{\theta})$

As noted in the paper, under assumed regularity conditions a. - g., the consistency of $\hat{\Lambda}_z$ is easily proved through the consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$, the continuous mapping theorem, and the Uniform Strong Law of Large Numbers (USLLN). With respect to asymptotic normality, we begin with the decomposition:

$$n^{1/2}(\hat{\Lambda}_z(t) - \Lambda_z(t)) = \hat{\alpha}_{z1}(t) + \hat{\alpha}_{z2}(t) + \hat{\alpha}_{z3}(t) + \hat{\alpha}_{z4}(t)$$

where

$$\hat{\alpha}_{z1}(t) = n^{1/2}\{\hat{\Lambda}_z(t; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}_z, \hat{\Lambda}^c_{0z}) - \Lambda_z(t; \boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}_z, \hat{\Lambda}^c_{0z})\}$$

$$\hat{\alpha}_{z2}(t) = n^{1/2}\{\hat{\Lambda}_z(t; \boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}_z, \hat{\Lambda}^c_{0z}) - \Lambda_z(t; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z}, \hat{\Lambda}^c_{0z})\}$$

$$\hat{\alpha}_{z3}(t) = n^{1/2}\{\hat{\Lambda}_z(t; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z}, \hat{\Lambda}^c_{0z}) - \Lambda_z(t; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z}, \Lambda^c_{0z})\}$$

$$\hat{\alpha}_{z4}(t) = n^{1/2}\{\hat{\Lambda}_z(t; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z}, \Lambda^c_{0z}) - \Lambda_z(t)\}$$

for $z \in \{0, 1\}$, with $\hat{\Lambda}_z(t) = \hat{\Lambda}_z(t; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\Lambda}^c_{0z})$ and

$$\hat{\Lambda}_z(t; \boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}_z, \hat{\Lambda}^c_{0z}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\hat{w}^c_i(s; \hat{\boldsymbol{\theta}}_z) w^e_i(\boldsymbol{\beta}_0) dN_{iz}(s)}{n^{-1}\sum_{l=1}^{n} \hat{w}^c_l(s; \hat{\boldsymbol{\theta}}_z) w^e_l(\boldsymbol{\beta}_0) Y_{lz}(s)}$$

$$\hat{w}^c_i(s; \hat{\boldsymbol{\theta}}_z) = \exp(\hat{\Lambda}_{0z}(s; \hat{\boldsymbol{\theta}}_z) \exp(\boldsymbol{X}_i^T \hat{\boldsymbol{\theta}}_z))$$

$$\hat{\Lambda}_z(s; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z}, \hat{\Lambda}^c_{0z}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\hat{w}^c_i(s; \boldsymbol{\theta}_{0z}) w^e_i(\boldsymbol{\beta}_0) dN_{iz}(s)}{n^{-1}\sum_{l=1}^{n} \hat{w}^c_l(t; \boldsymbol{\theta}_{0z}) w^e_l(\boldsymbol{\beta}_0) Y_{lz}(s)}$$

$$\hat{w}^c_i(s; \boldsymbol{\theta}_{0z}) = \exp(\hat{\Lambda}^c_{0z}(s; \boldsymbol{\theta}_{0z}) \exp(\boldsymbol{X}_i^T \boldsymbol{\theta}_{0z}))$$

$$\hat{\Lambda}_z(t; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z}, \Lambda^c_{0z}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{w^c_i(s; \boldsymbol{\theta}_{0z}) w^e_i(\boldsymbol{\beta}_0) dN_{iz}(s)}{n^{-1}\sum_{l=1}^{n} w^c_l(s; \boldsymbol{\theta}_{0z}) w^e_l(\boldsymbol{\beta}_0) Y_{lz}(s)}$$

$$w^c_i(s; \boldsymbol{\theta}_{0z}) = \exp(\Lambda^c_{0z}(s) \exp(\boldsymbol{X}_i^T \boldsymbol{\theta}_{0z}))$$

We can express $\hat{\alpha}_{z1}(t)$ as follows:

$$\hat{\alpha}_{z1}(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t \left\{ \frac{w_i^e(\hat{\boldsymbol{\beta}})}{B(s; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}_z)} - \frac{w_i^e(\boldsymbol{\beta}_0)}{B(s; \boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}_z)} \right\} \hat{w}_i^c(s; \hat{\boldsymbol{\theta}}_z) dN_{iz}(s), \tag{B.4}$$

where $B(s; \boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}_z) = n^{-1} \sum_{i=1}^{n} \hat{w}_i^c(s; \hat{\boldsymbol{\theta}}_z) w_i^e(\boldsymbol{\beta}_0) Y_{iz}(s)$, so that $\hat{\alpha}_{z1}(t) = \hat{\alpha}_{z11}(t) + \hat{\alpha}_{z12}(t)$, where:

$$\hat{\alpha}_{z11}(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t \left\{ \frac{w_i^e(\hat{\boldsymbol{\beta}}) - w_i^e(\boldsymbol{\beta}_0)}{B(s; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}_z)} \right\} \hat{w}_i^c(s; \hat{\boldsymbol{\theta}}_z) dN_{iz}(s)$$

$$\hat{\alpha}_{z12}(t) = \sum_{i=1}^{n} \int_0^t w_i^e(\boldsymbol{\beta}_0) \left\{ \frac{1}{B(s; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}_z)} - \frac{1}{B(s; \boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}_z)} \right\} \hat{w}_i^c(s; \hat{\boldsymbol{\theta}}_z) dN_{iz}(s).$$

With respect to $\hat{\alpha}_{z11}(t)$, by a linear Taylor series expansion, we have:

$$n^{1/2} \{ w_i^e(\hat{\boldsymbol{\beta}}) - w_i^e(\boldsymbol{\beta}_0) \} = \boldsymbol{a}_i^T(\boldsymbol{\beta_0})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{B.5}$$

where, as defined in the paper, $\boldsymbol{a}_i^T(\boldsymbol{\beta_0}) = \frac{\partial w_i^e(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta_0}}$. Further, from maximum likelihood theory:

$$n^{1/2} \{ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0} \} = \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1) \tag{B.6}$$

Combining the above results, we have:

$$n^{1/2} \{ w_i^e(\hat{\boldsymbol{\beta}}) - w_i^e(\boldsymbol{\beta_0}) \} = \boldsymbol{a}_i^T(\boldsymbol{\beta_0}) \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1) \tag{B.7}$$

Using the above result, then applying SLLN and continuity, as $n \to \infty$, we can re-express $\hat{\alpha}_{z11}(t)$ as:

$$\hat{\alpha}_{z11}(t) = \boldsymbol{h}_z^T(t) \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1) \tag{B.8}$$

$$\boldsymbol{h}_z(t) = E\Big[\int_0^t \frac{\hat{w}_i^c(s;\hat{\boldsymbol{\theta}}_z)\boldsymbol{a}_i(\boldsymbol{\beta_0})}{E[\hat{w}_i^c(s;\hat{\boldsymbol{\theta}}_z)w_i^e(\boldsymbol{\beta_0})Y_{iz}(s)]}dN_{iz}(s)\Big] \tag{B.9}$$

With respect to $\hat{\alpha}_{z12}(t)$, combining a Taylor expansion with $B.3$

$$n^{1/2}\{B^{-1}(s;\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}_z) - B^{-1}(s;\boldsymbol{\beta_0},\hat{\boldsymbol{\theta}}_z)\} = -\frac{\boldsymbol{B^{\beta}}(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z)}{B(s;\boldsymbol{\beta_0},\hat{\boldsymbol{\theta}}_z)^2}\boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0})n^{-1/2}\sum_{i=1}^n \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1) \tag{B.10}$$

where we define:

$$\boldsymbol{B^{\beta}}(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z) = \frac{\partial B(s;\boldsymbol{\beta},\hat{\boldsymbol{\theta}}_z)}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta_0}} = \frac{1}{n}\sum_{i=1}^n Y_{iz}(t)w_i^e(\boldsymbol{\beta}_0)\hat{w}_i^c(s;\hat{\boldsymbol{\theta}}_z)\boldsymbol{a}_i(\boldsymbol{\beta}_0), \tag{B.11}$$

which converges almost surely to $\boldsymbol{b^{\beta}}(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z) = E[Y_{iz}(t)w_i^e(\boldsymbol{\beta}_0)\hat{w}_i^c(s;\hat{\boldsymbol{\theta}}_z)\boldsymbol{a}_i(\boldsymbol{\beta}_0)]$. Then, applying SLLN and continuity:

$$\hat{\alpha}_{z12}(t) = \boldsymbol{d}_z^T(t)\boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0})n^{-1/2}\sum_{i=1}^n \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1) \tag{B.12}$$

$$\boldsymbol{d}_z(t) = \int_0^t \frac{\boldsymbol{b^{\beta}}(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z)}{b(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z)}dN_{iz}(s), \tag{B.13}$$

where $b(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z) = E[\hat{w}_i^c(s;\hat{\boldsymbol{\theta}}_z)w_i^e(\boldsymbol{\beta_0})Y_{iz}(s)]$ Combining the above results,

$$\hat{\alpha}_{z1}(t) = \{\boldsymbol{h}_z(t) + \boldsymbol{d}_z(t)\}^T\boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0})n^{-1/2}\sum_{i=1}^n \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1). \tag{B.14}$$

As we did for $\hat{\alpha}_{z1}(t)$, we can write $\hat{\alpha}_{z2}(t) = \hat{\alpha}_{z21}(t) + \hat{\alpha}_{z22}(t)$, where

$$\hat{\alpha}_{z21}(t) = n^{-1/2}\sum_{i=1}^n \int_0^t w_i^e(\boldsymbol{\beta}_0)B(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z)^{-1}\{\hat{w}_i^c(s;\hat{\boldsymbol{\theta}}_z) - \hat{w}_i^c(s;\boldsymbol{\theta}_{0z})\}dN_i(s) \tag{B.15}$$

$$\hat{\alpha}_{z22}(t) = \sum_{i=1}^n \int_0^t w_i^e(\boldsymbol{\beta}_0)\Big\{\frac{1}{B(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z)} - \frac{1}{B(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})}\Big\}\hat{w}_i^c(s;\boldsymbol{\theta}_{0z})dN_{iz}(s). \tag{B.16}$$

Through standard partial likelihood theory (Fleming and Harrington, 1991),

$$n^{1/2}(\hat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_{0z}) = \boldsymbol{\Omega}_c^{-1}(\boldsymbol{\theta}_{0z}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi}_i^c(\boldsymbol{\theta}_{0z}) + o_p(1) \tag{B.17}$$

$$\boldsymbol{\Omega}_c(\boldsymbol{\theta}) = \int_0^\tau \boldsymbol{v}_c(t,\boldsymbol{\theta}) b_c(t,\boldsymbol{\theta}) d\Lambda_0^c(t), \tag{B.18}$$

$$\boldsymbol{\psi}_i^c(\boldsymbol{\theta}) = \int_0^\tau \{\boldsymbol{X}_i - \bar{\boldsymbol{x}}_c(t,\boldsymbol{\theta})\} dM_i^c(t). \tag{B.19}$$

Using a Taylor expansion, the SLLN and continuity,

$$n^{1/2}\{\hat{w}_i^c(s;\hat{\boldsymbol{\theta}}) - \hat{w}_i^c(s;\boldsymbol{\theta}_{0z})\} = \hat{w}_i^c(s;\boldsymbol{\theta}_{0z}) \boldsymbol{k}_i^T(s,\boldsymbol{\theta}_{0z}) \boldsymbol{\Omega}_c^{-1}(\boldsymbol{\theta}_{0z}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi}_i^c(\boldsymbol{\theta}_{0z}) + o_p(1) \tag{B.20}$$

$$\boldsymbol{k}_i^T(s,\boldsymbol{\theta}_{0z}) = \int_0^s \exp\{\boldsymbol{X}_i^T \boldsymbol{\theta}_{0z} - \bar{\boldsymbol{x}}_c(u,\boldsymbol{\theta}_{0z})\} d\Lambda_{0z}^c(u). \tag{B.21}$$

Substituting the above expression into $\hat{\alpha}_{z21}(t)$, then applying SLLN,

$$\hat{\alpha}_{z21}(t) = \boldsymbol{g}_z^T(t) \boldsymbol{\Omega}_c^{-1}(\boldsymbol{\theta}_{0z}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi}_i^c(\boldsymbol{\theta}_{0z}) + o_p(1) \tag{B.22}$$

$$\boldsymbol{g}_z^T(t) = E\left[\int_0^t w_i^e(\boldsymbol{\beta}_0), \hat{w}_i^c(s;\boldsymbol{\theta}_{0z}) b_z(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})^{-1} \boldsymbol{k}_i(s) dN_{iz}(s) \right]. \tag{B.23}$$

With respect to $\hat{\alpha}_{z22}(t)$, through another Taylor series expansion,

$$n^{1/2}\{B^{-1}(s;\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}}_z) - B^{-1}(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})\} = -\frac{\boldsymbol{B}^{\boldsymbol{\theta}}(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})}{B(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})^2} \boldsymbol{\Omega}_{\boldsymbol{c}}^{-\mathbf{1}}(\boldsymbol{\theta}_{\mathbf{0z}}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi}_{\boldsymbol{i}}^{\boldsymbol{c}}(\boldsymbol{\theta}_{\mathbf{0z}}) + o_p(1) \tag{B.24}$$

where we define:

$$\boldsymbol{B}^{\boldsymbol{\theta}}(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z}) = \frac{\partial B(s;\boldsymbol{\beta}_0,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}_{\mathbf{0z}}} = \frac{1}{n} \sum_{i=1}^{n} Y_{iz}(s) w_i^e(\boldsymbol{\beta}) \hat{w}_i^c(s;\boldsymbol{\beta},\boldsymbol{\theta}_{0z}) \boldsymbol{k}_i(s), \tag{B.25}$$

which converges almost surely to $\boldsymbol{b}^{\boldsymbol{\theta}}(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z}) = E[Y_{iz}(s)w_i^e(\boldsymbol{\beta}_0)\hat{w}_i^c(s;\boldsymbol{\beta},\boldsymbol{\theta}_{0z})\boldsymbol{k}_i(s)]$. Substituting into expression for $\hat{\alpha}_{z22}(t)$, and again applying SLLN and using continuity:

$$\hat{\alpha}_{z22}(t) = \boldsymbol{f}_z^T(t)\boldsymbol{\Omega}_c^{-1}(\boldsymbol{\theta}_{0z})n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\psi}_i^c(\boldsymbol{\theta}_{0z}) + o_p(1) \tag{B.26}$$

$$\boldsymbol{f}_z(t) = \int_0^t \frac{\boldsymbol{b}^{\boldsymbol{\theta}}(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})}{b(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})}d\Lambda_z(s). \tag{B.27}$$

Combining the above results,

$$\hat{\alpha}_{z2}(t) = \{\boldsymbol{g}_z(t) + \boldsymbol{f}_z(t)\}^T\boldsymbol{\Omega}_{\boldsymbol{c}}^{-\mathbf{1}}(\boldsymbol{\theta}_{\mathbf{0}\boldsymbol{z}})n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\psi}_{\boldsymbol{i}}^{\boldsymbol{c}}(\boldsymbol{\theta}_{\mathbf{0}\boldsymbol{z}}) + o_p(1). \tag{B.28}$$

With respect to $\hat{\alpha}_{z3}(t)$ we can write:

$$\hat{\alpha}_{z3}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t w_i^e(\boldsymbol{\beta}_0)B(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)^{-1}\{\hat{w}_i^c(s;\boldsymbol{\theta}_{0z}) - w_i^c(s;\boldsymbol{\theta}_{0z})\}dN_i(s). \tag{B.29}$$

Applying the Functional Delta Method,

$$n^{1/2}\{\hat{w}_i^c(s;\boldsymbol{\theta}_0) - w_i^c(s;\boldsymbol{\theta}_{0z})\} = w_i^c(s;\boldsymbol{\theta}_{0z})n^{1/2}\{\hat{\Lambda}_i^c(s;\boldsymbol{\theta}_{0z}) - \Lambda_i^c(s)\} \tag{B.30}$$

$$= w_i^c(s;\boldsymbol{\theta}_{0z})n^{-1/2}\sum_{l=1}^{n}\int_0^s \frac{\exp\{\boldsymbol{\theta}_{0z}^T\boldsymbol{X}_i\}}{B_c(u,\boldsymbol{\theta}_{0z})}dM_l^C(u) = o_p(1). \tag{B.31}$$

Substituting this expression into the above expression for $\hat{\alpha}_{z3}(t)$, changing the orders of integration and summation, then applying the SLLN,

$$\hat{\alpha}_{z3}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t q_z(s,t)b(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})^{-1}dM_i^C(s) \tag{B.32}$$

$$q_z(s,t) = E\left[\exp\{{\boldsymbol{\theta}_{0z}}^T\boldsymbol{X}_i\}\int_u^t w_i^e(\boldsymbol{\beta}_0), w_i^C(s;\boldsymbol{\theta}_{0z})b(s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_{0z})^{-1}dN_{iz}(s)\right]. \tag{B.33}$$

Finally, the quantity, $\hat{\alpha}_{z4}(t)$ can be written as

$$\hat{\alpha}_{z4}(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t w_i^e(\boldsymbol{\beta}_0), w_i^C(s; \boldsymbol{\theta}_{0z}) b(s; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z})^{-1} dM_{iz}(s). \tag{B.34}$$

Thus, combining the above results,

$$n^{1/2}\{\hat{\Lambda}_z(t) - \Lambda(t)\} = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\Phi}_{iz}(t) + o_p(1) \tag{B.35}$$

where $\boldsymbol{\Phi}_{iz}(t)$ as defined in Theorem 1 is:

$$\begin{aligned}
\boldsymbol{\Phi}_{iz}(t) = \{\boldsymbol{h}_z(t) + \boldsymbol{d}_z(t)\}^T \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) \\
+\{\boldsymbol{g}_z(t) + \boldsymbol{f}_z(t)\}^T \boldsymbol{\Omega_C^{-1}}(\boldsymbol{\theta_{0z}}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i^C}(\boldsymbol{\theta_{0z}}) \\
+n^{-1/2} \sum_{i=1}^{n} \int_0^t q_z(s,t) b(s; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z})^{-1} dM_i^C(s) \\
+n^{-1/2} \sum_{i=1}^{n} \int_0^t w_i^e(\boldsymbol{\beta}_0), w_i^C(s; \boldsymbol{\theta}_{0z}) b(s; \boldsymbol{\beta}_0, \boldsymbol{\theta}_{0z})^{-1} dM_{iz}(s).
\end{aligned}$$

As $n \to \infty$, $n^{1/2}\{\hat{\Lambda}_z(t) - \Lambda(t)\}$ behaves like asume of independent and identically distributed mean 0 random variates. Therefore, by the multivariate central limit theorem, for any finite set of (say k) time points, the vector $n^{1/2}[\{\hat{\Lambda}_z(t_1) - \Lambda(t_1)\}, ..., \{\hat{\Lambda}_z(t_k) - \Lambda(t_k)\}]$ converges to a mean zero multivariate normal distribution. Using techniques in Bilias, Gu and Ying (1997), $n^{1/2}\{\hat{\Lambda}_z(t) - \Lambda(t)\}$ can be shown to be tight (Pollard (1990), van der Vaart and Wellner (1996)); so that $n^{1/2}\{\hat{\Lambda}_z(t) - \Lambda(t)\}$ converges to a mean zero Gaussian process with covariance function $E\{\boldsymbol{\Phi}_{iz}(s)\boldsymbol{\Phi}_{iz}(t)\}$ for any pair of time points $(s,t) \in [0,\tau] \times [0,\tau]$.

Proof of THEOREM 2

With respect to asymptotic normality, we begin with the decomposition:

$$n^{1/2}(\hat{\delta}(t) - \delta(t)) = \hat{\gamma}_1(t) + \hat{\gamma}_2(t)$$

where

$$\hat{\gamma}_1(t) = n^{1/2}\{\hat{\delta}(t; \hat{\mu}_{A,1}, \hat{\mu}_{A,0}, \hat{\mu}_{T,1}, \hat{\mu}_{T,0}) - \hat{\delta}(t; \mu_{A,1}, \mu_{A,0}, \hat{\mu}_{T,1}, \hat{\mu}_{T,0})\}$$
$$\hat{\gamma}_2(t) = n^{1/2}\{\hat{\delta}(t; \mu_{A,1}, \mu_{A,0}, \hat{\mu}_{T,1}, \hat{\mu}_{T,0}) - \delta(t)\}$$

We can express $\hat{\gamma}_1(t)$ as follows:

$$\hat{\gamma}_1(t) = n^{-1/2} \sum_{i=1}^{n} [R^{-1}(\hat{\boldsymbol{\beta}}) - R^{-1}(\boldsymbol{\beta_0})]\{\hat{\mu}_{T,1}(t) - \hat{\mu}_{T,0}(t)\} \tag{B.36}$$

where

$$R(\hat{\boldsymbol{\beta}}) = \mu_{A,1}(\hat{\boldsymbol{\beta}}) - \mu_{A,0}(\hat{\boldsymbol{\beta}})$$
$$R(\boldsymbol{\beta_0}) = \mu_{A,1}(\boldsymbol{\beta_0}) - \mu_{A,0}(\boldsymbol{\beta_0})$$
$$\mu_{A,z}(\hat{\boldsymbol{\beta}}) = \frac{\sum_{i=1}^{n} w_i^e(\hat{\boldsymbol{\beta}}) A_i I(Z_i = z)}{\sum_{i=1}^{n} w_i^e(\hat{\boldsymbol{\beta}}) I(Z_i = z)}$$
$$\mu_{A,z}(\boldsymbol{\beta_0}) = \frac{\sum_{i=1}^{n} w_i^e(\boldsymbol{\beta_0}) A_i I(Z_i = z)}{\sum_{i=1}^{n} w_i^e(\boldsymbol{\beta_0}) I(Z_i = z)}$$

With respect to the expression for $\hat{\gamma}_1(t)$, by a linear Taylor series expansion:

$$n^{1/2}\{R^{-1}(\hat{\boldsymbol{\beta}}) - R^{-1}(\boldsymbol{\beta_0})\} = -\frac{\boldsymbol{R^{\beta}}(\boldsymbol{\beta_0})}{R(\boldsymbol{\beta_0})^2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}) + o_p(1) \tag{B.37}$$

where, we define:

$$\boldsymbol{R^{\beta}}(\boldsymbol{\beta_0}) = \frac{\partial R(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta_0}}$$
$$= \sum_{z\in\{0,1\}} \left[\frac{\sum_{i=1}^{n} I(Z_i = z) I(A_i = 1) \boldsymbol{a}_i^e(\boldsymbol{\beta_0})}{\sum_{i=1}^{n} I(Z_i = z) w_i^e} - \frac{\sum_{i=1}^{n} W_z I(Z_i = z) I(A_i = 1) \boldsymbol{a}_i^e(\boldsymbol{\beta_0})}{(\sum_{i=1}^{n} I(Z_i = z) w_i^e)^2} \right] \tag{B.38}$$

where $W_z = \frac{\sum_{i=1}^{n} I(Z_i=z) I(A_i=1) w_i^e}{\sum_{i=1}^{n} I(Z_i=z) w_i^e}$ and from maximum likelihood theory:

$$n^{1/2}\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}\} = \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1) \tag{B.39}$$

Combining the above results, we have:

$$n^{1/2}\{R^{-1}(\hat{\boldsymbol{\beta}}) - R^{-1}(\boldsymbol{\beta_0})\} = -\frac{\boldsymbol{R^{\beta}}(\boldsymbol{\beta_0})}{R(\boldsymbol{\beta_0})^2} \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1) \tag{B.40}$$

Using the above result, and by application of strong law of large numbers and continuity, as $n \to \infty$, we can write $\hat{\gamma}_1(t)$ as:

$$\hat{\gamma}_1(t) = \boldsymbol{Q}^T \boldsymbol{\Omega^{-1}}(\boldsymbol{\beta_0}) n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + o_p(1) \tag{B.41}$$

where $\boldsymbol{Q} = E[-\frac{\boldsymbol{R^{\beta}}(\boldsymbol{\beta_0})}{R(\boldsymbol{\beta_0})^2}]$.

We can write $\hat{\gamma}_2(t)$ as:

$$\hat{\gamma}_2(t) = n^{1/2}[\{\hat{\mu}_{T,1}(t) - \hat{\mu}_{T,0}(t)\} - \{\mu_{T,1}(t) - \mu_{T,0}(t)\}] \frac{1}{\mu_{A,1}(t) - \mu_{A,0}(t)}$$
$$= n^{1/2} \left[\left\{ \int_0^t \hat{S}_1(u) - S_1(u) dt \right\} - \left\{ \int_0^t \hat{S}_0(u) - S_0(u) dt \right\} \right] \frac{1}{\mu_{A,1}(t) - \mu_{A,0}(t)}$$

Note that, as $\hat{\Lambda}_z(t)$ converges almost surely to $\Lambda_z(t)$ (Theorem 1), through continuity, $\hat{S}_z(t)$ converges almost surely to $S_z(t)$. This implies that $\int_0^t \hat{S}_z(u) du \xrightarrow{a.s} \int_0^t S_z(u) du$, which in turn implies the almost surely convergence of $\hat{\delta}(t)$ to $\delta(t)$ in $t \in [0, \tau]$. Fur-

ther, by functional Delta Method, $n^{1/2}(\hat{S}_z(u) - S_z(u)) = -S_z(u)n^{1/2}\{\hat{\Lambda}_z(u) - \Lambda_z(u)\}$. By integrating, we get that $n^{1/2}\{\int_0^t \hat{S}_z(u) - S_z(u)dt\} = -n^{1/2}\int_0^t S_z(u)\{\hat{\Lambda}_z(u) - \Lambda_z(u)\}$. Switching the order of integration, and substituting $\mu_{T,z}(t) = \int_0^t S_z(u)dt$, we obtain:

$$n^{1/2}\{\hat{\mu}_{T,z}(t) - \mu_{T,z}(t)\} = -n^{1/2}\int_0^t \{\mu_{T,z}(t) - \mu_{T,z}(u)\}d\{\hat{\Lambda}_z(u) - \Lambda_z(u)\} \tag{B.42}$$

Using results from Theorem 1, this can be written as:

$$n^{1/2}\{\hat{\mu}_{T,z}(t) - \mu_{T,z}(t)\} = -n^{-1/2}\sum_{i=1}^{n}\int_0^t \{\mu_{T,z}(t) - \mu_{T,z}(u)\}d\boldsymbol{\Phi}_{iz}(u) + o_p(1), \tag{B.43}$$

such that, asymptotically,

$$\hat{\gamma}_2(t) = \{\mu_{A,1}(t) - \mu_{A,0}(t)\}^{-1}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\Sigma}_i(t);$$

$$\boldsymbol{\Sigma}_i(t) = \int_0^t \{\mu_{T,0}(t) - \mu_{T,0}(u)\}d\boldsymbol{\Phi}_{i0}(u) - \int_0^t \{\mu_{T,1}(t) - \mu_{T,1}(u)\}d\boldsymbol{\Phi}_{i1}(u)$$

Thus, combining the above results,

$$n^{1/2}(\hat{\delta}(t) - \delta(t)) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_i(t) + o_p(1)$$

where, $\boldsymbol{\xi}_i(t)$, as defined in Theorem 2 is given by:

$$\boldsymbol{\xi}_i(t) = \boldsymbol{Q}^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta_0})\sum_{i=1}^{n}\boldsymbol{\psi_i}(\boldsymbol{\beta_0}) + \{\mu_{A,1}(t) - \mu_{A,0}(t)\}^{-1}\sum_{i=1}^{n}\boldsymbol{\Sigma}_i(t)$$

Following results from Theorem 1, $n^{1/2}(\hat{\delta}(t) - \delta(t))$ behaves like a sum of independent and identically distributed mean 0 random variates. Hence, by multivariate central limit theorem $n^{1/2}[\{\hat{\delta}(t_1) - \delta(t_1)\}, ..., \{\hat{\delta}(t_k) - \delta(t_k)\}]$. Further, using arguments analogous to those used in the proof of Theorem 1, $n^{1/2}(\hat{\delta}(t) - \delta(t))$ converges to with

covariance function $E\{\boldsymbol{\xi}_i(s)\boldsymbol{\xi}_i(t)\}$ for any pair of time points $(s,t) \in [0,\tau] \times [0,\tau]$.

BIBLIOGRAPHY

BIBLIOGRAPHY

Andersen, P., Klein, J. P., and Rosthj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* **90,** 15-27.

Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.

Ash A.S., Fienberg S.E., Louis T.A., Normand S.T., Stukel T.A., and Utts J. (2012). Statistical issues in assessing hospital performance. White paper, Committee of Presidents of Statistical Societies.

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

Balan T., Boonk S.E., Vermeer M.H., and Putter H. (2016). Score test for association between recurrent events and a terminal event. *Statistics in medicine* **35(18)**, 3037–3048.

Bandeen-Roche, K. and Liang, K.-Y. (2002). Modeling multivariate failure time associations in the presence of a competing risk. *Biometrika* **89,** 299 - 314.

Bilias, Y., Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for the Cox model with staggered entry. The Annals of Statistics 25, 662682.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

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

Chan, K. C. (2016). Instrumental variable additive hazards models with exposure dependent censoring. *Biometrics* **72**, 1003–1005.

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

Cheng, Y. and Fine, J. P. (2008). Nonparametric estimation of cause-specic cross hazard ratio with bivariate competing risks data. *Biometrika* **95,** 233-240.

Cheng, Y., Fine, J. P. and Bandeen-Roche, K. (2010). Association analyses of clustered competing risks data via cross hazard ratio. *Biostatistics* **11,** 82-92.

Cheng, Y., Fine, J. P. and Kosorok, M. R. (2007). Nonparametric analysis of bivariate competing risks data. *Journal of the American Statistical Association* **102,** 1407-1416.

Choi, S. and Huang, X. (2014). Maximum likelihood estimation of semiparametric mixture component models for competing risks data. *Biometrics* **70,** 588-598.

Cox D.R. (1959). The analysis of exponentially distributed lifetimes with two types of failure. *Journal of Royal Statistical Society, Series B* **21**, 411–421.

Crowder M.J. (2001) *Classical competing risks*. London: Chapman and Hall/CRC.

Cuzick, J., Sasieni, P., Myles, J., and Tyler, J. (2007). Estimating the Effect of Treatment in a Proportional Hazards Model in the Presence of Non-compliance and Contamination. *Journal of the Royal Statistical Society, Series B (Methodological)* **69**, 565–588.

Do Ha I., Christian N.J., Jeong J.H., Park J., and Lee Y. (2014). Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. *Statistical methods in medical research*. Published Online. DOI: 10.1177/0962280214526193.

Elashoff, R., Li, G., Zhou, Y. (2012). Nonparametric inference for assessing treatment efficacy in randomized clinical trials with a time-to-event outcome and all-or-none compliance. *Biometrika*. **99(2)**, 393–404.

Fan L., and Schaubel D.E. (2016). Comparing center-specific cumulative incidence functions. *Lifetime data analysis* **22(1)**, 1–21.

Fenton, S.A., Schaubel, D.E., Desmeules, M., Morrison, H.I., Mao, Y., Copleston, P., Jeffery, J.R., and Kjellstrand, C.M. (1997). Hemodialysis versus peritoneal dialysis: a comparison of adjusted mortality rates. *American Journal of Kidney Diseases*, **30(3)**, 334–342.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94,** 496-509.

Frolich, M. (2007). Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics* **139**, 35–75.

Gail M.H. (1975). A review and critique of some models used in competing risk analysis. *Biometrics* **31**, 209–222.

Gorfine M., and Hsu L. (2011). FrailtyBased Competing Risks Model for Multivariate Survival Data. *Biometrics* **67(2)**, 415–426.

Gorfine, M., Hsu, L., Zucker, D. M., and Parmigiani, G. (2014). Calibrated predictions for multivariate competing risks models. *Lifetime Data Analysis* **20(2)**, 234–251.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics* **16,** 1141-1154.

Gray R.J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognoses. *Journal of the American Statistical Association* **87**, 942–951.

Hart A., Smith J. M., Skeans M. A., Gustafson S. K., Stewart D. E., Cherikh W. S., Wainright J. L., Boyle G., Snyder J. J., Kasiske B. L. and Israni A. K. (2016). OPTN/SRTR Annual Data Report 2014: Kidney. *American Journal of Transplant* **16 (Suppl 2)**, 11–46.

He K., and Schaubel D.E. (2014a). Methods for comparing centerspecific survival outcomes using direct standardization. *Statistics in medicine* **33(12)**, 2048–2061.

He K., and Schaubel D.E. (2014b). Standardized Mortality Ratio for Evaluating Center-Specific Mortality: Assessment and Alternative. *Statistics in Biosciences* **7(2)**, 1–26.

Heaf, J.G., Lkkegaard, H., and Madsen, M. (2002). Initial survival advantage of peritoneal dialysis relative to haemodialysis. *Nephrology Dialysis Transplantation* **17(1)**, 112–117.

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

Imbens, G.W., and Angrist, J.D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62**, 467–475.

Jaar, B.G., Coresh, J., Plantinga, L.C., Fink, N.E., Klag, M.J., Levey, A.S., Levin, N.W., Sadler, J.H., Kliger, A., and Powe, N.R. (2005). Comparing the risk for death with peritoneal dialysis and hemodialysis in a national cohort of patients with chronic kidney disease. *Annals of internal medicine* **143(3)**, 174–183.

Jeong, J.-H. and Fine, J. P. (2007). Parametric regression on cumulative incidence function. *Biostatistics* **8,** 184-196.

Johnson, D.W., Dent, H., Hawley, C.M., McDonald, S.P., Rosman, J.B., Brown, F.G., Bannister, K., Wiggins, K.J. (2009). Association of dialysis modality and cardiovascular mortality in incident dialysis patients. *Clinical Journal of the American Society of Nephrology* **4**, 1620–1628.

Kalbfleisch J.D., and Prentice R.L. (2002). *The statistical analysis of failure time data*, 2nd Edition. New York: Wiley.

Kalbfleisch J.D., and Wolfe R. (2013). On monitoring outcomes of medical providers. *Statistics in Biosciences* **5**, 286–302.

Katsahian S., and Boudreau C. (2011). Estimating and testing for center effects in competing risks. *Statistics in medicine* **30(13)**, 1608–1617.

Katsahian, S., RescheRigon, M., Chevret, S. and Porcher, R. (2006). Analysing multicenter competing risks data with a mixed proportional hazards model for the subdistribution. *Statistics in Medicine* **25,** 4267-4278.

Kim, H., Kim, K.H., Park, K., Kang, S-W., Yoo, T-H., Ahn, S.V., Ahn, H.S., Hann, H.J., Lee, S., Ryu, J-H., et al. (2014) A population-based approach indicates an overall higher patient mortality with peritoneal dialysis compared to hemodialysis in Korea. *Kidney international* **86(5)**, 991–1000.

Kim, H., Kim, K.H., Ahn, S.V., Kang, S.W., Yoo, T.H., Ahn, H.S., Hann, H.J., Lee, S., Ryu, J.H., Yu, M., Kim, S.J., Kang, D.H., Choi, K.B., Ryu, D.R. (2015). Risk of major cardiovascular events among incident dialysis patients: A Korean national population-based study. *International Journal of Cardiology* **198**, 95–101.

Kjaersgaard, M.I., and Parner, E.T.(2016) Instrumental variable method for time-to-event data using a pseudo-observation approach. *Biometrics* **72(2)**, 463–472.

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

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

Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society Series C* **34,** 201-211.

Li, F., Morgan, K.L., and Zaslavsky, A.M. (2016). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association to appear.*

Li, J., Fine, J.P. and Brookhart, M.A. (2015). Instrumental variable additive hazards models. *Biometrics* **71**, 122–130.

Li, L., and Greene, T. (2013). A Weighting Analogue to Pair Matching in Propensity Score Analysis. *International Journal of Biostatistics* **9**, 1–20.

Loeys, T. and Goetghebeur, E. (2003). A Causal Proportional Hazards Estimator for the Effect of Treatment Actually Received in a Randomized Trial with All-or-nothing Compliance. *Biometrics* **59**, 100–105.

Lu, W. and Peng, L. (2008). Semiparametric analysis of mixture regression models with competing risks data. *Lifetime Data Analysis* **14,** 231-252.

Maller, R. A. and Zhou, X. (2002). Analysis of parametric models for competing risks. *Statistica Sinica* **12,** 725-750.

Mark, S.D., Robins, J.M. (1993). Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model. *Statistics in Medicine* **12**, 1605–1628.

Nie, H., Cheng, J., Small, D. (2011). Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics* **67**, 1397–1405.

Ohlssen D.I., Sharples L.D., and Spiegelhalter D.J. (2006). A hierarchical modelling framework for identifying unusual performance in health care providers. *Journal of the Royal Statistical Society, Series A* **170**, 865–890.

Pollard, D. (1990). Empirical processes: Theory and Applications. Hayward: Institute of Mathematical Statistics.

Prentice, R. L., Kalbfleisch, J. D., Peterson, JR, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34,** 541-554.

Putter H., Fiocco M., and Geskus R. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine* **26**, 2389–2430.

Richardson, A., Hudgens, M., Fine, J.P., and Brookhart, M.A. (2017). Nonparametric binary instrumental variable analysis of competing risks data. *Biostatistics* **18**, 48–61.

Ripatti S., and Palmgren J. (2000). Estimation of multivariate frailtymodels using penalized partial likelihood. *Biometrics* **56**, 1016–1022.

Robins, J.M., Hernan, M.A., Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560.

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

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

Scheike, T. H., Sun, Y., Zhang, M. J. and Jensen, T. K. (2010). A semiparametric random eects model for multivariate competing risks data. *Biometrika* **97,** 133-145.

Scheike, T. H., Zhang, M. J. and Gerds, T. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika* **95,** 205-220.

Scheike, T. and Sun, Y. (2012). On cross-odds ratio for multivariate competing risks data. *Biostatistics* **13,** 680-694.

Sjölander A., Lichtenstein P., Larsson H, Pawitan Y. (2013). Between-within models for survival analysis. *Statistics in Medicine* **32**, 3067–3076.

Spiegelhalter D., Sherlaw-Johnson C., Bardsley M., Blunt I., Wood C., and Grigg O. (2012). Statistical methods for healthcare regulation: rating, screening and surveillance. *Journal of Royal Statistical Society, Series A* **175(1)**, 1–47.

Sun L, Liu J, Sun J and Zhang M (2006). Modeling the subdistribution of a competing risk. *Statistica Sinica* **16,** 1367-1385.

Tan, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association* **101**, 1607–1618.

Tchetgen Tchetgen, E.J., Walter, S., Vansteelandt, S., Martinussen, T. and Glymour, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology* **26**, 402–410.

Therneau T. (2015). Package 'coxme'. Mixed Effects Cox Models. R Package version 2.2–5.

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

VanRompaye B., Goetghebeur E., and Jaffar S. (2010). Design and testing for clinical trials faced with misclassified causes of death. *Biostatistics* **11**, 546–558.

Van Rompaye B., Eriksson M., and Goetghebeur E. (2015). Evaluating hospital performance based on excess causespecific incidence. *Statistics in medicine* **34(8)**, 1334–1350.

Varewyck M., Goetghebeur E., Eriksson M., and Vansteelandt S. (2014). On shrinkage and model extrapolation in the evaluation of clinical center performance. *Biostatistics* **15(4)**, 651–664.

Weinhandl, E.D., Foley, R.N., Gilbertson, D.T., Arneson, T.J., Snyder, J.J., and Collins, A.J. (2010). Propensity-matched mortality comparison of incident hemodialysis and peritoneal dialysis patients. *Journal of the American Society of Nephrology* **21(3)**, 499–506.

Yu, W., Chen, K., Sobel, M., Ying, Z. (2015). Semiparametric transformation models for causal inference in time to event studies with all-or-nothing compliance. *Journal of the Royal Statistical Society, Series B (Methodological)* **77(2)**, 397–415.

Zhao, L., Shi, J., Shearon, T. H., and Li, Y. (2015). A Dirichlet process mixture model for survival outcome data: assessing nationwide kidney transplant centers. *Statistics in medicine* **34(8)**, 1404–1416.

Zheng, C., Dai, R., Hari, P. N., and Zhang, M-J. (2017) Instrumental variable with competing risk model. *Statistics in Medicine* **36**, 1240–1255.

Zhou, B., Fine, J., Latouche, A. and Labopin, M. (2012). Competing risks regression for clustered data. *Biostatistics* **13,** 371-383.