SEMIPARAMETRIC METHODS FOR ESTIMATING CUMULATIVE TREATMENT EFFECTS IN THE PRESENCE OF NON-PROPORTIONAL HAZARDS AND DEPENDENT CENSORING

by

Guanghui Wei

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2008

Doctoral Committee:

Associate Professor Douglas E. Schaubel, Chair
Professor John D. Kalbfleisch
Associate Professor Susan Murray
Assistant Professor Panduranga S. Rao

ACKNOWLEDGEMENTS

I would like to express my deepest and sincere gratitude to all those who helped me to complete this thesis. I especially want to thank for my advisor, Doug, who gave me the most important guidance during my PhD study. He set me an excellent example of a researcher. I also want to thank the Canadian Organ Replacement Register and the Scientific Registry for Transplant Recipients for giving me permission to use their data. Furthermore, I want to thank for my committee members, Doug, Jack, Susan and Pandu, who offered valuable inputs to my thesis. These people devoted their time unselfishly to my dissertation. Finally, I would like to give my special thanks to my parents. Their love and support enabled me to complete this work. To them I dedicate this thesis.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER I

Introduction

In clinical studies of time to event data, non-proportional hazards are very common. The Cox model is frequently used assuming that the treatment effect is either constant over time or a specific function of time. However, it is difficult to determine whether the form chosen for the treatment effect is correct. Even if the correct form is chosen, the cumulative treatment effect is often preferred over treatment effect on the hazard function in many applications. For example, clinicians are often interested in comparing the 5-year survival between the treatment groups. In the presence of non-proportional hazards, survival or cumulative hazard curves can be compared using non-parametric estimators such as the Nelson-Aalen (Nelson, 1972; Aalen, 1978) or Kaplan-Meier (Kaplan and Meier, 1958) estimator. These estimators will lead to biased treatment comparisons in the presence of confounders, as is often the case in observational studies.

Several methods have been proposed to estimate cumulative treatment effects. Doksum and Song (1989) introduced a relative change function in terms of cumulative hazards to compare two groups. Confidence bands were constructed under a proportional hazards assumption. Parzen, Wei and Ying (1997) examined the difference between two survival functions and constructed confidence bands based on

simulation techniques. McKeague and Zhao (2002) proposed the ratio of survival functions as a treatment effect measure and constructed simultaneous confidence bands using empirical likelihood. Kalbfleisch and Prentice (1981) estimated an average hazard ratio, averaged with respect to a weight function.

Each of the above listed methods are not suitable for observational data since they do not accommodate the adjustment of potential confounders. In the context of semiparametric models which adjust for covariates, Schemper (1992) examined the estimation of the average hazard ratio for two populations through a weighted Cox model. Xu and O'Quigley (2000) estimated the average regression effect under a non-proportional hazards model with time-varying regression coefficients, with a weighted score equation used to obtain a consistent estimator.

In this dissertation, we develop three novel methods to estimate cumulative treatment effects. In Chapter 2, we utilize a stratified Cox model, but with treatment groups serving as strata. The ratio of cumulative hazards between each treatment and the reference group is estimated. Through the use of a treatment-stratified Cox model, the functional form of the treatment effect does not need to be specified, while adjustment for potential confounders is achieved through a model. With time-constant adjustment covariates, the proposed measure can be reduced to a ratio of baseline cumulative hazards, which can be estimated by the method of Breslow (1972). When proportional hazards holds for the treatment effect, our measure converges to the commonly used hazard ratio. Our proposed estimator is proved to follow a Gaussian process, with a variance estimator in an explicit form derived using the theory of empirical processes.

The method proposed in Chapter 2 assumes that the adjustment covariate effects follow proportional hazards. When proportional hazards does not hold, the

functional forms of these effects need to be specified. In Chapter 3, we propose an inverse probability of treatment weighting (IPTW) method to balance the distribution of confounders among treatment groups. Three measures are developed to quantify the cumulative treatment effects, contrasting treatment-specific cumulative hazards, probability of death, and restricted mean lifetime. These measures were estimated using non-parametric estimators derived from weighted Nelson-Aalen estimators (Nelson 1972, Aalen, 1978). The probability of treatment assignment given the potential confounders is estimated through a generalized logit model. After applying IPTW, pusedo populations are created among treatment categories. The distribution of confounders for each of these populations is the same as the entire population. The weighted measures contrast the scenario wherein the treatment is applied to the entire population, to the scenario where the reference treatment is applied. The proposed estimators are proved to follow a Gaussian process with explicit variance estimators derived using empirical process theory.

The IPTW method has been applied in many applications. Brumback, Hernan & Robins (2000) estimated the causal effect of time-dependent exposure adjusting for time-dependent confounders using marginal structural models. In their setting, there exists a time-dependent covariate that is a risk factor for mortality and also predicts subsequent exposure, and past exposure is predictive of this covariate. A inverse weight was applied in order to obtain the true causal effect of exposure. The inverse weight is time-dependent, depending on the treatment and confounder history. Hernan, Brumback & Robins (2000) extended the above work to analyze time to event data. A proportional hazard Cox model was assumed, adjusting for the covariates at baseline. A time-dependent inverse weight was applied to obtain the causal effect of treatment. Xie and Liu (2005) developed an IPTW Kaplan-Meier

estimator, with a weighted log rank test proposed to examine the treatment effects.

Current methods in estimating the cumulative effect in the context of time to event data usually focus on the survival or cumulative hazard function. However, mean lifetime is more relevant in many areas of medicine, particularly for organ failure patients. For example, in the U.S., a proposed liver allocation system would rank the patients on the waiting list by difference in 5-year restricted mean lifetime. Chen & Tsiatis (2001) compared the restricted mean lifetime between two treatment groups with Cox proportional hazards models assumed for each treatment group. The survival function for each group is estimated by explicitly averaging over all subjects in the sample. We also propose a difference in restricted mean lifetime in Chapter 3, but using weighted non-parametric estimators. Unlike the method by Chen & Tsiatis (2001), we balance the distribution of confounders through IPTW so that functional forms of the confounding effects do not need to be specified.

The methods proposed in Chapter 2 and Chapter 3 were illustrated through the comparison of peritoneal dialysis and hemodialysis therapy for patients with end-stage renal disease (ESRD). The cumulative treatment effect is examined using each of the above proposed measures.

Censoring times are assumed to be independent of event times given treatment in Chapter 3. When censoring time depends on risk factors for the event, event and censoring times will be correlated through such factors. If these factors are time-dependent and they are not only risk factors for the event but also affected by treatment, standard methods adjusting for these time-dependent factors may produce biased treatment effects. However, if baseline values instead of time-dependent factors are adjusted, dependent censoring may be an issue. Standard hazard regression methods, such as the Cox model, generally assume that event and censoring are

independent given the adjustment covariates.

The inverse probability of censoring weighing (IPCW) method has been proposed to handle the dependent censoring. This method was originally developed by Robins and Rotnitzky (1992) and Robins (1993). A Cox proportional hazards model is assumed for the event time. An inverse probability of censoring weight is applied in the estimating equation for the effect parameters. This weight is the inverse of the survival function for censoring. After applying the inverse censoring weight, the estimators converge to the same measures as in the case of independent censoring. Robins and Finkelstein (2000) applied IPCW to overcome dependent censoring in an AIDS clinical trial. Matsuyama and Yamaguchi (2008) estimated the marginal survival time in the presence of dependent competing risks, using IPCW to handle the dependent censoring. Yoshida, Matsuyama and Ohashi (2007) estimated the treatment effect using a Cox proportional hazards model, again applying IPCW.

In Chapter 4, we extend the methods proposed in Chapter 3 to accommodate dependent censoring. We develop estimators which combine IPTW (to balance the treatment groups with respect to baseline confounders) and IPCW (to handle the dependent censoring induced by time-dependent variates not captured by IPTW). Comparing our works to that of Yoshida, Matsuyama and Ohashi (2007), our methods do not need to assume proportional hazards for the adjustment covariates. Our weighted estimators are proved to converge to Gaussian processes and closed-form covariance function estimators are developed. Our methods proposed in Chapter 4 are applied to the comparison of wait-list survival between race groups (Caucasian vs. African American) for patients with end-stage renal disease.

CHAPTER II

Estimating Cumulative Treatment Effects In The Presence Of Non-proportional Hazards

ABSTRACT: Often in medical studies of time to an event, the treatment effect is not constant over time. In the context of Cox regression modeling, the most frequent solution is to apply a model that assumes the treatment effect is either piece-wise constant or varies smoothly over time; i.e., the Cox non-proportional hazards model. This approach has at least two major limitations. First, it is generally difficult to assess whether the parametric form chosen for the treatment effect is correct. Second, in the presence of non-proportional hazards, investigators are usually more interested in the cumulative than the instantaneous treatment effect (e.g., determining if and when the survival functions cross). Therefore, we propose an estimator for the aggregate treatment effect in the presence of non-proportional hazards. Our estimator is based on the treatment-specific baseline cumulative hazards estimated under a stratified Cox model. No functional form for the non-proportionality need be assumed. Asymptotic properties of the proposed estimators are derived, and the finite-sample properties are assessed in simulation studies. Pointwise and simultaneous confidence bands of the estimator can be computed. The proposed method is applied to data from a national organ failure registry.

KEY WORDS: Confidence bands; Cumulative hazards; Observational studies; Stratification; Survival analysis; Time-dependent effect.

2.1 Introduction

In medical studies featuring survival time data, non-proportional hazards are very common. In Cox (1972) regression modeling, the most frequent solution is to apply a model that assumes that the treatment effect is either piecewise constant or varies smoothly over time. However, it is generally difficult to assess whether the parametric form chosen for the treatment effect is correct. Even if the correct form is chosen, investigators are usually more interested in the cumulative than the instantaneous treatment effect. This is particularly true in settings where the hazard ratio changes direction over time, in which case researchers are often interested in if and when the two survival curves cross. Therefore, we propose an estimator of the cumulative treatment effect under non-proportional hazards. Under our proposed method, the treatment effect is viewed as a process that unfolds over time and is measured by the ratio of cumulative hazards; no functional form need be assumed for the non-proportionality.

The analysis that motivated our research aims to compare survival of end-stage renal disease patients on two dialysis methods: hemodialysis (HD) and peritoneal dialysis (PD). Peritoneal dialysis is less expensive than HD, but newer and hence less established; PD has long been suspected of providing reduced survival relative to HD. The debate over PD versus HD is one of the most contentious issues in medicine and, helping to fuel the debate, previous studies have produced conflicting results (Bloembergen et al., 1995; Fenton et al., 1997). Fenton et al. (1997) compared PD to HD using non-proportional hazards models assuming a piece-wise constant hazard

ratio. The authors found that hazard ratios (PD versus HD) is significantly decreased early in the follow-up period, but that the effect changed direction later on. Since the cumulative effect was not evaluated, one cannot tell which therapy is better in terms of survival based on their results. Applying our method to national registry data, we compare PD and HD covariate-adjusted survival, without assuming proportional hazards. We can estimate the time-dependent cumulative effect of PD relative to HD on mortality without assuming any functional form for that effect. The treatment effect is viewed as a process over time, which is reflected by our inference procedures.

Several methods have been proposed for the comparison of survival or cumulative hazard functions in nonparametric settings. Dabrowska, Doksum and Song (1989) introduced a relative change function involving the survival functions for two populations and constructed pointwise confidence intervals. Simultaneous confidence bands for this function were constructed under a proportional hazards assumption. Parzen, Wei and Ying (1997) constructed simulation-based confidence bands for the difference of survival functions. McKeague and Zhao (2002) derived simultaneous confidence bands for ratios of survival functions based on empirical likelihood. Kalbfleisch and Prentice (1981) estimated an average hazard ratio using a weight function. Since each of the above methods was designed for nonparametric settings, they would be suitable for randomized clinical trials but would generally not apply to observational data where covariate adjustment is required. In the context of covariate adjustment, Schemper (1992) suggested the estimation of average hazard ratio of the two populations through a weighted Cox model. Xu and O'Quigley (2000) estimated the average regression effect through weighted score equation, under a non-proportional hazard model with time-varying regression coefficients.

In this chapter, we propose an estimator based on the treatment-specific baseline

cumulative hazards estimated under a stratified Cox model. The treatment effect is viewed as a process that unfolds over time, and can be related directly to the treatment-specific survival functions. Pointwise confidence intervals and simultaneous confidence bands of our estimator are constructed.

The remainder of this chapter is organized as follows. In the next section, the proposed measure and its estimator are described. We develop the asymptotic properties of the proposed estimator in Section 2.3. Section 2.4 evaluates the applicability of the derived asymptotic results to finite samples through simulation. In Section 2.5, we apply our proposed method to compare survival on hemodialysis and peritoneal dialysis using data from a national organ failure registry. We provide some discussion of the proposed and related methods in Section 2.6.

2.2 Proposed methods

We first set up the notation used throughout the article. Let $J+1$ be the number of treatment groups (numbered $j = 0, 1, \ldots, J$), where the first group (j=0) represents a reference category to which the remaining treatment groups are compared. The total number of subjects is denoted by n. Let T_i be the survival time for subject i. The survival time of a subject is potentially right censored, with censoring time given by C_i. The observation time and observed event indicator are given by $X_i = T_i \wedge C_i$ and $\Delta_i = I(T_i \leq C_i)$, respectively, where $a \wedge b = \min\{a, b\}$ and $I(A)$ is an indicator function taking the value 1 when condition A holds and 0 otherwise. The event counting processes are defined as $N_i(t) = \Delta_i I(X_i \leq t)$. The risk indicators are denoted by $Y_i(t) = I(X_i \geq t)$. Let G_i denote the treatment group for subject i and $G_{ij} = I(G_i = j)$. Correspondingly, we set $Y_{ij}(t) = Y_i(t)G_{ij}$ and $dN_{ij}(t) = dN_i(t)G_{ij}$. The observed data consist of n independent vectors, $(X_i, \Delta_i, G_i, \mathbf{Z}_i)$, where $\mathbf{Z}_i$ is a

vector of adjustment covariates.

We assume that T_i follows a stratified Cox model, with hazard function

$$\lambda_{ij}(t) = \lambda_i(t|G_i = j) = \lambda_{0j}(t)\exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_i\}, \tag{2.1}$$

where $\lambda_{0j}(t)$ is an unspecified treatment-specific baseline hazard function, and $\boldsymbol{\beta}_0$ is an unknown parameter vector. Under (2.1), proportionality of the hazard functions is not assumed to hold across treatment groups, but is assumed with respect to the adjustment covariates. Note that in the set-up we consider, the adjustment covariates vector is treated as time-constant. We revisit the issue of time-dependent covariates in Section 2.6.

The partial likelihood (Cox, 1975) estimator of $\boldsymbol{\beta}_0$ is denoted by $\widehat{\boldsymbol{\beta}}$, and is given by the solution to $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$ where $\mathbf{0}$ is a vector of zeros and

$$\begin{aligned} \mathbf{U}(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\sum_{j=0}^{m}\int_0^{\tau}\left\{\mathbf{Z}_i - \overline{\mathbf{Z}}_j(t,\boldsymbol{\beta})\right\} dN_{ij}(t), \\ \overline{\mathbf{Z}}_j(t,\boldsymbol{\beta}) &= \mathbf{S}_j^{(1)}(t,\boldsymbol{\beta})/S_j^{(0)}(t,\boldsymbol{\beta}), \end{aligned}$$

with $\mathbf{S}_j^{(d)}(t,\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} Y_{ij}(t)\mathbf{Z}_i^{\otimes d}\exp\{\boldsymbol{\beta}^T\mathbf{Z}_i\}$ for $d = 0, 1, 2$, where $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for a vector $\mathbf{a}$. The quantity τ satisfies $P(X_i > \tau) > 0$ and would ordinarily be set to the maximum observation time such that all observed events are included in the analysis.

To compare each treatment group to the reference group, we propose the following measure,

$$\theta_j(t) = \frac{\Lambda_{0j}(t)}{\Lambda_{00}(t)}, \quad \text{for } j = 1, \cdots, J, \tag{2.2}$$

where $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(s)ds$ is the cumulative baseline hazard for treatment group j. Under (2.1), $\theta_j(t)$ can be used as a measure of the aggregate treatment effect across the $(0, t]$ interval.

Let $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s)ds$. Note that, under model (2.1),

$$\frac{\Lambda_{ij}(t|\mathbf{Z}_i = \mathbf{z})}{\Lambda_{i0}(t|\mathbf{Z}_i = \mathbf{z})} = \theta_j(t).$$

That is, contrasting patients who have the same covariate pattern but receive different treatments, the ratio of cumulative hazards and ratio of baseline cumulative hazards are equal. Note also that the proposed cumulative hazard ratio reduces to the hazard ratio if proportionality holds. That is, if proportionality holds across the treatment groups, such that the model $\lambda_{ij}(t) = \lambda_0(t)\exp\{\rho_j + \boldsymbol{\beta}_0^T\mathbf{Z}_i\}$ applies, then

$$\frac{\Lambda_{ij}(t|\mathbf{Z}_i = \mathbf{z})}{\Lambda_{i0}(t|\mathbf{Z}_i = \mathbf{z})} = \exp\{\rho_j\}.$$

In this light, one could view the proposed ratio of cumulative hazards as a generalization of the familiar hazard ratio.

The proposed cumulative effect measure, $\theta_j(t)$, can be estimated by

$$\widehat{\theta}_j(t) = \frac{\widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}})}{\widehat{\Lambda}_{00}(t,\widehat{\boldsymbol{\beta}})}, \quad \text{for } j = 1,\cdots,J,\ t \in [t_L, t_U], \tag{2.3}$$

where t_L is chosen sufficiently large to avoid the situation where $\widehat{\Lambda}_{00}(t_L,\widehat{\boldsymbol{\beta}}) = 0$, while t_U is chosen to avoid well-known instability that exists in the tail of the observation time distribution. The cumulative baseline hazards can be estimated through the Breslow (1972) estimator, $\widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}})$, where

$$\widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{dN_{ij}(s)}{S_j^{(0)}(s,\widehat{\boldsymbol{\beta}})}.$$

In the next section, we derive the asymptotic properties of the proposed estimator.

2.3 Asymptotic properties

To derive the large-sample properties of $\widehat{\theta}_j(t)$, we assume the following regularity conditions for $i = 1,\cdots,n$ and $j = 0,\cdots,J$.

(a) $(X_i, \Delta_i, G_i, \mathbf{Z}_i)$ are independent and identically distributed random vectors.

(b) Z_{ik} have bounded total variation; i.e., $|Z_{ik}| < \kappa$ for all $i = 1, \cdots, n$ and $k = 1, \cdots, p$, where κ is a constant and Z_{ik} is the kth component of Z_i.

(c)$\int_0^\tau \lambda_0(t)dt < \infty$ where τ is a pre-specified time point.

(d) Continuity of the following functions:

$$\mathbf{s}_j^{(1)}(t, \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{s}_j^{(0)}(t, \boldsymbol{\beta}),\ \mathbf{s}_j^{(2)}(t, \boldsymbol{\beta}) = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \mathbf{s}_j^{(0)}(t, \boldsymbol{\beta}),$$

where $\mathbf{s}_j^{(d)}(t, \boldsymbol{\beta})$ is the limiting value of $\mathbf{S}_j^{(d)}(t, \boldsymbol{\beta})$ for $d = 0, 1, 2$, with $\mathbf{s}_j^{(1)}(t, \boldsymbol{\beta})$ and $\mathbf{s}_j^{(2)}(t, \boldsymbol{\beta})$ bounded and $s_j^{(0)}(t, \boldsymbol{\beta})$ bounded away from 0 for $t \in [0, \tau]$ and $\boldsymbol{\beta}$ in an open set.

(e) Positive-definiteness of the matrix $\boldsymbol{\Omega}(\boldsymbol{\beta})$ where

$$\begin{aligned} \boldsymbol{\Omega}(\boldsymbol{\beta}) &= \sum_{j=0}^{m} \int_0^\tau \mathbf{v}_j(t, \boldsymbol{\beta}) s_j^{(0)}(t, \boldsymbol{\beta}) \lambda_{0j}(t) dt, \\ \mathbf{v}_j(t, \boldsymbol{\beta}) &= \mathbf{s}_j^{(2)}(t, \boldsymbol{\beta}) / s_j^{(0)}(t, \boldsymbol{\beta}) - \overline{\mathbf{z}}_j(t, \boldsymbol{\beta})^{\otimes 2}, \end{aligned} \tag{2.4}$$

and $\overline{\mathbf{z}}_j(t, \boldsymbol{\beta}) = \mathbf{s}_j^{(1)}(t, \boldsymbol{\beta}) / s_j^{(0)}(t, \boldsymbol{\beta})$ is the limiting value of $\overline{\mathbf{Z}}_j(t, \boldsymbol{\beta})$.

(f) $P(G_{ij} = 1) > 0$.

The asymptotic behavior of our estimator is summarized by the following two theorems.

THEOREM 1. *Under conditions (a) to (f), $\widehat{\theta}_j(t)$ converges to $\theta_j(t)$ almost surely and uniformly for $t \in [\tau_L, \tau_U]$.*

The consistency of $\widehat{\theta}_j(t)$ follows from the uniform consistency of $\widehat{\Lambda}_{0j}(t, \widehat{\boldsymbol{\beta}}), \widehat{\Lambda}_{00}(t, \widehat{\boldsymbol{\beta}})$, and $\widehat{\boldsymbol{\beta}}$ as well as the Functional Delta Method (Pollard, 1990) and various results from empirical processes theory (Bilias, Gu and Ying, 1997).

THEOREM 2. *Under conditions (a) to (f), $n^{1/2}[\widehat{\theta}_j(t) - \theta_j(t)]$ converges asymptotically to a zero-mean Gaussian process with covariance function $\sigma_j(s, t) = E[\xi_{ij}(s, \boldsymbol{\beta}_0)\xi_{ij}(t, \boldsymbol{\beta}_0)]$,*

where:

$$\xi_{ij}(t,\boldsymbol{\beta}) = \frac{1}{\Lambda_{00}(t)}\Phi_{ij}(t,\boldsymbol{\beta}) - \frac{\Lambda_{0j}(t)}{\Lambda_{00}(t)^2}\Phi_{i0}(t,\boldsymbol{\beta}), \tag{2.5}$$

$$\Phi_{ij}(t,\boldsymbol{\beta}) = \mathbf{h}_j{}^T(t,\boldsymbol{\beta})\boldsymbol{\Omega}(\boldsymbol{\beta})^{-1}\Psi_i(\boldsymbol{\beta}) + \int_0^t s^{(0)}(s,\boldsymbol{\beta})^{-1} dM_{ij}(s,\boldsymbol{\beta}), \tag{2.6}$$

$$\mathbf{h}_j(t,\boldsymbol{\beta}) = -\int_0^t \overline{\mathbf{z}}_j(s,\boldsymbol{\beta}) d\Lambda_{0j}(s), \tag{2.7}$$

$$\Psi_i(\boldsymbol{\beta}) = \sum_{j=0}^{m} \int_0^\tau \{\mathbf{Z}_i - \overline{\mathbf{z}}_j(t,\boldsymbol{\beta})\}\, dM_{ij}(t,\boldsymbol{\beta}), \tag{2.8}$$

$$dM_{ij}(t,\boldsymbol{\beta}) = dN_{ij}(t) - Y_{ij}(t)\exp\{\boldsymbol{\beta}^T\mathbf{Z}_i\} d\Lambda_{0j}(t). \tag{2.9}$$

The covariance function can be consistently estimated by $\widehat{\sigma}_j(s,t,\widehat{\boldsymbol{\beta}})$ where:

$$\widehat{\sigma}_j(s,t,\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\xi}_{ij}(s,\widehat{\boldsymbol{\beta}})\widehat{\xi}_{ij}(t,\widehat{\boldsymbol{\beta}}), \tag{2.10}$$

with $\widehat{\xi}_{ij}(t,\widehat{\boldsymbol{\beta}})$ obtained by replacing all limiting values in $\xi_{ij}(t,\boldsymbol{\beta}_0)$ with their empirical counterparts.

The asymptotic normality of $n^{1/2}[\widehat{\theta}_j(t) - \theta_j(t)]$ can be proved by first writing $\{\widehat{\theta}_j(t) - \theta_j(t)\}$ as

$$\frac{1}{\Lambda_{00}(t)}\left\{\widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}}) - \Lambda_{0j}(t)\right\} + \widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}})\left\{\frac{1}{\widehat{\Lambda}_{00}(t,\widehat{\boldsymbol{\beta}})} - \frac{1}{\Lambda_{00}(t)}\right\}.$$

The quantity $\{\widehat{\Lambda}_{00}(t,\widehat{\boldsymbol{\beta}})^{-1} - \Lambda_{00}(t)^{-1}\}$ can be written as a function of $\{\widehat{\Lambda}_{00}(t,\widehat{\boldsymbol{\beta}}) - \Lambda_{00}(t)\}$ by using the Functional Delta Method. The proof involves decomposing $\{\widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}}) - \Lambda_{0j}(t)\}$ into $\{\widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}}) - \widehat{\Lambda}_{0j}(t,\boldsymbol{\beta}_0)\} + \{\widehat{\Lambda}_{0j}(t,\boldsymbol{\beta}_0) - \Lambda_{0j}(t)\}$. The Central Limit Theorem and various results from the theory of empirical processes are applied in the proof, which is outlined in the Web Appendix A.

Some comments on model misspecification are in order. If model (2.1) is misspecified, Lin and Wei (1989) demonstrated that $\widehat{\boldsymbol{\beta}}$ converges to a vector $\boldsymbol{\beta}^* \neq \boldsymbol{\beta}_0$. Further, if the true model is $\lambda_{ij}(t) = \lambda_{0j}(t)\exp\{\boldsymbol{\beta}_0^T f(\mathbf{Z}_i)\}$, while the assumed model is $\lambda_{ij}(t) = \lambda_{0j}^*(t)\exp\{\boldsymbol{\beta}^T g(\mathbf{Z}_i)\}$, where $f(\mathbf{Z}_i)$ and $g(\mathbf{Z}_i)$ are functions of covariates

$\mathbf{Z}_i$, under a misspecified model, $\widehat{\Lambda}_{0j}(t)$ (Gerds and Schumacher, 2001) converges to $\Lambda^*_{0j}(t) \neq \Lambda_{0j}(t)$. We examine this issue numerically in Section 2.4.

In certain situations, investigators will want to estimate $\theta_j(t)$ at a pre-specified value, $t = t_0$ (e.g., 1 year, 5 years, etc). In these cases, inference could be based on a Wald-type test since $n^{1/2}[\widehat{\theta}_j(t_0) - \theta_j(t_0)]\sigma_j(t_0)^{-1}$ will asymptotically follow a standard normal distribution, with $\sigma_j^2(t) \equiv \sigma_j(t,t)$. However, in many practical applications, it makes more sense to view $\theta_j(t)$ as a process over time, and this view should be captured by the corresponding inference procedures. For instance, in our motivating example, based on analyses reported in the literature, we anticipate that the effect of PD (vs. HD) will depend on time and there is no single specific time point at which we wish to conduct our inference. Lin, Fleming and Wei (1994) proposed a method to construct simultaneous confidence bands for survival curve under the Cox model. We extend this to our estimator. The idea is to approximate the normalized distribution of $\widehat{Q}(t) = n^{1/2}[\widehat{\theta}_j(t) - \theta_j(t)]$ for $t \in [t_L, t_U]$ by a zero-mean Gaussian process $\widetilde{Q}(t) = n^{-1/2}\sum_{i=1}^{n} \widehat{\xi}_{ij}(t, \widehat{\boldsymbol{\beta}})R_i$, where R_i is a standard normal random variable. The distribution of $\widehat{Q}(t)$ is generated through simulation by repeatedly generating independent standard normal random samples $R_i (i = 1, \cdots, n)$. To avoid the resulting lower bound of the band being negative, we consider a log-transformed process, $n^{1/2}\left[\log\{\widehat{\theta}_j(t)\} - \log\{\theta_j(t)\}\right]$, whose distribution can be approximated by $\widehat{Q}(t)/\widehat{\theta}_j(t)$ after applying the Functional Delta Method. In addition, a weight function, $w(t)$, is chosen to adjust the width of the band at different time points. By using weight function, $w(t) = \widehat{\theta}_j(t)/\widehat{\sigma}(t)$, suggested by Nair (1984) and the previously-described simulation method, we may obtain an approximate $100(1-\alpha)\%$ empirical quantile,

$\widehat{q}_\alpha$, satisfying

$$Pr\left\{\sup_{t\in[t_L,t_U]}\left|n^{-1/2}w(t)\widehat{\theta}_j(t)^{-1}\sum_{i=1}^{n}\widehat{\xi}_{ij}(t,\widehat{\boldsymbol{\beta}})R_i\right|>\widehat{q}_\alpha\right\}=\alpha.$$

With the log-transformation, a $100(1-\alpha)\%$ simultaneous confidence band for $\theta_j(t)$ over $[t_L, t_U]$ is given by $\widehat{\theta}_j(t)\exp\left\{\pm n^{-1/2}\widehat{q}_\alpha/w(t)\right\}$.

2.4 Simulation study

The finite sample properties of the proposed estimator were evaluated through a series of simulation studies. For convenience, we consider two treatment groups. Death times were generated as

$$T_i=\left\{-\log(U_i)/\left[\alpha_j\exp\{\boldsymbol{\beta}_0^T\mathbf{Z}_i\}\right]\right\}^{1/\gamma_j},$$

for $i=1,\cdots,n$ and $j=0,1$, where U_i is a Uniform(0,1) random variable, $\boldsymbol{\beta}_0=0.5$, and $\mathbf{Z}_i$ is a bivariate vector with each element following a Bernoulli (0.5) distribution. This set-up implies that T_i follows a Weibull model with hazard function

$$\lambda_{ij}(t)=\lambda_i(t|G_i=j)=\alpha_j\gamma_j t^{\gamma_j-1}\exp\{\beta_0^T Z_i\}.$$

Non-proportionality of the hazard functions for groups 0 and 1 is induced when $\gamma_1\neq\gamma_0$. Various values of γ_j were used to make the hazard ratio constant, decrease, and increase through time. Censoring times were generated from a Uniform$(\tau/2,\tau)$ distribution with $\tau=5$. Different values of α_j were used to vary the percent of censoring (denoted by C%). For each data configuration, the no-censoring setting was also examined. We varied the sample size as n=50, 100, 200, 500, and each data configuration was replicated 1,000 times. We compared the ratio of cumulative hazard to its true value at the $75th$ percentile of the observation time distribution, which we denote by $t_{0.75}$. Results are shown in Table 2.1 and Table 2.2.

The proposed estimator generally performs well in finite samples, n=100, 200, 500 (Table 2.1). Even in the presence of a very high proportion of censoring, the empirical mean of $\widehat{\theta}_1(t)$ is approximately unbiased for sample sizes of n=500 and n=200, and almost all simulations with size of n=100. In general, the bias is reduced as the number of subjects in each treatment group increases. The average asymptotic standard error (ASE) is generally close to the empirical standard deviation (ESD), while the empirical coverage probabilities (CP) are consistent with the nominal value of 0.95.

For smaller sample sizes (e.g., $n = 50$), the bias of $\widehat{\theta}_1$ is relatively large and the coverage probabilities are notably lower than the nominal value of 0.95 (Table 2.2). However, if $\log \widehat{\theta}_1(t_{0.75})$ is considered, the bias reduces dramatically and the coverage probability is quite good. Therefore, when the sample size is very small (e.g., $n = 50$), inference should be based on $\log \theta_1(t)$.

We also looked at the performance of our estimator at various percentiles of the observation time distribution. The scenario where the hazard ratio increases with time and sample size n=500 are considered. We find that our estimator is approximately unbiased even at the $10th$ and $90th$ percentiles of the observation distribution and that the coverage probabilities are close to the nominal value of 0.95 (Table 2.3).

We explored the performance of our estimator and its variance under models with functional misspecification and incorrect covariate adjustment (results not shown). We find that under a misspecified model, the proposed estimator is biased, although the bias is relatively small; as well, the ASE is close to the ESD.

2.5 Analysis of dialysis data

We applied our proposed methods to compare patient survival on hemodialysis (HD) and peritoneal dialysis (PD). Hemodialysis served as the reference category (j=0), while PD was labeled as j=1. Hence, the parameter of interest is $\theta_1(t)$, which contrasts the cumulative hazard for PD relative to HD. Data were obtained from the Canadian Organ Replacement Register (CORR), a nation-wide and population-based organ failure registry which is maintained by the Canadian Institute for Health Information. The mortality hazard on dialysis was investigated for End Stage Renal Disease (ESRD) patients who were either on HD or PD at the time of renal replacement therapy initiation. The dialysis method is inherently time-dependent since a patient may switch therapies. We carried out two separate analyses. The first analysis, in the spirit of an intent-to-treat (ITT) analysis, classified patients based on first method of dialysis; that is, the type of dialysis received at the initiation of renal replacement therapy. The ITT analysis compares the risk of death between patients initially placed on PD (vs. HD) knowing that patients may switch therapies. The second analysis censored the follow-up time at the first dialysis therapy switch (CAFS). The CAFS analysis compares the risk of death for patients who stay on PD to patients who remain on HD.

The study population included n=23,254 registered patients aged 18 and above who initiated dialysis between 1990 and 1998. Patients began follow-up at the date of dialysis initiation and were followed until the earliest of death, loss to follow-up, kidney transplantation or the end of the observation period (December 31, 1998). For the ITT analysis, approximately 38% of HD patients (n_0=17,766) were observed to die, while 36% of patients on PD (n_1=5,488) died. For the CAFS analysis, the

proportion observed to die for patients on HD was approximately 30% and 25% for patients on PD. Approximately 17% of patients initially placed on HD and 27% of patients initially placed on PD switched therapy at least once.

Cox regression was employed, stratified by dialysis modality, and adjusting for age, gender, race, underlying renal diagnosis, region and various comorbid illnesses (cerebrovascular accident, cardiovascular disease, chronic obstructive pulmonary disease (COPD), malignancy, peripheral vascular disease, other illnesses). Through stratification, a distinct baseline mortality hazard is allowed for HD and PD, which allows the effect of dialysis method to be non-proportional and assumes no specific functional form for the non-proportionality. The resulting 95% pointwise and simultaneous confidence bands of $\theta_1(t)$ in time interval $[1, 90]$ months are given in Figure 2.1 for the ITT analysis and Figure 2.2 for the CAFS analysis. This time interval is chosen to avoid imprecision at the beginning of follow-up due to too few deaths occurring in the HD (reference) group, and instability at the tail of the observation time distribution. Based on the ITT analysis (treating dialysis method as fixed at t=0), relative to HD, patients initially placed on PD had significantly increased covariate-adjusted survival probability over the [1, 29] months interval with $\widehat{\theta}_1(t)$ ranging from a low of $\widehat{\theta}_1(t) = 0.33$ at $t = 1$ month, to a high of $\widehat{\theta}_1(t) = 0.90$ at $t = 29$ months. Survival was not significantly different for patients on PD relative to HD during the (29, 80] months interval. Long-term survival was significantly reduced for patients on PD after approximately 80 months with $\widehat{\theta}_1(t) \geq 1.17$. For the CAFS analysis (censored at first therapy switch), survival probability is higher for patients on PD than HD for approximately the first 31 months, while the survival was not significantly different after that point (Figure 2.2).

Comparing the ITT and CAFS analyses, as is evident from Figures 2.1-2.2, the

ITT analysis is more precise since deaths following therapy switches are not censored. In the short term, PD patients have significantly better survival under either analysis. In the long run, PD survival is not significantly different from that of HD under the CAFS analysis, but significantly lower under the ITT analysis. Supplementary analysis revealed that both $\widehat{\Lambda}_0(t)$ and $\widehat{\Lambda}_1(t)$ were greater for the ITT than the CAFS analysis (results not shown), implying that switching therapies (in either direction) is associated with increased mortality hazard. Since PD patients were more likely than HD patients to switch, it would make sense that PD would be viewed more favorably under a CAFS (relative to ITT) approach.

2.6 Discussion

In the survival analysis of biomedical studies, non-proportional hazards are frequently encountered. In this manuscript, we introduce a measure of the cumulative treatment effects when the proportional hazards assumption does not hold across the treatment groups. No functional form for the non-proportionality need be assumed for our proposed estimator. In cases where hazards are in fact proportional, the proposed measure reduces to the well-known hazard ratio. Simulation studies provide evidence that the proposed estimator is approximately unbiased, while the estimated standard errors are quite accurate. Applying our method to CORR dialysis data, we found that long-term survival (after approximately 80 months) is significantly reduced for patients initially placed on PD relative to HD (intent-to-treat analysis). The difference in long-term survival is non-significant after approximately the first 31 months based on the analysis with censoring at first therapy switch.

Since dialysis modality was not randomized, our results must be interpreted with caution. We did find that patients who were initially put on PD are healthier than

those were put on HD in terms of comorbidity profile. This does imply that selection bias due to unmeasured covariates may be an issue.

Various methods previously proposed to account for non-proportional hazards in a Cox regression model have featured a time-varying regression coefficient, $\boldsymbol{\beta}(t)$ (e.g., Sleeper and Harrington, 1990; Zucker and Karr, 1990; Murphy and Sen, 1991; Sargent, 1997; Gustafson, 1998; Xu and O'Quigley, 2000; Martinussen, Scheike and Skovgard, 2002; Scheike and Martinussen, 2004). A limitation of these and related approaches is that the estimator represents an instantaneous metric and, in the presence of non-proportional hazards, investigators are usually more interested in the cumulative than the instantaneous effect. The quantity $\int_0^t \boldsymbol{\beta}(s)ds$ is often proposed to estimate the cumulative effect. Despite its utility, a drawback of this approach is that the integral cannot generally be connected back to the treatment-specific cumulative hazard and hence survival functions. For example, in comparing treatment (G_i=1) and placebo (G_i=0), $\int_0^t \boldsymbol{\beta}(s)ds = 0$ generally will not imply $S_0(t) = S_1(t)$ and usually it would not be straightforward without further assumptions to determine b_0 such that $\int_0^t \boldsymbol{\beta}(s)ds = b_0$ implies equal survival. Our proposed approach does not consider the estimation of the instantaneous treatment effect, but proposes a direct measure for the cumulative effect. In terms of the survival function, equal survival at time t among the treatments being compared is implied by $\theta_j(t) = 1$. In the situation where researchers are interested in whether and when two survival curves cross, our method is preferable. In addition, an advantage of the method proposed in this manuscript is that it is computationally straightforward.

We derived the variance for the proposed estimator using the modern theory of empirical processes, instead of the Martingale Central Limit Theorem (Fleming and Harrington, 1991). Although the asymptotic results are easier to derive using

Martingale theory, the sandwich-type asymptotic variance derived through empirical processes should be more robust to model misspecification, such as missing covariate information, covariate measurement error and mis-modeling of adjustment covariates. In addition, the proposed variance could be easily extended to recurrent event setting, wherein the event of interest can be experienced more than once per subject. When proportionality does not hold across the treatment groups, we could fit a stratified version of the proportional means model (Lin et al., 2000), $E[N_{ij}(t)] \equiv \mu_{ij}(t) = \mu_{0j}(t)\exp\{\boldsymbol{\beta}_0^T\mathbf{Z}_i\}$, for $i = 1, \cdots, n$, where $\mu_{0j}(t)$ is unspecified baseline mean function for the jth treatment group. Among the methods available for recurrent event data (e.g., see Cai and Schaubel, 2004), the marginal means approach of Lin et al. (2000) would be considered a suitable method for comparing treatments. To compare treatment group $j(>0)$ to the reference group (j=0), one could use the ratio of the mean numbers of events, $\theta_j^*(t) = \mu_{0j}(t)/\mu_{00}(t)$ as a metric for the cumulative treatment effect. The estimate for $\theta_j^*(t)$ has same expression as in the univariate survival case, but with $N_{ij}(t)$ representing the number of events in $(0, t]$ instead of a time-dependent observed death indicator. The asymptotic results would be essentially the same after adding the condition that $N_{ij}(t) < \eta < \infty$. The asymptotic variance of $\widehat{\theta}_j^*(t)$ could be consistently estimated by that based on Theorem 2 of the current report, upon replacing $\Lambda_{0j}(t)$ with $\mu_{0j}(t)$.

In this chapter, our focus has been on the treatment effect. When the proportional hazard assumption does not hold for an adjustment covariate, traditional methods can be applied to remedy the non-proportionality; e.g., interactions with t.

Note that our proposed estimation procedure considers the case where the adjustment covariate vector is assumed to be time-independent. This is not a limitation for at least two important reasons. First, the assumption of time-independent adjust-

ment covariates matches the reality in most cases, such as the application in Section 5. Second, in settings where $\mathbf{Z}_i(t) \neq \mathbf{Z}_i$, it would be preferable to use $\mathbf{Z}_i = \mathbf{Z}_i(0)$ (i.e., the baseline covariate value) anyway, due to interpretation issues. For ease of illustration, suppose treatment is fixed at $t = 0$ but that the adjustment covariate, $Z_i(t)$, varies over time; $Z_i(0)$, as opposed to $Z_i(t)$, is included as an adjustment covariate. Consider two cases: (i) $Z_i(t)$ is uncorrelated with treatment (ii) $Z_i(t)$ is correlated with treatment. In case (i), $\widehat{\theta}_1(t)$ would be estimating the same quantity whether or not the adjustment covariate was coded as time-dependent, rendering the use of $Z_i(t)$ (in place of Z_i) unnecessary. In case (ii), $\widehat{\theta}_1(t)$ could be substantially biased towards 1 if the adjustment covariate was coded as time-dependent in the model. If $Z_i(t)$ is correlated with treatment after adjusting for $Z_i(0)$, it is much more likely that treatment is at least in part causing the variation in $Z_i(t)$, directly or indirectly, than the other way around; i.e., considering the temporality. Take the dialysis data in Section 5 as an example. We adjust for comorbid conditions, which are coded at time t=0. In the CORR database, serial comorbidity data are not available. But, even if they were, we would prefer to compare PD and HD only adjusting for time 0 comorbidity. It is quite plausible that, in addition to affecting the mortality hazard, dialysis method has other intermediate consequences relating to (for example) hospitalizations and the incidence of comorbid conditions. Suppose that PD (relative to HD) reduces mortality and decreases the incidence of cardiovascular disease (CVD), and that CVD onset increases mortality risk. If we adjust for time-dependent CVD, then we end up, essentially, comparing PD and HD patients of similar prognosis, therefore underestimating the magnitude of the difference between therapies with respect to mortality. In understanding this phenomenon, it helps to think of time-dependent covariates as intermediate end-points. It is well known in

survival analysis that adjusting for components of the causal pathway is inappropriate; as is made clear in survival-related causal inference approaches; e.g., Robins and Greenland (1994), Hernan, Brumback and Robins (2001), who proposed marginal structural models for use when adjustment covariates are time-dependent. If time-dependent comorbidity data were available, they could perhaps be incorporated by a marginal structural-type extension of the methods proposed in this article.

If proportionality holds across the treatment groups, such that the model $\lambda_{ij}(t) = \lambda_0(t)\exp\{\rho_j + \boldsymbol{\beta}_0^T\mathbf{Z}_i\}$ applies, then our proposed measure $\theta_j(t)$ equals $\exp\{\rho_j\}$ for $t \in [t_L, t_U]$. To test for proportionality, we have $H_0 : \theta_j(t) = \exp\{\rho_j\}$ and $H_1 : \theta_j(t) \neq \exp\{\rho_j\}$. Similar to the proof of normality of $n^{1/2}\left[\widehat{\theta}_j(t) - \theta_j(t)\right]$, we can obtain that $n^{1/2}\left[\exp(\widehat{\rho}_j) - \exp(\rho_j)\right] = n^{-1/2}\sum_{i=1}^n \Psi_{ij} + o_p(1)$, where Ψ_{ij} is an mean 0 variate. Therefore, $n^{1/2}\left\{[\widehat{\theta}_j(t) - \theta_j(t)] - [\exp(\widehat{\rho}_j) - \exp(\rho_j)]\right\} = n^{-1/2}\sum_{i=1}^n [\xi_{ij}(t, \boldsymbol{\beta}) - \Psi_{ij}]$ is asymptotically a Gaussian process with mean 0. The quantity $n^{-1/2}R_i\,[\xi_{ij}(t, \boldsymbol{\beta}) - \Psi_{ij}]$ would have the same distribution as $n^{1/2}[\widehat{\theta}_j(t) - \exp(\widehat{\rho}_j)]$ under H_0, where R_i, $i = 1, \cdots, n$, follows a standard normal distribution. Therefore, using techniques similar to those used to derive the confidence band, we can obtain the distribution of $\sup_{t\in[t_L,t_U]} n^{1/2}\left|\widehat{\theta}_j(t) - \exp(\widehat{\rho}_j)\right|$ under H_0. A test for the constancy of the treatment effect could then be based on this statistic.

2.7 Tables and Figures

n	γ_0	γ_1	α_0	α_1	C%	$\theta_1(t_{0.75})$	BIAS	ASE	ESD	CP
500	1.4	1.2	0.4	0.35	0%	0.739	0.002	0.084	0.087	0.94
200							0.008	0.132	0.139	0.94
100							0.020	0.188	0.206	0.92
500	1.4	1.2	0.4	0.35	10%	0.741	0.006	0.084	0.084	0.95
200							0.014	0.134	0.132	0.96
100							0.030	0.191	0.199	0.94
500	1.4	1.2	0.1	0.07	0%	0.468	0.005	0.055	0.058	0.93
200							-0.001	0.085	0.088	0.93
100							0.011	0.121	0.125	0.93
500	1.4	1.2	0.1	0.07	54%	0.536	0.001	0.081	0.077	0.95
200							0.014	0.130	0.133	0.94
100							0.017	0.184	0.188	0.93
500	1	1.5	0.5	0.3	0%	0.926	0.004	0.104	0.109	0.94
200							0.012	0.165	0.170	0.93
100							0.029	0.235	0.238	0.94
500	1	1.5	0.5	0.3	10%	0.912	0.007	0.103	0.103	0.95
200							0.019	0.165	0.170	0.94
100							0.015	0.230	0.238	0.92
500	1	1.5	0.2	0.1	0%	1.123	0.014	0.128	0.131	0.95
200							0.023	0.203	0.214	0.94
100							0.042	0.289	0.308	0.94
500	1	1.5	0.2	0.1	40%	0.933	0.006	0.121	0.118	0.95
200							0.014	0.192	0.197	0.94
100							0.016	0.272	0.287	0.93
500	1.5	1.5	0.4	0.2	0%	0.500	0.003	0.058	0.058	0.95
200							0.004	0.091	0.093	0.95
100							0.018	0.131	0.138	0.94
500	1.5	1.5	0.4	0.2	10%	0.500	0.003	0.058	0.056	0.96
200							0.002	0.091	0.091	0.94
100							0.018	0.131	0.138	0.94
500	1.5	1.5	0.1	0.05	0%	0.500	0.002	0.058	0.057	0.95
200							0.010	0.092	0.092	0.94
100							0.016	0.130	0.134	0.93
500	1.5	1.5	0.1	0.05	52%	0.500	0.009	0.075	0.076	0.94
200							0.014	0.120	0.121	0.94
100							0.014	0.169	0.177	0.93

Table 2.1: Simulation results: Examination of bias and accuracy of estimated standard error

	γ_0	γ_1	α_0	α_1	C%	$\theta_1(t_{0.75})$	BIAS	ASE	ESD	CP
	1.4	1.2	0.4	0.35	0%	0.739	0.042	0.266	0.294	0.91
					10%	0.741	0.034	0.265	0.292	0.92
	1.4	1.2	0.1	0.07	0%	0.468	0.023	0.169	0.182	0.91
					54%	0.536	0.048	0.269	0.300	0.90
	1.0	1.5	0.5	0.3	0%	0.926	0.048	0.331	0.347	0.92
					10%	0.912	0.068	0.334	0.381	0.91
$\widehat{\theta}_1$	1.0	1.5	0.2	0.1	0%	1.123	0.086	0.415	0.480	0.92
					40%	0.933	0.083	0.403	0.461	0.90
	1.5	1.5	0.4	0.2	0%	0.500	0.028	0.182	0.205	0.92
					10%	0.500	0.031	0.183	0.202	0.92
	1.5	1.5	0.1	0.05	0%	0.500	0.027	0.182	0.199	0.91
					52%	0.500	0.038	0.244	0.271	0.90
	γ_0	γ_1	α_0	α_1	C%	$\log\theta_1(t_{0.75})$	BIAS	ASE	ESD	CP
	1.4	1.2	0.4	0.35	0%	-0.303	-0.010	0.342	0.364	0.94
					10%	-0.300	-0.020	0.342	0.365	0.94
	1.4	1.2	0.1	0.07	0%	-0.759	-0.021	0.347	0.376	0.93
					54%	-0.624	-0.034	0.470	0.502	0.95
	1.0	1.5	0.5	0.3	0%	-0.077	-0.009	0.339	0.352	0.94
					10%	-0.092	0.001	0.341	0.381	0.93
$\log(\widehat{\theta}_1)$	1.0	1.5	0.2	0.1	0%	0.117	0.003	0.342	0.373	0.92
					40%	-0.069	-0.008	0.398	0.436	0.94
	1.5	1.5	0.4	0.2	0%	-0.693	-0.018	0.348	0.387	0.93
					10%	-0.693	-0.009	0.347	0.379	0.93
	1.5	1.5	0.1	0.05	0%	-0.693	-0.016	0.347	0.377	0.92
					52%	-0.693	-0.042	0.461	0.488	0.94

Table 2.2: Simulation results: Examination of bias and accuracy of estimated standard error for small sample size (n=50)

γ_0	γ_1	α_0	α_1	C%	t	$\theta_1(t)$	BIAS	ASE	ESD	CP
1	1.5	0.5	0.3	0%	$t_{0.10}$	0.287	0.013	0.104	0.110	0.93
					$t_{0.25}$	0.449	0.003	0.089	0.091	0.94
					$t_{0.50}$	0.665	0.006	0.089	0.087	0.95
					$t_{0.75}$	0.926	0.004	0.104	0.109	0.94
					$t_{0.90}$	1.130	0.016	0.137	0.141	0.94
1	1.5	0.5	0.3	10%	$t_{0.10}$	0.306	0.005	0.100	0.102	0.94
					$t_{0.25}$	0.469	0.004	0.089	0.086	0.95
					$t_{0.50}$	0.671	0.004	0.089	0.091	0.94
					$t_{0.75}$	0.912	0.007	0.103	0.103	0.95
					$t_{0.90}$	1.060	0.003	0.126	0.133	0.94
1	1.5	0.2	0.1	0%	$t_{0.10}$	0.387	0.011	0.123	0.129	0.94
					$t_{0.25}$	0.610	0.011	0.110	0.119	0.94
					$t_{0.50}$	0.869	0.009	0.111	0.114	0.94
					$t_{0.75}$	1.123	0.014	0.128	0.131	0.95
					$t_{0.90}$	1.440	0.004	0.178	0.182	0.94
1	1.5	0.2	0.1	40%	$t_{0.10}$	0.397	0.005	0.120	0.123	0.92
					$t_{0.25}$	0.594	0.003	0.109	0.114	0.94
					$t_{0.50}$	0.819	0.005	0.110	0.106	0.95
					$t_{0.75}$	0.933	0.006	0.121	0.118	0.95
					$t_{0.90}$	1.040	0.020	0.153	0.156	0.94

Note: $t_q = q'th$ percentile of observation time distribution, C%=percent censored, ASE=average asymptotic standard error, ESD=empirical standard deviation, CP=coverage probability.

Table 2.3: Simulation results: Examination of bias and accuracy of estimated standard error at different time points (n=500)

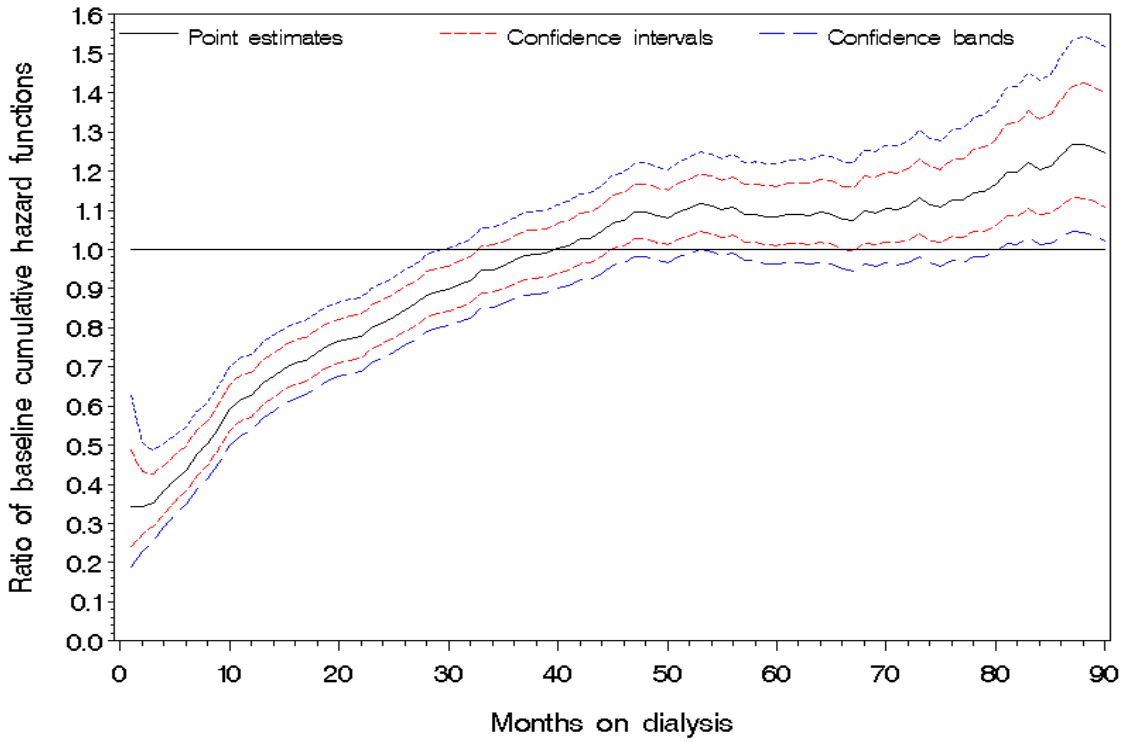

Figure 2.1: Analysis of dialysis data: Estimator and 95% pointwise confidence intervals and simultaneous confidence bands for the ratio of cumulative hazard functions (PD/HD), $\theta_1(t)$, for the ITT analysis.

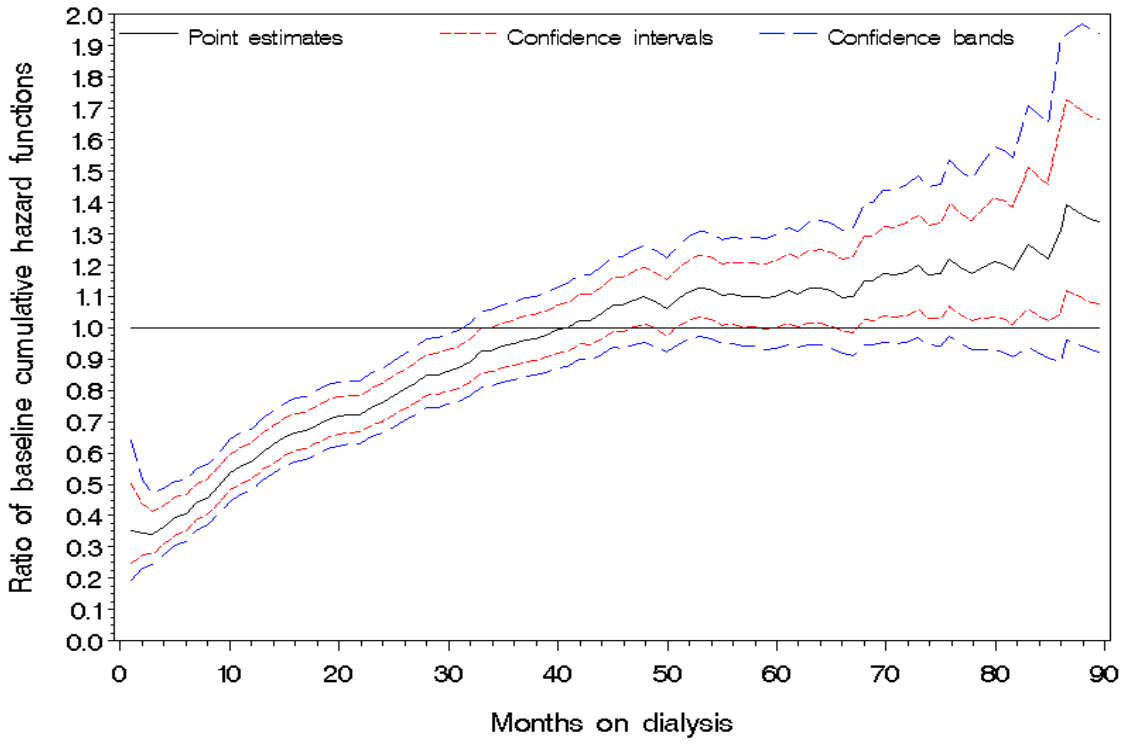

Figure 2.2: Analysis of dialysis data: Estimator and 95% pointwise confidence intervals and simultaneous confidence bands for the ratio of cumulative hazard functions (PD/HD), $\theta_1(t)$, for the CAFS analysis.

2.8 Appendix

Proof of Theorem 1

One can write:

$$n^{1/2}[\widehat{\theta}_j(t) - \theta_j(t)] = n^{1/2}\left[\frac{1}{\Lambda_{00}(t)}\left\{\widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}}) - \Lambda_{0j}(t)\right\} + \widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}})\left\{\frac{1}{\widehat{\Lambda}_{00}(t,\widehat{\boldsymbol{\beta}})} - \frac{1}{\Lambda_{00}(t)}\right\}\right].$$

By a Taylor series expansion,

$$\frac{1}{\widehat{\Lambda}_{00}(t,\widehat{\boldsymbol{\beta}})} - \frac{1}{\Lambda_{00}(t)} = -\frac{1}{\Lambda_{00}^2(t)}\left\{\widehat{\Lambda}_{00}(t,\widehat{\boldsymbol{\beta}}) - \Lambda_{00}(t)\right\} + o(n^{-1/2}).$$

For $j = 0, \cdots, m$, set $\widehat{\phi}_j(t) = \widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}}) - \Lambda_{0j}(t) = \widehat{\phi}_{j1}(t) + \widehat{\phi}_{j2}(t)$, where:

$$\begin{aligned}\widehat{\phi}_{j1}(t) &= \widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}}) - \widehat{\Lambda}_{0j}(t,\boldsymbol{\beta}_0)\\ \widehat{\phi}_{j2}(t) &= \widehat{\Lambda}_{0j}(t,\boldsymbol{\beta}_0) - \Lambda_{0j}(t).\end{aligned}$$

By the triangle inequality,

$$\sup_{t\in[0,\tau]}|\widehat{\phi}_j(t)| \leq \sup_{t\in[0,\tau]}|\widehat{\phi}_{j1}(t)| + \sup_{t\in[0,\tau]}|\widehat{\phi}_{j2}(t)|. \tag{2.11}$$

We can write

$$\widehat{\phi}_{j1}(t) = n^{-1}\sum_{i=1}^{n}\int_0^t \left\{S_j^{(0)}(s,\widehat{\boldsymbol{\beta}})^{-1} - S_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1}\right\} dN_{ij}(s). \tag{2.12}$$

Through another Taylor expansion,

$$S_j^{(0)}(s,\widehat{\boldsymbol{\beta}})^{-1} - S_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1} = -\frac{\overline{\mathbf{Z}}_j(s,\boldsymbol{\beta}_0)^T}{S_j^{(0)}(s,\boldsymbol{\beta}_0)}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o(n^{-1/2}).$$

Substituting this expression back into (2.12), one can obtain $\widehat{\phi}_{j1}(t) = \widehat{\mathbf{h}}_j(t,\boldsymbol{\beta}_0)^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o(n^{-1/2})$, where:

$$\widehat{\mathbf{h}}_j(t,\boldsymbol{\beta}) = -n^{-1}\sum_{i=1}^{n}\int_0^t \frac{\overline{\mathbf{Z}}_j(s,\boldsymbol{\beta})}{S_j^{(0)}(s,\boldsymbol{\beta})} dN_{ij}(s).$$

Using the facts that $N_{ij}(s)$, $S_j^{(1)}(s,\boldsymbol{\beta})$ are bounded, and $S_j^{(0)}(s,\boldsymbol{\beta})$ is bounded away from 0 as $n \to \infty$, $\widehat{\mathbf{h}}_j(t,\boldsymbol{\beta}_0)$ is bounded for sufficiently large n. This, combining with the fact that $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$ (Andersen and Gill, 1982) as $n \to \infty$, gives

$$\sup_{t\in[0,\tau]} |\widehat{\phi}_{j1}(t)| \xrightarrow{a.s.} 0. \tag{2.13}$$

The quantity $\widehat{\phi}_{j2}(t)$ can be written as

$$\widehat{\phi}_{j2}(t) = n^{-1}\sum_{i=1}^{n}\int_0^t S_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1} dM_{ij}(s,\boldsymbol{\beta}_0). \tag{2.14}$$

Since $S_j^{(0)}(s,\boldsymbol{\beta}_0)$ is bounded away from 0, and $n^{-1}\sum_{i=1}^n M_{ij}(t,\boldsymbol{\beta}_0) \xrightarrow{a.s.} 0$ as $n \to \infty$ for $t \in [0,\tau]$ by the Strong Law of Large Numbers (SLLN), we have

$$\sup_{t\in[0,\tau]} |\widehat{\phi}_{j2}(t)| \xrightarrow{a.s.} 0. \tag{2.15}$$

Combining (2.11), (2.13) and (2.15), give the consistency of $\widehat{\phi}_j(t)$ for $j = 0, \cdots, m$ and $t \in [0,\tau]$. This, combining with the fact that $\widehat{\Lambda}_{0j}(t,\widehat{\boldsymbol{\beta}}) \xrightarrow{a.s.} \Lambda_{0j}(t)$ (Anderson and Gill, 1982) concludes the proof of the uniform consistency of $\widehat{\theta}_j(t)$.

Proof of Theorem 2

We now consider the convergence of $n^{1/2}[\widehat{\theta}_j(t) - \theta_j(t)]$. One can write

$$\begin{aligned} n^{1/2}\widehat{\phi}_{j1}(t) &= n^{1/2}\widehat{\mathbf{h}}_j(t,\boldsymbol{\beta}_0)^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1) \\ n^{1/2}\widehat{\phi}_{j2}(t) &= n^{-1/2}\sum_{i=1}^{n}\int_0^t S_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1} dM_{ij}(s,\boldsymbol{\beta}_0). \end{aligned}$$

Conditions (a) to (f) and the SLLN give $\widehat{\mathbf{h}}_j(t,\boldsymbol{\beta}_0) \xrightarrow{a.s.} \mathbf{h}_j(t,\boldsymbol{\beta}_0)$ as $n \to \infty$, where $\mathbf{h}_j(t,\boldsymbol{\beta}_0)$ is defined as in (2.7). The partial likelihood score equation can be written

as

$$\begin{aligned}\mathbf{U}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \widehat{\Psi}_i(\boldsymbol{\beta}),\\ \widehat{\Psi}_i(\boldsymbol{\beta}) &= \sum_{j=0}^{m} \int_0^{\tau} \left\{ Z_i - \overline{Z}_j(t,\boldsymbol{\beta}) \right\} dM_{ij}(t,\boldsymbol{\beta}).\end{aligned}$$

Since $\overline{\mathbf{Z}}_j(t,\boldsymbol{\beta}) \xrightarrow{a.s.} \overline{\mathbf{z}}_j(t,\boldsymbol{\beta})$ as $n \to \infty$, $\widehat{\Psi}_i(\boldsymbol{\beta})$ converges to $\Psi_i(\boldsymbol{\beta})$ defined as in (2.8). By a Taylor series expansion, one can write

$$\begin{aligned}\mathbf{U}(\boldsymbol{\beta}) &= n\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o(n^{-1/2}),\\ \widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta}) &= n^{-1} \sum_{i=1}^{n} \sum_{j=0}^{m} \int_0^{\tau} [\mathbf{S}_j^{(2)}(s,\boldsymbol{\beta})/S_j^{(0)}(s,\boldsymbol{\beta}) - \overline{\mathbf{Z}}_j(s,\boldsymbol{\beta})^{\otimes 2}] dN_{ij}(s).\end{aligned}$$

As $n \to \infty$, $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})$ converges to $\boldsymbol{\Omega}(\boldsymbol{\beta})$, a positive definite matrix defined as in (2.4). These results yield

$$n^{1/2}\widehat{\phi}_{j1}(t) = \mathbf{h}_j(t,\boldsymbol{\beta}_0)^T \boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1} n^{-1/2} \sum_{i=1}^{n} \Psi_i(\boldsymbol{\beta}_0) + o_p(1). \tag{2.16}$$

With respect to $\widehat{\phi}_{j2}(t)$, using the fact that $S_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1}$ converges to $s_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1}$ as $n \to \infty$ and various results from empirical processes (Bilias, Gu and Ying, 1997; Lin et al., 2000), it can be shown that as $n \to \infty$,

$$n^{-1/2} \sum_{i=1}^{n} \int_0^{t} \left\{ S_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1} - s_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1} \right\} dM_{ij}(s,\boldsymbol{\beta}_0) \xrightarrow{a.s.} 0 \text{ for } t \in [0,\tau].$$

Therefore, we have

$$n^{1/2}\widehat{\phi}_{j2}(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^{t} s_j^{(0)}(s,\boldsymbol{\beta}_0)^{-1} dM_{ij}(s,\boldsymbol{\beta}_0) + o_p(1). \tag{2.17}$$

Combining (2.16) and (2.17), one can write $n^{1/2}\widehat{\phi}_j(t) = n^{-1/2} \sum_{i=1}^{n} \Phi_{ij}(t,\boldsymbol{\beta}_0) + o_p(1)$ for $j = 0, \cdots, m$, with $\Phi_{ij}(t,\boldsymbol{\beta}_0)$ defined as in (2.6). This, in conjunction with the fact that $\widehat{\Lambda}_{0j}(t) \xrightarrow{a.s.} \Lambda_{0j}(t)$, gives

$$n^{1/2}[\widehat{\theta}_j(t) - \theta_j(t)] = n^{-1/2} \sum_{i=1}^{n} \xi_{ij}(t,\boldsymbol{\beta}_0),$$

with $\xi_{ij}(t, \boldsymbol{\beta}_0)$ defined as in (2.5), which is a sum of n independent and identically distributed random variables. By the Central Limit Theorem, $n^{1/2}[\widehat{\theta}_j(t) - \theta_j(t)]$ converges to a mean-zero normal distribution. To prove the weak convergence, we show tightness of $n^{1/2}[\widehat{\theta}_j(t) - \theta_j(t)]$, which can be demonstrated by the manageability of $\xi_{ij}(t, \boldsymbol{\beta}_0)$ for $i = 1, \cdots, n$. Since $\Phi_{ij}(t, \boldsymbol{\beta}_0)$ and $\Phi_{i0}(t, \boldsymbol{\beta}_0)$ are differences of functions monotone in t, $\Phi_{ij}(t, \boldsymbol{\beta}_0)$ and $\Phi_{i0}(t, \boldsymbol{\beta}_0)$ are manageable. Hence, $\xi_{ij}(t, \boldsymbol{\beta}_0) = \Lambda_{00}(t)^{-1}\Phi_{ij}(t, \boldsymbol{\beta}_0) - \Lambda_{00}(t)^{-2}\Lambda_{0j}(t)\Phi_{i0}(t, \boldsymbol{\beta}_0)$ is manageable (Bilias et al., 1997; Pollard, 1990).

2.9 References

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

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**, 662-682.

Bloembergen, W.E., Port, F.K., Mauger, E.A., and Wolfe, R.A. (1995). A comparison of mortality between patients treated with hemodialysis and peritoneal dialysis. *Journal of American Society of Nephrology* **6**, 177-183.

Breslow, N. (1972). Contribution to the discussion of the paper by D.R. Cox. *Journal of the Royal Statistical Society, Series B* **34**, 187-220.

Cai, J. and Schaubel, D.E. (2004). Analysis of Recurrent Event Data. *Handbook of Statistics* **23**, 603-623.

Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34**, 187-202.

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

Dabrowska, D.M., Doksum, K.A. and Song, J. (1989). Graphical comparison of cumulative hazards for two populations. *Biometrika* **76**, 763-773.

Fenton, S.S., 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**, 334-342.

Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Gerds, T.A. and Schumacher, M. (2001). On functional misspecification of covariates in the Cox regression model. *Biometrika* **88**, 572-580.

Gustafson, P. (1998). Flexible bayesian modeling for survival data. *Lifetime Data Analysis* **4**, 281-299.

Hernan, M.A., Brumback, B. and Robins, J.M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* **96**, 440-448.

Kalbfleisch, J.D. and Prentice, R.L. (1981). Estimation of the average hazard ratio. *Biometrika* **68**, 105-112.

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

Lin, D.Y. and Wei, L.J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84**, 1074-1078.

Lin, D.Y., Fleming, T.R. and Wei, L.J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73-81.

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

Martinussen, T., Scheike, T.H. and Skovgaard, I.M. (2002). Efficient Estimation of Fixed and Time-varying covariate effects in multiplicative intensity models. *Scandi-*

navian Journal of Statistics **29** 57-74.

McKeague, I.W. and Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statistics & Probability Letters* **60**, 405-415.

Murphy, S.A. and Sen, P.K. (1991). Time dependent coefficients in a Cox-type regression model. *Stochastic Processes and Their Applications* **39**, 153-180.

Nair, V.N. (1984). Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26**, 265-275.

Parzen, M.I., Wei, L.J. and Ying, Z. (1997). Simultaneous confidence intervals for the difference of the two survival functions. *Scandinavian Journal of Statistics* **24**, 309-314.

Pollard, D. (1990). *Empirical processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward: Institute of Mathematical Statistics.

Robins, J.M. and Greenland, S. (1994). Adjusting for differential rates of PCP prophylaxis in high- versus low-dose AZT Treatment Arms in an AIDS Randomized Trial. *Journal of the American Statistical Association* **89**, 737-749.

Sargent, D.J. (1997). A flexible approach to time-varying coefficients in the Cox regression setting. *Lifetime Data Analysis* **3**, 13-25.

Scheike, T.H. and Martinussen, T. (2004). On Estimation and tests of time-varying effects in the proportional hazards model. *Scandinavian Journal of Statistics* **31**, 51-62.

Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. *The Statistician* **41**, 455-465.

Sleeper, L.A. and Harrington, D.P. (1990). Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association* **85**, 941-949.

Xu, R. and O'Quigley, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics* **1**, 423-439.

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

CHAPTER III

Weighted Estimation Of Treatment Effects In The Presence Of Non-Proportional Hazards

ABSTRACT: In medical studies featuring survival time data, it is very common that treatment-specific hazards are non-proportional. A frequently used method of dealing with non-proportional hazards is to apply a Cox model that assumes that the treatment effect is a specific function of time. However, it is often difficult to choose the correct parametric form for the treatment effect. Even if the correct form is chosen, investigators are usually more interested in the cumulative effect than the instantaneous effect in the presence of non-proportional hazards. We propose methods for estimating the cumulative treatment effect when proportional hazards does not hold. Three measures are proposed: ratio of cumulative hazards, relative risk and difference in restricted mean lifetime. No functional form need be assumed for the effects of treatment or the adjustment covariates. The proposed measures are estimated through non-parametric procedures after using inverse probability of treatment weighting (IPTW) to balance the treatment-specific covariate distributions. Asymptotic properties of the proposed estimators are derived, with finite-sample properties assessed in simulation studies. The proposed methods are applied to end stage renal disease (ESRD) data.

KEY WORDS: Cumulative hazard; Inverse weighting; Nonparametric estimator;

Relative risk; Restricted mean lifetime; Survival analysis

3.1 Introduction

In survival analysis, the most frequently used method is the Cox proportional hazards model (Cox 1972, 1975). When the proportional hazards assumption does not hold, the effect of covariates is often assumed to follow a pre-specified function of time. However, it is usually hard to assess whether the chosen functional form is correct. Moreover, when proportional hazards does not hold, investigators are usually more interested in the cumulative effect than the instantaneous effect. For example, in the presence of a treatment effect which changes direction over time, clinicians are often more interested in knowing when the survival probabilities (as opposed to hazard functions) are same for the two treatment groups.

As an alternative to the Cox non-proportional hazards model, one could compare survival or cumulative hazard curves using Nelson-Aalen (Nelson, 1972; Aalen, 1978) or Kaplan-Meier (Kaplan and Meier, 1958) estimators. In fact, several methods have been proposed to estimate cumulative treatment effects. Doksum and Song (1989) constructed confidence intervals and simultaneous confidence bands for a relative change function, expressed in terms of cumulative hazard functions. Parzen, Wei and Ying (1997) compared two survival functions and constructed confidence bands based on simulation techniques. McKeague and Zhao (2002) proposed the ratio of survival functions and constructed simultaneous confidence bands using empirical likelihood. The above listed methods are designed for nonparametric settings and do not adjust for potential confounders. It is well known that these estimators may lead to biased results when the treatments are not randomly assigned, as is the case in observational studies. We propose methods to estimate the cumulative

treatment effect by comparing the cumulative hazard, risk of death and restricted mean lifetime between treatment groups. We apply inverse probability of treatment weighting (IPTW) to adjust for potential confounders. If a subject has a higher (lower) probability of being in a certain treatment group, then the subject is given a lower (higher) weight.

This work is motivated by the continued desire to compare survival of end-stage renal disease patients on two dialysis methods: hemodialysis (HD) and peritoneal dialysis (PD). Peritoneal dialysis is less expensive, but newer and less established. Previous studies comparing HD and PD mortality have produced conflicting results. For example, using non-proportional hazards models assuming a piece-wise constant hazard ratio, Fenton et al. (1997) found that hazard ratios (PD/HD) significantly decreased early in the follow-up period, but that the effect changed direction later on, while Bloembergen et al. (1995) reached very different conclusions based on U.S. national data. Wei and Schaubel (2008) compared the cumulative hazard of death for patients treated by PD to those treated by HD using a stratified Cox model. The authors assumed that the effect of each adjustment covariate was constant over time. Applying our methods to Canadian Organ Replacement Register (CORR) data, we now seek to compare PD and HD without assuming proportional hazards for either the treatment or the adjustment covariates. The average cumulative hazard, risk of death, and restricted mean lifetime are compared between PD and HD.

Inverse probability of treatment weighting (IPTW; Hernan, Brumback & Robins, 2000; Robins, Hernan & Brumback, 2000) has been used to estimate the causal effect of time-dependent exposure and adjust for time-dependent confounders. Hernan et al. (2000) assumed proportional hazards for the exposure and adjustment covariate effects. In the case that there exist time-dependent confounders which are affected by

previous exposure and predict the subsequent exposure, the authors argued that the standard Cox partial likelihood estimator is a unbiased estimate of the association between the exposure and failure time, but is a biased estimate of the causal effect if only baseline covariates are adjusted in the Cox model. To estimate the causal effect, an inverse weight involving each subject's probability of receiving the exposure actually received at time t given previous exposure and covariates history (including the time-dependent confounders which are affected by exposure) was applied to each subject. The model adjusted for time-constant but not the time-dependent confounders. Our proposed methods differ from those of Hernan et al. (2000) in at least three important ways. First, we estimate the marginal treatment effect, i.e., the effect of assigning treatment to a patient from an intent-to-treat perspective. Second, we do not assume proportionality with respect to the treatment or the adjustment covariates. Third, we estimate the cumulative effect of treatment.

Current methods for estimating the cumulative effect in the context of censored data usually focus on the survival or cumulative hazard function. However, mean lifetime is often the more relevant quantity. For example, patients usually ask "How long will I live?", not "What is my 5-year survival probability?". This is true in many areas of medicine and is particularly true for organ failure patients. For example, in the U.S., the liver allocation system is currently being restructured such that patients will be ranked on the waiting list by difference in 5-year restricted mean lifetime. We propose estimating the cumulative treatment effect through IPTW-based estimators for the cumulative hazard, risk of death and restricted mean lifetime. Essentially, treatment-specific pseudo populations with the same confounder distribution are created by applying IPTW. Since the weighted treatment groups have the same adjustment covariate distribution, confounding is eliminated and estimation then

proceeds nonparametrically. Therefore, functional forms for neither the treatment nor the adjustment covariate effects need to be specified.

The remainder of this Chapter is organized as follows. In the next section, we describe our proposed methods. Asymptotic properties of the proposed estimators are developed in Section 3.3. We evaluate the performance of our estimators for finite samples in Section 3.4. In Section 3.5, we apply our proposed method to compare survival on hemodialysis and peritoneal dialysis using data from a national organ failure registry. We provide some discussion of the proposed methods in Section 3.6.

3.2 Proposed methods

Suppose that a total of n subjects are included in the study. Let T_i be the event time and C_i be the censoring time for subject i. Let $X_i = \min\{T_i, C_i\}$ and $\delta_i = I(T_i \leq C_i)$ represent the observation time and event indicattor, respectively, where $I(A)$ is an indicator function taking the value 1 when condition A holds and 0 otherwise. The at-risk indicator is denoted by $Y_i(t) = I(X_i \geq t)$ and the observed death counting processes is defined as $N_i(t) = \delta_i I(X_i \leq t)$. Let $J+1$ be the number of treatment groups (numbered $j = 0, 1, \ldots, J$), where the first group (j=0) represents a reference category to which the remaining treatment groups are compared. Let G_i denote the treatment group for subject i and set $G_{ij} = I(G_i = j)$. We let $Y_{ij}(t) = Y_i(t)G_{ij}$ and $dN_{ij}(t) = dN_i(t)G_{ij}$. The observed data consist of n vectors, $(X_i, \delta_i, G_i, \mathbf{Z}_i^T)^T$ assumed to be independent and identically distributed, where $\mathbf{Z}_i$ is a $p \times 1$ vector of covariates. We assume that C_i and T_i are independent given G_i and that there are no unmeasured confounders.

Our objective is compare the average survival that would result if treatment j was applied to the entire population to that if treatment 0 was applied to the entire pop-

ulation. The survival probability at time t if all subjects were assigned to treatment j is denoted by $S_j(t) = E_{\mathbf{Z}_i}[S(t|G_i = j, \mathbf{Z}_i)]$, where $S(t|G_i = j, \mathbf{Z}_i)$ is the survival probability given treatment j and the adjustment covariates. The expectation is with respect to the marginal distribution of $\mathbf{Z}_i$, such that same averaging is done across all $J + 1$ groups. Note that the quantity $\int_0^t S_j(s)ds$ equals the average restricted mean lifetime considered by Chen & Tsiatis (2001). The averaging we propose is done through inverse weighting, while Chen & Tsiatis (2001) explicitly averaged the fitted values from a treatment stratified proportional hazards model.

We assume that treatment assignment follows a generalized logit model,

$$\log\left\{\frac{p_{ij}(\boldsymbol{\beta}_0)}{p_{i0}(\boldsymbol{\beta}_0)}\right\} = \boldsymbol{\beta}_0^T \mathbf{X}_{ij},$$

where $p_{ij}(\boldsymbol{\beta}_0)$ is the probability that subject i is assigned to treatment j and $\mathbf{X}_{ij} = [\mathbf{0}_{1\times(j-1)(p+1)}, 1, \mathbf{Z}_i^T, \mathbf{0}_{1\times(J-j)(p+1)}]^T$ for $j = 1, \cdots, J$, with $\mathbf{0}_{1\times(j-1)(p+1)}$ a 1 by $(j-1)(p+1)$ matrix with elements 0. The estimate for $\boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\beta}}$ is obtained by maximum likelihood with score function defined as

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} \mathbf{X}_{ij}\left[G_{ij} - p_{ij}(\boldsymbol{\beta})\right]. \tag{3.1}$$

Let $\Lambda_j(t) = -\log\{S_j(t)\}$ denote the cumulative hazard if treatment j was assigned to the entire population. The probability that the event occurs by time t is denoted by $F_j(t) = 1 - S_j(t)$ and the restricted mean lifetime through $(0, t]$ is denoted by $e_j(t) = \int_0^t S_j(u)du$. To compare treatment j to the reference treatment, three measures are proposed:

(i) Ratio of cumulative hazards

$$\phi_j(t) = \frac{\Lambda_j(t)}{\Lambda_0(t)} \tag{3.2}$$

(ii) Relative risk

$$RR_j(t) = \frac{F_j(t)}{F_0(t)} \tag{3.3}$$

(iii) Difference in restricted mean lifetime

$$\Delta_j(t) = e_j(t) - e_0(t). \tag{3.4}$$

The measure $\phi_j(t)$ relates to the hazard ratio usually used when proportional hazards holds. The measure $RR_j(t)$ is a process version of relative risk, a quantity used regularly in epidemiologic studies. The quantity $\Delta_j(t)$ measures the area between the treatment j and treatment 0 survival curves.

Measures (3.2), (3.3) and (3.4) can be estimated using the inverse probability of treatment weighting (IPTW) method. Their corresponding estimators are

$$\begin{aligned}
\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}}) &= \frac{\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})}{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})}, \\
\widehat{RR}_j(t,\widehat{\boldsymbol{\beta}}) &= \frac{\widehat{F}_j(t,\widehat{\boldsymbol{\beta}})}{\widehat{F}_0(t,\widehat{\boldsymbol{\beta}})}, \\
\widehat{\Delta}_j(t,\widehat{\boldsymbol{\beta}}) &= \widehat{e}_j(t,\widehat{\boldsymbol{\beta}}) - \widehat{e}_0(t,\widehat{\boldsymbol{\beta}}),
\end{aligned}$$

where $\widehat{F}_j(t,\widehat{\boldsymbol{\beta}}) = 1 - \exp\{-\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})\}$, $\widehat{e}_j(t,\widehat{\boldsymbol{\beta}}) = \int_0^t \exp\{-\widehat{\Lambda}_j(s,\widehat{\boldsymbol{\beta}})\}ds$, and

$$\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{w_{ij}(\widehat{\boldsymbol{\beta}})}{R_j^{(0)}(s,\widehat{\boldsymbol{\beta}})} dN_{ij}(s),$$

where $R_j^{(0)}(s,\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} Y_{ij}(s)w_{ij}(\boldsymbol{\beta})$ and $w_{ij}(\widehat{\boldsymbol{\beta}}) = G_{ij}/p_{ij}(\widehat{\boldsymbol{\beta}})$ is the estimator for $w_{ij}(\boldsymbol{\beta}_0) = G_{ij}/p_{ij}(\boldsymbol{\beta}_0)$. The quantity $\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})$ is a weighted version of the Nelson-Aalen estimator. Measures (3.2) and (3.3) are considered in a time interval $[t_L, t_U]$, and (3.4) is considered in a time interval $(0, t_U]$, where t_L is chosen to avoid instances in which $\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}}) = 0$ and t_U is chosen to avoid the instability in tail of the observation time distribution.

After applying IPTW, $J+1$ pseudo populations are created, where the distribution of $\mathbf{Z}_i$ is balanced among the treatment groups and is same as that of the entire population. For example, (3.4) compares the restricted mean lifetime through $(0, t]$, if treatment j was assigned to the entire population to that if the reference treatment was assigned to the entire population.

3.3 Asymptotic properties

To derive the large-sample properties of estimators proposed in preceding section, we assume the following regularity conditions for $i = 1, \cdots, n$ and $j = 0, \cdots, J$.

(a) Z_{ik} have bounded total variation; i.e., $|Z_{ik}| < \kappa$ for $k = 1, \cdots, p$, where κ is a constant and Z_{ik} is the kth component of $\mathbf{Z}_i$.

(b) $\Lambda_j(\tau) < \infty$ where τ is a pre-specified time point satisfying $Pr(Y_{ij}(\tau) = 1) > 0$.

(c) Continuity of $r_j^{(0)}(t, \boldsymbol{\beta})$, where $r_j^{(0)}(t, \boldsymbol{\beta})$ is the limiting value of $R_j^{(0)}(t, \boldsymbol{\beta})$ with $r_j^{(0)}(t, \boldsymbol{\beta})$ bounded away from 0 for $t \in [0, \tau]$ for $\boldsymbol{\beta}$ in an open set.

(d) Positive-definiteness of the matrix $\boldsymbol{\Omega}(\boldsymbol{\beta})$ where

$$\boldsymbol{\Omega}(\boldsymbol{\beta}) = E\left\{\sum_{j=1}^{J} p_{ij}(\boldsymbol{\beta})\mathbf{X}_{ij}\left[\mathbf{X}_{ij}^T - \sum_{k=1}^{J}\mathbf{X}_{ik}^T p_{ik}(\boldsymbol{\beta})\right]\right\} \tag{3.5}$$

(e) $Pr(G_{ij} = 1) > 0$.

Asymptotic properties of the estimators are summarized by the following three theorems:

THEOREM 1. *Under conditions (a) to (e),* $\widehat{\phi}_j(t, \widehat{\boldsymbol{\beta}})$ *converges almost surely and uniformly to* $\phi_j(t)$*, and* $n^{1/2}[\widehat{\phi}_j(t, \widehat{\boldsymbol{\beta}}) - \phi_j(t)]$ *converges asymptotically to a zero-mean Gaussian process with covariance function* $\sigma_j^{\phi}(s, t) = E[\xi_{ij}^{\phi}(s, \boldsymbol{\beta}_0)\xi_{ij}^{\phi}(t, \boldsymbol{\beta}_0)]$*, for* $j = 1, \cdots, J$ *and* $t \in [t_L, t_U]$*, where*

$$
\begin{aligned}
\xi_{ij}^{\phi}(t,\boldsymbol{\beta}) &= \frac{1}{\Lambda_0(t)}\Phi_{ij}(t,\boldsymbol{\beta}) - \frac{\Lambda_j(t)}{\Lambda_0(t)^2}\Phi_{i0}(t,\boldsymbol{\beta}) \\
\Phi_{ij}(t,\boldsymbol{\beta}) &= h_j(t)^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta})\sum_{j=1}^{J}\mathbf{X}_{ij}\left[G_{ij} - p_{ij}(\boldsymbol{\beta})\right] + \int_0^t \frac{w_{ij}(\boldsymbol{\beta})}{r_j^{(0)}(s,\boldsymbol{\beta})}dM_{ij}(s) \\
dM_{ij}(s) &= dN_{ij}(s) - Y_{ij}(s)d\Lambda_j(s) \\
h_j(t) &= E\left[\int_0^t \left\{\frac{a_{ij}(\boldsymbol{\beta})}{r_j^{(0)}(s,\boldsymbol{\beta})} - \frac{E[a_{ij}(\boldsymbol{\beta})Y_{ij}(s)]w_{ij}(\boldsymbol{\beta})}{r_j^{(0)}(s,\boldsymbol{\beta})^2}\right\}dN_{ij}(s)\right] \\
a_{ij}(\boldsymbol{\beta}) &= (1-G_{i0})\left[\frac{\sum_{k=1}^{J}\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ik}\}\mathbf{X}_{ik}}{\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ij}\}} - \mathbf{X}_{ij}p_{ij}^{-1}(\boldsymbol{\beta})\right] + G_{i0}\sum_{j=1}^{J}\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ij}\}\mathbf{X}_{ij}.
\end{aligned}
$$

The consistency of $\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}})$ is proved by the Strong Law of Large Numbers (SLLN) and continuous mapping theorem as well as the consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})$. The estimator $\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})$ is proved to be consistent by using the SLLN and the fact that $\widehat{\boldsymbol{\beta}}$ is consistent. The quantities $n^{1/2}[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})-\Lambda_j(t)]$ and $n^{1/2}[\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})-\Lambda_0(t)]$ can each be written as sums of independent and identically distributed variates. The convergence of $n^{1/2}[\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}}) - \phi_j(t)]$ involves writing the quantity as functions of $n^{1/2}[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}}) - \Lambda_j(t)]$ and $n^{1/2}[\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})-\Lambda_0(t)]$, then applying the Central Limit theorem and results from empirical process (Bilias, Gu and Ying, 1997; Pollard, 1990). The quantity in $\xi_{ij}^{\phi}(t,\boldsymbol{\beta}_0)$ is consistently estimated by its empirical counterparts.

THEOREM 2. *Under conditions (a) to (e), $\widehat{RR}_j(t,\widehat{\boldsymbol{\beta}})$ converges almost surely and uniformly to $RR_j(t)$, and $n^{1/2}[\widehat{RR}_j(t,\widehat{\boldsymbol{\beta}})-RR_j(t)]$ converges asymptotically to a zero-mean Gaussian process with covariance function $\sigma_j^R(s,t) = E[\xi_{ij}^R(s,\boldsymbol{\beta}_0)\xi_{ij}^R(t,\boldsymbol{\beta}_0)]$, for $j = 1,\cdots,J$ and $t \in [t_L, t_U]$, where*

$$
\xi_{ij}^{R}(t,\boldsymbol{\beta}) = \frac{S_j(t)}{F_0(t)}\Phi_{ij}(t,\boldsymbol{\beta}) - \frac{F_j(t)S_0(t)}{F_0(t)^2}\Phi_{i0}(t,\boldsymbol{\beta}), \tag{3.6}
$$

with $\Phi_{ij}(t,\boldsymbol{\beta})$ defined as in Theorem 1.

In the next theorem, we consider the asymptotic behavior of our restricted mean lifetime difference estimator.

THEOREM 3. *Under conditions (a) to (e),* $\widehat{\Delta}_j(t,\widehat{\boldsymbol{\beta}})$ *converges almost surely and uniformly to* $\Delta_j(t)$*, and* $n^{1/2}[\widehat{\Delta}_j(t,\widehat{\boldsymbol{\beta}}) - \Delta_j(t)]$*, converges asymptotically to a zero-mean Gaussian process with covariance function* $\sigma_j^{\Delta}(s,t) = E[\xi_{ij}^{\Delta}(s,\boldsymbol{\beta}_0)\xi_{ij}^{\Delta}(t,\boldsymbol{\beta}_0)]$*, for* $j = 1, \cdots, J$ *and* $t \in (0, \tau_U]$*, where*

$$\xi_{ij}^{\Delta}(t,\boldsymbol{\beta}) = -\int_0^t S_j(s)\Phi_{ij}(t,\boldsymbol{\beta})ds + \int_0^t S_0(s)\Phi_{i0}(t,\boldsymbol{\beta})ds. \tag{3.7}$$

The proof for Theorem 2 and Theorem 3 proceeds much like that of Theorem 1.

3.4 Simulation study

We evaluated the finite sample properties of the proposed estimator through a series of simulation studies. A covariate Z_{i1} was generated from a Uniform (0,1) distribution and a covariate Z_{i2} was generated as a binary variable (0 or 1) with $Pr(Z_{i2} = 1) = 0.5$. We set up two treatments, with the treatment indicator, G_i, generated from a Bernoulli distribution with parameter $p_{i1}(\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 Z_{i1} + \beta_2 Z_{i2})/\{1 + \exp(\beta_0 + \beta_1 Z_{i1} + \beta_2 Z_{i2})\}$. We chose $\beta_0 = 0.1$, $\beta_1 = 0.2$ and $\beta_2 = -0.2$. We generated the death times from a Weibull model with hazard function

$$\lambda_{ij}(t) = \lambda_i(t|G_i = j, Z_{i1}, Z_{i2}) = \alpha_j\gamma_j t^{\gamma_j - 1}\exp\{\eta_1 Z_{i1} + \eta_2 Z_{i2}\} \quad \text{for } j = 0, 1.$$

Censoring times were generated from a Uniform (2.5, 5) distribution. Various values of γ_j were used to make the hazard ratio constant or vary through time. Censoring percentages ranged from 18% to 36%. Values of $\eta_1 = \pm 0.5$ and $\eta_2 = \pm 0.5$ were employed in the Cox model. We chose sample sizes $n = 200$ and $n = 100$. A total of 1000 simulations were used for each setting. For the first two measures, $\phi_j(t)$ and

$RR_j(t)$, we employed the log transformation, such that confidence intervals would always be in a valid range. Each of the three estimators were evaluated at time points $t = 1$, $t = 2$ and $t = 3$.

The proposed estimators generally perform well (Table 3.2 and Table 3.3). The empirical mean of our estimators is approximately equal to the true value. The average asymptotic standard errors (ASE) are generally close to the empirical standard deviations (ESD), while the coverage probabilities (CP) are close to the nominal value of 0.95 (Table 3.4).

We also checked the impact of misspecification of the treatment assignment probability model on estimates of our measures. The treatment indicator, G_i, was generated from a Bernoulli distribution with

$$Pr(G_i = 1|Z_{i1}, Z_{i2}) = \frac{\exp(0.2 - 0.5 \times Z_{i1} + Z_{i2} + Z_{i1} \times Z_{i2})}{1 + \exp(0.2 - 0.5 \times Z_{i1} + Z_{i2} + Z_{i1} \times Z_{i2})},$$

where Z_{i1} was generated from a Uniform (0,1) distribution and Z_{i2} was generated from a Bernoulli (0.4) distribution. Death times, T_i, were generated from a Cox model with hazard function

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp\{0.5 \times Z_{i1} + 0.5 \times Z_{i2}\},$$

where $\lambda_{00}(t) = 0.2$ and $\lambda_{01}(t) = 0.25$. We evaluated the performance of our estimators when some covariate terms were omitted or misspecified in the logistic model. When some covariates are not included in the logistic model, the estimates of our measures are biased, depending on the degree of misspecification (Table 3.5). For example, when Z_{i2} has a strong effect on treatment assignment, the estimates are very biased when Z_{i2} is excluded from the fitted logistic model. Results for a model containing only Z_{i1} were comparable to the intercept-only model.

3.5 Analysis of dialysis data

We applied our proposed methods to compare the survival of end stage renal disease (ESRD) patients by dialysis therapy. Data were obtained from the Canadian Organ Replacement Register (CORR) which is a nation-wide and population-based organ failure registry. Our analysis included patients who were either put on hemodialysis (HD) or peritonital dialysis (PD) at the time of renal replacement therapy initiation. We carried out an intent-to-treat analysis which classified patients based on the type of dialysis received at the initiation of renal replacement therapy. Hemodialysis served as the reference category (j=0).

This study included n=23,254 ESRD patients aged 18 and above who initiated HD or PD between 1990 and 1998. Patients began follow-up at the date of dialysis initiation and were followed until the earliest of death, loss to follow-up, kidney transplantation or the end of the observation period (December 31, 1998). There were 17,766 patients initially placed on HD, and approximately 38% of them died, while 36% of patients on PD (5,488) were observed to die. Potential confounders included in the data set are patient's age, gender, race, underlying renal diagnosis, region and various comorbid illnesses (cerebrovascular accident, cardiovascular disease, chronic obstructive pulmonary disease (COPD), malignancy, peripheral vascular disease, other illnesses). The probability of initiating dialysis on PD (HD) was calculated using a logistic model adjusting for the above covariates, with interaction terms included when significant. Interaction terms are selected by a backward elimination algorithm which started with a model that contained all main effects and all pair-wise interaction terms, forcing the main effects to be included.

We computed our proposed estimators over the (0, 96] month interval. Figure 3.1

shows that the cumulative hazard of death is significantly lower comparing PD to HD during the first 32 months with point estimate ranges from 0.38 to 0.94, while after month 45, the cumulative hazard of death is significantly higher for PD relative to HD with the cumulative hazard ratio ranges from 1.06 to 1.22. The mortality risk ratio comparing PD to HD ranges from 0.38 to 0.95 for the first 32 months, with the risk of death is significantly lower for PD relative to HD (Figure 3.2). The risk of death for PD is significantly higher after month 45 with the risk ratio ranging from 1.04 to 1.08. The restricted mean lifetime (in months) for the first 84 months is greater for PD compared to HD, although the difference is only significant in the (0,66] months time interval (Figure 3.3). The maximum difference is 1.4 months, statistically significant, although not likely of clinical importance.

3.6 Discussion

In this Chapter, we propose methods to estimate the cumulative treatment effect. The proposed methods would be most useful in observational studies due to the potential for bias in the absence of accounting for confounders. Our methods do not require that the proportional hazards assumption hold for either the treatments or the adjustment covariates. The proposed measures compare patient cumulative hazard of death, risk of death and restricted mean lifetime if the treatment was assigned to the entire population to that if the reference treatment was assigned to the entire population. Simulation studies show that the proposed estimators are approximately unbiased and that the estimated standard errors are quite accurate.

Our proposed methods do not require specifying the functional form of the effects of treatment or adjustment covariates. However, unlike a typical Cox model, we do need to model the treatment assignment probability and the consistency of our

proposed estimators is based on the correct estimate for this probability. Although we essentially trade one model for another, a binary (multinominal) outcome is typically easier to model than a hazard function due to the absence of the time dimension.

Various methods in the existing literature are related to those we propose. For example, a method proposed by Anstrom and Tsiatis (2001) can be used to compare treatment-specific survival funcations at a fixed time point, but was not designed to measure the treatment effect as a process over time. Xie and Liu (2005) developed an inverse probability of treatment weighted Kaplan-Meier estimator to estimate the survival function. This estimator reduced the bias of unweighted nonparametric Kaplan-Meier estimator when confounders for treatment exist. A weighted log rank statistic was proposed to test for the difference between the two survival curves. The authors did not propose a measure to estimate the treatment effect along with time while our proposed measures give a view of the time-dependent treatment effect. Moreover, we compared restricted mean lifetime which interests many researchers.

Chen & Tsiatis (2001) compared the restricted mean lifetime between two treatment groups. A Cox model was fitted for each treatment group such that the authors assumed that the effects of adjustment covariates follow proportional hazards. The survival function for each group was estimated by explicitly averaging over all subjects in the sample. The treatment-specific survival function (and hence, restricted mean lifetime) estimators we propose converge to the same values as the estimators proposed by Chen & Tsiatis (2001) under their specified model. One advantage of our estimators is that those of Chen & Tsiatis (2001) require explicitly averaging over all observed covariate patterns which would be taxing computationally, particulary for large data sets. Another advantage of our estimators is that we need not specify a functional form for the effects of adjustment covariates. Conversely, Chen & Tsiatis

(2001) do not need to assume that the censoring is conditionally independent of the adjustment covariates.

In this chapter, we constructed confidence intervals of proposed measures. If we want to look at the measures as a process over time, confidence bands can be constructed as in Wei and Schaubel (2008). However, confidence intervals are often preferred over confidence bands, since researchers are usually interested in the cumulative effect at some specific time point or set of time points. For example, surgeons want to know the restricted mean lifetime at 1, 3 and 5 years. In this case, point-wise confidence intervals are more appropriate.

Applying our methods to the CORR dialysis data using an intent-to-treat analysis, we found that cumulative hazard and risk of death are significantly lower for PD relative to HD for the first 32 months while the opposite was found after approximately 45 months from dialysis initiation. We found that the restricted mean life is significant longer comparing PD to HD based on the first 66 months while the difference is non-significant after that. We assumed that there is no selection bias due to unmeasured covariates which may be an issue since the assignment of dialysis may depend on patient characteristics for which no data are available to CORR. As such, our results must be interpreted with some caution.

Our proposed methods assume that censoring does not depend on confounders for treatment, although sub-analysis revealed that the censoring hazard does in fact depend on certain adjustment covariates. However, $\widehat{\phi}_1(t)$ (ratio of cumulative hazards) was similar to what we found in Chapter 2 which does not require that censoring time is independent of the covariates. This gives evidence that, even if the C_i and $\mathbf{Z}_i$ independence assumption is violated, the results were not very biased. As a sensitivity analysis to further address the dependent censoring issue, we applied an inverse cen-

soring weight (Robins and Rotnitzky, 1992; Robins, 1993) and IPTW concurrently to our estimators. The modified weight is given by

$$w_{ij}(t) = \frac{Y_{ij}(t)}{Pr(G_i = j|\mathbf{Z}_i)Pr(C_i > t|G_i = j, \mathbf{Z}_i)}.$$

The point estimates based on the modified estimators were quite similar to what we reported in last section. We develop a generalization of this method in the next chapter.

3.7 Tables and Figures

Setting	α_0	α_1	γ_0	γ_1	η_1	η_2
I	0.35	0.25	1	1	0.5	0.5
II	0.35	0.25	1	1	-0.5	0.5
III	0.35	0.25	1	1	0.5	-0.5
IV	0.2	0.35	1.2	1	0.5	0.5
V	0.2	0.35	1.2	1	-0.5	0.5
VI	0.2	0.35	1.2	1	0.5	-0.5

Table 3.1: Simulation study: Parameter settings

		t=1		t=2		t=3	
	Setting	$\log \phi_1(t)$	BIAS	$\log \phi_1(t)$	BIAS	$\log \phi_1(t)$	BIAS
	I	-0.329	0.001	-0.323	0.004	-0.317	0.009
	II	-0.332	-0.005	-0.328	-0.011	-0.324	-0.007
	III	-0.332	-0.003	-0.328	-0.001	-0.324	0.003
$\log \widehat{\phi}_1(t)$	IV	0.549	0.007	0.404	0.001	0.320	-0.002
	V	0.553	0.003	0.411	-0.007	0.327	-0.009
	VI	0.553	0.023	0.411	0.010	0.327	0.003
	Setting	$\log RR_1(t)$	BIAS	$\log RR_1(t)$	BIAS	$\log RR_1(t)$	BIAS
	I	-0.254	0.001	-0.190	0.001	-0.142	0.003
	II	-0.284	-0.005	-0.239	-0.008	-0.201	-0.005
	III	-0.284	-0.004	-0.239	-0.002	-0.201	0.002
$\log \widehat{RR}_1(t)$	IV	0.434	0.008	0.243	0.002	0.143	-0.001
	V	0.481	0.004	0.303	-0.005	0.203	-0.005
	VI	0.481	0.021	0.303	0.009	0.203	0.003
	Setting	$\Delta_1(t)$	BIAS	$\Delta_1(t)$	BIAS	$\Delta_1(t)$	BIAS
	I	0.060	-0.001	0.173	-0.003	0.287	-0.005
	II	0.041	-0.001	0.136	0.001	0.250	0.002
	III	0.041	-0.001	0.136	-0.001	0.250	-0.002
$\widehat{\Delta}_1(t)$	IV	-0.106	0.001	-0.263	0.001	-0.392	0.003
	V	-0.072	0.001	-0.201	0.005	-0.327	0.009
	VI	-0.072	-0.001	-0.201	-0.003	-0.327	-0.004

Table 3.2: Simulation results: Evaluation of bias (n=200)

		t=1		t=2		t=3	
	Setting	$\log \phi_1(t)$	BIAS	$\log \phi_1(t)$	BIAS	$\log \phi_1(t)$	BIAS
	I	-0.329	-0.023	-0.323	0.001	-0.317	0.002
	II	-0.332	0.026	-0.328	0.004	-0.324	0.005
	III	-0.332	-0.007	-0.328	-0.001	-0.324	-0.008
$\log \widehat{\phi}_1(t)$	IV	0.549	0.021	0.404	0.005	0.320	0.014
	V	0.553	0.024	0.411	0.008	0.327	-0.001
	VI	0.553	0.020	0.411	0.003	0.327	0.006
	Setting	$\log RR_1(t)$	BIAS	$\log RR_1(t)$	BIAS	$\log RR_1(t)$	BIAS
	I	-0.254	-0.020	-0.190	-0.001	-0.142	-0.001
	II	-0.284	0.022	-0.239	0.003	-0.201	0.003
	III	-0.284	-0.007	-0.239	-0.001	-0.201	-0.006
$\log \widehat{RR}_1(t)$	IV	0.434	0.022	0.243	0.007	0.143	0.011
	V	0.481	0.025	0.303	0.009	0.203	0.003
	VI	0.481	0.021	0.303	0.005	0.203	0.006
	Setting	$\Delta_1(t)$	BIAS	$\Delta_1(t)$	BIAS	$\Delta_1(t)$	BIAS
	I	0.060	0.002	0.173	0.002	0.287	-0.002
	II	0.041	-0.004	0.136	-0.009	0.250	-0.013
	III	0.041	-0.001	0.136	-0.003	0.250	-0.005
$\widehat{\Delta}_1(t)$	IV	-0.106	0.001	-0.263	0.001	-0.392	-0.001
	V	-0.072	0.001	-0.201	0.002	-0.327	0.004
	VI	-0.072	0.002	-0.201	0.004	-0.327	0.005

Table 3.3: Simulation results: Evaluation of bias (n=100)

	Setting	$\log \phi_1(t)$	BIAS	ASE	ESD	CP
	I	-0.323	0.004	0.182	0.187	0.94
	II	-0.328	-0.011	0.211	0.213	0.95
	III	-0.328	-0.001	0.211	0.211	0.96
$\log \widehat{\phi}_1(t)$	IV	0.405	0.001	0.186	0.191	0.94
	V	0.411	-0.007	0.216	0.224	0.94
	VI	0.411	0.010	0.217	0.219	0.95
	Setting	$\log RR_1(t)$	BIAS	ASE	ESD	CP
	I	-0.190	0.001	0.109	0.111	0.95
	II	-0.239	-0.008	0.155	0.157	0.95
	III	-0.239	-0.002	0.155	0.155	0.96
$\log \widehat{RR}_1(t)$	IV	0.244	0.002	0.114	0.118	0.94
	V	0.303	-0.005	0.162	0.168	0.94
	VI	0.303	0.009	0.162	0.164	0.95
	Setting	$\Delta_1(t)$	BIAS	ASE	ESD	CP
	I	0.173	-0.003	0.098	0.101	0.95
	II	0.136	0.001	0.092	0.092	0.95
	III	0.136	-0.001	0.093	0.094	0.95
$\widehat{\Delta}_1(t)$	IV	-0.264	0.002	0.095	0.099	0.95
	V	-0.201	0.005	0.088	0.089	0.94
	VI	-0.201	-0.003	0.088	0.092	0.95

Table 3.4: Simulation results: Accuracy of standard error estimator (t=2, n=200)

Covariates in logistic model	$\log \phi_1(t)$	BIAS	ASE	ESD	CP
$Z_1, Z_2, Z_1 \times Z_2$	0.216	0.005	0.223	0.233	0.94
$Z_1^\dagger, Z_2, Z_2 \times Z_1^\dagger$	0.216	0.004	0.223	0.233	0.94
Z_1, Z_2	0.216	0.009	0.225	0.229	0.95
Z_2	0.216	0.001	0.225	0.228	0.95
Z_1	0.216	0.166	0.213	0.217	0.88
None	0.216	0.159	0.214	0.217	0.88

Covariates in logistic model	$\log RR_1(t)$	BIAS	ASE	ESD	CP
$Z_1, Z_2, Z_1 \times Z_2$	0.149	0.006	0.157	0.164	0.94
$Z_1^\dagger, Z_2, Z_2 \times Z_1^\dagger$	0.149	0.006	0.158	0.161	0.94
Z_1, Z_2	0.149	0.009	0.158	0.161	0.95
Z_2	0.149	0.003	0.158	0.160	0.95
Z_1	0.149	0.117	0.151	0.154	0.90
None	0.149	0.113	0.151	0.153	0.89

Covariates in logistic model	$\Delta_1(t)$	BIAS	ASE	ESD	CP
$Z_1, Z_2, Z_1 \times Z_2$	-0.099	0.001	0.106	0.111	0.93
$Z_1^\dagger, Z_2, Z_2 \times Z_1^\dagger$	-0.099	0.002	0.105	0.111	0.93
Z_1, Z_2	-0.099	-0.001	0.107	0.109	0.94
Z_2	-0.099	0.003	0.107	0.109	0.94
Z_1	-0.099	-0.072	0.097	0.101	0.86
None	-0.099	-0.069	0.098	0.101	0.87

$Z_1^\dagger = 1$, if $Z_1 < 0.5$; $Z_1^\dagger = 0$, if $Z_1 \geq 0.5$

Table 3.5: Simulation results: Effect of model misspecification (t=2, n=200)

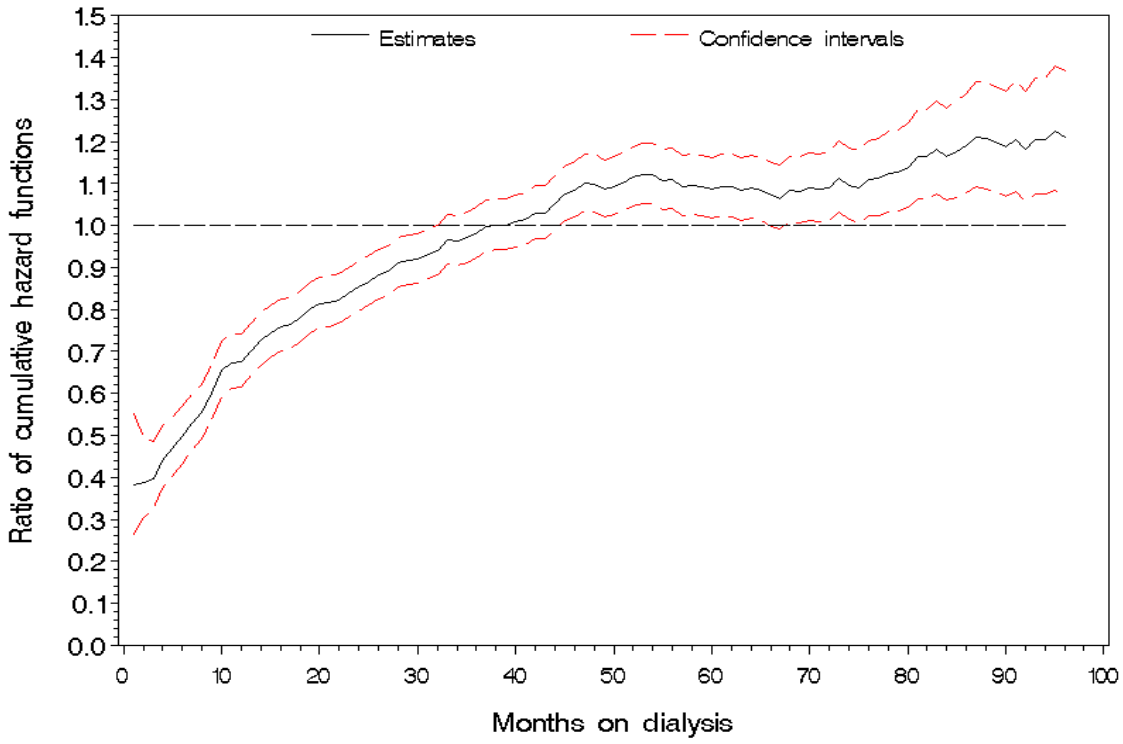

Figure 3.1: Analysis of dialysis data: Estimator and 95% pointwise confidence intervals for the ratio of cumulative hazard functions (PD/HD), $\phi_1(t)$.

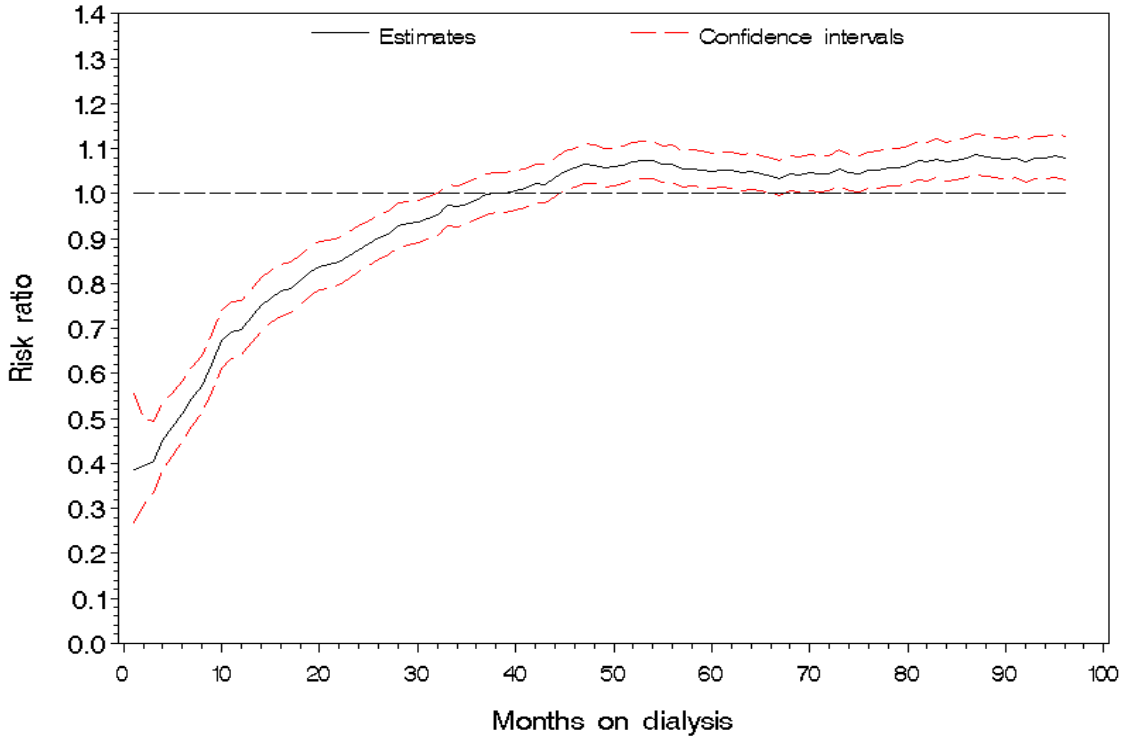

Figure 3.2: Analysis of dialysis data: Estimator and 95% pointwise confidence intervals for the risk ratio (PD/HD), $RR_1(t)$.

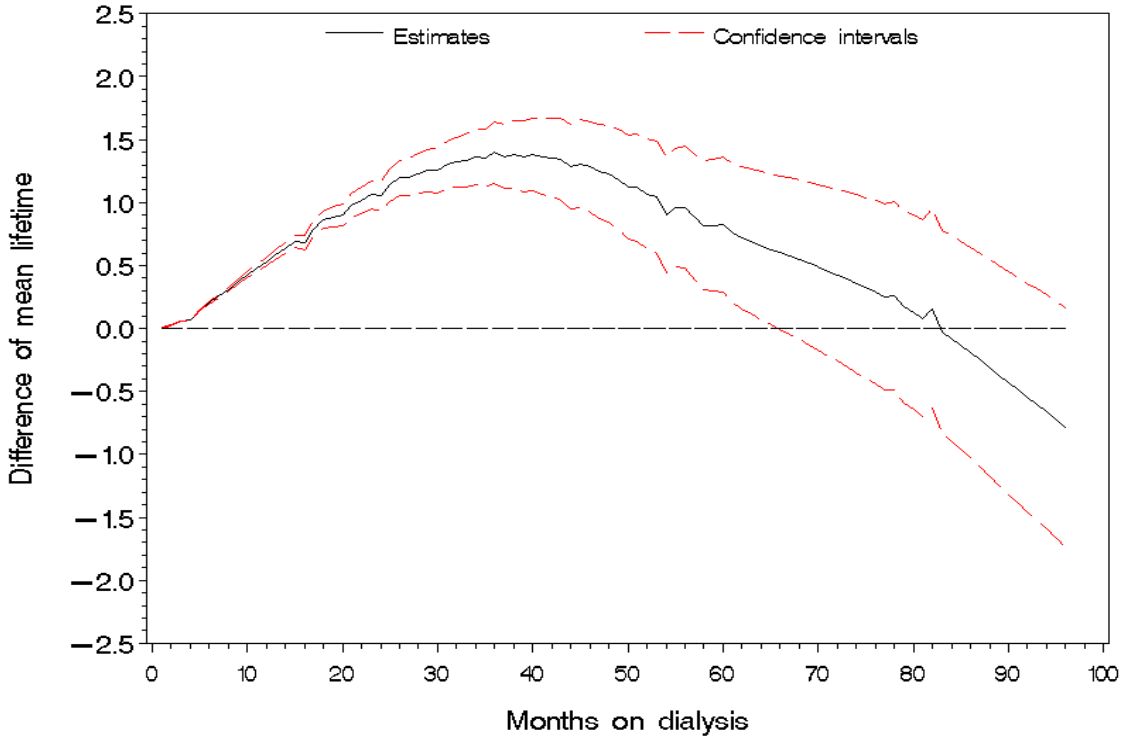

Figure 3.3: Analysis of dialysis data: Estimator and 95% pointwise confidence intervals for the difference in restricted mean lifetime (PD-HD), $\Delta_1(t)$.

3.8 Appendix

Proof of Theorem 1

Consistency:

The strong consistency of $\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}})$, $j=1,\cdots,J$, can be proved by first proving the consistency of $\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})$ for $j=0,\cdots,J$,where

$$\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})=\frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{w_{ij}(\widehat{\boldsymbol{\beta}})}{n^{-1}\sum_{i=1}^{n}Y_{ij}(s)w_{ij}(\widehat{\boldsymbol{\beta}})}dN_{ij}(s).$$

Using the fact that $w_{ij}(\widehat{\boldsymbol{\beta}}) \xrightarrow{a.s.} w_{ij}(\boldsymbol{\beta}_0)$ as $n\to\infty$, and the Strong Law of Large Numbers (SLLN), one can obtain that $n^{-1}\sum_{i=1}^{n}Y_{ij}(s)\widehat{w}_{ij}(\widehat{\boldsymbol{\beta}})$ converges almost surely to

$$\begin{aligned}
& E\left[Y_{ij}(s)w_{ij}(\boldsymbol{\beta}_0)\right] \\
= \;& E\left\{E\left[Y_{ij}(s)w_{ij}(\boldsymbol{\beta}_0)|\mathbf{Z}_i\right]\right\} \\
= \;& E\left\{Pr^{-1}(G_i=j|\mathbf{Z}_i)E\left[Y_{ij}(s)|\mathbf{Z}_i\right]\right\} \\
= \;& E\left[Pr^{-1}(G_i=j|\mathbf{Z}_i)Pr(T_i>s|G_i=j,\mathbf{Z}_i)Pr(C_i>s|G_i=j,\mathbf{Z}_i)Pr(G_i=j|\mathbf{Z}_i)\right] \\
= \;& Pr(C_i>s|G_i=j)E\left[Pr(T_i>s|G_i=j,\mathbf{Z}_i)\right] \\
= \;& Pr(C_i>s|G_i=j)E\left[S(s|G_i=j,\mathbf{Z}_i)\right] \\
= \;& Pr(C_i>s|G_i=j)S_j(s)
\end{aligned}$$

Using similar techniques, one can obtain that $n^{-1}\sum_{i=1}^{n}w_{ij}(\widehat{\boldsymbol{\beta}})dN_{ij}(s)$ converges almost surely to $Pr(C_i>s|G_i=j)dF_j(s)$ as $n\to\infty$. The above listed results give $\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}}) \xrightarrow{a.s.} \Lambda_j(t)$ for $j=0,\cdots,J$. Therefore, $\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}})$ converges to $\phi_j(t)$ almost surely as $n\to\infty$ using the continuous mapping theorem.

Asymptotic normality:

One can write:

$$n^{1/2}\left[\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}})-\phi_j(t)\right]=\frac{1}{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})}n^{1/2}[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})-\Lambda_j(t)]+\Lambda_j(t)n^{1/2}\left[\frac{1}{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})}-\frac{1}{\Lambda_0(t)}\right]$$

The quantity $n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})-\Lambda_j(t)\right]$ can be decomposed into $\widehat{\alpha}_{j1}(t)+\widehat{\alpha}_{j2}(t)$, where

$$\begin{aligned}\widehat{\alpha}_{j1}(t) &= n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})-\widehat{\Lambda}_j(t,\boldsymbol{\beta}_0)\right]\\ \widehat{\alpha}_{j2}(t) &= n^{1/2}\left[\widehat{\Lambda}_j(t,\boldsymbol{\beta}_0)-\Lambda_j(t)\right].\end{aligned}$$

The quantity $\widehat{\alpha}_{j1}(t)$ is written as

$$\begin{aligned}\widehat{\alpha}_{j1}(t) &= n^{-1/2}\sum_{i=1}^{n}\int_0^t\frac{w_{ij}(\widehat{\boldsymbol{\beta}})dN_{ij}(s)}{R_j^{(0)}(s,\widehat{\boldsymbol{\beta}})}-n^{-1/2}\sum_{i=1}^{n}\int_0^t\frac{w_{ij}(\boldsymbol{\beta}_0)dN_{ij}(s)}{R_j^{(0)}(s,\boldsymbol{\beta}_0)}\\ &= n^{-1/2}\sum_{i=1}^{n}\int_0^t\frac{[w_{ij}(\widehat{\boldsymbol{\beta}})-w_{ij}(\boldsymbol{\beta}_0)]dN_{ij}(s)}{R_j^{(0)}(s,\widehat{\boldsymbol{\beta}})}\\ &\quad +n^{-1/2}\sum_{i=1}^{n}\int_0^t w_{ij}(\boldsymbol{\beta}_0)\left[\frac{1}{R_j^{(0)}(s,\widehat{\boldsymbol{\beta}})}-\frac{1}{R_j^{(0)}(s,\boldsymbol{\beta}_0)}\right]dN_{ij}(s),\end{aligned}$$

By Taylor Series expansions, $\widehat{\alpha}_{j1}(t)$ can be written as

$$\widehat{\alpha}_{j1}(t)=\widehat{h}_j(t)^T n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(1), \tag{3.8}$$

where

$$\widehat{h}_j(t)=\frac{1}{n}\sum_{i=1}^{n}\int_0^t\left\{\frac{a_{ij}(\boldsymbol{\beta}_0)}{R_j^{(0)}(s,\widehat{\boldsymbol{\beta}})}-\frac{w_i(\boldsymbol{\beta})[n^{-1}\sum_{i=1}^{n}a_{ij}(\boldsymbol{\beta}_0)Y_{ij}(s)]}{R_j^{(0)}(s,\boldsymbol{\beta}_0)^2}\right\}dN_{ij}(s)$$

and

$$\begin{aligned}a_{ij}(\boldsymbol{\beta}) &= \frac{\partial w_{ij}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\\ &= (1-G_{i0})\left[\frac{\sum_{k=1}^{J}\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ik}\}\mathbf{X}_{ik}}{\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ij}\}}-\mathbf{X}_{ij}p_{ij}^{-1}(\boldsymbol{\beta})\right]+G_{i0}\sum_{j=1}^{J}\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ij}\}\mathbf{X}_{ij}.\end{aligned}$$

By the Strong Law of Large Numbers (SLLN), $R_j^{(0)}(s,\widehat{\boldsymbol{\beta}})$ and $R_j^{(0)}(s,\boldsymbol{\beta}_0)$ converge to $r_j^{(0)}(s,\boldsymbol{\beta}_0)$, and $n^{-1}\sum_{i=1}^{n}a_{ij}(\boldsymbol{\beta}_0)Y_{ij}(s)$ converges to $E[a_{ij}(\boldsymbol{\beta}_0)Y_{ij}(s)]$. Therefore, by

the SLLN, $\widehat{h}_j(t)$ converges to $h_j(t)$ where

$$h_j(t) = E\left[\int_0^t \left\{ \frac{a_{ij}(\boldsymbol{\beta})}{r_j^{(0)}(s,\boldsymbol{\beta})} - \frac{E[a_{ij}(\boldsymbol{\beta})Y_{ij}(s)]w_{ij}(\boldsymbol{\beta})}{r_j^{(0)}(s,\boldsymbol{\beta})^2} \right\} dN_{ij}(s)\right]$$

By a Taylor expansion, one can obtain that

$$\begin{aligned} n^{-1/2}U(\boldsymbol{\beta}_0) &= \widehat{\Omega}(\boldsymbol{\beta}_0)n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1) \\ \widehat{\Omega}(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\mathbf{X}_{ij}p_{ij}(\boldsymbol{\beta})\left[\mathbf{X}_{ij}^T - \sum_{k=1}^{J}\mathbf{X}_{ik}^T p_{ik}\right]. \end{aligned}$$

By using the SLLN, $\widehat{\Omega}(\boldsymbol{\beta}_0)$ converges to $\Omega(\boldsymbol{\beta}_0)$ defined as in (3.5). Since $U(\boldsymbol{\beta}_0)$ can be written as in (3.1), we obtain that

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \Omega(\boldsymbol{\beta}_0)^{-1}n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{J}\mathbf{X}_{ij}\left[G_{ij} - p_{ij}(\boldsymbol{\beta}_0)\right] + o_p(1) \tag{3.9}$$

Using results (3.8), (3.9) and the fact that $\widehat{h}_j(t) \xrightarrow{a.s.} h_j(t)$, we obtain that

$$\widehat{\alpha}_{j1}(t) = n^{-1/2}\sum_{i=1}^{n} h_j(t)^T\Omega(\boldsymbol{\beta}_0)^{-1}\sum_{j=1}^{J}\mathbf{X}_{ij}\left[G_{ij} - p_{ij}(\boldsymbol{\beta}_0)\right] + o_p(1). \tag{3.10}$$

The quantity $\widehat{\alpha}_{j2}(t)$ can be written as

$$\widehat{\alpha}_{j2}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{w_{ij}(\boldsymbol{\beta}_0)}{R_j^{(0)}(s,\boldsymbol{\beta}_0)}dM_{ij}(s)$$

where $dM_{ij}(s) = dN_{ij}(s) - Y_{ij}(s)d\Lambda_j(s)$. Using the fact that $R_j^{(0)}(s,\boldsymbol{\beta}_0) \xrightarrow{a.s} r_j^{(0)}(s,\boldsymbol{\beta})$, we obtain that

$$\widehat{\alpha}_{j2}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{w_{ij}(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0)}dM_{ij}(s) + o_p(1) \tag{3.11}$$

Based on (3.10) and (3.11), one can obtain that

$$\begin{aligned} & n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}}) - \Lambda_j(t)\right] \\ = \; & n^{-1/2}\sum_{i=1}^{n}\left[h_j(t)^T\Omega(\boldsymbol{\beta}_0)^{-1}\sum_{j=1}^{J}\mathbf{X}_{ij}\left[G_{ij} - p_{ij}(\boldsymbol{\beta}_0)\right] + \int_0^t \frac{w_{ij}(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0)}dM_{ij}(s)\right] + o_p(1) \end{aligned}$$

By a Taylor series expansion, we obtain that

$$n^{1/2}\left[\frac{1}{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})}-\frac{1}{\Lambda_0(t)}\right]=-n^{1/2}\frac{1}{\Lambda_0(t)^2}\left[\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})-\Lambda_0(t)\right]+o_p(1).$$

Therefore, one can write

$$n^{1/2}\left[\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}})-\phi_j(t)\right]=n^{-1/2}\sum_{i=1}^{n}\left[\frac{1}{\Lambda_0(t)}\Phi_{ij}(t,\boldsymbol{\beta}_0)-\frac{\Lambda_j(t)}{\Lambda_0(t)^2}\Phi_{i0}(t,\boldsymbol{\beta}_0)\right]+o_p(1),$$

where

$$\Phi_{ij}(t,\boldsymbol{\beta}_0)=h_j(t)^T\Omega(\boldsymbol{\beta}_0)^{-1}\sum_{j=1}^{J}\mathbf{X}_{ij}\left[G_{ij}-p_{ij}(\boldsymbol{\beta}_0)\right]+\int_0^t\frac{w_{ij}(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0)}dM_{ij}(s).$$

Since $E[\Phi_{ij}(t,\boldsymbol{\beta}_0)]=0$ for $j=0,\cdots,J$, $n^{1/2}\left[\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}})-\phi_j(t)\right]$ is a scaled sum of n independent and identically distributed mean 0 random variates and therefore converges to a mean 0 Normal distribution, for fixed t. Furthermore, since $\left[\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}})-\phi_j(t)\right]$ is tight, $n^{1/2}\left[\widehat{\phi}_j(t,\widehat{\boldsymbol{\beta}})-\phi_j(t)\right]$ converges to a zero mean Gaussian process with covariance function $E[\xi_{ij}^{\phi}(s,\boldsymbol{\beta}_0)\xi_{ij}^{\phi}(t,\boldsymbol{\beta}_0)]$ for a pair (s,t), where

$$\begin{aligned}\xi_{ij}^{\phi}(t,\boldsymbol{\beta}) &= \frac{1}{\Lambda_0(t)}\Phi_{ij}(t,\boldsymbol{\beta})-\frac{\Lambda_j(t)}{\Lambda_0(t)^2}\Phi_{i0}(t,\boldsymbol{\beta})\\ \Phi_{ij}(t,\boldsymbol{\beta}) &= h_j(t)^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta})\sum_{j=1}^{J}\mathbf{X}_{ij}(G_{ij}-p_{ij})+\int_0^t\frac{w_{ij}(\boldsymbol{\beta})}{r_j^{(0)}(s,\boldsymbol{\beta})}dM_{ij}(s).\end{aligned}$$

Proof of Theorem 2

Since $\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})$ converges almost surely to $\Lambda_j(t)$, by the continuous mapping theorem, $\widehat{RR}_j(t,\widehat{\boldsymbol{\beta}})$ converges almost surely to $RR_j(t)$. With respect to asymptotic normality, one can write

$$\begin{aligned}n^{1/2}[\widehat{RR}_j(t,\widehat{\boldsymbol{\beta}})-RR_j(t)] &= n^{1/2}\left\{\frac{F_j(t,\widehat{\boldsymbol{\beta}})}{\widehat{F}_0(t,\widehat{\boldsymbol{\beta}})}-\frac{F_j(t)}{F_0(t)}\right\}\\ &= n^{1/2}\frac{1}{\widehat{F}_0(t,\widehat{\boldsymbol{\beta}})}\left[S_j(t)-\widehat{S}_j(t,\widehat{\boldsymbol{\beta}})\right]+n^{1/2}\left[\frac{F_j(t)}{\widehat{F}_0(t,\widehat{\boldsymbol{\beta}})}-\frac{F_j(t)}{F_0(t)}\right]\end{aligned}$$

By Functional Delta Method,

$$\begin{aligned} n^{1/2}\left[S_j(t)-\widehat{S}_j(t,\widehat{\boldsymbol{\beta}})\right] &= S_j(t)n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})-\Lambda_j(t)\right]+o_p(1) \\ n^{1/2}\left[\frac{1}{\widehat{F}_0(t,\widehat{\boldsymbol{\beta}})}-\frac{1}{F_0(t)}\right] &= \frac{-S_0(t)}{[F_0(t)]^2}n^{1/2}\left[\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})-\Lambda_0(t)\right]+o_p(1). \end{aligned}$$

Above listed results give

$$\begin{aligned} & n^{1/2}[\widehat{RR}_j(t,\widehat{\boldsymbol{\beta}})-RR_j(t)] \\ = \; & \frac{S_j(t)}{F_0(t)}n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})-\Lambda_j(t)\right]-\frac{F_j(t)S_0(t)}{[F_0(t)]^2}n^{1/2}\left[\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}})-\Lambda_0(t)\right]+o_p(1), \end{aligned}$$

which is a sum of mean 0 variates, using the fact that $n^{1/2}[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})-\Lambda_j(t)] = n^{-1/2}\sum_{i=1}^{n}\Phi_{ij}(t,\boldsymbol{\beta}_0)+o_p(1)$ for $j=0,\cdots,J$, shown in the proof of Theorem 1. Multivariate Central Limit Theorem and results from empirical process can be used, as in the proof of Theorem 1, to demonstrate that $n^{1/2}\left[\widehat{RR}_j(t,\widehat{\boldsymbol{\beta}})-RR_j(t)\right]$ converges asymptotically to a mean 0 Gaussian process.

Proof of Theorem 3

Using the fact that $\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}})$ converges almost surely to $\Lambda_j(t)$ and the continuous mapping theorem, we obtain that $\widehat{\Delta}_j(t,\widehat{\boldsymbol{\beta}})$ converges almost surely to $\Delta_j(t)$. With respect to the asymptotic normality, one can write

$$\begin{aligned} & n^{1/2}\left[\widehat{\Delta}_j(t,\widehat{\boldsymbol{\beta}})-\Delta_j(t)\right] \\ = \; & n^{1/2}\int_0^t\left[\widehat{S}_j(s,\widehat{\boldsymbol{\beta}})-\widehat{S}_0(s,\widehat{\boldsymbol{\beta}})\right]ds-n^{1/2}\int_0^t[S_j(s)-S_0(s)]\,ds \\ = \; & n^{1/2}\int_0^t\left[\widehat{S}_j(s,\widehat{\boldsymbol{\beta}})-S_j(s)\right]ds-n^{1/2}\int_0^t\left[\widehat{S}_0(s,\widehat{\boldsymbol{\beta}})-S_0(s)\right]ds. \end{aligned}$$

By Taylor expansions and similar techniques as in the proof of Theorem 1, one can

write

$$\begin{aligned} & n^{1/2}\left[\widehat{\Delta}_j(t,\widehat{\boldsymbol{\beta}})-\Delta_j(t)\right] \\ = \; & n^{1/2}\int_0^t -S_j(s)\left[\widehat{\Lambda}_j(s,\widehat{\boldsymbol{\beta}})-\Lambda_j(s)\right]+S_0(s)\left[\widehat{\Lambda}_0(s,\widehat{\boldsymbol{\beta}})-\Lambda_0(s)\right]ds+o_p(1) \\ = \; & n^{-1/2}\sum_{i=1}^{n}\int_0^t -S_j(s)\Phi_{ij}(s)+S_0(s)\Phi_{i0}(s)ds+o_p(1), \end{aligned}$$

which is a sum of n independent and identically distributed mean 0 random variables. Similar to the proof of Theorem 1, one can obtain that $n^{1/2}[\widehat{\Delta}_j(\widehat{\boldsymbol{\beta}},t)-\Delta_j(t)]$ converges to a mean 0 Gaussian process with covariance function $E[\xi_{ij}^{\Delta}(s,\boldsymbol{\beta}_0)\xi_{ij}^{\Delta}(t,\boldsymbol{\beta}_0)]$ defined as in (3.7).

3.9 References

Aalen O.O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics* **6**, 701-726.

Anstrom K.J., Tsiatis, A.A. (2001). Utilizing propensity scores to estimate causal treatment effects with censored time-lagged data. *Biometrics* **57**, 1207-1218.

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**, 662-682.

Bloembergen, W.E., Port, F.K., Mauger, E.A., and Wolfe, R.A. (1995). A comparison of mortality between patients treated with hemodialysis and peritoneal dialysis. *Journal of American Society of Nephrology* **6**, 177-183.

Chen, P. and Tsiatis A.A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57** 1030-1038.

Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34**, 187-202.

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

Dabrowska, D.M., Doksum, K.A. and Song, J. (1989). Graphical comparison of cumulative hazards for two populations. *Biometrika* **76**, 763-773.

Fenton, S.S., 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**, 334-342.

Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Hernan, M.A., Brumback, B. and Robins, J.M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* **11**, 561-570.

Hernan, M.A., Brumback, B. and Robins, J.M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* **96**, 440-448.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481.

Lin, D.Y., Fleming, T.R. and Wei, L.J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73-81.

McKeague, I.W. and Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statistics & Probability Letters* **60**, 405-415.

Nelson, W.B. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945-965.

Parzen, M.I., Wei, L.J. and Ying, Z. (1997). Simultaneous confidence intervals for the difference of the two survival functions. *Scandinavian Journal of Statistics* **24**, 309-314.

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

Robins, J.M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology – Methodological Issues*, Jewell N, Dietz K, Farewell V (eds). Birkhauser: Boston, MA, 297-331.

Robins, J.M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association,* 24-33.

Robins J.M., Hernan, M.A. and Brumback, B. (2000). Marginal Structrural Models and Causal Inference in Epidemiology. *Epidemiology* **11**, 550-560.

Wei, G. and Schaubel D.E. (2008). Estimating cumulative treatment effects in the presence of non-proportional hazards. *Biometrics* **64**, 724 - 732.

Xie, J. and Liu, C. (2005). Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine* **24**, 3089-3110.

CHAPTER IV

Estimating Cumulative Treatment Effects In The Presence Of Non-Proportional Hazards And Dependent Censoring

ABSTRACT: In medical studies of time to event data, non-proportional hazards and dependent censoring are very common issues when estimating the treatment effect. A traditional method for dealing with time-dependent treatment effects is to model the time-dependence parametrically. Limitations of this approach include the difficulty to verify the correctness of the specified functional form and the fact that, in the presence of a treatment effect that varies over time, investigators are usually interested in the cumulative as opposed to instantaneous treatment effect. In many applications, censoring time is not independent of event time. Therefore, we propose methods for estimating the cumulative treatment effect in the presence of non-proportional hazards and dependent censoring. As in Chapter 3, the three proposed measures include the ratio of cumulative hazards, relative risk and difference in restricted mean lifetime. For each measure, we propose a double-inverse-weighted estimator, constructed by first using inverse probability of treatment weighting (IPTW) to balance the treatment-specific covariate distributions, then using inverse probability of censoring weighting (IPCW) to handle the dependent censoring. The proposed estimators are consistent and aymptotically normal. We study their finite-sample properties through simulation. The proposed methods are used to compare kidney

wait list mortality by race.

KEY WORDS: Cumulative hazard; Dependent censoring; Inverse weighting; Relative Risk; Restricted mean lifetime; Survival analysis; Treatment effect.

4.1 Introduction

In clinical and epidemiologic studies of survival data, it is very common that the treatment effect is not constant over time. In the presence of non-proportional hazards, the Cox (1972) model is often modified such that the treatment effect is assumed to vary as a specified function of time. However, the functional form chosen may not be correct. Moreover, researchers are usually more interested in the cumulative treatment effect in settings where the treatment effect is time-dependent. Without specifying the functional form of the treatment effect, one can compare survival or cumulative hazard curves using the Nelson-Aalen (Nelson, 1972; Aalen, 1978) estimator or Kaplan-Meier (Kaplan and Meier, 1958) estimator. These estimators may lead to biased results when confounders for treatment exist, as is often the case in observational studies. When censoring time depends on factors predictive of the event, the event and censoring time are correlated through these factors. If these prognostic factors are time-dependent and if they are not only risk factors for the event but also affected by treatment, standard methods of covariate adjustment (such as Cox regression) may produce biased treatment effects. If baseline values instead of time-dependent factors are adjusted, standard methods are still invalid since the event and censoring times will be dependent through their mutual correlation with the time-dependent factors.

The investigation which motivated our proposed research involves comparing wait-list survival for patients with end-stage renal disease. The effect of race (Caucasian

vs African American) on survival is of interest and may vary over time. A patient's hospitalization history is a predictor of wait-list mortality and also affects transplantation probability, since patients with more hospitalizations are less likely to receive a kidney transplant. Although a patient's death may be observed following kidney transplantation, receipt of a transplant does censor their wait-list mortality. Therefore, the mortality and censoring will be correlated unless the model adjusts for hospitalization history. However, one would not want to adjust for hospitalization history, since doing so may result in the marginal effect of race being under or over estimated. Therefore, we need to handle dependent censoring in this analysis. In addition, there are some time constant covariates, such as age and diagnosis, for which adjustment is necessary.

Current methods usually focus on the survival or cumulative hazard function when estimating cumulative treatment effects in the presence of censored data. However, mean lifetime is often relevant since patients usually want to know how long they will live. Chen & Tsiatis (2001) compared restricted mean lifetime between two treatment groups using treatment-specific Cox proportional hazard models. The survival function for each group was estimated by averaging over all subjects in the sample. Their proposed model requires that proportionality holds for the adjustment covariates.

Without specifying the functional form for the effects of adjustment covariates, inverse probability of treatment weighting (IPTW) can be applied to balance the distribution of confounders among the treatment groups. Hernan, Brumback & Robins (2000, 2001) and Robins, Hernan & Brumback (2000) used marginal structural models to estimate the causal effect of a time-dependent exposure. Inverse weighting was applied to adjust for time-dependent confounders that are affected by previous

treatment. In the context of survival analysis, the authors assumed Cox proportional hazard models; i.e., the effect of treatment is assumed to be constant. With respect to the related nonparametric methods, Xie and Liu (2005) developed an adjusted Kaplan-Meier curve using inverse weighting to handle potential confounders, assuming that the event time and censoring time are independent.

Inverse probability of censoring weighting (IPCW) has been applied in many applications to overcome dependent censoring. This method is originally proposed by Robins and Rotnitzky (1992). A Cox proportional hazard model is assumed for the event time, while an inverse probability of censoring weight is applied in the estimating equation for the effect parameters. This weight is the inverse of the survival function for censoring, which is estimated by non-parametric or semiparametric Kaplan-Meier estimators from a Cox model. Robins and Finkelstein (2000) applied IPCW to handle dependent censoring in an AIDS clinical trial. The IPCW method has been applied in various other settings (Matsuyama & Yamaguchi, 2008; Yoshida, Matsuyama & Ohashi, 2007).

We propose three cumulative treatment effect measures: ratio of cumulative hazards, relative risk, and difference in restricted mean lifetime. The proposed estimators are computed by double inverse weighting, wherein inverse probability of treatment weighting (IPTW; Hernan, Brumback & Robins, 2000, 2001; Robins, Hernan & Brumback, 2000) is used to balance the treatment-specific baseline adjustment covariate distributions and IPCW (Robins and Rotnitzky 1992; Robins 1993) is concurrently applied to handle the dependent censoring due to time-varying factors. After applying the double inverse weight to the observed data, estimation of the cumulative treatment effects proceeds nonparametrically, negating the need to specify functional forms for the effect of either the treatment or the adjustment covariates.

The remainder of this chapter is organized as follows. We describe our proposed methods in the next section. In Section 4.3, we derive the asymptotic properties of our proposed estimators. We evaluate the performance of our estimators for finite samples in Section 4.4. In Section 4.5, we apply the methods to kidney wait list data. Discussion is provided in Section 4.6.

4.2 Proposed methods

Suppose that n subjects are included in the data set. Let D_i be the event time and C_i be the censoring time for subject i. Let $X_i = \min\{D_i, C_i\}$ and $\delta_i = I(D_i \leq C_i)$ where $I(A)$ is an indicator function taking the value 1 when condition A holds and 0 otherwise. The observed event counting processes is defined as $N_i^D(t) = \delta_i I(X_i \leq t)$ and the observed censoring counting processes is defined as $N_i^C(t) = (1-\delta_i)I(X_i \leq t)$. The risk indicator is denoted by $Y_i(t) = I(X_i \geq t)$. Let j $(j = 0, 1, \ldots, J)$ be the index for treatment group, with group $j = 0$ representing a reference category to which the remaining treatment groups are compared. Let G_i denote the treatment group for subject i and set $G_{ij} = I(G_i = j)$. Correspondingly, we set $Y_{ij}(t) = Y_i(t)G_{ij}$, $dN_{ij}^D(t) = dN_i^D(t)G_{ij}$ and $dN_{ij}^C(t) = dN_i^C(t)G_{ij}$. The observed data consist of n independent and identically distributed vectors, $(X_i, \delta_i, G_i, \widetilde{\mathbf{Z}}_i^T(X_i))^T$, where $\widetilde{\mathbf{Z}}_i(t) = \{\mathbf{Z}_i(s); s \in [0, t]\}$ and $\mathbf{Z}_i(t)$ is a $p \times 1$ vector of covariates which may contain some time-dependent elements. We let $\mathbf{Z}_i(0)$ denote the covariate values at baseline.

In the case where $\mathbf{Z}_i(t)$ not only affects the event time but also affects censoring, event and censoring are dependent unless the effect of $\mathbf{Z}_i(t)$ on the event is modeled explicitly. However, we usually would prefer adjust for $\mathbf{Z}_i(0)$, instead of $\mathbf{Z}_i(t)$ when $\mathbf{Z}_i(t)$ is affected by treatment. It is of interest to compare the average survival that would result if treatment j was assigned to the entire population to that if

reference treatment was assigned to the entire population. The average survival function if treatment j was assigned to is given by $S_j(t) = E[S(t|G_i = j, \mathbf{Z}_i(0))]$. The expectation is with respect to the marginal distribution of $\mathbf{Z}_i(0)$, such that the same averaging is done across all $J+1$ treatment groups.

To compare the cumulative effect of treatment j to the reference treatment, three measures are proposed. The first proposed measure is the ratio of cumulative hazards,

$$\phi_j(t) = \frac{\Lambda_j(t)}{\Lambda_0(t)}, \tag{4.1}$$

where $\Lambda_j(t) = -\log\{S_j(t)\}$ is the cumulative hazard at time t. The measure $\phi_j(t)$ is equal to the hazard ratio we usually use when proportional hazard holds. The second proposed measure is the ratio of cumulative distribution function,

$$RR_j(t) = \frac{F_j(t)}{F_0(t)}, \tag{4.2}$$

where $F_j(t) = 1 - S_j(t)$ is the probability of death by time t. The $RR_j(t)$ is a process version of relative risk, a quantity which is frequently estimated in epidemiologic studies. The third proposed measure is the difference in restricted mean lifetime,

$$\Delta_j(t) = e_j(t) - e_0(t), \tag{4.3}$$

where $e_j(t) = \int_0^t S_j(u)du$ is the area under the survival curve (restricted mean lifetime) through $(0, t]$. The $\Delta_j(t)$ measure equals the area between the survival curves that would result if treatment j versus treatment 0 was assigned to all subjects in the population.

Let $\mathbf{X}_{ij} = [\mathbf{0}_{1\times(j-1)(p+1)}, 1, \mathbf{Z}_i^T(0), \mathbf{0}_{1\times(J-j)(p+1)}]^T$ where $\mathbf{0}_{1\times(j-1)(p+1)}$ is a 1 by $(j-1)(p+1)$ matrix with elements 0, for $j = 1, \cdots, J$. We assume that treatment assignment follows a generalized logit model,

$$\log\left\{\frac{p_{ij}(\boldsymbol{\beta}_0)}{p_{i0}(\boldsymbol{\beta}_0)}\right\} = \boldsymbol{\beta}_0^T \mathbf{X}_{ij},$$

where $p_{ij}(\boldsymbol{\beta}_0) = Pr(G_i = j|\mathbf{Z}_i(0))$. The model could be extended to include interaction terms. The maximum likelihood estimator of $\boldsymbol{\beta}_0$, denoted by $\widehat{\boldsymbol{\beta}}$, is the root of $U_G(\boldsymbol{\beta}) = 0$, where

$$U_G(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} \mathbf{X}_{ij}\left[G_{ij} - p_{ij}(\boldsymbol{\beta})\right]. \tag{4.4}$$

Since it is preferred to adjust for $\mathbf{Z}_i(0)$ instead of $\mathbf{Z}_i(t)$, the event and censoring processes are dependent through their mutual association with $\mathbf{Z}_i(t)$ for $t > 0$. We apply an inverse probability of censoring weight to handle the dependent censoring. We assume that C_i follows a Cox model with hazard function defined as:

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

where $\mathbf{Z}_i^C(t)$ contains terms representing G_i and $\mathbf{Z}_i(t)$. The inverse censoring weight at time t is denoted by $w_i^C(t,\boldsymbol{\theta}_0) = Y_i(t)\exp\{\Lambda_i^C(t)\}$, where $\Lambda_i^C(t) = \int_0^t \lambda_i^C(s)ds$. The quantity $w_i^C(t,\boldsymbol{\theta}_0)$ is estimated by $\widehat{w}_i^C(t,\widehat{\boldsymbol{\theta}}) = Y_i(t)\exp\{\widehat{\Lambda}_i^C(t,\widehat{\boldsymbol{\theta}})\}$, where

$$\begin{aligned}
\widehat{\Lambda}_i^C(t,\boldsymbol{\theta}) &= \int_0^t Y_i(s)\exp\{\boldsymbol{\theta}^T\mathbf{Z}_i^C(s)\}d\widehat{\Lambda}_0^C(s,\boldsymbol{\theta}),\\
\widehat{\Lambda}_0^C(t,\boldsymbol{\theta}) &= \frac{1}{n}\int_0^t \frac{dN_i^C(s)}{R_C^{(0)}(s,\boldsymbol{\theta})},
\end{aligned}$$

and $\mathbf{R}_C^{(d)}(t,\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n Y_i(t)\mathbf{Z}_i^C(t)^{\otimes d}\exp\{\boldsymbol{\theta}^T\mathbf{Z}_i^C(t)\}$ for $d = 0,1,2$ with $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for a vector $\mathbf{a}$. The quantity $\Lambda_0^C(t,\widehat{\boldsymbol{\theta}})$ is the Nelson-Aalen estimator for $\Lambda_0^C(t)$. The parameter $\boldsymbol{\theta}_0$ is estimated through partial likelihood (Cox, 1975) by $\widehat{\boldsymbol{\theta}}$, the root of score equation $U_C(\boldsymbol{\theta}) = 0$, where

$$\begin{aligned}
\mathbf{U}_C(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\int_0^\tau \left\{\mathbf{Z}_i^C(t) - \overline{\mathbf{Z}}^C(t,\boldsymbol{\theta})\right\} dN_i^C(t), \qquad (4.5)\\
\overline{\mathbf{Z}}^C(t,\boldsymbol{\theta}) &= \frac{\mathbf{R}_C^{(1)}(t,\boldsymbol{\theta})}{R_C^{(0)}(t,\boldsymbol{\theta})}.
\end{aligned}$$

Note that if proportional hazards does not hold for treatment effect, a stratified Cox model can be applied to obtain treatment-specific survival functions for the censoring distribution.

The measure $\phi_j(t)$ is estimated by

$$\widehat{\phi_j}(t) = \frac{\widehat{\Lambda}_j(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}}))}{\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}}))},$$

where $\widehat{w}^C(t, \boldsymbol{\theta}) = \{\widehat{w}_i^C(s, \boldsymbol{\theta}), i = 1, \cdots, n, s \in [0, t]\}$ and

$$\begin{aligned}
\widehat{\Lambda}_j(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}})) &= \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{\widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) w_{ij}^G(\widehat{\boldsymbol{\beta}})}{R_j^{(0)}(s, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(s, \widehat{\boldsymbol{\theta}}))} dN_{ij}^D(s), \\
R_j^{(0)}(s, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(s, \widehat{\boldsymbol{\theta}})) &= \frac{1}{n} \sum_{i=1}^{n} Y_{ij}(s) \widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) w_{ij}^G(\widehat{\boldsymbol{\beta}}),
\end{aligned}$$

with $w_{ij}^G(\boldsymbol{\beta}) = G_{ij}/p_{ij}(\boldsymbol{\beta})$. The relative risk measure, $RR_j(t)$, is estimated by

$$\widehat{RR_j}(t) = \frac{\widehat{F}_j(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}}))}{\widehat{F}_0(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}}))},$$

where $\widehat{F}_j(t, \boldsymbol{\beta}, \widehat{w}^C(t, \boldsymbol{\theta})) = 1 - \widehat{S}_j(t, \boldsymbol{\beta}, \widehat{w}^C(t, \boldsymbol{\theta}))$. The estimator for difference in restricted mean lifetime, $\Delta_j(t)$, is given by

$$\widehat{\Delta}_j(t) = \widehat{e}_j(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}})) - \widehat{e}_0(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}})),$$

where $\widehat{e}_j(t, \boldsymbol{\beta}, \widehat{w}^C(t, \boldsymbol{\theta})) = \int_0^t \widehat{S}_j(s, \boldsymbol{\beta}, \widehat{w}^C(t, \boldsymbol{\theta}))ds$.

The measures $\phi_j(t)$ and $RR_j(t)$ are considered on a time interval $[t_L, t_U]$, and $\Delta_j(t)$ is considered in a time interval $[0, t_U]$, where t_L is chosen to avoid division by 0 and t_U is chosen to avoid the well known instability in the tail of the observation time distribution.

The IPTW weight, $w_{ij}^G(\boldsymbol{\beta}_0)$, is used for balancing the covariate distribution among the treatment groups. After applying $w_{ij}^G(\boldsymbol{\beta}_0)$ to our estimators, $J + 1$ pseudo-populations are created with treatment-specific $\mathbf{Z}_i(0)$ distribution equals to that of the entire population. For example, for the restricted mean lifetime $e_j(t) = \int_0^t E[S(s|G_i = j, \mathbf{Z}_i(0))]ds$, the expectation is with respect to the marginal distribution of $\mathbf{Z}_i(0)$, such that same averaging is done across all $J + 1$ treatment groups.

The IPCW weight, $w_i^C(t, \boldsymbol{\theta}_0)$, is applied to handle the dependent censoring. After applying the proposed double inverse weighting, the weighted versions of each of the proposed measures converges to the same true values listed in Chapter 3.

4.3 Asymptotic properties

To derive the large-sample properties of estimators proposed in last section, we assume the following regularity conditions for $i = 1, \cdots, n$ and $j = 0, \cdots, J$.

(a) $\Lambda_j(\tau) < \infty$ and $\Lambda_i^C(\tau) < \infty$, where τ is a pre-specified time point.

(b) $Z_{ik}(t)$ are bounded for $t \in [0, \tau]$; i.e., $|Z_{ik}(t)| < \kappa$ for $k = 1, \cdots, p$, where κ is a constant and $Z_{ik}(t)$ is the kth component of $\mathbf{Z}_i(t)$.

(c) Continuity of the following functions:

$$\mathbf{r}_C^{(1)}(t, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} r_C^{(0)}(t, \boldsymbol{\theta}),\ \mathbf{r}_C^{(2)}(t, \boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} r_C^{(0)}(t, \boldsymbol{\theta}),$$

where $\mathbf{r}_C^{(d)}(t, \boldsymbol{\theta})$ is the limiting value of $\mathbf{R}_C^{(d)}(t, \boldsymbol{\theta})$ for $d = 0, 1, 2$, with $r_C^{(0)}(t, \boldsymbol{\theta})$ bounded away from 0 for $t \in [0, \tau]$ and $\boldsymbol{\theta}$ in an open set.

(d) For $(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T$ in an open set, the quantity $r_j^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\theta})$ is continuous and bounded away from 0 in $[0, \tau]$, where $r_j^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\theta})$ is the limiting value of $R_j^{(0)}(t, \boldsymbol{\beta}, w^C(t, \boldsymbol{\theta}))$.

(e) Positive-definiteness of the matrices $\boldsymbol{\Omega}_G(\boldsymbol{\beta})$ and $\boldsymbol{\Omega}_C(\boldsymbol{\theta})$, where

$$\boldsymbol{\Omega}_G(\boldsymbol{\beta}) = E\left\{\sum_{j=1}^{J} p_{ij}(\boldsymbol{\beta}) \mathbf{X}_{ij} \left[\mathbf{X}_{ij}^T - \sum_{k=1}^{J} \mathbf{X}_{ik}^T p_{ik}(\boldsymbol{\beta})\right]\right\}, \tag{4.6}$$

$$\begin{aligned} \boldsymbol{\Omega}_C(\boldsymbol{\theta}) &= \sum_{j=0}^{m} \int_0^{\tau} \mathbf{v}(t, \boldsymbol{\theta}) r_C^{(0)}(t, \boldsymbol{\theta}) \lambda_0^C(t) dt, \\ \mathbf{v}(t, \boldsymbol{\theta}) &= \mathbf{r}_C^{(2)}(t, \boldsymbol{\theta}) / r_C^{(0)}(t, \boldsymbol{\theta}) - \overline{\mathbf{z}}^C(t, \boldsymbol{\theta})^{\otimes 2}, \end{aligned} \tag{4.7}$$

and $\overline{\mathbf{z}}^C(t, \boldsymbol{\theta}) = \mathbf{r}_C^{(1)}(t, \boldsymbol{\theta}) / r_C^{(0)}(t, \boldsymbol{\theta})$ is the limiting value of $\overline{\mathbf{Z}}^C(t, \boldsymbol{\theta})$.

(f) $Pr(G_{ij} = 1) > 0$.

We summarize the asymptotic properties of the proposed estimators in the following theorems.

THEOREM 1. *Under conditions (a) to (f),* $\widehat{\phi}_j(t)$ *converges almost surely and uniformly to* $\phi_j(t)$ *for* $t \in [t_L, t_U]$, *and* $n^{1/2}[\widehat{\phi}_j(t) - \phi_j(t)]$ *converges asymptotically to a zero-mean Gaussian process with covariance function* $\sigma_j^{\phi}(s,t) = E[\xi_{ij}^{\phi}(s, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)\xi_{ij}^{\phi}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)]$, *where*

$$
\begin{aligned}
\xi_{ij}^{\phi}(t,\boldsymbol{\beta},\boldsymbol{\theta}) &= \frac{1}{\Lambda_0(t)}\Phi_{ij}(t,\boldsymbol{\beta},\boldsymbol{\theta}) - \frac{\Lambda_j(t)}{\Lambda_0(t)^2}\Phi_{i0}(t,\boldsymbol{\beta},\boldsymbol{\theta}) \\
\Phi_{ij}(t,\boldsymbol{\beta},\boldsymbol{\theta}) &= \Phi_{ij1}(t) + \Phi_{ij2}(t) + \Phi_{ij3}(t) + \Phi_{ij4}(t) \\
\Phi_{ij1}(t) &= h_j(t)^T \Omega_G^{-1}(\boldsymbol{\beta})^{-1} \sum_{j=1}^{J} \mathbf{X}_{ij}\left[G_{ij} - p_{ij}(\boldsymbol{\beta})\right] \\
h_j(t) &= \int_0^t \frac{E[w_i^C(s,\boldsymbol{\theta})a_{ij}(\boldsymbol{\beta})dN_{ij}^D(s)]}{r_j^{(0)}(s,\boldsymbol{\beta},\boldsymbol{\theta})} \\
a_{ij}(\boldsymbol{\beta}) &= I(G_i \neq 0)\left[\frac{\sum_{k=1}^{J}\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ik}\}\mathbf{X}_{ik}}{\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ij}\}} - \mathbf{X}_{ij}p_{ij}^{-1}(\boldsymbol{\beta})\right] + G_{i0}\sum_{k=1}^{J}\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ik}\}\mathbf{X}_{ik} \\
\Phi_{ij2}(t) &= g_j(t)^T\Omega_C^{-1}(\boldsymbol{\theta})\int_0^{\tau}[\mathbf{Z}_i^C(t) - \overline{\mathbf{z}}_i^C(t)]dM_i^C(t) \\
dM_i^C(t) &= dN_i^C(t) - Y_i(t)d\Lambda_i^C(t) \\
g_j(t) &= \int_0^t \frac{E\left[b_i(s,\boldsymbol{\theta})w_{ij}^G(\boldsymbol{\beta})dN_{ij}^D(s)\right]}{r_j^{(0)}(s,\boldsymbol{\beta},\boldsymbol{\theta})} \\
b_i(s,\boldsymbol{\theta}) &= w_i^C(s,\boldsymbol{\theta})\int_0^s Y_i(u)\exp\{\boldsymbol{\theta}^T Z_i^C(u)\}\left[Z_i^C(u) - \overline{\mathbf{z}}^C(t)\right]d\Lambda_0^C(u) \\
\Phi_{ij3}(t) &= \int_0^t E\left[\frac{\exp\{\boldsymbol{\theta}^T\mathbf{Z}_k^C(u)\}}{r_C^{(0)}(u,\boldsymbol{\theta})}\int_u^t \frac{w_k^C(s,\boldsymbol{\theta})w_{kj}^G(\boldsymbol{\beta})}{r_j^{(0)}(s,\boldsymbol{\beta},\boldsymbol{\theta})}dN_{kj}^D(s)\right]dM_i^C(u) \\
\Phi_{ij4}(t) &= \int_0^t \frac{w_i^C(s,\boldsymbol{\theta})w_{ij}^G(\boldsymbol{\beta})}{r_j^{(0)}(s,\boldsymbol{\beta},\boldsymbol{\theta})}dM_{ij}^D(s) \\
dM_{ij}^D(s) &= dN_{ij}^D(s) - Y_{ij}(s)d\Lambda_j(s)
\end{aligned}
$$

THEOREM 2. *Under conditions (a) to (f),* $\widehat{RR}_j(t)$ *converges almost surely and uniformly to* $RR_j(t)$ *for* $t \in [t_L, t_U]$, *and* $n^{1/2}[\widehat{RR}_j(t,\widehat{\boldsymbol{\theta}}) - RR_j(t)]$ *converges asymptotically to a zero-mean Gaussian process with covariance function* $\sigma_j^R(s,t) =$

$E[\xi_{ij}^R(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\xi_{ij}^R(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]$, *where*

$$\xi_{ij}^R(t,\boldsymbol{\beta},\boldsymbol{\theta}) = \frac{S_j(t)}{F_0(t)}\Phi_{ij}(t,\boldsymbol{\beta},\boldsymbol{\theta}) - \frac{F_j(t)S_0(t)}{F_0(t)^2}\Phi_{i0}(t,\boldsymbol{\beta},\boldsymbol{\theta}), \tag{4.8}$$

with $\Phi_{ij}(t,\boldsymbol{\beta},\boldsymbol{\theta})$ *defined as in Theorem 1.*

THEOREM 3. *Under conditions (a)-(f),* $\widehat{\Delta}_j(t)$ *converges almost surely and uniformly to* $\Delta_j(t)$ *for* $t \in [0, t_U]$, *and* $n^{1/2}[\widehat{\Delta}_j(t) - \Delta_j(t)]$, *converges asymptotically to a zero-mean Gaussian process with covariance function* $\sigma_j^{\Delta}(s,t) = E[\xi_{ij}^{\Delta}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\xi_{ij}^{\Delta}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]$, *where*

$$\xi_{ij}^{\Delta}(t,\boldsymbol{\beta},\boldsymbol{\theta}) = \int_0^t \{S_0(s)\Phi_{i0}(s,\boldsymbol{\beta},\boldsymbol{\theta}) - S_j(s)\Phi_{ij}(s,\boldsymbol{\beta},\boldsymbol{\theta})\}\, ds, \tag{4.9}$$

with $\Phi_{ij}(t,\boldsymbol{\beta},\boldsymbol{\theta})$ *defined as in Theorem 1.*

The consistency of $\widehat{\phi}_j(t)$, $\widehat{RR}_j(t)$ and $\widehat{\Delta}_j(t)$ is proved by the consistency of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$, the continuous mapping theorem, and the Uniform Strong Law of Large Numbers (SLLN). The proof of asymptotic normality involves decomposing $\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}})) - \Lambda_j(t)\right]$ into $\widehat{\alpha}_{j1}(t) + \widehat{\alpha}_{j2}(t) + \widehat{\alpha}_{j3}(t) + \widehat{\alpha}_{j4}(t)$, where

$$\begin{aligned}
\widehat{\alpha}_{j1}(t) &= \left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}})) - \widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))\right] \\
\widehat{\alpha}_{j2}(t) &= \left[\widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,\widehat{w}^C(t,\widehat{\boldsymbol{\theta}})) - \widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,w^C(t,\widehat{\boldsymbol{\theta}}))\right] \\
\widehat{\alpha}_{j3}(t) &= \left[\widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,w^C(t,\widehat{\boldsymbol{\theta}})) - \widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,w^C(t,\boldsymbol{\theta}_0))\right] \\
\widehat{\alpha}_{j4}(t) &= \left[\widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,w^C(t,\boldsymbol{\theta}_0)) - \Lambda_j(t)\right].
\end{aligned}$$

The quantity $n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}})) - \Lambda_j(t)\right]$ then can be written as sum of independent and identically distributed mean 0 variates. Using the Functional Delta Method, we can write each of $n^{1/2}[\widehat{\phi}_j(t) - \phi_j(t)]$, $n^{1/2}[\widehat{RR}_j(t) - RR_j(t)]$ and $n^{1/2}[\widehat{\Delta}_j(t) - \Delta_j(t)]$ as functions of $n^{1/2}[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}})) - \Lambda_j(t)]$. Therefore, $n^{1/2}[\widehat{\phi}_j(t) - \phi_j(t)]$,

$n^{1/2}[\widehat{RR}_j(t) - RR_j(t)]$ and $n^{1/2}[\widehat{\Delta}_j(t) - \Delta_j(t)]$ can be written as sum of independent and identically distributed mean 0 variates. The Multivariate Central Limit Theorem and various results from the theory of empirical processes (Pollard, 1990; Bilias, Gu &Ying, 1997) are applied in the proof. The covariance function can be consistently estimated by replacing all limiting values with their empirical counterparts. For large data set, variance computation can be computationally intense, in which case the bootstrap is a useful alternative.

4.4 Simulation study

We evaluated the finite sample properties of the proposed estimators through a series of simulation studies. Due to the computation of derived asymptotic variance, we evaluated the bootstrap variances. For each of the n subjects, a covariate Z_{i1} was generated as a binary variable with values 0 or 1 and $Pr(Z_{i1} = 1) = 0.5$. The treatment indicator, G_i, was generated from a Bernoulli distribution with parameter $p_{i1}(\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 Z_{i1})/\left[1 + \exp(\beta_0 + \beta_1 Z_{i1})\right]$. We chose $\beta_0 = \log(1/3)$ and $\beta_1 = \log(9)$ such that $Pr(Z_{i1} = 1|G_i = 1) = 0.75$ and $Pr(Z_{i1} = 1|G_i = 0) = 0.25$. When $G_i = 1$, we generated a variable Z_{i2} as piece-wise constant with probabilities $P(Z_{i2} = k) = P(Z_{i2} = k + 1) = 0.5$ across time interval $(k, k + 1]$, for $k = 0, \cdots, 4$. When $G_i = 0$, Z_{i2} was generated as a binary variable (0 or 1) with $Pr(Z_{i2} = 1) = 0.5$. Event times were generated from a Cox model with hazard function

$$\lambda_i(t) = \lambda_0(t) \exp\{\eta_1 G_i + \eta_2 Z_{i1} + \eta_3 Z_{i2}(t)\},$$

while censoring times were generated from a Cox model with hazard function

$$\lambda_i^C(t) = \lambda_0^C(t) \exp\{\theta Z_{i2}(t)\}.$$

Various values of (η_1, η_2, η_3), and θ were employed for the Cox models. For each set of parameters, several percentages of censoring were investigated by varying the baseline death and censoring hazards. Censoring times were truncated at $t = 5$.

Sample sizes of $n = 500$ and $n = 200$ were examined, and a total of 1000 simulations were used for each simulation setting. For the first two measures, we employed the log transform to ensure that the confidence interval bounds were in a valid range. To assess the finite-sample performance of our proposed method, the bias of each of the three estimators was evaluated at time points $t = 1$, $t = 2$ and $t = 3$. The bootstrap standard errors were evaluated at $t = 2$ with sample size $n = 200$, and 100 bootstrap resamples per simulation.

For $n = 200$, our estimators appear to be approximately unbiased in general (Table 4.1). The bias is reduced when sample size increases to $n = 500$ (Table 4.2). The average bootstrap standard errors (ASE) are genearlly close to the empirical standard deviations (ESD) for sample size $n = 200$ (Table 4.3) and, correspondingly, the empirical coverage probabilities (CP) are fairly close to the nominal value of 0.95.

4.5 Data analysis

We applied the proposed methods to analyze wait-list survival for patients with end-stage renal disease, where the effect of race (Caucasian vs. African American) was of interest. Data were obtained from the Scientific Registry of Transplant Recipients (SRTR) and collected by the Organ Procurement and Transplant Network (OPTN). Hospitalization data were obtained from the Center for Medicare Sciences (CMS). Only patients whose primary payer was Medicare were included in the analysis. The data included n=7110 Caucasian and African American patients who were placed on kidney transplant waiting list in calender year 2000. Among the 2975

African Americans, 27% died and 45% received a kidney transplant. Among the 4135 Caucasians, 27% died and 54% got transplanted. Patients were followed from the time of placement on the kidney transplant waiting list to the earliest of death, transplantation, loss to follow-up or end of study (Dec 31, 2005).

It has been reported that African Americans have lower kidney wait list mortality rate than Caucasians. However, Caucasians also have a higher kidney transplant rate than African Americans. Unlike liver, lung and heart transplantation, poor patient health is a contra-indication for kidney transplantation. Although donor kidneys are not specially directed towards healthier patients, it is generally felt that patients in poorer health are less likely to receive a kidney transplant. It is quite possible that the healthiest patients are transplanted off the wait list at a greater rate for Caucasians than for African Americans. Therefore, we suspect that dependent censoring exists of kidney wait list mortality via kidney transplantation. We use time-dependent hospitalization history as a surrogate for patient health. Note that hospitalization history is inappropriate as an adjustment covariate for patient wait list survival. Patients with a greater number of previous hospitalizations have a greater mortality hazard and hospital admissions can be viewed as intermediate end points along the path from wait listing to death. Previous comparisons of wait list mortality by race did not adjust for dependent censoring. Moreover, most previous comparisons of Caucasians and African Americans assumed that effect of race is constant over time.

Logistic regression was used to model the probability that a patient is Caucasian given age, gender, diagnosis (diabetes, hypertension, Glaucoma, polycystic kidney disease and other), body mass index and chronic obstructive lung disease (yes or no). A stratified Cox model (stratified by race) was fitted to estimate the inverse probability censoring weight adjusting for the covariates listed above, as well as

time-dependent number of hospitalizations. The transplant hazard is significantly decreased by 8% for each additional hospitalization (Table 4.4). The IPTW weight ranged from 1.04 to 21.39 and the IPCW weight ranged from 1.04 to 33.07.

Due to the size of the data set, standard errors of the estimators were based on the m of n bootstrap to reduce computation time. The idea is to sample with replacement m subjects from all n subjects in the sample. The standard error estimator is then the bootstrap standard error multiplied by $\sqrt{m/n}$. We sampled 1000 subjects from the 7110 subjects for each bootstrap resample and 5000 bootstrap samples were drawn.

We evaluated the race effect over the [0,70] month interval. Within approximately one month after wait listing, Caucasians have significantly lower cumulative hazard of death than African Americans (Figure 4.1). The cumulative hazard is lower at the very beginning of wait-listing, while it is significantly higher comparing Caucasians to African Americans after approximately 11 months with ratio of cumulative hazards ranged from $\widehat{\phi}_1(t) = 1.18$ to $\widehat{\phi}_1(t) = 1.47$. The pattern of the estimated relative risk is similar as ratio of cumulative hazards (Figure 4.2). Figure 4.3 shows that Caucasians have shorter restricted mean lifetime than African Americans based on the first 8 months after wait-listing. The difference in restricted mean lifetime is significant after 18 months of wait-listing, with the estimated difference ranging from $\widehat{\Delta}_1(t) = -0.17$ months to $\widehat{\Delta}_1(t) = -3.39$ months comparing Caucasians to African Americans.

We compared the results above to those without applying inverse censoring weight. The double inverse weighting estimates for ratio of cumulative hazards and relative risk are lower than those without applying IPCW, especially after 22 months of wait-listing. When dependent censoring is ignored, the estimated difference of restricted mean lifetime is smaller than the estimates based on double inverse weighting, espe-

cially towards the end of follow-up period.

4.6 Discussion

In this chapter, we proposed measures to estimate the cumulative treatment effect when the proportional hazards assumption does not hold. The proposed estimators adjust for discrepancies in treatment-specific baseline covariate distribution and overcome dependent censoring due to time-dependent covariates by applying double inverse weighting. Simulation studies show that the proposed estimators are approximately unbiased and the estimated standard errors are accurate.

Applying our methods to kidney wait list survival data, we found that the effect of race (Caucasian vs. African American) is time dependent. The cumulative hazard and risk of death are significantly higher for Caucasians relative to African Americans 11 months after wait listing, and Caucasians have significantly shorter (3.39 months shorter) restricted mean lifetime based on the first 70 months after wait listing. The difference in restricted mean lifetime is significant in a long term, although not of clinical importance.

In Chapter 3, we found that misspecification of the logistic model (model for inverse treatment weight) may bring bias to our proposed estimators. The estimates of the proposed measures in this article also depend on the estimate for inverse treatment weight. Therefore, misspecification of inverse treatment weight model may bring bias to the treatment effect. The accuracy of the estimates of our proposed measures do depend on the accuracy of the estimate of inverse probability of censoring weight. When very few subjects are censored in the sample, it may not worthwhile to use IPCW since the estimate for inverse probability of censoring weight may be biased due to too small sample size for censoring. Accordingly, the estimates for our

proposed measures may be biased in such settings.

For a very large data set, the computation of asymptotic variance of our proposed estimators is intensive. The bootstrap method can be applied when sample size is too large, with the m of n bootstrap method being a practical means of reducing the computation time.

4.7 Tables and Figures

			t=1		t=2		t=3	
	Setting	C%	$\log\phi_1(t)$	BIAS	$\log\phi_1(t)$	BIAS	$\log\phi_1(t)$	BIAS
	I	23%	0	0.004	0.283	0.017	0.584	0.008
		40%	0	0.019	0.283	0.022	0.584	-0.007
$\log\widehat{\phi}_1(t)$	II	13%	0.496	0.013	0.778	0.004	1.076	0.009
		33%	0.496	0.037	0.778	0.022	1.076	0.017
	III	28%	0.496	0.038	0.778	0.018	1.076	0.012
		46%	0.496	0.058	0.778	0.024	1.076	0.006
	Setting	C%	$\log RR_1(t)$	BIAS	$\log RR_1(t)$	BIAS	$\log RR_1(t)$	BIAS
	I	23%	0	0.004	0.238	0.015	0.429	0.008
		40%	0	0.019	0.238	0.019	0.429	-0.004
$\log\widehat{RR}_1(t)$	II	13%	0.451	0.014	0.619	0.007	0.715	0.011
		33%	0.451	0.036	0.619	0.021	0.715	0.016
	III	28%	0.451	0.037	0.619	0.017	0.715	0.011
		46%	0.451	0.056	0.619	0.024	0.715	0.011
	Setting	C%	$\Delta_1(t)$	BIAS	$\Delta_1(t)$	BIAS	$\Delta_1(t)$	BIAS
	I	23%	0	0.001	-0.036	-0.002	-0.171	-0.002
		40%	0	-0.001	-0.036	-0.004	-0.171	-0.003
$\widehat{\Delta}_1(t)$	II	13%	-0.042	0.001	-0.197	0.005	-0.502	0.009
		33%	-0.042	-0.001	-0.197	-0.002	-0.502	0.001
	III	28%	-0.042	-0.001	-0.197	-0.002	-0.502	0.001
		46%	-0.042	-0.001	-0.197	-0.001	-0.502	0.005

I: $\eta_1 = 0,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 1$
II: $\eta_1 = 0.5,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 1$
III:$\eta_1 = 0.5,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 0.25$

Table 4.1: Simulation results: Examination of bias at different time points (n=200)

			t=1		t=2		t=3	
	Setting	C%	$\log \phi_1(t)$	BIAS	$\log \phi_1(t)$	BIAS	$\log \phi_1(t)$	BIAS
	I	23%	0	0.009	0.283	0.001	0.584	-0.006
		40%	0	0.005	0.283	-0.001	0.584	-0.009
$\log \widehat{\phi}_1(t)$	II	13%	0.496	0.007	0.778	0.001	1.076	-0.011
		33%	0.496	0.007	0.778	0.010	1.076	-0.011
	III	28%	0.496	0.006	0.778	-0.007	1.076	-0.012
		46%	0.496	0.019	0.778	0.018	1.076	0.007
	Setting	C%	$\log RR_1(t)$	BIAS	$\log RR_1(t)$	BIAS	$\log RR_1(t)$	BIAS
	I	23%	0	0.009	0.238	0.001	0.429	-0.004
		40%	0	0.004	0.238	0.001	0.429	-0.007
$\log \widehat{RR}_1(t)$	II	13%	0.451	0.007	0.619	0.001	0.715	-0.007
		33%	0.451	0.007	0.619	0.009	0.715	-0.006
	III	28%	0.451	0.006	0.619	-0.004	0.715	-0.008
		46%	0.451	0.018	0.619	0.015	0.715	0.005
	Setting	C%	$\Delta_1(t)$	BIAS	$\Delta_1(t)$	BIAS	$\Delta_1(t)$	BIAS
	I	23%	0	-0.001	-0.036	-0.001	-0.171	0.001
		40%	0	-0.001	-0.036	-0.001	-0.171	0.001
$\widehat{\Delta}_1(t)$	II	13%	-0.042	0.001	-0.197	0.001	-0.502	0.005
		33%	-0.042	0.001	-0.197	-0.001	-0.502	0.003
	III	28%	-0.042	0.001	-0.197	0.003	-0.502	0.008
		46%	-0.042	-0.002	-0.197	-0.005	-0.502	-0.006

I: $\eta_1 = 0,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 1$
II: $\eta_1 = 0.5,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 1$
III:$\eta_1 = 0.5,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 0.25$

Table 4.2: Simulation results: Examination of bias at different time points (n=500)

	Setting	$C\%$	$\log \phi_1(t)$	BIAS	ASE	ESD	CP
	I	23%	0.283	0.017	0.331	0.317	0.96
		40%		0.022	0.344	0.337	0.96
$\log \widehat{\phi}_1(t)$	II	13%	0.778	0.004	0.309	0.320	0.94
		33%		0.022	0.323	0.332	0.95
	III	28%	0.778	0.017	0.321	0.326	0.95
		46%		0.024	0.347	0.356	0.95
	Setting	$C\%$	$\log RR_1(t)$	BIAS	ASE	ESD	CP
	I	23%	0.238	0.015	0.283	0.270	0.96
		40%		0.019	0.293	0.286	0.96
$\log \widehat{RR}_1(t)$	II	13%	0.619	0.007	0.255	0.264	0.94
		33%		0.021	0.267	0.274	0.95
	III	28%	0.619	0.017	0.265	0.269	0.96
		46%		0.024	0.288	0.297	0.95
	Setting	$C\%$	$\Delta_1(t)$	BIAS	ASE	ESD	CP
	I	23%	-0.036	-0.002	0.087	0.087	0.95
		40%		-0.004	0.089	0.090	0.94
$\widehat{\Delta}_1(t)$	II	13%	-0.196	0.005	0.095	0.100	0.93
		33%		-0.002	0.096	0.099	0.93
	III	28%	-0.196	-0.002	0.096	0.099	0.94
		46%		-0.001	0.099	0.102	0.94

I: $\eta_1 = 0,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 1$
II: $\eta_1 = 0.5,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 1$
III:$\eta_1 = 0.5,\ \ \eta_2 = 0.2,\ \ \eta_3 = 0.5,\ \ \theta = 0.25$

Table 4.3: Simulation results: Examination of bootstrap standard errors (t=2, n=200)

Covariate	Covariate value	Hazard ratio	P-value
Hospitalizations		0.92	<0.001
Gender (reference=Male)		0.95	0.170
Age	18-29	1	–
	30-40	0.85	0.015
	40-50	0.75	<0.001
	50-60	0.76	<0.001
	≥60	0.66	<0.001
Lung disease (reference=No)		0.79	0.168
BMI	0-20	0.90	0.085
	20-25	1	–
	25-30	0.89	0.004
	30-35	0.94	0.215
	>35	0.76	<0.001
Diagnosis	Diabetes	1	–
	Glomerulo nephritis	1.11	0.120
	Polycystic kidney disease	1.19	0.013
	Hypertension	0.99	0.920
	Other	1.05	0.250

Table 4.4: Analysis of wait list mortality by race: Parameter estimates for censoring model

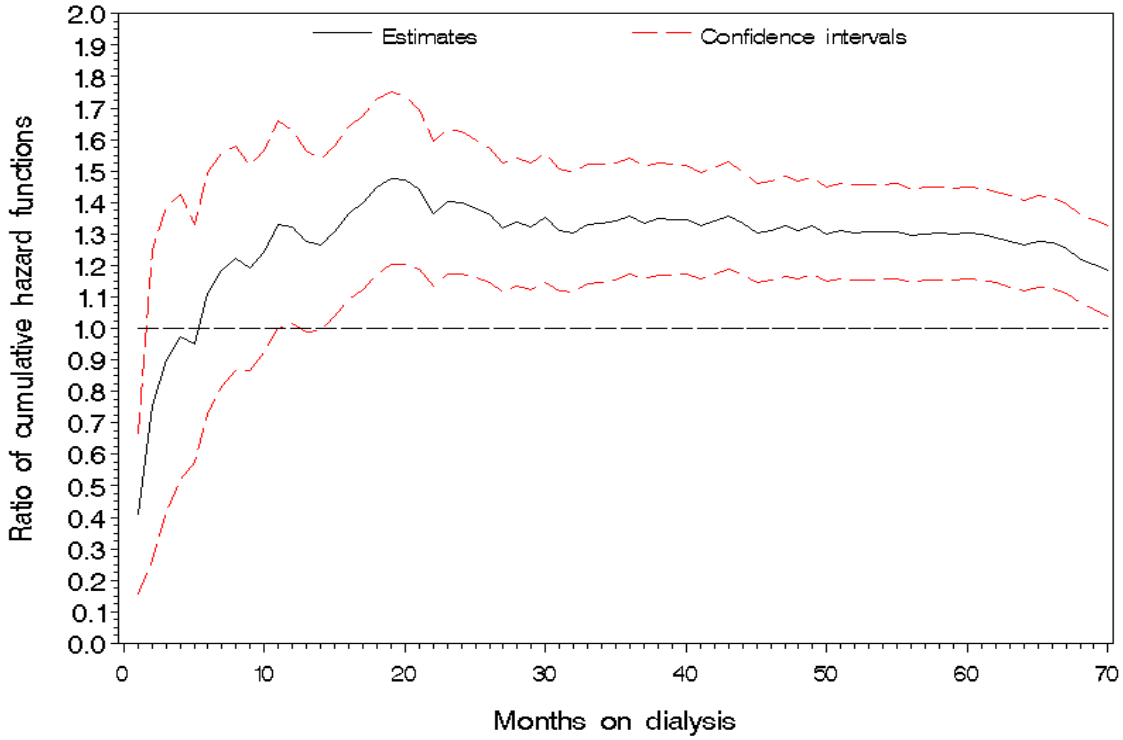

Figure 4.1: Analysis of wait list mortality by race: Estimator and 95% pointwise confidence intervals for the ratio of cumulative hazard functions (Caucasian/African American), $\phi_1(t)$.

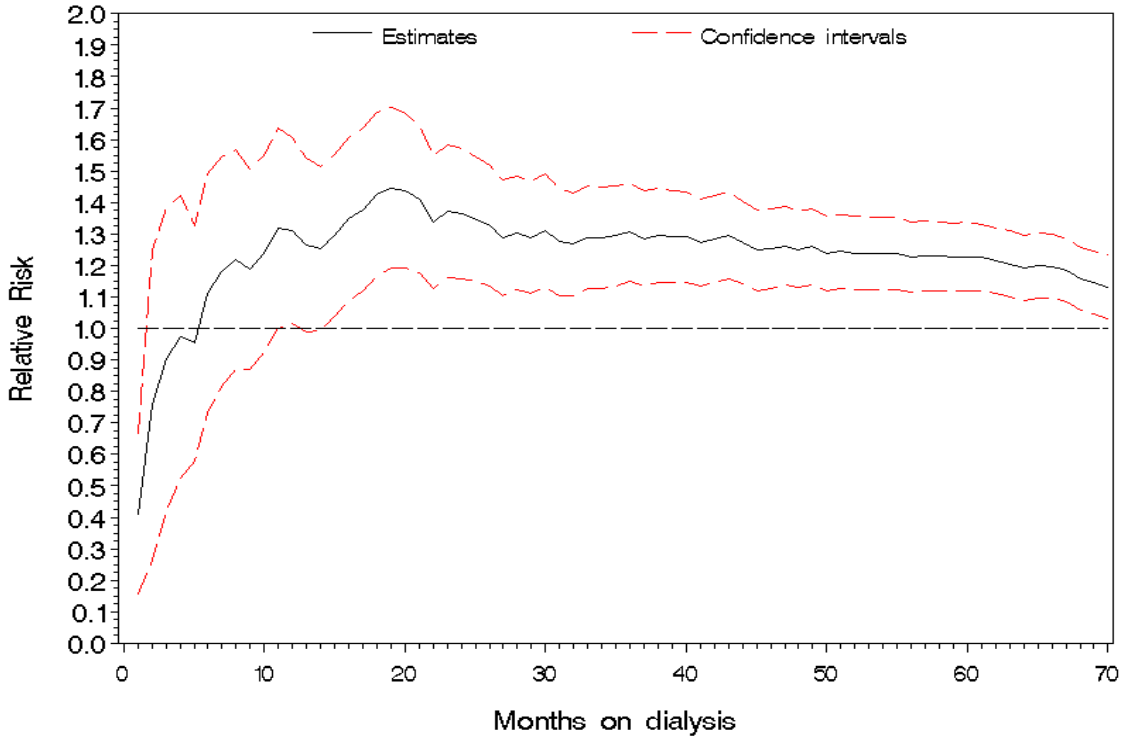

Figure 4.2: Analysis of wait list mortality by race: Estimator and 95% pointwise confidence intervals for the risk ratio (Caucasian/African American), $RR_1(t)$.

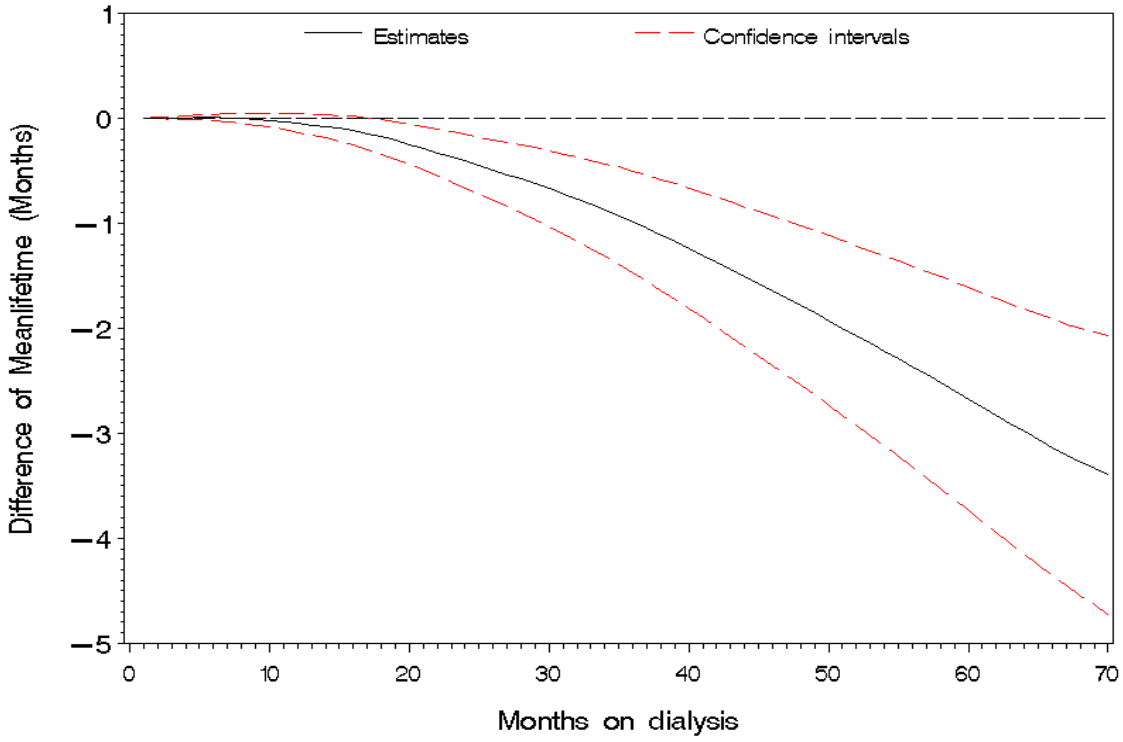

Figure 4.3: Analysis of wait list mortality by race: Estimator and 95% pointwise confidence intervals for the difference in restricted mean lifetime (Caucasian-African American), $\Delta_1(t)$.

4.8 Appendix

Proof of Theorem 1

Consistency:

The strong consistency of $\widehat{\phi}_j(t)$ can be proved by proving the strong consistency of $\widehat{\Lambda}_j(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}}))$ and $\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}}))$, where

$$\widehat{\Lambda}_j(t, \widehat{\boldsymbol{\beta}}, \widehat{w}^C(t, \widehat{\boldsymbol{\theta}})) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) w_{ij}^G(\widehat{\boldsymbol{\beta}})}{n^{-1}\sum_{i=1}^n Y_{ij}(s)\widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) w_{ij}^G(\widehat{\boldsymbol{\beta}})} dN_{ij}^D(s).$$

Since $\widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) \xrightarrow{a.s.} w_i^C(s, \boldsymbol{\theta}_0)$ and $w_{ij}^G(\widehat{\boldsymbol{\beta}}) \xrightarrow{a.s.} w_{ij}^G(\boldsymbol{\beta}_0)$, by Strong Law of Large Numbers (SLLN, Pollard, 1990), $n^{-1}\sum_{i=1}^n Y_{ij}(s)\widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) w_{ij}^G(\widehat{\boldsymbol{\beta}})$ converges almost surely to $E\left[Y_{ij}(s) w_i^C(s, \boldsymbol{\theta}_0) w_{ij}^G(\boldsymbol{\beta}_0)\right]$. Let $\mathbf{Z}_{i0} = \mathbf{Z}_i(0)$, $\mathbf{Z}_{is} = \{\mathbf{Z}_i(u); u \in (0, s]\}$, $w_{ij}^C(s, \boldsymbol{\theta}_0) = Y_i(s) Pr(C_i > s | G_i = j, \widetilde{\mathbf{Z}}_i(s))^{-1}$, and $f(\widetilde{\mathbf{z}}_i(s))$ be the density function of $\widetilde{\mathbf{Z}}_i(s)$. For ease of presentation, we assume that $\widetilde{\mathbf{Z}}_i(s)$ is continuous, in the development that follows. We obtain that

$$\begin{aligned}
& E\left[Y_{ij}(s) w_i^C(s, \boldsymbol{\theta}_0) w_{ij}^G(\boldsymbol{\beta}_0)\right] \\
= \; & E\left\{E\left[Y_{ij}(s) w_{ij}^C(s, \boldsymbol{\theta}_0) w_{ij}^G(\boldsymbol{\beta}_0) | \widetilde{\mathbf{Z}}_i(s)\right]\right\} \\
= \; & E\left\{Pr^{-1}(C_i > s | G_i = j, \widetilde{\mathbf{Z}}_i(s)) Pr^{-1}(G_i = j | \mathbf{Z}_{i0}) E\left[Y_{ij}(s) | \widetilde{\mathbf{Z}}_i(s)\right]\right\} \\
= \; & E\left[Pr^{-1}(G_i = j | \mathbf{Z}_{i0}) Pr(T_i > s | G_i = j, \widetilde{\mathbf{Z}}_i(s)) Pr(G_i = j | \widetilde{\mathbf{Z}}_i(s))\right] \\
= \; & E_{\widetilde{\mathbf{Z}}_i(s)}\left[Pr^{-1}(G_i = j | \mathbf{Z}_{i0}) Pr(T_i > s, G_i = j, \widetilde{\mathbf{Z}}_i(s)) f(\widetilde{\mathbf{Z}}_i(s))^{-1}\right] \\
= \; & \int_{\widetilde{\mathbf{Z}}_i(s)} Pr(T_i > s, G_i = j, \widetilde{\mathbf{Z}}_i(s)) Pr^{-1}(G_i = j | \mathbf{Z}_{i0}) d\widetilde{\mathbf{Z}}_i(s) \\
= \; & \int_{\mathbf{z}_{i0}}\int_{\mathbf{z}_{is}} Pr(T_i > s, G_i = j, \mathbf{Z}_{i0} = \mathbf{z}_{i0}, \mathbf{Z}_{is} = \mathbf{z}_{is}) Pr^{-1}(G_i = j | \mathbf{z}_{i0}) d\mathbf{z}_{is} d\mathbf{z}_{i0} \\
= \; & \int_{\mathbf{z}_{i0}} Pr(T_i > s | G_i = j, \mathbf{Z}_{i0} = \mathbf{z}_{i0}) f(\mathbf{z}_{i0}) d\mathbf{z}_{i0} \\
= \; & E_{\mathbf{Z}_{i0}}[S(s | G_i = j, \mathbf{Z}_{i0})] \\
\equiv \; & S_j(s).
\end{aligned}$$

Similarly, it can be shown that $n^{-1}\sum_{i=1}^{n}\widehat{w}_i^C(s,\widehat{\boldsymbol{\theta}})w_{ij}^G(\widehat{\boldsymbol{\beta}})dN_{ij}^D(s)$ converges almost surely to $E\left[dF_j(s)\right]$. Combining results above, and using continuity, we obtain that $\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))$ converges to

$$\Lambda_j(t)=\int_0^t\frac{E\left[dF_j(s)\right]}{E\left[S_j(s)\right]}.$$

Exploiting the continuity of $\phi_j(t)$ as a map of $\Lambda_j(t)$ and $\Lambda_0(t)$, we obtain that $\widehat{\phi}_j(t)$ converges almost surely to $\phi_j(t)$.

Asymptotic normality:

One can write:

$$\begin{aligned}n^{1/2}\left\{\widehat{\phi}_j(t)-\phi_j(t)\right\} &= \frac{1}{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))}n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))-\Lambda_j(t)\right]\\ &\quad+\Lambda_j(t)n^{1/2}\left[\frac{1}{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))}-\frac{1}{\Lambda_0(t)}\right]\end{aligned}$$

By a Taylor Expansion, we obtain that

$$\left[\frac{1}{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))}-\frac{1}{\Lambda_0(t)}\right]=-\frac{1}{\Lambda_0(t)^2}\left[\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))-\Lambda_0(t)\right].$$

This result gives

$$\begin{aligned}&n^{1/2}\left\{\widehat{\phi}_j(t)-\phi_j(t)\right\} \qquad (4.10)\\ &= \frac{1}{\Lambda_0(t)}n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))-\Lambda_j(t)\right]-\frac{\Lambda_j(t)}{\Lambda_0(t)^2}n^{1/2}\left[\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}}))-\Lambda_0(t)\right]\end{aligned}$$

For $j=0,\cdots,J$, one can decompose $\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))-\Lambda_j(t)\right]$ into $\widehat{\alpha}_{j1}(t)+\widehat{\alpha}_{j2}(t)+\widehat{\alpha}_{j3}(t)+\widehat{\alpha}_{j4}(t)$, where

$$\begin{aligned}\widehat{\alpha}_{j1}(t) &= \widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))-\widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))\\ \widehat{\alpha}_{j2}(t) &= \widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,\widehat{w}^C(t,\widehat{\boldsymbol{\theta}}))-\widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,\widehat{w}^C(t,\boldsymbol{\theta}_0))\\ \widehat{\alpha}_{j3}(t) &= \widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,\widehat{w}^C(t,\boldsymbol{\theta}_0))-\widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,w^C(t,\boldsymbol{\theta}_0))\\ \widehat{\alpha}_{j4}(t) &= \widehat{\Lambda}_j(t,\boldsymbol{\beta}_0,w^C(t,\boldsymbol{\theta}_0))-\Lambda_j(t).\end{aligned}$$

The quantity $n^{1/2}\widehat{\alpha}_{j1}(t)$ is written as

$$n^{1/2}\widehat{\alpha}_{j1}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \left\{ \frac{\widehat{w}_i^C(s,\widehat{\boldsymbol{\theta}})w_{ij}^G(\widehat{\boldsymbol{\beta}})dN_{ij}^D(s)}{R_j(s,\widehat{\boldsymbol{\beta}},\widehat{w}^C(s,\widehat{\boldsymbol{\theta}}))} - \frac{\widehat{w}_i^C(s,\widehat{\boldsymbol{\theta}})w_{ij}^G(\boldsymbol{\beta}_0)dN_{ij}^D(s)}{R_j(s,\boldsymbol{\beta}_0,\widehat{w}^C(s,\widehat{\boldsymbol{\theta}}))} \right\}$$

Since $R_j(s,\widehat{\boldsymbol{\beta}},\widehat{w}^C(s,\widehat{\boldsymbol{\theta}}))$ and $R_j(s,\boldsymbol{\beta}_0,\widehat{w}^C(s,\widehat{\boldsymbol{\theta}}))$ converge almost surely to $r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ as $n\to\infty$, we obtain that

$$n^{1/2}\widehat{\alpha}_{j1}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{\widehat{w}_i^C(s,\widehat{\boldsymbol{\theta}})[w_{ij}^G(\widehat{\boldsymbol{\beta}}) - w_{ij}^G(\boldsymbol{\beta}_0)]dN_{ij}^D(s)}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)}$$

By a Taylor Expansion, one can write $[w_{ij}^G(\widehat{\boldsymbol{\beta}}) - w_{ij}^G(\boldsymbol{\beta}_0)] = a_{ij}^T(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(1)$, where

$$\begin{aligned} a_{ij}(\boldsymbol{\beta}_0) &= \frac{\partial w_{ij}^G(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}|\boldsymbol{\beta}_0 \\ &= I(G_i\neq 0)\left[\frac{\sum_{k=1}^J \exp\{\boldsymbol{\beta}^T\mathbf{X}_{ik}\}\mathbf{X}_{ik}}{\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ij}\}} - \mathbf{X}_{ij}p_{ij}^{-1}(\boldsymbol{\beta}_0)\right] + G_{i0}\sum_{k=1}^{J}\exp\{\boldsymbol{\beta}^T\mathbf{X}_{ik}\}\mathbf{X}_{ik}. \end{aligned}$$

Therefore, we obtain that

$$\begin{aligned} n^{1/2}\widehat{\alpha}_{j1}(t) &= \widehat{h}_j(t)^T n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(1) \\ \widehat{h}_j(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\widehat{w}_i^C(s,\boldsymbol{\theta})a_{ij}(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)}dN_{ij}^D(s), \end{aligned}$$

Using the SLLN, one can obtain that $\widehat{h}_j(t)$ converges to $h_j(t)$, where

$$h_j(t) = E\left[\int_0^t \frac{w_i^C(s,\boldsymbol{\theta}_0)a_{ij}(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)}dN_{ij}^D(s)\right]$$

One can write $n^{-1/2}U_G(\boldsymbol{\beta}_0)$ as

$$\begin{aligned} n^{-1/2}U_G(\boldsymbol{\beta}_0) &= -\frac{1}{n}\frac{\partial U_G(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T}\Big|_{\boldsymbol{\beta}_0} n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(1) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J}\left\{p_{ij}(\boldsymbol{\beta}_0)\mathbf{X}_{ij}[\mathbf{X}_{ij}^T - \sum_{k=1}^{J}\mathbf{X}_{ik}^T p_{ik}(\boldsymbol{\beta}_0)]\right\} n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0). \end{aligned}$$

Using the SLLN, one can obtain that $n^{-1}\sum_{i=1}^n\sum_{j=1}^J\left\{p_{ij}(\boldsymbol{\beta}_0)\mathbf{X}_{ij}[\mathbf{X}_{ij}^T - \sum_{k=1}^J\mathbf{X}_{ik}^T p_{ik}(\boldsymbol{\beta}_0)]\right\}$ converges almost surely to $\boldsymbol{\Omega}_G(\boldsymbol{\beta}_0)$ which is defined as in (4.6). These results along

with (4.4) give

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{\Omega}_G^{-1}(\boldsymbol{\beta}_0) n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{J} \mathbf{X}_{ij} \left[G_{ij} - p_{ij}(\boldsymbol{\beta}_0) \right]$$

Using the results above, we can obtain that

$$n^{1/2}\widehat{\alpha}_{j1}(t) = n^{-1/2} h_j^T(t) \boldsymbol{\Omega}_G^{-1}(\boldsymbol{\beta}_0) \sum_{i=1}^{n} \sum_{j=1}^{J} \mathbf{X}_{ij} \left[G_{ij} - p_{ij}(\boldsymbol{\beta}_0) \right] + o_p(1). \tag{4.11}$$

Since $R_j(s, \boldsymbol{\beta}_0, \widehat{w}^C(s, \widehat{\boldsymbol{\theta}}))$ and $R_j(s, \boldsymbol{\beta}_0, \widehat{w}^C(s, \boldsymbol{\theta}_0))$ converge to $r_j^{(0)}(s, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ as $n \to \infty$, one can obtain

$$\begin{aligned}
& n^{1/2}\widehat{\alpha}_{j2}(t) \\
= \; & n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) w_{ij}^G(\boldsymbol{\beta}_0) dN_{ij}^D(s)}{R_j(s, \boldsymbol{\beta}_0, \widehat{w}^C(s, \widehat{\boldsymbol{\theta}}))} - n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\widehat{w}_i^C(s, \boldsymbol{\theta}_0) w_{ij}^G(\boldsymbol{\beta}_0) dN_{ij}^D(s)}{R_j(s, \boldsymbol{\beta}_0, \widehat{w}^C(s, \boldsymbol{\theta}_0))} \\
= \; & n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{[\widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) - \widehat{w}_i^C(s, \boldsymbol{\theta}_0)] w_{ij}^G(\boldsymbol{\beta}_0)}{r_j^{(0)}(s, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)} dN_{ij}^D(s)
\end{aligned}$$

By a Taylor Expansion, one can obtain that $n^{1/2}[\widehat{w}_i^C(s, \widehat{\boldsymbol{\theta}}) - \widehat{w}_i^C(s, \boldsymbol{\theta}_0)] = \widehat{b}_i(s, \boldsymbol{\theta}_0) n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, where

$$\begin{aligned}
\widehat{b}_i(s, \boldsymbol{\theta}_0) &= \left. \frac{\partial \widehat{w}_i^C(s, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \\
&= \widehat{w}_i^C(s, \boldsymbol{\theta}_0) \int_0^s Y_i(u) \exp\{\boldsymbol{\theta}_0{}^T \mathbf{Z}_i^C(u)\} \left[\mathbf{Z}_i^C(u) - \overline{\mathbf{Z}}^C(u) \right] d\widehat{\Lambda}_0^C(u).
\end{aligned}$$

By the SLLN, $\widehat{b}_i(s, \boldsymbol{\theta}_0)$ converges to $b_i(s, \boldsymbol{\theta}_0)$,where

$$b_i(s, \boldsymbol{\theta}_0) = w_i^C(s, \boldsymbol{\theta}_0) \int_0^s Y_i(u) \exp\{\boldsymbol{\theta}_0{}^T \mathbf{Z}_i^C(u)\} \left[\mathbf{Z}_i^C(u) - \overline{\mathbf{z}}^C(u) \right] d\Lambda_0^C(u).$$

Using the results above, we obtain that

$$\begin{aligned}
n^{1/2}\widehat{\alpha}_{j2}(t) &= n^{1/2} \widehat{g}_j^T(t)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\
\widehat{g}_j(t) &= \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{b_i(s, \boldsymbol{\theta}_0) w_{ij}^G(\boldsymbol{\beta}_0)}{r_j^{(0)}(s, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)} dN_{ij}^D(s) + o_p(1).
\end{aligned}$$

By the SLLN, one can obtain that $\widehat{g}_j(t)$ converges to $g_j(t)$, where

$$g_j(t) = \int_0^t \frac{E\left[b_i(s,\boldsymbol{\theta}_0) w_{ij}^G(\boldsymbol{\beta}_0) dN_{ij}^D(s)\right]}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)}.$$

Using another Taylor expansion and (4.5), we obtain that

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \boldsymbol{\Omega}_C^{-1}(\boldsymbol{\theta}_0) n^{-1/2} \sum_{i=1}^n \int_0^\tau \left\{\mathbf{Z}_i^C(t) - \overline{\mathbf{z}}^C(t,\boldsymbol{\theta}_0)\right\} dM_i^C(t) + o_p(1)$$

Therefore, one can write

$$n^{1/2}\widehat{\alpha}_{j2}(t) = n^{-1/2} g_j^T(t) \Omega_C^{-1}(\boldsymbol{\theta}_0) \sum_{i=1}^n \int_0^\tau \left\{\mathbf{Z}_i^C(t) - \overline{\mathbf{z}}^C(t,\boldsymbol{\theta}_0)\right\} dM_i^C(t) + o_p(1).$$

Since $R_j(s,\boldsymbol{\beta}_0,\widehat{w}^C(s,\boldsymbol{\theta}_0))$ and $R_j(s,\boldsymbol{\beta}_0,w^C(s,\boldsymbol{\theta}_0))$ converge to $r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$, one can write

$$n^{1/2}\widehat{\alpha}_{j3}(t) = n^{-1/2} \sum_{i=1}^n \int_0^t \frac{[\widehat{w}_i^C(s,\boldsymbol{\theta}_0) - w_i^C(s,\boldsymbol{\theta}_0)] w_{ij}^G(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)} dN_{ij}^D(s).$$

Applying the Functional Delta Method,

$$\begin{aligned} n^{1/2}[\widehat{w}_i^C(s,\boldsymbol{\theta}_0) - w_i^C(s,\boldsymbol{\theta}_0)] &= w_i^C(s,\boldsymbol{\theta}_0) n^{1/2}[\widehat{\Lambda}_i^C(s,\boldsymbol{\theta}_0) - \Lambda_i^C(s)] \\ &= w_i^C(s,\boldsymbol{\theta}_0) n^{-1/2} \sum_{k=1}^n \int_0^s \frac{\exp\{\boldsymbol{\theta}_0^T Z_i^C(u)\}}{R_C^{(0)}(u,\boldsymbol{\theta}_0)} dM_k(u) + o_p(1) \end{aligned}$$

Therefore, $n^{1/2}\widehat{\alpha}_{j3}(t)$ can be written as

$$\begin{aligned} & n^{1/2}\widehat{\alpha}_{j3}(t) \\ = \; & n^{-1/2} \sum_{i=1}^n \int_0^t \frac{w_{ij}^G(\boldsymbol{\beta}_0) w_i^C(s,\boldsymbol{\theta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)} \frac{1}{n} \sum_{k=1}^n \int_0^s \frac{\exp\{\boldsymbol{\theta}_0^T Z_i^C(u)\}}{R_C^{(0)}(u,\boldsymbol{\theta}_0)} dM_k(u) dN_{ij}^D(s) + o_p(1) \\ = \; & n^{-1/2} \sum_{i=1}^n \int_0^t \frac{1}{n} \sum_{k=1}^n \frac{\exp\{\boldsymbol{\theta}_0^T Z_k^C(u)\}}{r_C^{(0)}(u,\boldsymbol{\theta}_0)} \int_u^t \frac{w_k^C(s,\boldsymbol{\theta}_0) w_{kj}^G(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)} dN_{kj}^D(s) dM_i^C(u) + o_p(1). \end{aligned}$$

Since $\left[n^{-1}\sum_{k=1}^n \int_u^t w_k(s,\boldsymbol{\beta}_0) w_{kj}^G(\boldsymbol{\beta}_0) r^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)^{-1} dN_{kj}^D(s) \exp\{Z_k^C(u)^T\boldsymbol{\theta}_0\}\right]$ converges almost surely to

$$E\left[\exp\{\boldsymbol{\theta}_0^T Z_k^C(u)\} \int_u^t \frac{w_k^C(s,\boldsymbol{\theta}_0) w_{kj}^G(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)} dN_{kj}^D(s)\right]$$

we obtain that

$$n^{1/2}\widehat{\alpha}_{j3}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t E\left[\frac{\exp\{\boldsymbol{\theta}_0^T Z_k^C(u)\}}{r_C^{(0)}(u,\boldsymbol{\theta}_0)}\int_u^t \frac{w_k^C(s,\boldsymbol{\theta}_0)w_{kj}^G(\boldsymbol{\beta}_0)}{r_j^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)}dN_{kj}^D(s)\right]dM_i^C(u) + o_p(1).$$

The quantity $n^{1/2}\widehat{\alpha}_{j4}(t)$ can be written as

$$n^{1/2}\widehat{\alpha}_{j4}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{w_{ij}^G(\boldsymbol{\beta}_0)w_i^C(s,\boldsymbol{\theta}_0)}{R_j(s,\boldsymbol{\beta}_0,w^C(s,\boldsymbol{\theta}_0)}dM_{ij}^D(s)$$

Since $R_j(s,\boldsymbol{\beta}_0,w^C(s,\boldsymbol{\theta}_0)$ converges to $r^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$, $n^{1/2}\widehat{\alpha}_{j4}(t)$ can be written as

$$n^{1/2}\widehat{\alpha}_{j4}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{w_i^T(\boldsymbol{\beta}_0)w_i^C(s,\boldsymbol{\theta}_0)}{r^{(0)}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)}dM_{ij}^D(s) + o_p(1) \tag{4.12}$$

Using expressions for $\widehat{\alpha}_{j1}(t)$, $\widehat{\alpha}_{j2}(t)$, $\widehat{\alpha}_{j3}(t)$ and $\widehat{\alpha}_{j4}(t)$, one can obtain that

$$n^{1/2}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}}) - \Lambda_j(t)\right] = n^{-1/2}\sum_{i=1}^{n}\Phi_{ij} + o_p(1),$$

where Φ_{ij} is defined as in the Theorem 1. Using (4.10), one can write

$$n^{1/2}\left\{\widehat{\phi}_j(t) - \phi_j(t)\right\} = n^{-1/2}\sum_{i=1}^{n}\left[\frac{1}{\Lambda_0(t)}\Phi_{ij}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) - \frac{\Lambda_j(t)}{\Lambda_0(t)^2}\Phi_{i0}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right],$$

which is a sum 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 $\left[n^{1/2}\left\{\widehat{\phi}_j(t_1) - \phi_j(t_1)\right\},\cdots,n^{1/2}\left\{\widehat{\phi}_j(t_k) - \phi_j(t_k)\right\}\right]$ converges to a mean zero multivariate normal distribution. Further, since $\widehat{\phi}_j(t) - \phi_j(t)$ is tight, the process $n^{1/2}\left\{\widehat{\phi}_j(t) - \phi_j(t)\right\}$ converges to a mean 0 Gaussian process with covariance function $\sigma_j^{\phi}(s,t) = E[\xi_{ij}^{\phi}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\xi_{ij}^{\phi}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]$ for any set pair of (s,t), where $\xi_{ij}^{\phi}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ is defined as in Theorem 1.

Proof of Theorem 2

Using the continuous mapping theorem, $\widehat{S}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}})) = \exp\{-\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}}))\}$ converges almost surely to $S_j(t) = E[S(t|G_i = j,\mathbf{Z}_i(0))]$, Therefore, $\widehat{RR}_j(t)$ converges almost surely to $RR_j(t)$.

Using the Functional Delta method, we obtain that

$$\begin{aligned} n^{1/2}\left\{\widehat{RR}_j(t) - RR_j(t)\right\} &= n^{1/2}\frac{S_j(t)}{F_0(t)}\left[\widehat{\Lambda}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}})) - \Lambda_j(t)\right] \\ &\quad -n^{1/2}\frac{F_j(t)S_0(t)}{F_0(t)^2}\left[\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}})) - \Lambda_0(t)\right] + o_p(1). \end{aligned}$$

Using similar techniques as in the proof of Theorem 1, we can obtain that

$$\begin{aligned} & n^{1/2}\left\{\widehat{RR}_j(t) - RR_j(t)\right\} \\ = \; & n^{-1/2}\sum_{i=1}^{n}\left[\frac{S_j(t)}{F_0(t)}\Phi_{ij}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) - \frac{F_j(t)S_0(t)}{F_0(t)^2}\Phi_{i0}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right] + o_p(1), \end{aligned}$$

which is a sum of independent and identically distributed mean 0 random variables. By the Multivariate Central Limit Theorem and the tightness of $\left\{\widehat{RR}_j(t) - RR_j(t)\right\}$, we obtain that $n^{1/2}\left\{\widehat{RR}_j(t) - RR_j(t)\right\}$ converges to a mean 0 Gaussian process with covariance function $\sigma_j^R(s,t) = E[\xi_{ij}^R(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\xi_{ij}^R(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]$, where $\xi_{ij}^R(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ is defined as in (4.8).

Proof of Theorem 3

Since $\widehat{S}_j(s,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}}))$ converges to $S_j(t)$, using continuous mapping theorem, we obtain that $\widehat{e}_j(t,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}})) = \int_0^t \widehat{S}_j(s,\widehat{\boldsymbol{\beta}},\widehat{w}^C(\widehat{\boldsymbol{\theta}}))ds$ converges to $e_j(t) = \int_0^t S_j(t)dt$. Accordingly, $\widehat{\Delta}_j(t)$ converges to $\Delta_j(t)$.

Similar to the proof of Theorem 1, we can write

$$\begin{aligned} n^{1/2}\left\{\widehat{\Delta}_j(t) - \Delta_j(t)\right\} &= n^{-1/2}\sum_{i=1}^{n}\int_0^t -S_j(s)\Phi_{ij}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)ds \\ &\quad +n^{-1/2}\sum_{i=1}^{n}\int_0^t S_0(s)\Phi_{i0}(s,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)ds + o_p(1), \end{aligned}$$

which is a sum of independent and identically distributed mean 0 random variables. Demonstration of normality and weak convergence to a zero mean Gaussian process is similar as that of Theorem 1.

4.9 References

Aalen O.O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics* **6**, 701-726.

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**, 662-682.

Chen, P. and Tsiatis A.A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57** 1030-1038.

Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34**, 187-202.

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

Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Hernan, M.A., Brumback, B. and Robins, J.M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* **11**, 561-570.

Hernan, M.A., Brumback, B. and Robins, J.M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* **96**, 440-448.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481.

Matsuyama, Y. and Yamaguchi, T. (2008). Estimation of the marginal survival time

in the presence of dependent competing risks using inverse probability of censoring weighted (IPCW) methods. *Pharmaceutical Statistics* **7**, 202 - 214.

Nelson, W.B. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945-965.

Pollard, D. (1990). *Empirical processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward: Institute of Mathematical Statistics.

Robins, J.M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology – Methodological Issues*, Jewell N, Dietz K, Farewell V (eds). Birkhauser: Boston, MA, 297-331.

Robins, J.M. and Rotnitzky, A. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association.* 24-33.

Robins J.M., Hernan, M.A. and Brumback, B. (2000). Marginal Structrural Models and Causal Inference in Epidemiology. *Epidemiology* **11**, 550-560.

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

Xie, J. and Liu, C. (2005). Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine* **24**, 3089-3110.

Yoshida, M., Matsuyama Y., and Ohashi, Y. (2007). Estimation of treatment effect adjusting for dependent censoring using the IPCW method: an application to a large primary prevention study for coronary events (MEGA study). *Clinical Trials* **4**, 318-328.

CHAPTER V

Conclusion

This dissertation proposes three novel methods for estimating the cumulative treatment effect for time to event data in the setting when the treatment-specific hazards are not proportional. These methods were motivated by research questions on organ failure data. Therefore, the contribution of this research is not only statistical but also clinical. Chapter 2 used the ratio of cumulative hazards to compare treatment categories. Chapter 3 considered the setting where, in addition to the treatment effect, the effect of the adjustment covariates may be non-proportional. In addition to accommodating non-proportionality and unbalanced covariate distribution among treatment groups, Chapter 4 handled the dependent censoring due to time-dependent covariates.

Through a stratified Cox model, the method in Chapter 2 proposed the ratio of cumulative hazards as a treatment effect measure suited to the setting where the treatment effect varies over time. The proposed measure has the familiar hazard ratio interpretation when proportional hazards holds. Methods proposed in both Chapter 3 and Chapter 4 applied inverse probability of treatment weighting (IPTW) to balance the distribution of adjustment covariates among treatment groups. An inverse probability of censoring weight (IPCW) was applied to deal with dependent

censoring in Chapter 4. Chapters 3 and 4 compared cumulative hazards, relative risk, and restricted mean lifetime between treatment categories.

Each method was applied to organ failure data. In Chapter 2 and Chapter 3, we found that patients with end-stage renal disease who were treated by peritoneal dialysis have a lower cumulative hazard at the beginning but higher hazard in the long term compared to those treated by hemodialysis. In Chapter 4, we found that Caucasians on the kidney transplant wait list have shorter restricted mean lifetime than African Americans in the long term. This result is consistent with the current literature.

The methods proposed in this dissertation could be extended in several directions. For example, the methods in each of Chapters 2 to 4 focused on a treatment assigned at baseline (time 0). It would be interesting to develop methods to estimate the cumulative effect of a time-dependent treatment. Additionally, the survival time was assured to be univariate. The extension of the proposed methods to accommodate multivariate failure time data (e.g., recurrent events) would be valuable.