HAZARD AND RATE REGRESSION IN THE PRESENCE OF DIFFERENTIAL SELECTION OR TERMINATION PROBABILITY

by

Qing Pan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2007

Doctoral Committee:

    Assistant Professor Douglas E. Schaubel, Chair
    Professor John D. Kalbfleisch
    Associate Professor Richard A. Hirth
    Associate Professor Susan Murray

ACKNOWLEDGEMENTS

When I accepted the offer of the Ph.D program in Biostatistics at the University of Michigan, I didn't realize that it was one of the best decisions I made in my whole life. Many people have helped me and made my four years of Ph.D study a wonderful journey both academically and spiritually. Among them, my advisor, Doug, gave me the most important guidance. I learned statistics from him, more importantly, he set up an excellent example of researcher and director, and helped me find the career path to pursue. My research assistantship at the Kidney Epidemiology and Costs Center provided me a nice environment to practise statistical techniques in real applications besides financial support. In the Prospective Payment System project, Richard, Jack and Marc have shown me how to contribute and collaborate in a diverse group. My committee members, Doug, Jack, Richard and Susan, offered valuable inputs to my dissertation. All these people devoted their precious time unselfishly in my graduation and job-hunting process, whose favor I hope to return by helping my students in the future. Finally, I owe all my achievements to my parents, who always have confidence in me and support me unconditionally.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

Introduction

The research in this dissertation focuses on the analysis of time to event data. Chapter 2 considers weighted proportional hazards models in the presence of biased sampling and estimated selection probabilities. Chapter 3 examines the degree of bias corrected by the method proposed in Chapter 2 and proposes a method for evaluating the performance of the model. Chapter 4 considers the study of recurrent events in the presence of a terminating event. The proposed procedures involve modelling the recurrent event rate and terminal event hazard sparately; then combining the models to estimate the treatment effect on the recurrent event mean.

In biomedical studies, investigators often encounter data which are a potentially biased (i.e., unrepresentative) sample of the target population. Of particular interest is the two-stage sampling set-up. In the first stage, a representative sample of the target population is obtained with complete information on certain auxiliary variables. In the second stage, a biased sample is selected from the representative sample and the outcome (e.g., failure time) variable is missing for unselected subjects. Because of the partially missing outcome, it is not possible to fit a model (e.g., a proportional hazards model) to the representative sample and make unbiased inferences applicable to the target population. The possibility for biased sampling makes fitting

an unweighted proportional hazards model to the subjects with complete information (complete-case analysis) potentially problematic. The estimated effect of the covariate of interest would be biased if there exist variables which play the role of "biasing factor", where the term "biasing factor" refers to a variable correlated with the covariate of interest which affects the failure time hazard and the second stage selection probability. The observed relationship between the outcome and the factor of interest among subjects selected at the second stage is systematically different from that in the target population, since the effect of the biasing factor (not being accounted for in the estimation procedure) on the outcome is partially reflected.

A general method for overcoming selection bias involves weighting each subject by the inverse of their second-stage selection probability. This procedure was developed by Horvitz and Thompson (1952) to estimate a population mean. Binder (1992) extended the Horvitz-Thompson estimator for application to proportional hazards models. The data structure of interest in the proposal by Binder (1992) has only one selection step (from the target population to the biased sample) with known selection probabilities; a very reasonable framework for survey studies or designed experiments. After weighting each subject by the inverse of their selection probability (p_i), each subject in the observed sample represents $1/p_i$ identical subjects in the target population. As such, one can write out an appropriate partial likelihood and corresponding score function which would apply to the target population.

Lin (2000) extended the one-step sampling framework in Binder's paper (1992) to two-stage sampling, in the sense that a representative sample was added between the target population and the biased sample. The underlying target population was assumed by Lin (2000) to be infinite. The selection probabilities at the second stage were still assumed to be known. The weighted score equations in Lin (2000)

had the same form as those in Binder (1992), and hence the regression coefficient estimators were also the same. The difference in the estimator of Lin (2000) was that additional variation was induced by the sampling step from the target population to the representative sample. Correspondingly, the variance estimator proposed by Lin (2000) had one extra component relative to that of Binder (1992) which accounted for the extra variation introduced by the first-stage sampling mechanism.

Both Binder (1992) and Lin (2000) assumed that the selection probabilities for each subject were known and hence could be treated as fixed. In observational studies, it is common to encounter data structures with two-stage sampling and unknown selection probabilities. That is, one observes the auxiliary information for the representative sample and the complete information for the biased sample without any foreknowledge of the selection probabilities. This data structure is the subject of the weighted proportional hazards model in Chapter 2.

For two-stage sample data with unknown selection probabilities, we propose two-stage methods to obtain appropriate parameter and variance estimates. In the first stage, a logistic model is fitted to estimate the selection probability from the representative sample to the biased sample, adjusting for auxiliary variables (assumed to represent the biasing factors). At the second stage, a weighted proportional hazards model is fitted using the inverse of the estimated selection probabilities as weights. Under the assumption of consistent selection probability estimators, the regression parameter and cumulative baseline hazard estimators are consistent for their respective true underlying values and hence converge to the same limiting values as their predecessors with fixed weights. However, the proposed variance estimators are different from those of either Binder (1992) or Lin (2000) since the weights themselves have variation. Large-sample properties of the proposed parameter estimators are de-

rived and their applicability to finite samples is examined through simulation studies. The proposed methods are then applied to a kidney transplant data set to estimate the effects of expanded criteria donor (ECD) kidneys on the post-transplant graft failure hazard. The increase in the hazard for patients receiving an ECD (relative to non-ECD) kidney is found to be greatly underestimated in the existing literature which has used the information from the transplanted ECD organs only.

There exist methods in the current literature which are related to those proposed in Chapter 2, such as inverse-probability-of-treatment-weighting (IPTW) (Robins, 2000; Hernan et al., 2000; Hernan et al., 2001), and propensity scoring (Rosenbaum and Rubin, 1983; Rosenbaum, 2002). These methods are either infeasible for our real data setup or answer questions different from those of interest, as we describe later in some detail.

In applying the method proposed in Chapter 2, a practical question to be answered is whether or not the second stage sampling mechanism actually does generate bias. The effect of interest has two true underlying values; one describing the relationship between the factor under study and the outcome in the target population, the other describing the corresponding relationship in a hypothetical population represented by the potentially biased sample selected in the second sampling stage. These two underlying parameters would be different if and only if biasing factors exist. The same can be said of the baseline hazard. To identify cases where the two parameters differ and an empirically weighted proportional hazards model is necessary, two conceptually simple yet comprehensive tests are proposed in Chapter 3. For the regression parameter and cumulative baseline hazard, test statistics are constructed as the difference of the weighted estimates proposed in Chapter 2 and their unweighted counterparts. The asymptotic distributions of each of the test statistics is derived

and explicit variance estimators are proposed. The empirical significance level and power are validated through simulation. Factors affecting the power are examined. It is observed that as long as the size of the second stage sample is not too small (e.g., $n \geq 20$), the achieved power is generally quite high. A similar test for the baseline hazard function is proposed. Finally, both tests are applied to another renal failure data set, from which we seek to draw inference for the wait-listed patients (target population) and only observe transplanted patients (potential biased sample) with hospitalization history serving as the assumed potential biasing factor.

In Chapter 4, we study recurrent events in the presence of a terminating event. Several approaches to the analysis of recurrent event data exist in literature. If the cumulative number of events are of interest, one can model the recurrent event rate/mean; the rate being the derivative of mean function (e.g., Pepe and Cai, 1993; Lawless and Nadeau, 1995; Lin, Wei, Yang and Ying, 2000). If the instantaneous event occurrence probability given the event history is of interest, an intensity model can be employed (e.g., Prentice et al., 1981; Andersen and Gill, 1982; Wei et al., 1989; Lee et al., 1992). Depending on the form of covariate effects, the recurrent event model can be either multiplicative or additive (e.g., Schaubel, Zeng and Cai, 2006).

Another characteristic of the data structure of interest is the coexistence of a terminating event which is also affected by the treatment and adjustment covariates. Commonly applied methods to handle the terminating event either model the marginal number of the recurrent events averaging over living and deceased subjects (e.g., Ghosh and Lin, 2000; Ghosh and Lin, 2002) or model the conditional recurrent event rate given survival (e.g., Cook and Lawless, 1997; Lin et al., 2000). Both types of methods have their own limitations. For the marginal method, the occurrence of

more recurrent events in a treatment group could result from longer survival or a higher recurrent event rate while subjects are alive, and it is difficult to differentiate the two possible causes. For the conditional method, the cumulative treatment effects over time cannot be estimated by integrating the instantaneous rate over time, since the integral is only interpretable in the unrealistic setting where subjects never die.

We propose fitting a proportional hazards model for the terminating event and an additive model for the recurrent event rate conditional on survival separately, then integrating over time to estimate the cumulative recurrent event means. In this way, the treatment effect is measured through a time varying process without assuming a functional form for the treatment effect. Two measures are proposed along these lines. The first compares the treatment effect on the recurrent events and factors out differences in survival distributions between the groups. In order to isolate the treatment effect on the recurrent events, a hypothetical scenario is considered in which treatment-specific survival is equal. The second incorporates the potentially distinct survival distributions in treatment groups and reflects the mean difference in treatment-specific marginal recurrent event means, again computed by combining the survival probability and the conditional recurrent event rate.

Asymptotic properties of the estimators for both measures are derived and evaluated in finite samples. In addition, these methods are found to perform reasonably well under misspecified models. Finally the proposed methods are applied to kidney transplantation data to study the difference between ECD and non-ECD transplantation with respect to the cumulative number of post-transplant hospitalizations. In spite of the fact that ECD recipients die significantly earlier, they experience significantly more hospitalizations, accounting for the fact that death precludes subsequent

admissions.

CHAPTER II

Proportional Hazards Models Based on Biased Samples and Estimated Selection Probabilities

Abstract: In non-randomized biomedical studies using the proportional hazards model, the observed data often constitute a biased (i.e., unrepresentative) sample of the underlying target population, resulting in biased regression coefficients. The bias can be avoided by weighting included subjects by the inverse of their respective selection probabilities, as proposed by Horvitz & Thompson (1952) and extended to the proportional hazards setting for use in surveys by Binder (1992) and Lin (2000). The weights can be treated as fixed in cases where they are known (e.g., chosen by the investigator) or based on voluminous data (e.g., a large-scale survey). However, in many practical applications, the weights are estimated and must be treated as such in order for the resulting inference to be accurate. We propose a two-stage weighted proportional hazards model in which, at the first stage, weights are estimated through a logistic regression model fitted to a representative sample from the target population. At the second stage, a weighted Cox model is fitted to the biased sample. We propose estimators for the regression parameter and cumulative baseline hazard. Asymptotic properties of the parameter estimators are derived, accounting for the difference in the variance introduced by the randomness of the weights. The accuracy of the asymptotic approximations in finite samples is evaluated through

simulation. Our method is illustrated in an analysis of renal transplant patients using data obtained from the Scientific Registry of Transplant Recipients (SRTR).

Key words and phrases: Hazard regression; observational studies; selection bias; survival analysis; weighted Cox model.

2.1 Introduction

Often in observational studies, the study population observed is a biased sample of the target population. For example, if the importance of a certain characteristic is of interest (e.g., diabetes), then the true effect of being diabetic (relative to non-diabetic) in the target population may be greatly understated by a regression model based on the study sample if the diabetics entered into the study are systematically healthier than those in the target population (i.e., in ways unmeasured by the remaining model covariates). We propose a method for estimating the parameters in the Cox proportional hazards model (Cox 1972), for settings in which the survival model is fitted to a biased sample. Unlike existing weighted versions of the Cox model (Binder 1992; Lin 2000), the proposed method uses weights which are estimated and treated as such in deriving the asymptotic properties of the parameter estimators.

Our proposed method is motivated by the study of the use of "expanded criteria donor" (ECD) organs by end-stage renal disease patients. Kidney transplantation is the preferred method of renal replacement therapy (the alternative being dialysis) in terms of patient survival (Schaubel et al. 1995; Wolfe et al. 1999). With the increasing demand for donor organs, kidneys from ECDs are increasingly being considered for patients on the wait-list, even though it has been demonstrated that ECD organs are associated with a significantly increased risk of graft failure (Port et

al. 2002), which is usually defined as the minimum of time of death and time when the transplanted organ ceases to function.

Our objective is to accurately estimate the magnitude of the increase in the graft failure hazard for ECD relative to non-ECD deceased-donor kidneys. Although the fact that ECD kidneys have an increased risk of graft failure can already be concluded based on previous reports (e.g., Port et al. 2002), the true magnitude of the increase requires further investigation. It is necessary that an accurate estimate of the relative hazard of graft failure of ECD organs be available. Because this estimate is targeted to provide information for clinicians and wait-listed patients facing the choice of accepting or rejecting a procured kidney, inference should be based on the hypothetical scenario where all procured kidneys (ECD or non-ECD) are transplanted. A Cox model fitted to the post-transplant population is likely to underestimate the true average ECD effect, since organs in the transplanted population represent a biased sample of all the donor kidneys available for transplantation. Kidneys from younger healthier ECDs have a higher chance of being transplanted, and donor's health conditions have a direct influence on the graft failure hazard. If we compare the survival function of the "best" ECD kidneys to that of "average" non-ECD kidneys without adjusting for selection bias, the negative effects of ECD organs on survival will be attenuated. In reality, there will also be selection of non-ECD organs. But, it is likely that a much greater fraction of ECD organs are discarded, reflecting a greater degree of selection from the ECD organ pool.

A natural idea to generate consistent estimators of population parameters in the presence of biased samples is to weight each subject by the reciprocal of their probability of being sampled. In the context of the transplant data, subjects refer to kidneys and selection probability refers to the probability of the organ be-

ing transplanted, as opposed to discarded. Horvitz & Thompson (1952) proposed such a method to provide an unbiased estimate of a population mean. The use of inverse-probability-of-selection weighting was extended to the Cox proportional hazards model by Binder (1992), then Lin (2000) and Boudreau & Lawless (2006). Binder (1992) developed his method primarily for use in survey sampling, with the weights being chosen by the investigator. Thus, the partial likelihood score equation used weights based on inclusion probabilities which were treated as fixed, which is reasonable when the sampling scheme is known or information of the whole target population is collected (e.g., United States census). Lin (2000) further studied the case when the biased sample is selected from a representative sample of the target population, resulting in extra variation in the parameter estimates. Boudreau & Lawless (2006) proposed stratified proportional hazards models that account for the fact that the data have been collected according to a complex survey design. In many practical settings, inclusion probabilities are not known and must be estimated empirically by auxiliary information from the representative sample. If the weights are estimated consistently and treated as fixed, the Cox model parameter estimates are still consistent. However, not accounting for the fact that the weights are trained by the data leads to variance estimators which overestimate the true variability, resulting in conservative confidence intervals and significance levels.

In this article, we propose a two-stage procedure for fitting a weighted proportional hazards model. At the first stage, selection probabilities are estimated from a logistic regression model applied to both selected and unselected subjects. The inverse of the estimated selection probabilities are used as weights. At the second stage, a weighted Cox model is fitted, with the randomness in the weight estimates accounted for in the proposed inference procedures. In the causal inference literature, inverse-probability-

of-treatment weighting (IPTW) (Robins 2000; Hernan, Brumback & Robins 2000; Hernan, Brumback & Robins 2001) and pair-matching or stratification by propensity scores (Rosenbaum & Rubin 1983; Rosenbaum 2002) are widely used methods for biased samples. Differences between IPTW, propensity scoring and the proposed method are explained in Section 2.

The remainder of this article is organized as follows. In Section 2, we describe the proposed estimation procedure and its connection to existing methods. In Section 3, the asymptotic properties of our parameter estimators are provided. In Section 4, the applicability of the asymptotic properties in finite samples is evaluated by numerical studies. In Section 5, the proposed method is applied to data from a national organ failure registry. In Section 6, a discussion and some concluding remarks are provided. Proofs of the main results listed in Section 3 are outlined in the Appendix.

2.2 Models And Methods

We begin by establishing the necessary notation. For subject i ($i = 1, \ldots, N$) with event time T_i and censoring time C_i, we define $\widetilde{T}_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$, $Y_i(t) = I(\tilde{T}_i \geq t)$ and $N_i(t) = I(\tilde{T}_i \leq t, \Delta_i = 1)$. There are a total of N subjects in the representative sample, with a biased study sample of size $n < N$.

In analyzing time to event data, proportional hazards are often assumed, such that the hazard function at time t for subject i with $p \times 1$ covariate vector $Z_i(t)$ is proportional to the baseline hazard function at time t,

$$\lambda_i(t) = \lambda_0(t) e^{\beta_0' Z_i(t)},$$

where β_0 is the true value of the regression parameter vector and the true baseline hazard function $\lambda_0(t)$ is unspecified. Ordinarily, one would estimate β_0 by $\widehat{\beta}$, the

solution to the partial likelihood (Cox 1975) score equation

$$(2.1) \qquad U(\beta) = \sum_{i=1}^{N} \int_{0}^{\tau} \{Z_i(t) - \overline{Z}(\beta, t)\} dN_i(t) = 0,$$

where $dN_i(t) = N_i(t + dt) - N_i(t)$, τ is a prespecified constant satisfying $Pr(\widetilde{T}_i > \tau) > 0$ and typically set to the end of the study to include all observed event times,

$$
\begin{aligned}
\overline{Z}(\beta, t) &= \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}, \\
S^{(0)}(\beta, t) &= n^{-1} \sum_{i=1}^{n} Y_i(t) e^{\beta' Z_i(t)}, \\
S^{(1)}(\beta, t) &= n^{-1} \sum_{i=1}^{n} Y_i(t) e^{\beta' Z_i(t)} Z_i(t).
\end{aligned}
$$

Under mild conditions, Andersen & Gill (1982) proved that $n^{1/2}(\widehat{\beta} - \beta_0)$ converges asymptotically to a zero-mean Gaussian process with covariance consistently estimated by $\widehat{A}(\widehat{\beta})^{-1}$ where $\widehat{A}(\widehat{\beta}) = -n^{-1} \partial U(\beta)/\partial \beta|_{\beta = \widehat{\beta}}$.

In instances where subjects may have unequal probability of being sampled, Binder (1992) & Lin (2000) proposed the weighted proportional hazards model. The parameter β_0 is estimated by incorporating weights into the score equation,

$$(2.2) \qquad U_N^w(\beta) = \sum_{i=1}^{N} \int_{0}^{\tau} w_i \{Z_i(t) - \overline{Z}_w(\beta, t)\} \, dN_i(t),$$

where $w_i = I_i/p_i$, I_i is the indicator for the ith subject being sampled, $p_i = Pr(I_i = 1)$ and

$$
\begin{aligned}
\overline{Z}_w(\beta, t) &= \frac{S_w^{(1)}(\beta, t)}{S_w^{(0)}(\beta, t)}, \\
S_w^{(0)}(\beta, t) &= N^{-1} \sum_{i=1}^{N} w_i Y_i(t) e^{\beta' Z_i(t)}, \\
S_w^{(1)}(\beta, t) &= N^{-1} \sum_{i=1}^{N} w_i Y_i(t) e^{\beta' Z_i(t)} Z_i(t).
\end{aligned}
$$

Lin (2000) proved that the solution to (3.4) is also asymptotically zero-mean normal with an explicit covariance matrix expression. Both Binder (1992) and Lin

(2000) studied the case when the weights are known and hence their estimating functions have the same form. Binder's method can be used when the biased study sample is selected directly from the target population with fixed probabilities; Lin's proposal further studied the settings where the biased study sample is selected from a representative sample of the underlying target population, with the target population being considered infinite and referred to as a super-population. The difference lies in the asymptotic variance of the parameter estimators. The proposed variance estimator in Lin (2000) is larger than that proposed by Binder (1992), owing to the additional variation attributable to the random selection of the representative sample from the target population.

In existing weighted Cox models, w_i is assumed known and treated as fixed. There are many practical settings where inclusion probability is not known and must be estimated empirically. We propose a two-stage method wherein the probabilities used in the weighted Cox model are estimated from a logistic regression model based on both selected and unselected subjects. Our proposed method uses the same setting as in Lin (2000), but with selection probabilities that are not fixed and are empirically estimated. We denote the true weight by:

$$w_i(\theta_0) = \frac{I_i}{p_i(\theta_0)},$$

where θ_0 is the true value of the $q \times 1$ parameter vector in the logistic regression and hence

$$p_i(\theta) = \frac{e^{\theta' X_i}}{1 + e^{\theta' X_i}},$$

and X_i is the vector of covariates in the logistic regression model. Ideally, X_i is the complete set of biasing factors. By biasing factors, we refer to factors predictive of selection probability and survival time and correlate with the factor of interests.

In the case of unmeasured biasing factors, consistent estimators of θ_0 and hence β_0 are unobtainable. The robustness of our estimators to unmeasured confounders is examined through simulation in Section 4. Note that internal covariates governing both p_i and $\lambda_i(t)$ would be included in both X_i and Z_i. When the Z_i covariate is time-dependent (i.e., $Z_i(t) \neq Z_i(0)$), it would make sense that $Z_i(0)$ be included in X_i, but not $Z_i(t)$ for $t > 0$, since selection into the biased sample is determined at $t = 0$.

Through maximum likelihood, θ_0 is estimated by $\widehat{\theta}$, the solution to $U_L(\theta) = 0$, with

$$
\begin{aligned}
U_L(\theta) &= \sum_{i=1}^{N} U_{Li}(\theta), \\
U_{Li}(\theta) &= X_i\{I_i - p_i(\theta)\}.
\end{aligned}
$$

Having estimated θ_0, β_0 is estimated by $\widehat{\beta}_w$, the solution to $U_N^w(\beta, \widehat{\theta}) = 0$, where

$$
U_N^w(\beta, \theta) = \sum_{i=1}^{N} \int_0^\tau w_i(\theta)\{Z_i(t) - \overline{Z}_w(\beta, \theta, t)\} \, dN_i(t).
$$

Having computed $\widehat{\beta}_w$, the cumulative hazard function $\Lambda_0(t)$ can be consistently estimated by $\widehat{\Lambda}_0^w(\widehat{\beta}_w, \widehat{\theta}; t)$ where

$$
\widehat{\Lambda}_0^w(\beta, \theta; t) = N^{-1} \sum_{i=1}^{N} \int_0^t \frac{w_i(\theta) dN_i(s)}{S_w^{(0)}(\beta, \theta; s)}.
$$

We describe the asymptotic properties of the proposed parameter estimators in the next section.

The proposed method is built around inverse probability of selection weighting (IPSW), an idea related to but distinct from inverse probability of treatment weighting (IPTW) (Robins & Greenland 1994; Robins, Greenland & Hu 1999; Robins 2000). To avoid confusion, it is worth clarifying some of the differences between IPSW and IPTW, and in doing so we restrict attention to the setting of interest; i.e., data to estimate selection probability are available from a super-population. In IPSW,

subjects are weighted by the inverse of their respective selection probabilities, which is the transplantation probability per kidney in the motivating example. In IPTW, subjects are weighted by the inverse of the probability they received their assigned treatment, which would be probability of receiving an ECD kidney per recipient. The purpose of IPSW is to mimic the composition of the representative sample, while the IPTW method aims to obtain an unconfounded comparison between the treatments. In IPSW, the logistic model uses information from all subjects in the representative sample. In IPTW, the logistic model would be fitted to only the biased sample. Note also that a logistic model would suffice for IPTW when the covariate of interest was binary. However, it would appear that estimation of the weight would become cumbersome if the treatment covariate was continuous or, even worse, multi-dimensional. In contrast, the IPSW weight is easily estimated, whether the weights are intended to correct bias with respect to a single covariate (e.g., ECD) or several (e.g., ECD, age, gender, race, diagnosis, etc).

It is also worth noting that, for the application of our interest in the current study, IPTW does not apply and does not even make sense in principle. For the kidney transplant data, the representative sample contains all kidneys recovered in $1999 - 2003$, while the biased sample consists of kidneys which are actually transplanted. The factor of interest is ECD status, while selection probability represents the probability that a recovered kidney is transplanted. For an IPTW analysis, one would estimate the probability of a transplanted organ being its actual status (ECD or non-ECD) given all donor characteristics, including those in the ECD definition. As such, in attempting to apply IPTW, one would weight subjects by $Pr(E_i = z|X_i)^{-1}$, where E_i is a binary indicator for ECD status and z is a realization of E_i. Since X_i contains the components of the ECD definition, $Pr(E_i = z|X_i)$

always equals 0 or 1.

The IPSW method is also related to but different from methods based on propensity scores (Rosenbaum & Rubin 1983; Rosenbaum 2002), as the latter involves scores constructed from the probability of being assigned the treatment (as opposed to placebo), given all confounders used for pair-matching or stratification.

2.3 Asymptotic Properties

In this section, we describe the large-sample properties of the model parameter estimators. Proof sketches are provided in the Appendix. Note that since $w_i(\theta)$ will contain elements not in the filtration pertaining to the assumed Cox model, the Martingale Central Limit Theorem (Fleming & Harrington 1991; Kalbfleisch & Prentice 2002) is not applicable here. We assume the following regularity conditions:

(a) $\{T_i, C_i, Z_i, X_i\}$ are independent and identically distributed for $i = 1, \ldots, N$.

(b) $Pr(\widetilde{T}_i > \tau) > 0$.

(c) $\int_0^\tau d\Lambda_0(t) < \infty$ and $\Lambda_0(t)$ is differentiable over time.

(d) $|X_{ik}| < \infty$; $|Z_{ik}(0)| + \int_0^\tau |Z_{ik}(s)|ds < \infty$ almost surely, where the second subscript refers to the kth element.

(e) Positive-definiteness of the matrices, $A^w(\beta, \theta)$ and $B(\theta)$, where

$$
A^w(\beta, \theta) = E\left[\int_0^\tau \{Z_i(t) - \bar{z}_w(\beta, \theta; t)\}^{\otimes 2} w_i(\theta)Y_i(t)e^{\beta' Z_i(t)}d\Lambda_0(t)\right],
$$

$$
s_w^{(1)}(\beta, \theta; t) = \lim_{N\to\infty} S_w^{(1)}(\beta, \theta; t),
$$

$$
s_w^{(0)}(\beta, \theta; t) = \lim_{N\to\infty} S_w^{(0)}(\beta, \theta; t),
$$

$$
\bar{z}_w(\beta, \theta; t) = \frac{s_w^{(1)}(\beta, \theta; t)}{s_w^{(0)}(\beta, \theta; t)},
$$

$$
B(\theta) = E[X_i p_i(\theta)\{1 - p_i(\theta)\}X_i']
$$

with $a^{\otimes 2} = aa'$.

(f) There exists a δ such that $p_i(\theta) > \delta > 0$ almost surely.

Condition (a) is employed in the application of the Functional Central Limit Theorem (Pollard 1990). Provided that the subjects are independent, the assumption applies quite generally. Independence would be violated if, for example, subjects are clustered by some factor related to the failure time of interest. Condition (b) is a standard identifiability requirement. Condition (c) leads to the boundedness of several quantities examined in the proofs of the theorems stated later in this section. The bounded covariate condition, (d), is not required but simplifies proofs of the asymptotic results and is applicable in most practical situations. Assumption (e) essentially requires that there are no linear dependencies among the covariates in either the survival or selection probability models. With respect to condition (f), selection probability is non-zero for all subjects of interest in the underlying target population, by definition. This condition guarantees that N and n go to ∞ at the same rate, such that $\sqrt{n/N}$ converges to a constant.

The proposed methods also assume implicitly that $Z_i \cup X_i$ consists of all factors affecting $\lambda_i(t)$; that is, the no-unmeasured confounders assumption. In addition, we assume that C_i is conditionally independent of T_i; specifically,

$$\lim_{\delta \to 0} \frac{1}{\delta} Pr(t \leq T_i < t + \delta | T_i \geq t, C_i \geq t, Z_i, X_i) = \lim_{\delta \to 0} \frac{1}{\delta} Pr(t \leq T_i < t + \delta | T_i \geq t, Z_i, X_i).$$

THEOREM 1. *Under conditions (a) to (f), $\widehat{\beta}_w$ converges almost surely to β_0.*

The proof of Theorem 1 is outlined in the Appendix. It proceeds through a Taylor Series expansion, repeated application of the Strong Law of Large Numbers (SLLN), followed by arguments from convex function theory.

THEOREM 2. *Under conditions (a) to (f), $N^{\frac{1}{2}}(\widehat{\beta}_w - \beta_0)$ is asymptotically zero-mean normal with covariance matrix $A^w(\beta_0, \theta_0)^{-1}\Sigma_w(\beta_0, \theta_0)A^w(\beta_0, \theta_0)^{-1}$ as $N \to \infty$, where $A^w(\beta_0, \theta_0)$ is as defined in condition (e),*

$$
\begin{aligned}
\Sigma_w(\beta, \theta) &= E\{\psi_i^w(\beta, \theta)^{\otimes 2}\}, \\
\psi_i^w(\beta, \theta) &= \int_0^\tau w_i(\theta)\{Z_i(t) - \overline{z}_w(\beta, \theta; t)\}dM_i(t) - G(\beta, \theta)B(\theta)^{-1}U_{Li}, \\
G(\beta, \theta) &= E\left[\int_0^\tau \{Z_i(t) - \overline{z}_w(\beta, \theta; t)\}X_i' e^{-\theta'X_i}dM_i(\beta; t)\right]
\end{aligned}
$$

with $dM_i(\beta; t) = dN_i(t) - Y_i(t)e^{\beta'Z_i}d\Lambda_0(t)$.

The matrix $\Sigma_w(\beta_0, \theta_0)$ can be consistently estimated by replacing β_0, θ_0 and all expectations by their empirical estimates; that is,

$$
\widehat{\Sigma}_w(\widehat{\beta}_w, \widehat{\theta}) = N^{-1} \sum_{i=1}^N \left[\int_0^\tau w_i(\widehat{\theta})\{Z_i(t) - \overline{Z}_w(\widehat{\beta}_w, \widehat{\theta}; t)\}d\widehat{M}_i(\widehat{\beta}, \widehat{\theta}; t) - \widehat{G}(\widehat{\beta}_w, \widehat{\theta})\widehat{B}(\widehat{\theta})^{-1}\widehat{U}_{Li}(\widehat{\theta})\right]^{\otimes 2},
$$

where

$$
\begin{aligned}
d\widehat{M}_i(\beta, \theta; t) &= dN_i(t) - Y_i(t)e^{\beta'Z_i}d\widehat{\Lambda}_0^w(\beta, \theta; t), \\
\widehat{G}(\beta, \theta) &= N^{-1}\sum_{i=1}^N \int_0^\tau \{Z_i(t) - \overline{Z}_w(\beta, \theta; t)\}X_i'e^{-\theta'X_i}d\widehat{M}_i(\beta; t), \\
\widehat{B}(\theta) &= -N^{-1}\frac{\partial U_L(\theta)}{\partial \theta'} = N^{-1}\sum_{i=1}^N X_i p_i(\theta)\{1 - p_i(\theta)\}X_i'.
\end{aligned}
$$

Similarly, the matrix $A^w(\beta_0, \theta_0)$ can be consistently estimated by $\widehat{A}^w(\widehat{\beta}_w, \widehat{\theta})$, where

$$
(2.3) \quad \widehat{A}^w(\beta, \theta) = N^{-1}\sum_{i=1}^N \int_0^\tau \{Z_i(t) - \overline{Z}_w(\beta, \theta; t)\}^{\otimes 2}w_i(\theta)Y_i(t)e^{\beta'Z_i(t)}d\widehat{\Lambda}_0^w(\beta, \theta; t).
$$

Therefore, the covariance matrix for $N^{\frac{1}{2}}\widehat{\beta}_w$ can be consistently estimated by $\widehat{A}^w(\widehat{\beta}_w, \widehat{\theta})^{-1}\widehat{\Sigma}_w(\widehat{\beta}_w, \widehat{\theta})\widehat{A}^w(\widehat{\beta}_w, \widehat{\theta})^{-1}$. The asymptotic distribution of $\widehat{\beta}_w$ can be derived by combining the Multivariate Central Limit Theorem (van der Vaart 2000) and

results from empirical process theory (Pollard 1990; Wellner 1996; Bilias, Gu & Ying 1997).

If the weights are treated as fixed, the variance estimator differs from the proposed one with respect to the ψ_i^w component. Under the setting of designed sampling over the target population with fixed selection probability, the variance estimator also has the sandwich form,

$$(2.4) \qquad \widehat{A}^w(\widehat{\beta}_w, \widehat{\theta})^{-1} \widehat{\Sigma}_f(\widehat{\beta}_w, \widehat{\theta}) \widehat{A}^w(\widehat{\beta}_w, \widehat{\theta})^{-1},$$

where $\widehat{A}^w(\widehat{\beta}_w, \widehat{\theta})$ is the same as that defined in (2.3) and

$$\begin{aligned}
\Sigma_f(\beta) &= E\{\psi_i^f(\beta)^{\otimes 2}\}, \\
\psi_i^f(\beta) &= \int_0^\tau w_i(\theta)\{Z_i(t) - \bar{z}_w(\beta, \theta; t)\} dM_i(t).
\end{aligned}$$

We summarize the essential asymptotic results for the baseline cumulative hazard estimator in the following two theorems.

THEOREM 3. *Under conditions (a) to (f), $\widehat{\Lambda}_0^w(\widehat{\beta}_w, \widehat{\theta}; t)$ converges uniformly to $\Lambda_0(t)$ for $t \in [0, \tau]$.*

The proof of Theorem 3 involves decomposing $\{\widehat{\Lambda}_0^w(\widehat{\beta}_w, \widehat{\theta}; t) - \Lambda_0(t)\}$ into $\{\widehat{\Lambda}_0^w(\widehat{\beta}_w, \widehat{\theta}; t) - \widehat{\Lambda}_0^w(\widehat{\beta}_w, \theta_0; t)\}$, $\{\widehat{\Lambda}_0^w(\widehat{\beta}_w, \theta_0; t) - \widehat{\Lambda}_0^w(\beta_0, \theta_0; t)\}$ and $\{\widehat{\Lambda}_0^w(\beta_0, \theta_0; t) - \Lambda_0(t)\}$, then applying the Uniform SLLN and various empirical process results.

THEOREM 4. *Under conditions (a) to (f), $N^{\frac{1}{2}}\{\widehat{\Lambda}_0^w(\widehat{\beta}_w, \widehat{\theta}; t) - \Lambda_0(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function:*

$$\Omega(s, t) = E\{\phi_i(\beta_0, \theta_0; s)\phi_i(\beta_0, \theta_0; t)\},$$

for $(s,t) \in [0,\tau] \times [0,\tau]$, where $\phi_i(\beta,\theta;t) = \phi_{i1}(\beta,\theta;t) + \phi_{i2}(\beta,\theta;t) + \phi_{i3}(\beta,\theta;t)$, with

$$
\begin{aligned}
b(\beta,\theta;t) &= E\{X_i e^{-\theta' X_i} Y_i(t) e^{\beta' Z_i(t)}\}, \\
k(\beta,\theta;t) &= -E\left\{ \int_0^t \frac{X_i e^{-\theta' X_i}}{s_w^{(0)}(\beta,\theta;s)} dN_i(s) + \int_0^t \frac{w_i(\theta) b(\beta,\theta;s)}{s_w^{(0)}(\beta,\theta;s)^2} dN_i(s) \right\}, \\
\phi_{i1}(\beta,\theta;t) &= k'(\beta,\theta;t) B(\theta)^{-1} U_{Li}(\theta), \\
h(\beta,\theta;t) &= -\int_0^t \bar{z}_w(\beta,\theta;s) d\Lambda_0(s), \\
\phi_{i2}(\beta,\theta;t) &= h'(\beta,\theta;t) A^w(\beta,\theta)^{-1} \psi_i^w(\beta,\theta), \\
\phi_{i3}(\beta,\theta;t) &= \int_0^t \frac{dM_i(\beta,\theta;s)}{s_w^{(0)}(\beta,\theta;s)}.
\end{aligned}
$$

The proof involves applications of the Central Limit Theorem and empirical process results. The quantity $\Omega(s,t)$ can be consistently estimated by replacing all unknown quantities by their empirical counterparts.

2.4 Simulation Study

In our simulation settings, there are three covariates, with $Z_i = (Z_{i1}, Z_{i2}, Z_{i3})'$. The covariate Z_{i1} is distributed as Bernoulli with equal probabilities of being 1 or 0; Z_{i2} is normally distributed with mean zero and variance 25; Z_{i3} is uniformly distributed on $(0,4)$. Time to event, T_i, follows an exponential distribution with hazard $\lambda_i(t) = \lambda_0 e^{\beta_0' Z_i}$, where $\lambda_0 = 0.02$, and $\beta_0' = (\beta_1, \beta_2, \beta_3) = (0.5, 0.1, 1.0)$. The censoring time, C_i, was generated as uniform on $(0,20)$ or $(0,40)$ to create data sets with different percentages of censoring (denoted by $C\%$); specifically 20% and 30%. We create representative samples with three different sizes (500, 1000 or 5000) and biased samples are created by selecting different percentages of subjects in each Z_{i1} and Z_{i3} combination. Specifically, subjects with $Z_{i1} = 0$ are always included

in the biased sample; for subjects with $Z_{i1} = 1$, probability of being included in the biased sample depends on Z_{i3} with selection probability being constant across $[0, 1], (1, 2], (2, 3], (3, 4]$. Each data configuration is replicated 1000 times.

Of primary interest are the marginal effects of Z_{i1} and Z_{i2}, while Z_{i3} represents a variable that affects $\lambda_i(t)$ but is not incorporated in the Cox model. Hence the Cox model fitted to each replicate is $\lambda_i(t) = \lambda_0(t)e^{\beta_1 Z_{i1} + \beta_2 Z_{i2}}$. The weights were based on the logistic model, $\log\{p_i(\theta)/(1 - p_i(\theta))\} = \theta' X_i$, where $X_i = (Z_{i1}, Z_{i3})'$. It is worth noting that very few failure time distributions (e.g. positive stable frailty) have proportional hazards for both the conditional and the marginal models. In cases where the proportional hazards assumption does not hold perfectly, the β_1 estimate can be understood as a time-weighted average Z_{i1} effect; albeit with weights for which no explicit form exists and which depend on the censoring distribution (Struthers & Kalbfleisch, 1986).

Although in the representative sample Z_{i3} is independent of Z_{i1}, in the biased sample Z_{i3} distributes differently for $Z_{i1} = 0$ and $Z_{i1} = 1$. Therefore, the marginal effects of Z_{i1} over Z_{i3} in the representative sample and in the biased sample are unequal. Thus, in the biased sample, omitting biasing factor Z_{i3} from the Cox model results in biased estimator of β_1 if we use an unweighted Cox model. The true β_1 is determined empirically by applying proportional hazards model to the representative sample. Take the following data configuration as an example: selection probability on the four increasing levels of Z_{i3} are 1, 0.5, 0.4 and 0.1, and the parameter estimates from the representative sample is $(\beta_1, \beta_2)' = (0.318, 0.064)'$. Note that there are two sets of true parameters for the representative sample here; the conditional regression coefficients $(\beta_1, \beta_2, \beta_3)' = (0.5, 0.1, 1.0)'$ and the marginal regression coefficients $(\beta_1, \beta_2)' = (0.318, 0.064)'$. Because our goal is to estimate the marginal effect,

| C%[1] | | $Pr(I_i = 1 \| Z_{1i} = 1, Z_{3i})$ | | | | $\widehat{\beta}_1$ | | | $\widehat{\beta}_{1w}$ | | | |
| | | | | | | | | | $w_i(\widehat{\theta})$: fixed [2] | | $w_i(\widehat{\theta})$: estimated [3] | |
	N	[0,1]	(1,2]	(2,3]	(3,4]	Bias	Bias	ESD	ASE	CP	ASE	CP
30	250	1.000	0.500	0.400	0.100	-0.47	-0.03	0.20	0.22	0.95	0.18	0.92
	500	1.000	0.500	0.400	0.100	-0.47	-0.00	0.14	0.16	0.97	0.13	0.94
	500	1.000	0.750	0.500	0.250	-0.34	-0.01	0.12	0.13	0.98	0.12	0.94
	500	0.500	0.500	0.500	0.500	-0.00	-0.00	0.13	0.14	0.97	0.13	0.94
	1000	0.500	0.250	0.200	0.050	-0.47	-0.01	0.11	0.14	0.99	0.11	0.94
	5000	0.100	0.050	0.040	0.010	-0.47	-0.01	0.10	0.14	0.99	0.09	0.93
	5000	0.200	0.100	0.075	0.025	-0.47	-0.01	0.07	0.09	1.00	0.07	0.95
20	250	1.000	0.500	0.400	0.100	-0.43	-0.03	0.18	0.19	0.96	0.16	0.93
	500	1.000	0.500	0.400	0.100	-0.42	-0.01	0.11	0.14	0.98	0.11	0.95
	500	1.000	0.750	0.500	0.250	-0.30	-0.01	0.11	0.12	0.95	0.11	0.95
	500	0.500	0.500	0.500	0.500	-0.00	-0.00	0.12	0.13	0.96	0.12	0.95
	1000	0.500	0.250	0.200	0.050	-0.42	-0.01	0.10	0.13	0.99	0.09	0.94
	5000	0.100	0.050	0.040	0.010	-0.42	-0.01	0.08	0.12	0.99	0.08	0.94
	5000	0.200	0.100	0.075	0.025	-0.42	-0.01	0.06	0.08	0.99	0.06	0.94

[1] $C\% = \widehat{E}(I(C_i < T_i))$.

[2] $w_i(\widehat{\theta})$: fixed : variance estimator which treats weights as fixed, as given in equation (2.4).

[3] $w_i(\widehat{\theta})$: estimated : proposed variance estimator, which treats the weights as estimated and is derived in Theorem 1.

Table 2.1: Simulation Results:Weighted and Unweighted Proportional Hazards Model.

$(0.318, 0.064)'$ are the target population parameters of interest. If we fit a proportional hazards model to the selected sample without any weight adjustment, serious bias is observed; the empirical mean of $\widehat{\beta}_1$ across the $1,000$ replicates computed as $\widehat{E}(\widehat{\beta}_1) = -0.152$. Note that $\widehat{\beta}_2$ is still approximately unbiased, with $\widehat{E}(\widehat{\beta}_2) = 0.067$, since Z_{i2} is not correlated with Z_{i3} and therefore not affected by the omission of Z_{i3} from the fitted Cox model.

Simulation results are listed in Table 1. For each data configuration, the bias of the unweighted and weighted β_1 estimates (denoted by $\widehat{\beta}_1$ and $\widehat{\beta}_{w1}$, respectively) are compared. We also list side by side the average asymptotic standard error (ASE) for

two different variance estimators; the first variance estimator treats the weights as fixed and is given by (2.4), while the second is the proposed variance estimator based on Theorem 1. In addition, both ASEs are compared to the empirical standard deviation (ESD). The comparison between the two variance estimators is also made with respect to empirical coverage probability (CP), which has nominal level 0.95. From Table 1, the unweighted β_1 estimates are obviously biased, while the corresponding weighted estimates are approximately unbiased. Treating the weights as fixed leads to over-estimation of $SE(\widehat{\beta}_{w1})$, with coverage probability being considerably larger than 95%. The proposed SE estimator is on average much closer to the ESD and correspondingly has empirical coverage probability more closely approximating the nominal 0.95.

Rather than increasing the variation of $\widehat{\beta}_w$, our proposed variance estimator actually has smaller variance than the variance estimator which treats the weights as fixed. The decrease in the variance can be understood heuristically by the fact that the weights are "trained" by the data, since subjects used to fit the weighted Cox model are also used to estimate θ_0. In row $6-9$, this difference in ASE and CP remains the same with sample sizes increasing from 500 to 5000.

We also examined the setting where the selection probabilities and, hence, weights were in fact known. In this case, with the weights appropriately treated as fixed in estimating the variance, the β_1 estimator is approximately unbiased, but with ESD considerably larger than that of $\widehat{\beta}_{w1}$ (data not tabulated).

Next, we examine the impact of misspecification of the selection probability model. The overall criteria for model performance is to minimize the sum of bias square and variance; that is, mean square error (MSE). The simulation results listed in Table 1 are generated using a correct logistic model which includes all covariates that

influence the selection probability, with each covariate modelled using its correct functional form. Additional simulations aimed at assessing the MSE of our estimator disclose that both bias and variance of $\widehat{\beta}_w$ increases with the degree to which the selection probability model is misspecified. In the most extreme case of model misspecification when the logistic model contains no covariates (i.e., an intercept-only model), $\widehat{\beta}_w$ and $\widehat{SE}(\widehat{\beta}_w)$ equal those from the unweighted model. The impact of misspecifying the selection probability model is exemplified in Table 2. In the last row of Table 2, the correct selection probability model is fitted. Other rows feature various degrees of model misspecification, with the misspecification generally increasing from bottom to top. Comparing the ASEs computed using Theorem 1 to the ESD, the proposed SE estimator agrees with the empirical SE even when the selection probability model is misspecified. The more accurate the selection probability model, the smaller the bias and the greater the efficiency gain associated with the proposed variance estimator relative to that which treats the weights as fixed.

2.5   Application

Of interest is the impact of Expanded Criteria Donor (ECD) kidneys on the graft failure hazard for renal transplant recipients. The time of graft failure is defined as the time between transplantation and the earliest of the time of death and the time at which the transplanted kidney ceases to function. By the definition of Port et al. (2002), ECDs are either (i) age ≥ 60, or (ii) age $50 - 59$ and with at least two of the following three characteristics: hypertensive, serum creatinine concentration $> 1.5\text{mg/dl}$, or death due to stroke. To the nephrology community, ECD is a well-accepted quality index for donated kidneys. In fact, patients who are willing to accept an ECD organ are essentially placed on a separate waiting list. It is well-known that

Covariates in Logistic Model	Bias		ASE		
	$\widehat{\beta}_1$	$\widehat{\beta}_{1w}$	$w_i(\widehat{\theta})$: fixed [4]	$w_i(\widehat{\theta})$: estimated [5]	ESD
various incorrect models:					
no covariates	-0.427	-0.427	0.101	0.101	0.100
categorical Z_3	-0.423	-0.429	0.099	0.099	0.098
Z_1	-0.427	-0.429	0.103	0.103	0.105
Z_1, linear Z_3	-0.423	-0.040	0.103	0.079	0.080
Z_1, categorical Z_3	-0.427	-0.015	0.120	0.077	0.081
correct model:					
Z_1, categorical Z_3, interaction of categorical Z_3 and Z_1	-0.428	-0.014	0.120	0.077	0.080

[4] $w_i(\widehat{\theta})$: fixed : variance estimator which treats weights as fixed, as given in equation (2.4).
[5] $w_i(\widehat{\theta})$: estimated : proposed variance estimator, which treats the weights as estimated and is derived in Theorem 1.

Table 2.2: Impact of selection probability model misspecification on bias and efficiency of proposed variance estimator.

ECD kidneys have higher risk of graft failure or death. However, we hypothesize that the magnitude of the increase is considerably greater than that currently reported in the nephrology literature. Suppose that among the ECD kidneys procured, only the "healthiest" are selected for transplantation (the remainder discarded). Suppose also that the same sort of selection process occurs for non-ECD organs, but to a much lesser extent. If we fit an unweighted model using transplanted kidneys only, we would be comparing the healthiest ECD kidneys with a closer-to-representative sample of non-ECD kidneys, and the negative impact of ECD on graft survival would therefore be underestimated. In order to remove the effect of the biased sampling, we apply the proposed weighted proportional hazards model. The covariate vector includes terms for recipients (age, gender, race, years on ESRD and diabetic status) and donor ECD status. Note that adjusting for the ECD components in the Cox model is undesirable since the ECD parameter of interest cannot be identified in the

presence of such adjustment.

Our treatment of the time axis warrants some discussion. First, since the time origin is the time of transplant, all patients begin observation at $t = 0$, such that left truncation does not occur. Second, there is the issue of length-biased sampling, discussed by Wang (1999). That is, patients with longer time survived on the waitlist would be over-represented in the data set. However, as previously indicated, "time on dialysis prior to transplant" is an element of Z_i. Finally, on a related note, induced dependent censoring potentially resulting from correlation between waitlist-to-transplant and transplant-to-death gap times (Schaubel & Cai 2004) is also mitigated; i.e., the second gap time is, essentially, modelled as a function of the first gap time.

The target population is all kidneys from deceased donors in U.S.. The representative sample consists of all such kidneys with initial referral calls to their Organ Procurement Organization (OPO) made during the 1999 to 2003 period. Demographic and clinical data on donors and recipients, dates of graft failure and death where applicable, as well as various clinical measures, are obtained from the Scientific Registry of Transplant Recipients (SRTR) and collected by the Organ Procurement and Transplant Network (OPTN). Kidneys transplanted to recipients under age 18 are excluded. Each recipient was followed until graft failure, loss to follow up or the conclusion of the observation period (December 31, 2004). In total, there were 57,213 donors with complete information. Of the $12,673$ ECD kidneys, $5,830$ (46%) got transplanted. In contrast, of the $44,540$ non-ECD kidneys procured, $37,227$ (84%) were transplanted. Altogether $43,057$ (75%) out of the procured kidneys got transplanted. Among the $43,057$ transplantations, $4,452$ had missing information (transplantation date unknown), $29,640$ recipients were alive with a functioning

transplant at the end of follow-up, and $8,965$ (23%) experienced graft failure (i.e., either died or had a transplant that stopped functioning) before the end of 2004.

Since the SRTR data include complete information of both transplanted and discarded kidneys, we fit a logistic model to estimate each procured kidney's transplantation probability, adjusting for donor characteristics, including: demographics (gender, age, race), cause of death (stroke, anoxia, tumor, others), lifestyle habits (alcohol, smoking, cocaine, other drug), disease history (cancer, diabetes, hypertension), serology results (hepatitis B antibody, hepatitis C antibody, cytomegalovirus antibody), cardiovascular disease, high creatinine level, blood urea nitrogen, presence of protein in urine, clinical infection, and having tattoos. Among these variables, donor age, hypertension history, high creatinine level, and stroke as cause of death are used in the ECD definition. The weight assigned to each recipient is the inverse of the estimated probability that their respective organ was transplanted as opposed to discarded.

Selected parameters from the selection probability model are listed in Table 3. All ECD-related variables have significant effects on the probability of being selected for transplantation. The model appears to fit fairly well in a general sense, as the proportion of concordant outcomes (C-statistic) equals 82%.

Comparison of the weighted and unweighted ECD coefficient estimates and different estimators for the standard error from the weighted Cox model is made in Table 4. We observe a great difference between the unweighted and weighted ECD coefficient estimates, with $\widehat{\beta} = 0.54$ versus $\widehat{\beta}_w = 0.79$. We can interpret these results as follows: among the kidneys transplanted, patients getting an ECD kidney have $e^{0.54} = 1.72$ times higher hazards for graft failure compared to patients getting a non-ECD kidney; while among all the kidneys recovered for transplantation, ECD

k	X_{ik}	$\widehat{\theta}_k$	$\widehat{SE}(\widehat{\theta}_k)$	p	$e^{\widehat{\theta}_k}$
1	High creatinine	-1.17	0.04	< 0.0001	0.31
2	Hypertension	-0.47	0.03	< 0.0001	0.62
3	Cause of death = stroke	-0.18	0.03	< 0.0001	0.84
4	Age 60 - 64	-0.58	0.06	< 0.0001	0.56
5	Age 65 - 69	-1.19	0.06	< 0.0001	0.31
6	Age $\geq$ 70	-2.59	0.07	< 0.0001	0.08

Table 2.3: Estimated regression parameter for ECD-related characteristics.

Quantity	Estimate	Formula	Comment
$\widehat{\beta}$	0.54	solution to (3.3)	unweighted parameter estimator
$e^{\widehat{\beta}}$	1.72		unweighted hazard ratio
$\widehat{\beta}_w$	0.79	solution to (3.4)	weighted parameter estimator
$e^{\widehat{\beta}_w}$	2.20		weighted hazard ratio
$\widehat{SE}$	0.152	equation (2.4)	treating $w_i(\widehat{\theta})$ as known
$\widehat{SE}$	0.149	Theorem 1	treating $w_i(\widehat{\theta})$ as estimated
p	< 0.0001		based on proposed method

Table 2.4: Analysis of kidney transplant data estimated regression parameter and standard error for ECD covariate.

status is associated with a $e^{0.79} = 2.20$ times higher graft failure hazard. If we treat the weights as fixed, the estimator for the standard error of the ECD element of $\widehat{\beta}_w$ is 0.152. Using the proposed variance estimator, the standard error is estimated at 0.149. In this example, whether we treat the weights either as fixed or as estimated, the ECD indicator is highly significant ($p < 0.0001$). However, it is easy to envision other cases where the overestimated SE leads to a deceptively large p value, a wider confidence interval and perhaps a type I error.

2.6   Discussion

Building on the weighted proportional hazards models proposed by Binder (1982), Lin (2000) and Boudreau & Lawless (2006), this paper examines the setting where weights are estimated from the data and thus should be treated as such for the purposes of inference. The estimating equations for the parameter estimators are similar to those of existing weighted Cox models, but the variance estimators have a different form. We prove the consistency and derive the asymptotic distribution of the proposed regression parameters and cumulative baseline hazard estimator. Through simulation, the asymptotic results were found to be applicable to finite samples. Finally, the proposed method was applied to a national kidney transplant data set to evaluate the effect of expanded criteria donor organs on graft survival.

The proposed method depends on the availability of additional information of the representative sample from which the biased sample is selected. If one can obtain only the information on a biased sample, the probability of being selected into the biased sample cannot be estimated and weights are unavailable. Survival information for the unselected subjects in the representative sample is missing, such that a Cox model based on the representative sample is inapplicable.

In our analysis of national kidney transplant data, the hazard ratio for ECD (versus non-ECD) was estimated at 2.20 based on the weighted Cox model and 1.72 based on the unweighted model. Thus, among the deceased-donor organs currently transplanted, ECD kidneys are associated with a 72% increase in graft failure hazard relative to non-ECDs. The 72% increase is quite consistent with the results of Port et al. (2002), from which the ECD definition was derived. The results of Ojo et al. (2001) indicate that the mortality hazard is lower for patients with an ECD trans-

plant than those on the waiting list. Moreover, Schaubel, Wolfe & Port (2006) report that, on average, patients have lower mortality if they accept an ECD compared to "standard therapy" (foregoing an ECD with the possibility of subsequently receiving a non-ECD transplant). In the interests of patients, current efforts to further expand the kidney donor pool by discarding fewer ECDs must be tempered by the fact that the results of Ojo et al. (2001) and Schaubel, Wolfe & Port (2006) are based not on ECD kidneys, but on ECDs currently transplanted. In particular, the results of Schaubel, Wolfe & Port (2006) demonstrate a mortality reduction (ECD versus standard therapy) which is quite modest and hence could easily be eliminated by utilizing ECD kidneys with a >2-fold increase in graft failure risk. The difference between our weighted (reflecting ECDs generally; HR=2.20) and unweighted (reflecting ECDs selected for transplantation; HR=1.72) analyses underscore the value of clinician judgement.

The proposed method should have broad applicability in biomedical research, particularly observational studies. For example, in liver transplant research, there is interest in a recently developed Donor Risk Index (DRI) on the survival of recipients (Feng et al. 2006). In this case, livers with high DRI values have higher probability of being discarded, which may result in under-estimating the importance of DRI on post-liver transplant outcomes.

Investigators may raise the question of whether using the weighted version of the Cox model is required for their particular application. In certain cases, this issue could be addressed indirectly without fitting the weighted model. That is, if there are no covariates (not already included in the Cox model) which predict selection probability, then the unweighted model should generate approximately the same parameter estimates as the weighted model. However, as our simulation results indi-

cate, weighting can improve efficiency (using the proposed variance estimator) even when the selection probability model is mis-specified. Formal tests of the equality of the weighted and unweighted parameters could prove very useful.

2.7   Appendix

Proof of Theorem 1. The weighted log partial likelihood is given by: $\ell_N^w(\beta) = \sum_{i=1}^N \int_0^\tau [\beta' Z_i(t) - \log\{S_w^{(0)}(\beta, \widehat{\theta}; t)\}] w_i(\widehat{\theta}) dN_i(t)$. Setting $D_N^w(\beta) = N^{-1}\{\ell_N^w(\widehat{\beta}_w) - \ell_N^w(\beta_0)\}$, we write $D_N^w(\beta) = \sum_{k=1}^4 D_{k:N}^w(\beta)$, where

$$D_{1:N}^w(\beta) = N^{-1} \sum_{i=1}^N \int_0^\tau \{w_i(\widehat{\theta}) - w_i(\theta_0)\}(\beta - \beta_0)' Z_i(t) dN_i(t)$$

$$D_{2:N}^w(\beta) = N^{-1} \sum_{i=1}^N \int_0^\tau w_i(\theta_0)(\beta - \beta_0)' Z_i(t) dN_i(t)$$

$$D_{3:N}^w(\beta) = N^{-1} \sum_{i=1}^N \int_0^\tau \{w_i(\widehat{\theta}) - w_i(\theta_0)\} \log\left\{\frac{S_w^{(0)}(\beta_0, \widehat{\theta}; t)}{S_w^{(0)}(\beta, \widehat{\theta}; t)}\right\} dN_i(t)$$

$$D_{4:N}^w(\beta) = N^{-1} \sum_{i=1}^N \int_0^\tau w_i(\theta_0) \log\left\{\frac{S_w^{(0)}(\beta_0, \widehat{\theta}; t)}{S_w^{(0)}(\beta, \widehat{\theta}; t)}\right\} dN_i(t).$$

By assumption (d), X_i and Z_i are bounded almost surely; the increment $dN_i(t)$ is either zero or one. We restrict attention to β in a compact set, $B_r = \{\beta :\| \beta - \beta_0 \|\leq r\}$, with boundary $\partial B_r = \{\beta :\| \beta - \beta_0 \|= r\}$. Since $\widehat{\theta}$ is the Maximum Likelihood Estimator (MLE) of θ_0, $\widehat{\theta} \xrightarrow{a.s.} \theta_0$ as $N \longrightarrow \infty$. Therefore, after applying a Taylor Series expansion and the Strong Law of Large Numbers (SLLN), $D_{1:N}^w(\beta) = N^{-1} \sum_{i=1}^N \int_0^\tau -I_i e^{-\theta_0' X_i}(\beta - \beta_0)' Z_i(t) dN_i(t) X_i'(\widehat{\theta} - \theta_0) + o_p(1)$ and converges almost surely to zero as $N \longrightarrow \infty$.

Under the assumed conditions,

$$(2.5) \qquad \sup_{t\in[0,\tau]} ||S_w^{(d)}(\beta, \theta; t) - s_w^{(d)}(\beta, \theta; t)|| \xrightarrow{a.s.} 0,$$

for $d = 0, 1,$ or 2 and any β in a compact set. Using the strong consistency of $\widehat{\theta}$ for θ_0, $S_w^{(d)}(\beta, \widehat{\theta}; t) \xrightarrow{a.s.} s_w^{(d)}(\beta, \theta_0; t)$ by the continuous mapping theorem, with $s_w^{(0)}(\beta, \theta; t)$

bounded away from zero for all β and θ. Applying the continuous mapping theorem, followed by arguments similar to those used for $D^w_{1:N}(\beta)$, $D^w_{3:N}(\beta) \xrightarrow{a.s.} 0$ as $N \longrightarrow \infty$.

Regarding $D^w_{2:N}(\beta)$, $\int_0^\tau (\beta - \beta_0)' Z_i(t) dN_i(t)$ is bounded. By condition (f), $p_i(\theta_0) > 0$ and hence $w_i(\theta_0)$ is also bounded. Using the SLLN, $D^w_{2:N}(\beta) \xrightarrow{a.s.} E\left\{\int_0^\tau w_i(\theta_0)(\beta - \beta_0)' Z_i(t) dN_i(t)\right\}$ $\equiv D^w_2(\beta)$. Similarly, $D^w_{4:N}(\beta) \xrightarrow{a.s.} E\left[\int_0^\tau w_i(\theta_0) \log\{s^{(0)}_w(\beta_0, \theta_0; t)/s^{(0)}_w(\beta, \theta_0; t)\} dN_i(t)\right] \equiv$ $D^w_4(\beta)$. Combining results given above regarding $D^w_{k:N}(\beta)$ ($k = 1, 2, 3, 4$), $D^w_N(\beta) \xrightarrow{a.s.}$ $D^w_2(\beta) + D^w_4(\beta) \equiv D^w(\beta)$, where

$$D^w(\beta) = E\left[\int_0^\tau w_i(\theta_0) \left\{(\beta - \beta_0)' Z_i(t) + \log\left(\frac{s^{(0)}_w(\beta_0, \theta_0; t)}{s^{(0)}_w(\beta, \theta_0; t)}\right)\right\} dN_i(t)\right].$$

Since β falls in a compact set, B_r, with B_r specifically set such that it contains β_0, then $D^w_N(\beta) \xrightarrow{a.s.} D^w(\beta)$ uniformly on $\beta \in B_r$. In addition, $\partial D^w(\beta)/\partial\beta \mid_{\beta=\beta_0} = 0_{p\times 1}$ and $\partial^2 D^w(\beta)/\partial\beta\partial\beta' = -A^w(\beta, \theta)$ is negative definite by condition (e). Therefore, $D^w(\beta)$ has a unique maximizer and is maximized at $\beta = \beta_0$. Since $\sup_{\beta \in B_r} |D^w_N(\beta) - D^w(\beta)| \xrightarrow{a.s.} 0$, when N is large enough, $D^w_N(\beta)$ will also have a unique maximizer. Then, given that $\partial D^w_N(\beta)/\partial\beta|_{\beta=\widehat{\beta}_w} = 0_{p\times 1}$, $\widehat{\beta}_w$ is the unique maximizer of $D^w_N(\beta)$. Setting r arbitrarily small, when $N \longrightarrow \infty$, $\widehat{\beta}_w \xrightarrow{a.s.} \beta_0$.

Proof of Theorem 2. Using a Taylor expansion around $\beta = \beta_0$, $N^{\frac{1}{2}}(\widehat{\beta}_w - \beta_0) =$ $A^w_N(\beta_*, \widehat{\theta})^{-1} N^{-\frac{1}{2}} U^w_N(\beta_0, \widehat{\theta})$ where β_* lies between $\widehat{\beta}$ and β_0 in $\mathcal{R}^p$. Using (2.5), the fact that $\widehat{\theta} \xrightarrow{a.s.} \theta_0$ and $\widehat{\beta}_w \xrightarrow{a.s.} \beta_0$, the continuous mapping theorem and the SLLN,

(2.6) $$\widehat{A}^w(\widehat{\beta}_w, \widehat{\theta}) \xrightarrow{a.s.} A^w(\beta_0, \theta_0).$$

Hence $N^{\frac{1}{2}}(\widehat{\beta}_w - \beta_0) = A^w_N(\beta_0, \theta_0)^{-1} N^{-\frac{1}{2}} U^w_N(\beta_0, \widehat{\theta}) + o_p(1)$. Through the definition of $S^{(d)}_w(\beta, \theta; t)$ and with some basic algebra,

$$U^w_N(\beta, \theta) = \sum_{i=1}^N \int_0^\tau w_i(\theta)\{Z_i(t) - \overline{Z}_w(\beta, \theta; t)\} dM_i(\beta, \theta; t).$$

We set $U_N^w(\beta, \theta) = U_{1:N}^w(\beta, \theta) + U_{2:N}^w(\beta, \theta)$, where

$$U_{1:N}^w(\beta, \widehat{\theta}) = \sum_{i=1}^N \int_0^\tau w_i(\theta_0)\{Z_i(t) - \overline{Z}_w(\beta, \widehat{\theta}; t)\}dM_i(\beta, \widehat{\theta}; t)$$

$$U_{2:N}^w(\beta, \widehat{\theta}) = \sum_{i=1}^N \int_0^\tau \left\{w_i(\widehat{\theta}) - w_i(\theta_0)\right\}\{Z_i(t) - \overline{Z}_w(\beta, \widehat{\theta}; t)\}dM_i(\beta, \widehat{\theta}; t).$$

By the continuous mapping theorem and the strong convergence of $\widehat{\theta}$ to θ_0,

$$N^{-\frac{1}{2}}U_{1:N}^w(\beta_0, \theta_0) = N^{-\frac{1}{2}}\sum_{i=1}^N \int_0^\tau w_i(\theta_0)\{Z_i(t) - \overline{z}_w(\beta_0, \theta_0; t)\}dM_i(\beta_0, \theta_0; t) + o_p(1),$$

using the fact that

$$\left\| N^{-\frac{1}{2}}\sum_{i=1}^N \int_0^\tau w_i(\theta_0)\{\overline{Z}_w(\beta_0, \theta_0; t) - \overline{z}_w(\beta_0, \theta_0; t)\}dM_i(\beta_0, \theta_0; t)\right\| \xrightarrow{P} 0_{p\times 1},$$

which can be demonstrated by employing various empirical process results. With respect to $U_{2:N}^w(\beta_0)$, apply the delta method, use (2.5), followed by a Taylor expansion,

$$N^{\frac{1}{2}}(\widehat{\theta} - \theta_0) = B(\theta_0)^{-1}N^{-\frac{1}{2}}U_L(\theta_0) + o_p(1),$$

Since, by the SLLN, $\widehat{B}(\theta) \xrightarrow{a.s.} B(\theta)$. Using arguments similar to those in the examination of $N^{-\frac{1}{2}}U_{1:N}^w(\beta_0)$, replacing $\widehat{\theta}$ with θ_0 and $\overline{Z}_w(\beta_0, \widehat{\theta}; t)$ with $\overline{z}_w(\beta_0, \theta_0; t)$, then switching the order of summation,

$$N^{-\frac{1}{2}}U_{2:N}^w(\beta_0) = N^{-\frac{1}{2}}\widehat{G}(\beta_0, \theta_0)B(\theta_0)^{-1}\sum_{i=1}^N U_{Li}(\theta_0) + o_p(1).$$

Again combining the SLLN and continuous mapping theorem, $\widehat{G}(\beta_0, \theta_0) \xrightarrow{a.s.} G(\beta_0, \theta_0)$. Combining results for $U_{1:N}^w(\beta_0)$ and $U_{2:N}^w(\beta_0)$,

$$N^{-\frac{1}{2}}U_N^w(\beta_0, \widehat{\theta}) = N^{-\frac{1}{2}}\sum_{i=1}^N \psi_i^w(\beta_0, \theta_0) + o_p(1).$$

Essentially, $U_N^w(\beta_0, \widehat{\theta})$ behaves asymptotically like the sum of independently and identically distributed mean-zero random vectors, and by Multivariate Central Limit

Theorem (van der Vaart, 2000), $N^{-\frac{1}{2}}U_N^w(\beta_0, \widehat{\theta}) \xrightarrow{D} N(0_{p\times 1}, \Sigma_w(\beta_0, \theta_0))$. Finally, we combine the asymptotic distribution of $U_N^w(\beta_0, \widehat{\theta})$ and (2.6), through Slutsky's Theorem (Sen and Singer, 1993), to complete the proof of Theorem 2.

Proof of Theorem 3. The proof of uniform consistency begins by decomposing $\alpha_N(t) = \widehat{\Lambda}_0^w(\widehat{\beta}_w, \widehat{\theta}; t) - \Lambda_0(t)$ into three parts, $\alpha_N(t) = \alpha_{1:N}(t) + \alpha_{2:N}(t) + \alpha_{3:N}(t)$ with

$$
\begin{aligned}
\alpha_{1:N}(t) &= \widehat{\Lambda}_0^w(\widehat{\beta}_w, \widehat{\theta}; t) - \widehat{\Lambda}_0^w(\widehat{\beta}_w, \theta_0; t) \\
\alpha_{2:N}(t) &= \widehat{\Lambda}_0^w(\widehat{\beta}_w, \theta_0; t) - \widehat{\Lambda}_0^w(\beta_0, \theta_0; t) \\
\alpha_{3:N}(t) &= \widehat{\Lambda}_0^w(\beta_0, \theta_0; t) - \Lambda_0(t).
\end{aligned}
$$

Applying a Taylor expansion about θ_0,

$$
\alpha_{1:N}(t) = N^{-1}\sum_{i=1}^{N}\int_0^t \left\{ \frac{-X_i' e^{-\theta_*' X_i}}{S_w^{(0)}(\widehat{\beta}_w, \theta_*; s)} + \frac{w_i(\theta_*)\widehat{b}'(\widehat{\beta}_w, \theta_*; s)}{S_w^{(0)}(\widehat{\beta}_w, \theta_*; s)^2} \right\} dN_i(s)(\widehat{\theta} - \theta_0),
$$

where θ_* lies between $\widehat{\theta}$ and θ_0 in $\mathcal{R}^q$ and $\widehat{b}(\beta, \theta; t) = N^{-1}\sum_{i=1}^{N} X_i e^{-\theta' X_i} Y_i(t) e^{\beta' Z_i(t)}$. Under assumptions (a) to (f), X_i, $S_w^{(0)}(\beta, \theta; s)$, $w_i(\theta)$ and $dN_i(s)$ are all bounded and $S_w^{(0)}(\beta, \theta; s)$ is bounded away from 0. Using these results, along with the strong convergence of $\widehat{\theta}$ to θ_0, $\mid \alpha_{1:N}(t) \mid \xrightarrow{a.s.} 0$. Similarly, through a Taylor Series expansion about β_0,

$$
\alpha_{2:N}(t) = -\int_0^t \overline{Z}_w'(\beta_*, \theta_0; s) d\Lambda_0^w(\beta_*, \theta_0; s)(\widehat{\beta} - \beta_0),
$$

where β_* lies between $\widehat{\beta}$ and β_0 in $\mathcal{R}^p$. Since the quantities $\overline{Z}_w(\beta_*, \theta_0; s)$ and $d\Lambda_0^w(\beta_*, \theta_0; s)$ are bounded, and since $\widehat{\beta}_w \xrightarrow{a.s.} \beta_0$, applying the SLLN to $\alpha_{2:N}(t)$, it follows that $\mid \alpha_{2:N}(t) \mid \xrightarrow{a.s.} 0$.

The last component, $\alpha_{3:N}(t)$, can be rewritten as $N^{-1}\sum_{i=1}^{N}\int_0^t S_w^{(0)}(\beta_0, \theta_0; s)^{-1} w_i(\theta_0) dM_i(\beta_0, \theta_0; s)$. By the Uniform Strong Law of Large

Numbers (USLLN; Pollard, 1990), $N^{-1}\sum_{i=1}^{N}\int_0^t dM_i^w(\beta_0,\theta_0;s) \xrightarrow{a.s.} 0$ for $t \in [0,\tau]$.
As $N \longrightarrow \infty$, $S_w^{(0)}(\beta_0,\theta_0;s) \longrightarrow s_w^{(0)}(\beta_0,\theta_0;s)$ which is bounded away from 0. There-
fore, $\mid \alpha_{3:N}(t) \mid \xrightarrow{a.s.} 0$. Combining results for $\alpha_{1:N}(t)$, $\alpha_{2:N}(t)$ and $\alpha_{3:N}(t)$ and the
triangle inequality,

$$\sup_{t\in[0,\tau]} \mid \widehat{\Lambda}_0^w(\widehat{\beta}_w,\widehat{\theta};t) - \Lambda_0(t) \mid \xrightarrow{a.s.} 0.$$

Proof of Theorem 4. Using the SLLN and the almost sure convergence of $\widehat{b}(\beta,\theta;s)$
to $b(\beta,\theta;s)$,

$$N^{-1}\sum_{i=1}^{N}\int_0^t \left\{ \frac{-X_i'e^{-\theta'X_i}dN_i(s)}{S_w^{(0)}(\beta,\theta;s)} + \frac{w_i(\theta)dN_i(s)\widehat{b}(\beta,\theta;s)}{S_w^{(0)}(\beta,\theta;s)^2} \right\} \xrightarrow{a.s.} k(\beta,\theta;s),$$

where $k(\beta,\theta;s)$ is as defined in Theorem 4. Then,

$$\begin{aligned}
N^{\frac{1}{2}}\alpha_{1:N}(t) &= k'(\widehat{\beta}_w,\widehat{\theta};t)N^{\frac{1}{2}}(\widehat{\theta}-\theta_0) + o_p(1) \\
&= N^{-\frac{1}{2}}\sum_{i=1}^{N}\left\{ k'(\beta_0,\theta_0;t)B(\theta_0)^{-1}U_{Li}(\theta_0)\right\} + o_p(1).
\end{aligned}$$

Using arguments similiar to those in Theorem 3 for $\alpha_{2:N}(t)$, $-\int_0^t \overline{Z}_w(\beta,\theta;s)d\widehat{\Lambda}_0^w(\beta,\theta;s) \xrightarrow{a.s.}$
$h(\beta,\theta;t)$ as $N \longrightarrow \infty$. Therefore, it can be shown that

$$\begin{aligned}
N^{\frac{1}{2}}\alpha_{2:N}(t) &= h'(\beta_*,\theta_0;t)N^{\frac{1}{2}}(\widehat{\theta}-\theta_0) + o_p(1) \\
&= h'(\beta_0,\theta_0;t)B(\theta_0)^{-1}N^{-\frac{1}{2}}\sum_{i=1}^{N}\psi_i^w(\theta_0) + o_p(1).
\end{aligned}$$

Regarding $\alpha_{3:N}(t)$,

$$\begin{aligned}
N^{\frac{1}{2}}\alpha_{3:N}(t) &= N^{-\frac{1}{2}}\sum_{i=1}^{N}\int_0^t \frac{w_i(\theta_0)dM_i(\beta_0,\theta_0;s)}{S_w^{(0)}(\beta_0,\theta_0;s)} \\
&= N^{-\frac{1}{2}}\sum_{i=1}^{N}\int_0^t s_w^{(0)}(\beta_0,\theta_0;s)^{-1}w_i(\theta_0)dM_i(\beta_0,\theta_0;s) + o_p(1).
\end{aligned}$$

where the last equality can be shown using the strong convergence of $S_w^{(0)}(\beta_0,\theta_0;s)$
to $s_w^{(0)}(\beta_0,\theta_0;s)$, the continuous mapping theorem and the USLLN.

Combining the above results, $N^{\frac{1}{2}}\alpha_N(t)$ is the sum of independent identically distributed mean-zero random variables. When N is large, by Multivariate CLT, the finite-dimensional distributions of $N^{\frac{1}{2}}\alpha_N(t)$ converge to those of a zero-mean multivariate normal with covariance $\Omega(s,t) = E\{\phi_i(\beta_0, \theta_0; s)\phi_i(\beta_0, \theta_0; t)\}$. Through the monotonicity of the various components of $\psi_i(\beta_0, \theta_0; t)$, it can be demonstrated that the $\psi_i(\beta_0, \theta_0; t)$ are manageable (Pollard, 1990) and hence that $N^{\frac{1}{2}}\alpha_N(t)$ is tight. Applying the Functional CLT (Pollard, 1990), $N^{\frac{1}{2}}\alpha_N(t)$ converges weakly to a zero-mean Gaussian process with covariance function $\Omega(s,t)$.

2.8   Acknowledgements

2.9   References

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

Y. Bilias, M. Gu & Z. Ying (1997). Towards a general asymptotic theory for the Cox model with staggered entry. *The Annals of Statistics*, 25, 662–682.

D. Binder (1992). Fitting Cox's proportional hazards models from survey data. *Biometrika*, 79, 139–147.

C. Boudreau & J. F. Lawless (2006). Survey analysis based on the proportional hazards model and survey data. *The Canadian Journal of Statistics*, 34, 203–216.

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

D. Cox (1975). Partial likelihood. *Biometrika*, 62, 262–276.

S. Feng, N. P. Goodrich, J. L. Bragg-Gresham, D. M. Dykstra, J. D. Punch, M. A. DebRoy, S. M. Greenstein & R. M. Merion (2006). Characteristics associated with liver graft failure: the concept of a donor risk index. *American Journal of Transplantation*, 6, 783–790.

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

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

M. A. Hernan, B. Brumback & J. M. Robins (2001). Marginal structural models to estimate the joint effect of non randomized treatments. *Journal of the American Statistical Association*, 96, 440–448.

D. G. Horvitz & D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the Royal Statistical Society Series B*, 47, 663–685.

J. D. Kalbfleisch & R. L. Prentice (2002). *The Statistical Analysis of Failure Time Data*. Wiley, Hoboken, New Jersey.

D. Y. Lin (2000). On fitting Cox's proportional hazards models to survey data. *Biometrika*, 87, 37–47.

A. O. Ojo, J. A. Hanson, H. Meier-Kriesche, C. N. Okechukwu, R. A. Wolfe, A. B. Leichtman, L. Y. Agodoa, B. Kaplan & F. K. Port (2001). Survival in recipients of marginal cadaveric donor kidneys compared with other recipients and wait-listed transplant candidates. *Journal of the American Society of Nephrology*, 12, 589–597.

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

F. K. Port, J. L. Bragg-Gresham, R. A. Metzger, D. M. Dykstra, B. W. Gillespie, E. W. Young, F. L. Delmonico, J. J. Wynn, R. M. Merion, R. A. Wolfe & P. J. Held (2002). Donor characteristics associated with reduced graft survival: An approach to expanding the pool of donor kidneys. *Transplantation*, 74, 1281–1286.

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

J. M. Robins & S. Greenland (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.

J. M. Robins, S. Greenland & F. Hu (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome (with discussion). *Journal of the American Statistical Association*, 94, 687–712.

P. R. Rosenbaum (2002). *Observational Studies*, Second Editioin. Springer-Verlag, New York.

P. R. Rosenbaum & D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

D. E. Schaubel & J. Cai (2004). Regression methods for gap time hazards of sequentially ordered multivariate failure time data. *Biometrika*, 91, 291–303.

D. E. Schaubel, M. Desmeules, Y. Mao, J. Jeffery & S. Fenton (1995). Survival experience among end-stage renal disease patients: A controlled comparison of transplantation and dialysis. *Transplantation*, 60, 1389–1394.

D. E. Schaubel, R. A. Wolfe & F. K. Port (2006). A sequential stratification method for estimating the effect of an time-dependent experimental treatment in observational studies. *Biometrics*, 62, 910–917.

P. K. Sen & J. M. Singer (1993). *Large Sample Methods in Statistics*. Chapman and Hall, New York.

C. A. Struthers & J. D. Kalbfleisch (1986). Misspecified proportional hazard models. *Biometrika*, 73, 363–369.

A. W. van der Vaart (2000). *Asymptotic Statistics*. Cambridge, Melbourne.

A. W. van der Vaart & J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

M. C. Wang (1999). Gap time bias in incident and prevalent cohorts. *Statistica Sinica*, 9, 999–1010.

R. A. Wolfe, V. B. Ashby, E. L. Milford, A. O. Ojo, R. E. Ettenger, L. Y. C. Agodoa, P. J. Held & F. K. Port (1999). Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *New England Journal of Medicine*, 341, 1725–1730.

CHAPTER III

Evaluating Bias Correction in Weighted Proportional Hazards Regression

Abstract: Often in observational studies of time to an event, the study population is a biased (i.e., unrepresentative) sample of the target population. In the presence of biased samples, it is common to weight subjects by the inverse of their respective selection probabilities. Pan and Schaubel (2007) recently proposed inference procedures for an inverse selection probability weighted (ISPW) Cox model, applicable when selection probabilities are not treated as fixed but estimated empirically. The proposed weighted regression parameter estimator is consistent for the target population parameter, while the unweighted estimator converges to a modification of the true value; the modification resulting from the potentially biased sampling mechanism. Similar statements apply to the weighted and unweighted cumulative hazard estimators. Although parameter estimation is consistent, computation is more intensive than that for an unweighted model. In this article, we propose methods for evaluating bias in the unweighted partial likelihood and Breslow-Aalen estimators. Asymptotic properties of the proposed test statistics are derived. The finite-sample significance level and power are evaluated through simulation. The proposed methods are then applied to data from a national organ failure registry to evaluate the bias in a post kidney transplant survival model.

Key words and phrases: Confidence bands; Inverse-selection-probability weights; Observational studies; Proportional hazards model; Selection bias; Wald test.

3.1   Introduction

In observational studies of time to an event, we often observe partial information. That is, some information is unavailable for all subjects sampled from the target population. There are several possibilities for the estimation of covariate effects in this setting. One method is a complete-case analysis, which assumes that the observed subjects are a representative sample of the target population; "representative" in the sense that the relationship of interest is not systematically distorted in the sample compared to that in the target population. In survival analysis, if there exists some auxiliary factor which is correlated with the covariate of interest and affects the hazard function and selection probability, the coefficient estimates for the covariate of interest (i.e., estimated without considering the auxiliary factor) will be systematically different in the selected sample and target population. Since our goal is to estimate covariate effects in the target population, estimates obtained based on only the selected sample are biased. We hereafter refer to a factor inducing such selection bias as a biasing factor.

We consider the data structure where a potentially biased sample is selected from a representative sample of the target population. In the presence of a biased sample, an alternative to a complete-case analysis is to weight each subject by the inverse of their probability of being selected into the sample, in order to reflect the composition of the target population. The inverse-selection-probability-weighting (ISPW) method was originally proposed by Horvitz and Thompson (1952). Binder (1992), Lin (2000) and Boudreau and Lawless (2006) extended the Horvitz-Thompson method to the

hazard regression setting in the context of large-scale surveys or designed experiments, where selection probability for each subject is set beforehand and thus can be treated as known. However, in many practical applications, the probability of being selected is not known. If appropriate auxiliary data are available (in particular, pertaining to the biasing factor) for the representative sample, selection probability may be estimated. Pan and Schaubel (2007) proposed a weighted proportional hazards model with empirical ISPW. Specifically, selection probabilities are estimated through a logistic model fitted to the representative sample of the underlying target population, assuming the availability of auxiliary data representing the biasing factor. Parameter estimators are consistent for their corresponding underlying target population quantities. In addition, since the estimated weights are, essentially, trained by the data, efficiency is gained relative to estimators obtained by treating the weights as if they were known. The elimination of bias and increase in precision come at the price of more complex modeling, longer computing time and the inability to use commercially available software packages directly. The methods proposed in Pan and Schaubel (2007) are necessary if and only if the potential biasing factors are (i) correlated with the covariate of interest (ii) independently affect the hazard function (iii) have conditional distributions (given the model covariates) which are different in the selected sample and target population. In the absence of biased selection, practitioners would prefer the simpler unweighted method. Whether or not one employs the weighted method depends on whether the criteria for biased selection are satisfied and the degree of distortion in the relationships of interest in the biased sample relative to that in the target population.

No method is currently available to identify cases where the IPSW method is necessary and, hence, when its advantages outrun the price of computing. In this ar-

ticle we propose statistics calculated as the scaled difference between the unweighted and empirically weighted estimates to measure the degree of bias. A Wald-type test is proposed for the regression parameter, while point-wise and process-based tests are developed for the cumulative baseline hazard. The unweighted estimators describe the relationship under study in the possibly biased sample, while the ISPW counterpart reflects the corresponding relationship applicable to the target population. Using the proposed tests, practitioners can make informed choices between the weighted and unweighted methods using formal procedures.

The proposed tests are motivated by kidney transplant data. At the time of this report, work is well underway to restructure the kidney allocation system. Currently, deceased-donor organs are allocated primarily by waiting time. Under a newly proposed allocation system, organs would be allocated based on the difference between predicted post-transplant and wait-list survival. Hence, there is great need for an accurate post-transplant survival model. Note, however, that the model will be applied to wait-listed patients, not transplanted patients. That is, the model will be used to predict the post-transplant lifetime of a wait-listed patient with a given covariate pattern; as opposed to patients with the same covariate pattern who were already selected (possibly through a biased selection mechanism) for transplantation. Therefore, the target population includes patients on the wait-list, not patients already selected off the wait-list to receive a transplant.

Naturally, not all wait-listed patients will receive a transplant. If we desire to develop a model to apply to wait-listed candidates, post-transplant information is missing for candidates who were not transplanted. The goal is to estimate the effects of patient characteristics on the post-transplant hazard. Patients on the kidney waiting list are not randomly selected for transplantation. Although patients on dialysis

who are not considered suitable candidates for transplantation are simply not placed on the wait list, the screening process does not end with the decision of whether or not to wait-list a patient. Certain patients are systematically bypassed on the list since they are felt to be inferior candidates for kidney transplantation, often due to the progression of concomitant illnesses which occur or further develop after the time of wait-listing. Therefore, patients transplanted with a specific covariate pattern are generally thought not to be representative of patients on the waiting list with the same covariate pattern. Hospitalization data represent a rich source of auxiliary data which could potentially account for the residual difference between patients selected for transplantation and those left on the waiting list. At the Scientific Registry of Transplant Recipients (SRTR), it is possible to merge the wait-list/transplant data obtained from the United Network for Organ Sharing (UNOS) and the hospitalization data obtained from the Centers for Medicare and Medicaid Services (CMS). Therefore, we are able to estimate selection probabilities, treating the hospitalization information as the potential biasing factor. It is important to note that UNOS will not have real-time access to hospitalization histories, meaning that a post-transplant survival model which used hospitalization history as covariates would be of no value. Our objective is to evaluate the degree of bias in a model which is intended to apply to wait-listed patients but is fitted only to transplanted patients and has no adjustment for the potentially biased selection. We can fit the estimated ISPW proportional hazards model of Pan and Schaubel (2007), using hospitalization history to predict transplant probabilities. The issue of whether or not the weighted model is necessary will be addressed by our proposed tests.

The remainder of this report is organized as follows. In Section 2, we propose test statistics for the regression parameter and cumulative baseline hazard and describe

their asymptotic properties (proved in the Web Appendix). In Section 3, the empirical significance level and power of the proposed test for the regression parameter are evaluated by simulation studies. In Section 4, each of the proposed procedures are applied to the kidney transplant data from a national organ failure registry. The proposed methods are further discussed in Section 5.

3.2 Proposed Methods and Asymptotic Properties

3.2.1 Set-up and Notation

We start by establishing the necessary notation. First, there are N subjects in the representative sample from the target population. We let I_i (0, 1) be a sampling indicator; i.e., $I_i = 1$ if the ith subject is selected from the representative sample into the possibly biased study sample. In total there are $n = \sum_{i=1}^{N} I_i$ subjects selected. For subject i with event time T_i and censoring time C_i, we define $\widetilde{T}_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$, $Y_i(t) = I(\widetilde{T}_i \geq t)$, $N_i(t) = I(\tilde{T}_i \leq t, \Delta_i = 1) = \int_0^t dN_i(s)$; with $dN_i(t) = N_i(t+dt) - N_i(t)$, where τ is a prespecified constant satisfying $P(\widetilde{T}_i > \tau) > 0$ and usually set to the end of the study to include all event times. The proportional hazards model applicable to the target population is as follows,

$$(3.1) \qquad \lambda_{iT}(t) = \lambda_{0T}(t)e^{\beta'_T Z_i(t)},$$

where $\lambda_{0T}(t)$ is an unspecified baseline hazard and $Z_i(t)$ is a $p \times 1$ covariate vector. The proportional hazards model which applies to the selected sample is given by

$$(3.2) \qquad \lambda_{iS}(t) = \lambda_{0S}(t)e^{\beta'_S Z_i(t)},$$

where $\lambda_{0S}(t)$ and β_S are the possibly biased version of the parameters of interest. Note that proportionality is assumed for both model (4.1) and model (3.2), but the constants of proportionality are potentially different, as too are the baseline hazards.

The hypothesis tests of interest in this report are (i) H_0: $\beta_T = \beta_S$ vs H_1: $\beta_T \neq \beta_S$ and (ii) H_0: $\lambda_{0T}(t) = \lambda_{0S}(t)$ vs H_1: $\lambda_{0T}(t) \neq \lambda_{0S}(t)$ for $t \in [0, \tau]$.

The selection probabilities are estimated through a logistic model:

$$p_i(\theta_0) = \frac{e^{\theta_0' X_i}}{1 + e^{\theta_0' X_i}},$$

where θ_0 is the true parameter vector and X_i is the corresponding $q \times 1$ vector of predictors. The weight is then given by $w_i(\theta_0) = I_i \, p_i(\theta_0)^{-1}$. Ideally, X_i is the set of all biasing factors; that is, all factors that are predictive of selection probability and survival time and correlate with the covariates of interests. If unmeasured biasing factors exist, consistent estimators of θ_0 and hence β_T are unobtainable.

3.2.2 Estimation: β_S and β_T

In addition to the regularity conditions listed in the Web Appendix A, the ISPW method also assumes that $Z_i \cup X_i$ consists of all factors affecting $\lambda_{iT}(t)$; that is, the "no-unmeasured-confounders" assumption. In addition, we assume that C_i is conditionally independent of T_i given $Z_i(t)$ and X_i; specifically,

$$\lim_{\delta \to 0} \frac{1}{\delta} P(t \leq T_i < t + \delta | T_i \geq t, C_i \geq t, Z_i(t), X_i) = \lim_{\delta \to 0} \frac{1}{\delta} P(t \leq T_i < t + \delta | T_i \geq t, Z_i(t), X_i).$$

The regression parameter estimator for model (3.2), denoted by $\widehat{\beta}_S$, is the solution to the partial likelihood (Cox 1975) score equation $U(\beta) = 0$, where

$$(3.3) \qquad U(\beta) = \sum_{i=1}^{N} \int_0^\tau I_i \{ Z_i(t) - \overline{Z}(t; \beta) \} dN_i(t).$$

Andersen and Gill (1982) proved that $n^{1/2}(\widehat{\beta}_S - \beta_S)$ converges asymptotically to a zero-mean Gaussian process with covariance consistently estimated by $\widehat{A}(\widehat{\beta}_S)^{-1}$. In addition, the cumulative baseline hazard, $\Lambda_{0S}(t) = \int_0^t \lambda_{0S}(u) du$, can be consistently estimated by the Breslow-Aalen estimator,

$$\widehat{\Lambda}_{0S}(t; \widehat{\beta}_S) = n^{-1} \sum_{i=1}^{N} \int_0^t I_i S^{(0)}(s; \widehat{\beta}_S)^{-1} dN_i(s).$$

The regression parameter estimator for the ISPW proportional hazards model is the root of the weighted score equation, $U_w(\beta, \widehat{\theta}) = 0$, where

$$(3.4) \qquad U_w(\beta, \theta) = \sum_{i=1}^{N} \int_0^\tau w_i(\theta)\{Z_i(t) - \overline{Z}_w(t; \beta, \theta)\} \, dN_i(t).$$

As proved in Pan and Schaubel (2007), the weighted estimator $\widehat{\beta}_T$ is strongly consistent for β_T, while $N^{\frac{1}{2}}(\widehat{\beta}_T - \beta_T)$ follows an asymptotic zero-mean normal distribution with a covariance matrix that can be consistently estimated by $\widehat{A}^w(\widehat{\beta}_T, \widehat{\theta})^{-1} \widehat{\Sigma}_w(\widehat{\beta}_T, \widehat{\theta}) \widehat{A}^w(\widehat{\beta}_T, \widehat{\theta})^{-1}$, where

$$\widehat{A}^w(\beta, \theta) = N^{-1} \sum_{i=1}^{N} \int_0^\tau \{Z_i(t) - \overline{Z}_w(t; \beta, \theta)\}^{\otimes 2} w_i(\theta) Y_i(t) e^{\beta' Z_i(t)} d\widehat{\Lambda}_0^w(t; \beta, \theta)$$

$$\widehat{\Sigma}_w(\widehat{\beta}_T, \widehat{\theta}) = N^{-1} \sum_{i=1}^{N} \widehat{\psi}_i^w(\widehat{\beta}_T, \widehat{\theta})^{\otimes 2}$$

$$\widehat{\psi}_i^w(\beta, \theta) = \int_0^\tau w_i(\theta)\{Z_i(t) - \overline{Z}_w(t; \beta, \theta)\} d\widehat{M}_i(t; \beta, \theta) - \widehat{G}(\beta, \theta) \widehat{B}(\theta)^{-1} \widehat{U}_{Li}(\theta)$$

$$d\widehat{M}_i(t; \beta, \theta) = dN_i(t) - Y_i(t) e^{\beta' Z_i} d\widehat{\Lambda}_0^w(t; \beta, \theta)$$

$$\widehat{G}(\beta, \theta) = N^{-1} \sum_{i=1}^{N} \int_0^\tau \{Z_i(t) - \overline{Z}_w(t; \beta, \theta)\} X_i' e^{-\theta' X_i} d\widehat{M}_i(t; \beta)$$

$$\widehat{B}(\theta) = -N^{-1} \frac{\partial U_L(\theta)}{\partial \theta'} = N^{-1} \sum_{i=1}^{N} p_i(\theta)\{1 - p_i(\theta)\} X_i^{\otimes 2},$$

where $a^{\otimes 2} = aa'$.

3.2.3 Test: Regression Parameter

The null hypotheses of interest are that the parameter estimators based on the potentially biased sampling design have the same limiting value as the corresponding values in the target population. These two effects will be different and hence the unweighted estimator will be biased if and only if the following conditions hold. First, there exists some biasing factor correlated with the effect of interest and not included in the proportional hazards model. Second, this factor affects the hazard function.

Third, the biasing factor affects selection probability. Testing the simultaneous existence of above three conditions is generally complicated and may in practice be quite tedious. For example, let $X_i = (X'_{i1}, X'_{i2})'$, where X_{i1} is captured by Z_i in the fitted Cox model and X_{i2} is not. Correspondingly, we partition θ_0 as $(\theta'_{01}, \theta'_{02})'$. To test each component of the biasing mechanism, one would test (i) $\theta_{01} = 0$, (ii) $\beta_2 = 0$ in the model $\lambda_i(t) = \lambda_{0T}(t) e^{\beta'_T Z_i + \beta'_2 X_{i2}}$ and (iii) test the hypothesis of zero pairwise association between each element of Z_i and each element of X_{i2}. It is possible that several of the hypotheses in (i), (ii) or (iii) could be rejected, but that the actual bias in $\widehat{\beta}_S$ is negligible. Conversely, it is also possible that only a small minority of the hypotheses are rejected, but that bias in at least some elements of $\widehat{\beta}_S$ is substantial. The bottom line is that explicitly testing the conditions under which bias in $\widehat{\beta}_S$ can occur is cumbersome and impractical if even a moderate number of covariates are involved. Similar arguments could apply to testing H_0: $\Lambda_{0T}(t) = \Lambda_{0S}(t)$ for $t \in (0, \tau]$, if testing were to proceed through first principles. Moreover, additional steps would be needed to calculate the overall significance level of the three sets of tests.

Here, we propose a single statistic to examine the degree of bias in $\widehat{\beta}_S$ without appealing to first principles. The proposed statistic is based on the quantity,

$$D_j = c'_j(\beta_T - \beta_S),$$

where c_j is a $p \times 1$ vector with $p - 1$ elements equal to 0 and the jth element equal to 1. The quantity D_j reflects the degree of selection bias in the jth element of $\widehat{\beta}_S$. Typically, the null hypothesis would be given by H_0: $D_j = 0$ with alternative H_1: $D_j \neq 0$, although a one-sided alternative hypothesis may be indicated depending on the nature of the specific application. The test statistic is given by $\widehat{D}_j^2 / \widehat{\text{Var}}(\widehat{D}_j)$, or by $\widehat{D}_j / \widehat{SE}(\widehat{D}_j)$ for single-sided H_1. In the remainder of this subsection, we demonstrate that $\widehat{D}_j / \widehat{SE}(\widehat{D}_j)$ follows a standard normal distribution asymptotically. As such,

$\widehat{D}_j^2/\widehat{\mathrm{Var}}(\widehat{D}_j)$ would follow a χ_1^2 distribution as $n \to \infty$. Further, if it was desired to test two or more elements of β_S for bias simultaneously, one could use a modification of the proposed statistic $D = C'(\beta_T - \beta_S)$, where C is a $p \times h$ matrix of constants (typically 0's and 1's) defined to extract the specific contrasts of interest. Each of the h columns for C is a vector for one parameter being tested simultaneously. For example, if we want to test the ith and jth parameter at the same time, $C = c_i \| c_j$, where $\|$ denotes horizontal concatenation. The test statistics could be given by $\widehat{D}'\widehat{\mathrm{Var}}(\widehat{D})^{-1}\widehat{D}$, which follows a χ_h^2 distribution under the null as $n \to \infty$.

In order to derive the large-sample properties of $\widehat{D}$, we assume the regularity conditions listed in the Web Appendix A.

Theorem 1. Under conditions (a) to (f), $\widehat{D}_j$ is a consistent estimator of D; that is, $\widehat{D}_j \xrightarrow{a.s.} D_j$, while $n^{\frac{1}{2}}(\widehat{D}_j - D_j)$ is asymptotically zero-mean normal with covariance matrix

$$\overline{p}c_j'A^w(\beta_T, \theta_0)^{-1}\Sigma_w(\beta_T, \theta_0)A^w(\beta_T, \theta_0)^{-1}c_j + c_j'A(\beta_S)^{-1}\Sigma(\beta_S)A(\beta_S)^{-1}c_j$$
$$- 2\overline{p}c_j'A(\beta_S)^{-1}\Sigma_\beta(\beta_S, \beta_T, \theta_0)A^w(\beta_T, \theta_0)^{-1}c_j, \text{ where } \overline{p}\equiv E(n/N).$$

The consistency of $\widehat{D}_j$ is proved using the consistency of $\widehat{\beta}_S$ (Andersen and Gill 1982) and β_T (Pan and Schaubel 2007) along with the continuous mapping theorem. To derive the covariance matrix of $n^{\frac{1}{2}}(\widehat{D}_j - D_j)$, the difference is decomposed into two parts,

$$(3.5) \qquad n^{\frac{1}{2}}(\widehat{D}_j - D_j) = \overline{p}^{\frac{1}{2}}N^{\frac{1}{2}}c_j'(\widehat{\beta}_T - \beta_T) - n^{\frac{1}{2}}c_j'(\widehat{\beta}_S - \beta_S).$$

Note that, because we assume $p_i(\theta) > 0$ for $i = 1, \ldots, N$, the average sampling probability, $\overline{p}$, converges to a constant between 0 and 1 and n, N go to ∞ at the same rate. The variances of $N^{\frac{1}{2}}c_j'\widehat{\beta}_T$ and $n^{\frac{1}{2}}c_j'\widehat{\beta}_S$ are the same as those of $N^{\frac{1}{2}}c_j'(\widehat{\beta}_T - \beta_T)$ and $n^{\frac{1}{2}}c_j'(\widehat{\beta}_S - \beta_S)$, and were derived by Pan and Schaubel (2007), along with consistent

estimators. The covariance between $N^{\frac{1}{2}}c_j'\widehat{\beta}_T$ and $n^{\frac{1}{2}}c_j'\widehat{\beta}_S$ can be written out in terms of the covariance between $N^{-\frac{1}{2}}c_j'A^w(\beta_T,\theta_0)^{-1}U_w(\beta_T,\theta_0)$ and $n^{-\frac{1}{2}}c_j'A(\beta_S)^{-1}U(\beta_S)$. Specifically,

$$\text{Cov}(N^{\frac{1}{2}}c_j'\widehat{\beta}_T, n^{\frac{1}{2}}c_j'\widehat{\beta}_S) = \bar{p}^{\frac{1}{2}}c_j'A(\beta_S)^{-1}\Sigma_\beta(\beta_S,\beta_T,\theta_0)A^w(\beta_T,\theta_0)^{-1}c_j,$$

where

$$\Sigma_\beta(\beta_1,\beta_2,\theta) = E\{\psi_i(\beta_1)\psi_i^w(\beta_2,\theta)'\}.$$

Here, $A(\beta_S)$, $A^w(\beta_T,\theta_0)$, $\psi_i(\beta_S)$ and $\psi_i^w(\beta_T,\theta_0)$ are defined in the calculation of the covariance matrix of $\widehat{\beta}_S$ and $\widehat{\beta}_T$. They can each be estimated by replacing β_S, β_T, Λ_{0S}, Λ_{0T} and θ_0 with their sample estimates. The quantity $c_j'\Sigma_\beta(\beta_S,\beta_T,\theta_0)c_j$ equals the covariance between $N^{-\frac{1}{2}}c_j'U_w(\beta_T,\theta_0)$ and $n^{-\frac{1}{2}}c_j'U(\beta_S)$. The (j,k) element in the matrix $\Sigma_\beta(\beta_S,\beta_T,\theta_0)$ is the covariance between the jth element of $N^{-\frac{1}{2}}U_w(\beta_T,\theta_0)$ and the kth element of $n^{-\frac{1}{2}}U(\beta_S)$. Both the weighted (3.4) and unweighted score functions (3.3) are zero at the estimated parameter values $\widehat{\beta}_S$, $\widehat{\beta}_T$ and $\widehat{\theta}$. Furthermore, when $N \to \infty$, each can be written as a sum of independent contributions from each subject of the selected sample. As such, an estimator of the covariance between the elements of $N^{-\frac{1}{2}}U_w(\beta_T,\theta_0)$ and the elements of $n^{-\frac{1}{2}}U(\beta_S)$ is the average of the outer product of the subject-specific contributions to the weighted score (3.4) function and unweighted score function (3.3). As such, the matrix $\Sigma_\beta(\beta_S,\beta_T,\theta_0)$ can hence be consistently estimated as

$$\widehat{\Sigma}_\beta(\widehat{\beta}_S,\widehat{\beta}_T,\widehat{\theta}) = n^{-1}\sum_{i=1}^{N} I_i\widehat{\psi}_i(\widehat{\beta}_S)\widehat{\psi}_i^w(\widehat{\beta}_T,\widehat{\theta})'.$$

3.2.4 Estimation: $\Lambda_{0S}(t)$ and $\Lambda_{0T}(t)$

Another quality of interest is the baseline hazard function, which is best viewed as a process over time. We use $\Lambda_{0T}(t)$ to denote the baseline hazard function in the

target population, which is consistently estimated by $\widehat{\Lambda}_0^w(t; \widehat{\beta}_T, \widehat{\theta})$, where

$\widehat{\Lambda}_0^w(t; \beta, \theta) = N^{-1} \sum_{i=1}^N \int_0^t w_i(\theta) S_w^{(0)}(s; \beta, \theta)^{-1} dN_i(s)$. A modified version of $\Lambda_{0T}(t)$ is $\Lambda_{0S}(t)$ with the modification resulting from the potentially biased sampling mechanism.

The unweighted version of the baseline hazard function estimator is $\widehat{\Lambda}_0(t; \widehat{\beta}_S)$ where

$$\widehat{\Lambda}_0(t; \beta) = n^{-1} \sum_{i=1}^N I_i \int_0^t S^{(0)}(s; \beta)^{-1} dN_i(s).$$

In the absence of biasing factors, $\widehat{\Lambda}_0(t) \xrightarrow{a.s.} \Lambda_{0T}(t)$ uniformly in $t \in [0, \tau]$ (Fleming and Harrington, 1991). In the presence of biasing factors, $\widehat{\Lambda}_0(t) \xrightarrow{a.s.} \Lambda_{0S}(t) \neq \Lambda_{0T}(t)$.

3.2.5 Test: Cumulative Baseline Hazard

The quantity $\Delta(t) = \Lambda_0^w(t) - \Lambda_0(t)$ reflects the magnitude of the bias in the unweighted baseline hazard function estimator. We summarize the essential asymptotic properties of an estimator of this quantity, $\widehat{\Delta}(t) = \widehat{\Lambda}_0^w(t) - \widehat{\Lambda}_0(t)$, in the following two theorems.

Theorem 3. Under conditions (a) to (f), $\widehat{\Delta}(t)$ converges uniformly to $\Delta(t)$ as $n \to \infty$ for $t \in [0, \tau]$.

The consistency of $\widehat{\Lambda}_0(t; \widehat{\beta}_S)$ to $\Lambda_{0S}(t)$ for $t \in (0, \tau]$ can be demonstrated by combining the Uniform Strong Law of Large Numbers (USLLN) and the Martingale Central Limit Theorem (Fleming and Harrington 1991). The consistency of $\widehat{\Lambda}_0^w(t; \widehat{\beta}_T, \widehat{\theta})$ for $\Lambda_{0T}(t)$ is proved in Pan and Schaubel (2007) through the USLLN and various empirical process results.

Theorem 4. Under conditions (a) to (f), $n^{\frac{1}{2}}\{\widehat{\Delta}(t) - \Delta(t)\}$ converges weakly to a

zero-mean Gaussian process with covariance function:

$$\bar{p}E\{\phi_i^w(s;\beta_T,\theta_0)\phi_i^w(t;\beta_T,\theta_0)\} + E\{\phi_i(s;\beta_S)\phi_i(t;\beta_S)\}$$

$$- \bar{p}\text{Cov}\{N^{\frac{1}{2}}\widehat{\Lambda}_0^w(s;\widehat{\beta}_T,\widehat{\theta}), n^{\frac{1}{2}}\widehat{\Lambda}_0(t;\widehat{\beta}_S)\} - \bar{p}\text{Cov}\{N^{\frac{1}{2}}\widehat{\Lambda}_0^w(t;\widehat{\beta}_T,\widehat{\theta}), n^{\frac{1}{2}}\widehat{\Lambda}_0(s;\widehat{\beta}_S)\},$$

for $(s,t) \in (0,\tau] \times (0,\tau]$

Since $\bar{p} = E(n/N)$, it would be exchangeable if we put $N^{\frac{1}{2}}$ as the scale factor instead of $n^{\frac{1}{2}}$. The proof is similar to that of Theorem 2, in the sense that we decompose $n^{\frac{1}{2}}\{\widehat{\Delta}(t) - \Delta(t)\}$ into $\bar{p}^{\frac{1}{2}}N^{\frac{1}{2}}\{\widehat{\Lambda}_0^w(t;\widehat{\beta}_T,\widehat{\theta}) - \Lambda_{0T}(t)\}$ and $n^{\frac{1}{2}}\{\widehat{\Lambda}_0(t;\widehat{\beta}_S) - \Lambda_{0S}(t)\}$. Through the Martingale Central Limit Theorem (Andersen and Gill 1982, Fleming and Harrington 1991), we obtain

$$\text{Cov}(n^{\frac{1}{2}}\widehat{\Lambda}_0(s;\widehat{\beta}_S), n^{\frac{1}{2}}\widehat{\Lambda}_0(t;\widehat{\beta}_S)) = E\{\phi_i(s;\beta_S)\phi_i(t;\beta_S)\},$$

where

$$\phi_i(t;\beta) = -\int_0^t \bar{z}'(s;\beta)d\Lambda_{0S}(s)(t;\beta)A(\beta)^{-1}\psi_i(\beta) + \int_0^t \frac{dM_i(s;\beta,\theta)}{s_w^{(0)}(s;\beta,\theta)}.$$

It is proved in Pan and Schaubel (2007) that

$$\text{Cov}(N^{\frac{1}{2}}\widehat{\Lambda}_0^w(s;\widehat{\beta}_T,\widehat{\theta}), N^{\frac{1}{2}}\widehat{\Lambda}_0^w(t;\widehat{\beta}_T,\widehat{\theta})) = E\{\phi_i^w(s;\beta_T,\theta_0)\phi_i^w(t;\beta_T,\theta_0)\}.$$

The covariance between $N^{\frac{1}{2}}\widehat{\Lambda}_0^w(s;\widehat{\beta}_T,\widehat{\theta})$ and $n^{\frac{1}{2}}\widehat{\Lambda}_0(s;\widehat{\beta}_S)$ results from the subject-specific contributions to each of the weighted and unweighted cumulative baseline hazard functions such that

$$\text{Cov}(N^{\frac{1}{2}}\widehat{\Lambda}_0^w(s;\widehat{\beta}_T,\widehat{\theta}), n^{\frac{1}{2}}\widehat{\Lambda}_0(t;\widehat{\beta}_S)) = \bar{p}^{\frac{1}{2}}E\{\phi_i^w(s;\beta_T,\theta_0)\phi_i(s;\beta_S)\}.$$

This covariance function can be consistently estimated by replacing all the unknown quantities by their empirical counterparts.

To address the question of whether or not the two samples are different in terms of the baseline hazards at a pre-specified time point, t_0, a pointwise confidence interval for $\Delta(t_0)$ can be calculated using the previously derived variance estimators. This interval is useful strictly when there are pre-specified time points of particular clinical interest. Another more general question is whether or not the two baseline hazard functions are equal over the entire follow up period; i.e., $H_0 : \Lambda_{0T}(t) = \Lambda_{0S}(t)$ for $t \in (0, \tau]$, implying the construction of a confidence band. Lin, Fleming and Wei (1994) provide a simulation-based method to estimate a confidence band for $\widehat{S}_0(t; \widehat{\beta}) = \exp\{-\widehat{\Lambda}_0(t; \widehat{\beta})\}$ based on an unweighted Cox model. We extend this general approach to our setting to generate the distribution of $\widehat{\delta} \equiv \sup_{t \in [0,\tau]} |n^{\frac{1}{2}}\{\widehat{\Delta}(t) - \Delta(t)\}Y(t)|$. We weight the $\widehat{\delta}$ values at each time point by $Y(t) = \sum_{i=1}^{N} I_i Y_i(t)$, the number of subjects at risk. Essentially, a zero-mean Gaussian process with the same covariance function as $n^{\frac{1}{2}}\widehat{\Delta}(t)$ is simulated to approximate the distribution of $n^{\frac{1}{2}}\widehat{\Delta}(t)$. Specifically, we replace $(n/N)^{\frac{1}{2}}N^{-\frac{1}{2}}\widehat{\psi}_i^w(t; \widehat{\beta}_T, \theta) - n^{-\frac{1}{2}}\widehat{\psi}_i(t; \widehat{\beta}_S)$, the contribution to $n^{\frac{1}{2}}\{\widehat{\Delta}(t) - \Delta(t)\}$ from each subject, with $\{(n/N)^{\frac{1}{2}}N^{-\frac{1}{2}}\widehat{\psi}_i^w(t; \widehat{\beta}_T, \theta) - n^{-\frac{1}{2}}\widehat{\psi}_i(t; \widehat{\beta}_S)\}H_i$, where H_i $(i = 1, \cdots, N)$ are independent standard normal variables. The simulated $\{(n/N)^{\frac{1}{2}}N^{-\frac{1}{2}}\widehat{\psi}_i^w(t; \widehat{\beta}_T, \theta) - n^{-\frac{1}{2}}\widehat{\psi}_i(t; \widehat{\beta}_S)\}H_i$ are still asymptotically independent among all subjects. Taking the sum of all subjects at each time point, we obtain

$$n^{\frac{1}{2}}\{\widetilde{\Delta}(t) - \Delta(t)\} = (n/N)^{\frac{1}{2}}N^{-\frac{1}{2}}\sum_{i=1}^{N}\widehat{\psi}_i^w(t; \widehat{\beta}_T, \theta)H_i - n^{-\frac{1}{2}}\sum_{i=1}^{N}I_i\widehat{\psi}_i(t; \widehat{\beta}_S)H_i$$

empirically. The aggregate difference in the baseline hazard estimators, $\widetilde{\delta} = \sup_{t \in [0,\tau]} |n^{\frac{1}{2}}\{\widetilde{\Delta}(t) - \Delta(t)\}Y(t)|$, has the same covariance matrix as the one proposed for $\widehat{\delta}$ on $[0, \tau]$. The supremum is taken over all time points at which an event is observed since $\widetilde{\delta}$ doesn't change between observed event times. The above simulation, with different sets of H_i $(i = 1, \cdots, N)$, is iterated a large number of times (e.g. 500). The obtained set of 500 $\widetilde{\delta}$ values, one from each iteration, form

an empirical sample of $\widetilde{\delta}$'s distribution. Finally, an approximate $(1-\alpha)$ confidence band for $\Delta(t)$ on $[0,\tau]$ is $\widehat{\Delta}(t) \pm n^{-\frac{1}{2}} \widehat{q}_\alpha / Y(t)$, where $\widehat{q}_\alpha$ denotes the empirical $100(1-\alpha)$ quantile satisfying $P\{\widetilde{\delta} > \widehat{q}_\alpha\} = \alpha$. The applicability of this confidence band procedure in finite samples has been validated in the unweighted proportional hazards setting through numerical studies by Lin, Fleming and Wei (1994).

3.3 Simulation Study

We simulated three covariates Z_{i1}, Z_{i2} and Z_{i3}, where Z_{i1} is distributed as Bernoulli(0.5), Z_{i2} is distributed as $N(0,25)$ and Z_{i3} is distributed as Uniform$(0,4)$. The event time, T_i, follows an exponential distribution with hazard $\lambda_i(t) = \lambda_0 e^{\beta_0' Z_i}$, where $\lambda_0 = 0.02$ and the vector of coefficients is $\beta_0' = (\beta_1, \beta_2, \beta_3) = (0.5, 0.1, 1.0)$. The censoring time, C_i, is uniform on $(0,40)$ such that the corresponding censoring percentage is approximately 20%. The representative samples (before selection) have sizes ranging from $N = 100$ to $N = 1000$, while the biased samples are created by selecting various percentages of subjects from the various Z_{i1} and Z_{i3} combinations. Specifically, subjects with $Z_{i1} = 0$ are always selected; for subjects with $Z_{i1} = 1$, probability of being included in the study sample depends on Z_{i3} with constant selection probabilities across $[0,1], (1,2], (2,3], (3,4]$. Each data configuration is iterated $1,000$ times.

Our goal is to estimate the regression coefficients when we fit a model with Z_{i1} and Z_{i2} only, such that the Cox model of interest is given by $\lambda_i(t) = \lambda_0(t) e^{\beta_1 Z_{i1} + \beta_2 Z_{i2}}$. The weights are estimated through the following logistic model, $\log\{p_i(\theta_0)/(1 - p_i(\theta_0))\} = \theta_0' X_i$, where $X_i = (Z_{i1}, Z_{i3})'$. Here, Z_{i3} works as a biasing factor, and the marginal effects of Z_{i1} over Z_{i3} in the target population and in the selected sample will be different in the presence of biasing factors. Thus, fitting an unweighted Cox model without Z_{i3} could potentially introduce bias in estimating the marginal regression

coefficients for Z_{i13}. The proposed difference $\widehat{D}_1$ is calculated by taking difference between $\widehat{\beta}_{1T}$ and $\widehat{\beta}_{1S}$ from the weighted and unweighted proportional hazards models, respectively. The variance of $\widehat{D}_1$ is estimated using the formula derived in Theorem 3 and the proposed test statistic is computed. The hypothesis H_0: $\beta_{T1} = \beta_{S1}$ is tested against H_1: $\beta_{T1} \neq \beta_{S1}$.

The performance of the $\widehat{D}$ statistic under H_0 is evaluated in Table 1. The quantity $E(n)$ is the expected sample size after selection from the representative sample of size N. Selection probability is equal across all levels of Z_{i3}. That is, all subjects with $Z_{i1} = 0$ have selection probability of 1; all subjects with $Z_{i1} = 1$, regardless of Z_{i3} values, have selection probability of 0.25 in row 1, 4 and 7; 0.5 in rows 2, 5, 8; 0.75 in rows 3, 6, 9. Although Z_{i3} affects $\lambda_i(t)$, it is not correlated with the factor of interest, Z_{i1}, nor does it predict selection probability. Therefore, the regression coefficient, β_{1T}, from the target population and β_{1S} from the selected sample are equal. For each data configuration, the unweighted and weighted β_1 estimates are obtained, with their average difference, $\widehat{E}(\widehat{D}_1)$, listed in column 5. We also list the empirical standard deviation (ESD) of $\widehat{D}$ in column 6. Two sets of variance estimators are calculated. In column 7 and 8, the estimated weights are inappropriately treated as fixed. Ignoring the extra variance induced by the estimation of the weights leads to incorrect asymptotic standard errors (ASE) and empirical significance levels (ESL) (i.e., artificially low ASEs and high ESLs). In columns 8 and 9, the variance estimator is based on Theorem 2. The ASEs using our proposed variance estimators are very close to the corresponding ESDs and the ESLs are always close to the nominal level of 5%, even when a relatively low percentage of subjects gets sampled.

We also evaluated the power of the proposed test under various departures from H_0: $\beta_{1T} = \beta_{1S}$. We set up different sampling percentages for subjects with different

row	N	$p_i(\theta_0)$	$E(n)$	$\widehat{E}(\widehat{D})$	ESD	$w_i(\widehat{\theta})$ treated as estimated ASE	ESL	$w_i(\widehat{\theta})$ treated as fixed ASE	ESL
1	100	0.25	62.5	-0.008	0.224	0.201	2.1%	0.112	28.4%
2		0.50	75	0.001	0.119	0.118	3.1%	0.051	38.3%
3		0.75	87.5	-0.003	0.064	0.065	3.4%	0.024	50.2%
4	500	0.25	312.5	-0.003	0.085	0.087	3.6%	0.029	55.3%
5		0.50	375	-0.001	0.047	0.046	4.0%	0.011	66.0%
6		0.75	437.5	0.000	0.026	0.026	4.8%	0.005	77.4%
7	1000	0.25	625	0.000	0.058	0.059	4.3%	0.015	63.5%
8		0.50	750	-0.001	0.032	0.033	4.4%	0.006	72.6%
9		0.75	875	-0.000	0.018	0.019	3.8%	0.002	80.0%

$p_i(\theta_0) = P\left(I_i = 1 | Z_{i1} = 1, Z_{i3} \in (0, 4]\right)$
ESD = empirical standard deviation
ASE = average asymptotic standard error
ESL = empirical significance level.

Table 3.1: Simulation Results: Performance of test for regression parameter under the null

Z_{i3} levels. The power is consistently high (and near 100%) for settings with moderate-size selected sample. The high power of the test, in part, results from the high correlation between $\widehat{\beta}_T$ and $\widehat{\beta}_S$, since the same data set is used in calculating both estimators.

An important factor that affects the power of our proposed test is the size of the selected sample. The power is poor when the representative sample before selection is small or when the representative sample is moderate-size but the biased sample is small. Figure 1 shows the power when the expected sample size of the potentially biased sample increases from 10 to 90 with the same sampling scheme. With moderate sampling probabilities (0.4, 0.2, 0.16, and 0.04 for increasing Z_{i3} levels), the power increases from 13.5% with original sample size 100 to 95.7% when the size before sampling is 500 and plateaus close to 100% thereafter.

Covariate measurement error reduces the power of the proposed test to detect selection bias, as shown by our simulation results. In the data configuration described in Section 4 with biased selection for different levels of Z_{i3} ($p_i(\theta_0) = 1, 0.5, 0.4, 0.1$
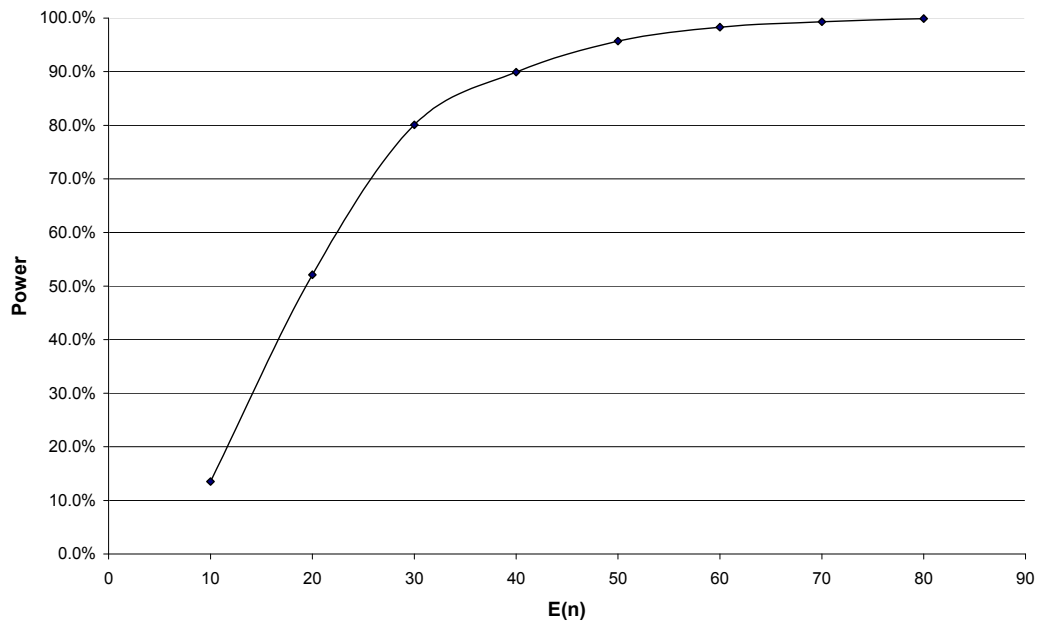
Figure 3.1: Simulation Results: Power of proposed test of H_0: $\beta_{1T} = \beta_{1S}$; note $n = \sum_{i=1}^{N} I_i$.
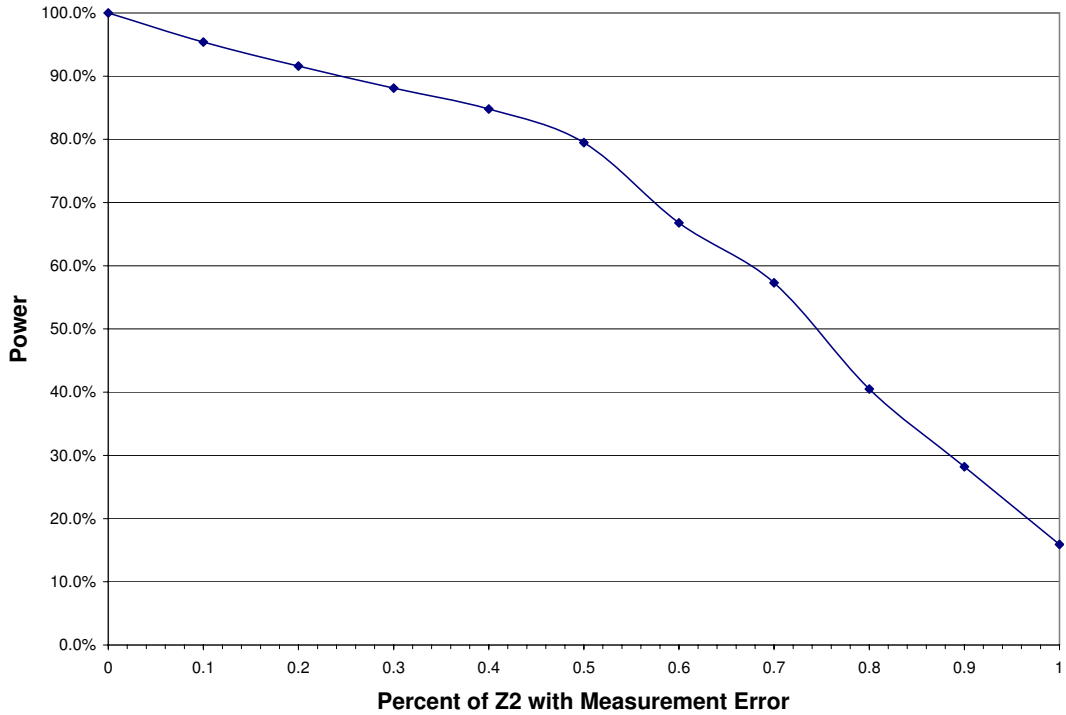
Figure 3.2: Simulation Results: Effect of measurement error on power of proposed test of H_0: $\beta_{1T} = \beta_{1S}$.

for increasing Z_{i3} levels), independent Normal$(0,3)$ random errors are added to some Z_{i3} while all Z_{i3} values with or without errors remain bounded within $[0,4]$. When the percentage of Z_{i3} with measurement errors increases, the power of our proposed test drops from 100% to 15.9% (Figure 2).

3.4   Application to Kidney Transplant Data

The target population consists of all patients wait-listed for primary kidney transplantation in the U.S.. Naturally, this hypothetical population is infinite. A representative sample was obtained by selecting a cross-section of all patients active on the wait-list as of January 1, 2000. In total, there were $13,627$ candidates on the

waiting list that day. Candidates under age 18 are excluded. A five-year time period $(01/01/2000 - 12/31/2004)$ was chosen over which kidney transplants among the cross-section subjects were observed. The decision for using a duration of five years can be justified by the plot of the cumulative incidence of transplantation (Web Figure 1), treating death and removal from the wait-list as competing risks (Kalbfleisch and Prentice 2002); Since the curve plateaus after approximately five years. On December 31, 2004, each patient had potential observation time on wait-list of at least five years. Those who did not receive a transplant either died or were removed from the waiting list due to illness or recovery of native renal function. Among the representative sample of $13,627$ patients, $6,425$ (47%) received a kidney transplant. The transplanted patients form the selected and potentially biased sample.

Demographic and clinical data on donors and recipients, dates of graft failure and death where applicable, as well as various clinical measures, were obtained from the Scientific Registry of Transplant Recipients (SRTR) and collected by the Organ Procurement and Transplant Network (OPTN). The potential biasing factors under consideration are represented in aggregate by the hospitalization history. For example, it is suspected that patients with fewer hospitalizations are preferred candidates for kidney transplantation, even after conditioning on all the covariates included in the proportional hazards model. The frequency and length of hospitalization is also correlated with many patient comorbidities, as well as the post-transplant mortality hazard.

Note that adjusting for hospitalization history through covariates in the post-transplant model is not an attractive option based on practical considerations. Specifically, hospitalization information on transplant candidates will not be available to UNOS at the time when an organ is to be allocated and hence when candidates must be ranked. The SRTR has access to both the OPTN (wait-list and transplant) and CMS (hospitalization history) databases and can link the sources by patient for those who pay for their health care expenses via Medicare. However, UNOS will not have access to updated CMS data for the purposes of ranking candidates.

We model the probability of being selected to receive a kidney transplant using

a binary end point: transplanted within five years $(0,1)$. Implicitly, we assume that the $\widehat{\theta}$ estimated through this approach converges to the θ_0 which governs transplant probability. In addition to hospitalization frequencies and total days hospitalized, other covariates adjusted for in the selection probability model include candidate demographics (gender, age, race), years with end stage renal dialysis (ESRD), various disease conditions (drug treated chronic obstructive pulminary disease (COPD), angina), primary diagnosis at time of listing (polycystic kidneys, diabetes, hypertension), functional status (from fully active daily living to severely disabled), and various interactions among these covariates. Most terms in the logistic model are significant predictors of transplant probability at the 0.05 level (Table 2). For each incidence of hospitalization, the covariate-adjusted odds of getting a transplant drops 4%, while the odds of receiving a transplant drops 1% for each additional day hospitalized holding all other factors constant (including number of hospitalizations). Of the $3,369$ patients with no previous hospitalization, $1,974$ (59%) got transplanted. In contrast, of the $10,258$ candidates with at least one hospitalization, $4,451(43\%)$ got transplants.

For the $6,425$ patients who received a transplant, each recipient was followed until death, loss to follow up or the conclusion of the observation period (06/30/2006). Among the $6,425$ transplant recipients, $5,025$ were alive at the end of follow-up, while $1,400$ (22%) died.

The predicted selection probabilities from the logistic model could get very close to zero and hence lead to unrealistically large weight values. These large values are very influential on $\text{Var}(\widehat{\beta}_T)$. Bounding weights at an arbitrary upper limit, u, reduces $\text{Var}\{\widehat{\beta}_T(u)\}$ at the price of larger bias. An optimal upper bound for weights should minimize mean square error (MSE) of $\widehat{\beta}_T(u)$, defined as the sum of $diag[\text{Var}\{\widehat{\beta}_T(u)\}]$ and $diag[\{\widehat{\beta}_T(u) - \widehat{\beta}_T\}^{\otimes 2}]$, where $diag(A)$ is a vector consisting of the main diagonal of A. In our real data analysis, the weights range from 1.21 to 54.22. Web Figure 2 shows the U-shape curve of the MSE for various upper bounds. For our application, we estimate that the minimum MSE is achieved when weights are bounded at 20.

Both weighted and unweighted proportional hazards models were fitted, adjusting

Covariate, X_{ik}	$\widehat{\theta}_k$	$\widehat{SE}(\widehat{\theta}_k)$	p	$e^{\widehat{\theta}_k}$
Hospitalization frequency	-0.03	0.008	< 0.0001	0.97
Days hospitalized	-0.007	0.001	< 0.0001	0.99
Female	-0.30	0.05	< 0.0001	0.74
Age 18 - 24	0.63	0.13	< 0.0001	1.87
Age 25 - 34	0.57	0.07	< 0.0001	1.77
Age 35 - 44	0.27	0.06	< 0.0001	1.31
Age 55 - 64	-0.26	0.05	< 0.0001	0.77
Age 65 - 70	-0.27	0.09	0.0021	0.76
Age $\geq$ 70	-0.69	0.11	< 0.0001	0.50
African American vs. Caucasian	-0.10	0.04	0.02	0.90
COPD	-0.35	0.17	0.04	0.70
Angina	-0.24	0.06	< 0.0001	0.78
Polycystic kidneys	0.29	0.09	0.0008	1.34
Diabetes	-1.17	0.21	< 0.0001	0.31
Hypertension	-0.11	0.05	0.02	0.90
Years on ESRD	-0.01	0.01	0.17	0.99
Functional status: minor disability	-0.16	0.07	0.02	0.85
Female$\times$Hospitalization frequency	0.02	0.01	0.02	1.02
Age 65-70$\times$Hospitalization frequency	-0.06	0.02	0.0006	0.94
Age$\times$Diabetes	0.01	0.004	0.0092	1.01

Table 3.2: Analysis of SRTR Data Estimated Regression Parameters from Selection Probability Model

for the same set of covariates: expanded criteria donor (ECD), recipients demographics (age, race, gender), years on ESRD prior to wait-listing, hepatitis C antibody (HCV) status, Chronic obstructive pulmonary disease (COPD), angina, primary renal diagnosis, functional status, and stay in the intensive care unit (ICU) at the time of wait-listing. All covariates had significant effects on the post-transplant mortality hazard.

Table 3 lists the difference between each of the weighted and unweighted regression coefficient estimates, as well as its corresponding standard error, Chi-square test statistic, and P-value. For almost all covariates, significant differences between the weighted and unweighted covariate effects are not detected. Age $35 - 44$ (compared to age $45 - 54$), years on ESRD and COPD (yes vs. no) have significantly different effects in the representative sample and in the selected sample. The hazard ratio for COPD is 2.40 in the representative sample and 1.55 in the transplanted sample (P= 0.008). For each additional year on ESRD, the post-transplant death hazard increases by a multiplier of 1.03 for the transplanted population and by a multiplier of 1.05 for the wait-listed population (P= 0.046). The hazard for age group $35 - 44$ is 0.79 times of that for the reference group $45 - 54$ using the weighted estimate, while the hazard ratio decreases to 0.67 with the unweighted estimate (P= 0.029).

Survival function estimates based on the proportional hazards model are calculated for both the representative sample and the transplanted sample, as well as estimated cumulative baseline hazard functions. The weighted and unweighted survival curves are plotted in Web Figure 3. Both the 95% pointwise confidence interval and confidence bands of the difference of the two estimated baseline hazards functions are calculated based on results from Theorem 4 and are together in Figure 3. It is clear from the figure that at most of the $0 - 5$ year post-transplant time interval, the point estimate of the baseline hazards for the wait-listed patients and that for the transplanted patients are significantly different on a 0.05 significance level. However, carrying out joint inference of the process across the follow up period, the two baseline functions appear not to be different.

Covariate, Z_{ik}	$\widehat{\beta}_T$	$\widehat{\beta}_S$	$\widehat{D}$	$\widehat{SE(\widehat{D})}$	Chi-square	P-value
Drug-treated-COPD	0.87	0.44	0.43	0.16	6.96	0.008
Years on ESRD	0.05	0.03	0.02	0.01	3.98	0.046
Age 18 to 24	-1.10	-1.13	0.03	0.20	0.02	0.883
Age 25 to 34	-0.56	-0.73	0.16	0.20	0.63	0.426
Age 35 to 44	-0.24	-0.40	0.17	0.08	4.77	0.029
Age 55 to 64	0.51	0.49	0.02	0.05	0.12	0.730
Age 65 to 70	0.71	0.73	-0.02	0.05	0.13	0.721
Age 70 up	1.08	1.01	0.07	0.08	0.69	0.405
HCV Positive	0.25	0.28	-0.02	0.11	0.09	0.770
In ICU	1.42	1.47	-0.05	0.15	0.11	0.737
Female	-0.06	-0.07	0.01	0.04	0.06	0.800
Angina	0.37	0.30	0.07	0.06	1.33	0.249
Polycystic kidney	-0.40	-0.32	-0.08	0.07	1.49	0.222
Diabetes	0.48	0.52	-0.03	0.04	0.53	0.465
Functional status: minor disability	0.26	0.21	0.05	0.10	0.26	0.613

Table 3.3: Analysis of SRTR Data Tests of Differences between Regression Coefficients estimated through models fitted to Wait-Listed Candidates ($\widehat{\beta}_T$) and Transplant Recipients ($\widehat{\beta}_S$)
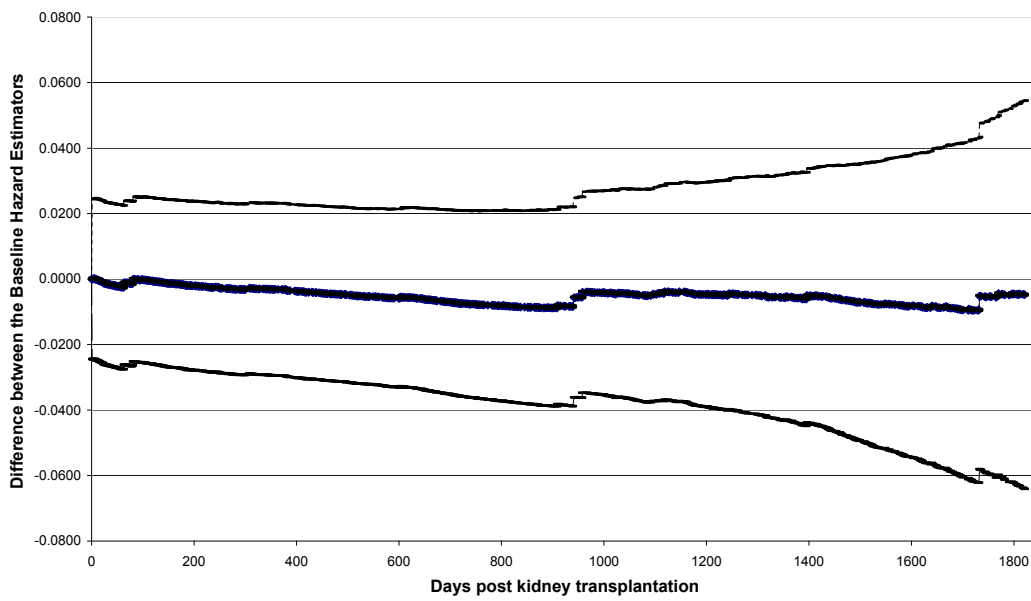
Figure 3.3: Analysis of SRTR data: Test of H_0: $\Lambda_{0T}(t) = \Lambda_{0S}(t)$. Middle line: $\widehat{\Lambda}_{0T}(t; \widehat{\beta}_T, \widehat{\theta}) - \widehat{\Lambda}_{0S}(t; \widehat{\beta}_S)$; Outer lines: 95% confidence band.

3.5   Discussion

Weighted proportional hazards models are an attractive option in survival analysis targeted at a larger underlying population, for the setting where only a possibly biased sample is observed. Pan and Schaubel proposed an inverse selection probability weighting proportional hazards model for the setting wherein sampling probabilities are unknown. In the presence of biasing factors, the weighted proportional hazards model with estimated ISPW yields parameter estimators which are consistent and which have reduced variance relative to those which would be estimated by treating the weights as fixed. These properties come at the expense of increased complexity and computing time. To identify cases where ISPW is required for a proportional hazards model, we propose tests for estimating the bias in the unweighted regression parameter and cumulative hazard estimators. The asymptotic properties of the proposed statistics are derived. The finite sample performance of a Wald test based on the proposed statistic is examined through simulation studies with various data configurations. In the case of no biasing factors, the empirical significance level is close to 0.05. In the presence of biasing factors, the proposed test is quite powerful. In cases where the size of the selected sample is insufficient (e.g., $n = 20$), the power drops to very low levels (e.g., 14%). A method for evaluating the difference between the baseline cumulative hazard functions, for the target population and that for the selected sample is also proposed, with the pertinent asymptotic distributions derived.

The application of the proposed test is very general. When practitioners encounter data with potential biasing factors, practitioners would be interested in quantifying the impact of the weighting on bias reduction. In the absence of such a test, choosing between the weighted and unweighted methods would be ad-hoc. The proposed testing methods can be used to formalize the decision to use a weighted or unweighted hazard regression model. The test proposed in this paper works as a statistical tool to decide which model (weighted or unweighted) to apply in fitting a proportional hazards model to a potentially biased sample. Similar tests could be derived for the data structures considered by Binder (1992), Lin (2000) and Boudreau and Lawless

(2006).

The proposed procedures are applied to national kidney transplant data, where hospitalization history represents the potential biasing factor. Age $35 - 44$, years on ESRD and COPD were each found to have significantly different regression coefficients for the representative sample of all wait-listed patients and the potentially biased sample of transplant recipients. The pointwise confidence interval for $\Delta(t)$ deviates from zero at some time points, but the confidence bands obtained by taking the supremum over all follow up period had the zero-difference line included throughout, indicating no significant difference. The test of H_0: $\beta_T = \beta_S$ is powerful in part due to the high correlation between $\widehat{\beta}_T$ and $\widehat{\beta}_S$. The correlation between $\widehat{\Lambda}_0^w(t; \widehat{\beta}_T)$ and $\widehat{\Lambda}_0(t; \widehat{\beta}_S)$ can thus also be expected to be very high and, by extension, $\widehat{\Delta}(t_1)$ and $\widehat{\Delta}(t_2)$ would be expected to be highly correlated. For this reason, the set of point-wise confidence intervals could give a misleading representation of the process $\{\widehat{\Delta}(t); t \in [0, \tau]\}$ since the interpretation of sets of point-wise confidence intervals makes sense only if $\widehat{\Delta}(t_1)$ is uncorrelated with $\widehat{\Delta}(t_2)$ for $t_1 \neq t_2$, which is clearly not the case.

Recently there is great interest in reconstructing the organ allocation system for various organs, including kidney. A newly proposed criterion by which to rank candidates for an available donor kidney is the difference between predicted median life with a transplant and that without a transplant. It is therefore of great value to obtain an accurate post-transplant survival model applicable to the patients which require ranking; namely patients wait-listed for a kidney transplant. An ISPW proportional hazards model results in estimators applicable to all wait-listed patients, but at the expense of increased computation and complexity. Based on the proposed test, the unweighted model appears to be sufficient. Among the twenty-one covariates, only three were found to be significantly different: age $35 - 44$, years on ESRD and presence of COPD. The majority of wait-listed patients are at least age 45 and are free of COPD. Moreover, no difference was detected with respect to the target and sample baseline cumulative hazard function.

The practicality of the proposed ISPW method depends on the availability of

information from a representative sample of the target population, as well as the availability of data on the potential biasing factors. Since the statistics, $\widehat{D}$ and $\widehat{\Delta}(t)$ are constructed by taking the difference between the ISPW parameter estimators and their unweighted counterparts, the same restriction applies to the test proposed in this chapter. If we are unable to fit a model to estimate selection probability, neither $\widehat{\beta}_T$, $\widehat{D}$ or $\widehat{\Delta}(t)$ are obtainable. In addition, appropriate upper bound can be set for the estimated weights, intended to achieve the weighted regression parameter with minimum MSE.

Although our simulation studies demonstrate high power for the proposed test, we only detect three covariates with significant differences in our real example of kidney transplant analysis. There are three possible reasons leading to the low detection number. First, we examine twenty-one recipient and donor characteristics in total. If we had information on more covariates affecting the post-transplant mortality hazard, it is quite possible that we would detect more covariates with significantly different effects for the representative sample and the transplanted sample. That is, more complete covariate adjustment may result in increased precision and, hence, power. Second, although the SRTR is an invaluable source of information on U.S. kidney transplant patients, covariate measurement error is unavoidable due to inaccurate recall on the part of the patients supplying the information and/or inaccurate interpretation or recording of information. From our simulation results (Figure 2), the power of the proposed tests drops in the presence of measurement error. Further study taking measurement error into account could be an interesting and valuable extension of our current work. Finally, the representative sample is assembled by taking a cross section on a single day, other random sampling methods might be applied to achieve a larger representative sample and thus a larger selected sample. With more subjects in our models, the power of the proposed test would increase and some parameter differences with borderline P-values could attain statistical significance (e.g., angina, initial diagnosis as polycystic kidney). Furthermore, if strong empirical or clinical prior knowledge exists to decide the direction of the difference between β_T and β_S, a one-sided test can be used. For example, if the effects of a

given comorbidity in the wait-listed population is always larger in magnitude than that in the transplanted population (e.g., since the selection among patients with the comorbidity is more stringent), then one would employ the one-sided alternative hypothesis, H_1: $D_j > 0$.

3.6 Acknowledgements

3.7 References

ANDERSEN, P. K., & GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-20.

BILIAS, Y., GU, M. & YING, Z. (1997). Towards a general asymptotic theory for the Cox model with staggered entry. *Ann. Statist.* **25**, 662-82.

BINDER, D. (1992). Fitting Cox's proportional hazards models from survey data. *Biometrika* **79**, 139-47.

BOUDREAU, C. & LAWLESS, J. F. (2006). Survey analysis based on the proportional hazards model and survey data. *The Canadian Journal of Statistics* **34**, 203-216.

Cox, D. (1975). Partial likelihood. *Biometrika* **62**, 262-76.

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

Horvitz, D. G. & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. R. Statist. Soc.* B **47**, 663-85.

Kalbfleisch, J. D & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Hoboken, NJ: Wiley.

Lin, D. Y. (2000). On fitting Cox's proportional hazards models to survey data. *Biometrika* **87**, 37-47.

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

Pan, Q. & Schaubel, D. E. (2007). Proportional Hazards Regression Based on Biased Samples and Estimated Selection Probabilities *Conditionally accepted by Canadian Journal of Statistics*

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

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Melbourne: Cambridge.

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

CHAPTER IV

Modeling Additive Treatment Effects on the Recurrent Event Mean in the Presence of a Terminating Event

Abstract: In this chapter, we study the situation where the event of interest can occur repeatedly for the same subject (i.e., a recurrent event; e.g., hospitalization) and may be ended by a terminating event (e.g., death). Methods in the existing literature for comparing treatment-specific event occurrences in the presence of a terminating event generally fall into one of two broad categories. Marginal methods compare the mean number of events, averaging over the survival experience; while conditional methods compare the conditional recurrent event rate given survival. Often, the difference between treatment-specific recurrent event means will not be constant over time, particularly when treatment-specific differences in survival exist. In such cases, it makes more sense to quantify treatment effect based on differences in the recurrent event (cumulative) means, as opposed to the (instantaneous) rate. We propose two methods that compare treatments by separately estimating the survival probabilities and recurrent event rate given survival, then integrating to get the mean number of events. Both methods combine an additive model for the conditional recurrent event rate and a proportional hazards model for the terminating event hazard. The first method factors out differences in the treatment-specific survival distributions by employing a common survival distribution (intended to serve as a standard) for both treatment groups. The second proposed method features treatment-specific survival distributions and generates an estimated difference in treatment-specific means. The example which motivates this research is the repeated occurrence of

hospitalization among kidney transplant recipients, where the effect of Expanded Criteria Donor (ECD) compared to non-ECD kidney transplantation on the mean number of hospitalizations is of interest.

Key words and phrases: Additive rates model, Competing risks, Marginal mean, Proportional hazards model, Recurrent event, Terminating event.

4.1 Introduction

Recurrent events are frequently of interest in clinical and epidemiologic studies. Examples include repeated infections among HIV patients and multiple hospitalizations in a health services utilization study. A large variety of semiparametric recurrent event models exist in the literature. These methods can generally be classified as intensity models or mean/rate models. Intensity models consider the instantaneous probability of event occurrence conditional on the event history (e.g., Prentice, Williams and Peterson, 1981; Andersen and Gill, 1982), while mean/rate (the rate being the derivative of mean) models consider the marginal mean number of events (e.g., Pepe and Cai, 1993; Lawless and Nadeau, 1995; Lin, Wei, Yang and Ying, 2000). Depending on the assumed form of the covariate effects, recurrent event models can be classified as multiplicative or additive. Proportional means/rates models assume covariate effects on a multiplicative scale, while additive models assume them on an additive scale.

Often in biomedical applications, the recurrent event sequence can be stopped permanently by a terminating event (e.g., death). Various approaches have been proposed for modelling recurrent events in the presence of a terminating event (e.g., Li and Lagakos, 1997; Cook and Lawless, 1997; Ghosh and Lin, 2000; Ghosh and Lin, 2002), and this area has attracted much attention recently (e.g., Liu, Wolfe and Huang, 2004; Huang and Wang, 2004; Ye, Kalbfleisch and Schaubel, 2007). The existing approaches generally fall into one of three categories. First, there are methods for modelling the marginal mean number of events (e.g., Ghosh and Lin, 2000; Ghosh and Lin, 2002). In this case, the mean averages over surviving and deceased subjects. In the second category of methods, one models the conditional

recurrent event rate given survival. This is among the approaches suggested by Cook and Lawless (1997), and was mentioned by Lin et al. (2000). This method has been applied quite frequently (e.g., Schaubel and Cai, 2005). A variation of this approach employs a latent (frailty) variable (e.g., Liu et al., 2004; Ye et al., 2007), conditional on which the recurrent and terminal events are assumed independent. The marginal and conditional methods explicitly acknowledge the fact that subjects no longer experience recurrent events after death and that time to death may differ between the treatment and placebo groups. The third approach considers the recurrent events to be a latent process, unobservable after death (e.g., Ghosh and Lin, 2003; Huang and Wang, 2004). Unlike the previously described marginal and conditional methods, the latent process approach does not explicitly acknowledge that a subject can not experience recurrent events subsequent to experiencing the terminating event.

In the application of the Wei, Lin and Weissfeld (WLW; 1989) model to recurrent event data, a separate proportional hazards model is fitted for each recurrent event number, k $(k = 1, \cdots, K)$, with the regression parameters and baseline hazard being k-specific. Li and Lagakos (1997) applied the WLW method to the recurrent/terminal event setting. The authors proposed various methods for handling death. One procedure treats death as a censoring of the recurrent event sequence, in which case the estimated intensity function was interpreted as conditional on the subject being alive. In another scenario, death was treated the same as a recurrent event; in which case the estimated treatment effects reflect the combined effects on both the recurrent event and death processes.

Cook and Lawless (1997) also studied the estimation of covariate effects on the recurrent event process conditional on survival. They modeled the recurrent event rates and means at time t conditional on either the death time equal time t or the death time no less than time t. Essentially, the comparison of treatment-specific recurrent event rates or means at time t considers subsets of subjects alive up to time t.

Ghosh and Lin (2000) proposed nonparametric tests for treatment-specific differences with respect to the marginal mean number of recurrent events. The Kaplan-

Meier estimator for survival probabilities and a Nelson-Aalen estimator of the conditional recurrent event rate function given survival are integrated to calculate the recurrent event means. Ghosh and Lin (2002) later proposed a semiparametric proportional means model for the marginal cumulative number of recurrent events. The recurrent event rate after death was still taken to be zero, and the marginal rates in the outcome are implicitly averaged over the recurrent event rates of living and deceased subjects. The authors modeled the same marginal mean as in Ghosh and Lin (2000), but allowed for covariate adjustment through a semiparametric proportional means model.

Often investigators are more interested in the absolute difference between recurrent event means (as opposed to their ratio) which implies an additive model. Schaubel, Zeng and Cai (2006) extended the method of Lin and Ying (1994) to develop an additive rates model, but it was not designed to handle terminating events. One could develop an extension of Schaubel et al. (2006) to accommodate terminating events, as did in Ghosh and Lin (2002) for the proportional means setting. However, treatment-specific differences in mean number of events are often not constant over time, particularly when the treatment-specific survival functions differ. That is, as follow-up time, t, increases, due to subjects dying, the composition of the study population will shift and the pattern of the shift will be treatment-specific if treatment affects survival. Therefore, treatment effects on the recurrent event mean would not be expected to be constant over time in the presence of treatment effects on survival. In cases where the treatment effect is not constant over time, the cumulative effect is of much greater interest for patient decision-making at time $t = 0$ than the instantaneous effect.

In this article, we propose two novel semiparametric methods for comparing treatment-specific recurrent event mean functions in the presence of a terminating event. Both methods estimate the treatment effect as a process over time and, in doing so, assume no functional form for the effect. The proposed methods combine a proportional hazards model for the terminating event and an additive model for the conditional recurrent event rate given survival. The first proposed method (here-

after referred to as Method I) employs a common survival distribution in computing treatment-specific recurrent event means, such that the treatment-specific differences in survival are factored out. The second method (Method II) employs treatment-specific survival functions, with the treatment effects measured by differences in treatment-specific means.

Under a marginal approach, an increase in the event mean could be due to patients surviving longer, or could be due to patients experiencing events at an increased rate while they survive, and there is no convenient way to distinguish these two phenomena. In an attempt to factor out differences in survival, Method I estimates what the difference in the recurrent event means would be if treatment-specific survival were equal. It measures treatment effects through a standardized difference in the marginal recurrent event means. That is, applying a standard survival distribution to compare treatment-specific expected numbers of recurrent events. The measure proposed in Method I is a hypothetical quantity. There may be occasions when investigators are interested in a difference in means which would directly incorporate (rather than factor out) treatment-specific difference in survival. Motivated by this consideration, we also propose Method II which features treatment-specific survival distributions and measures treatment effects through the estimated difference in treatment-specific marginal recurrent event means.

The remainder of this chapter is organized as follows. Section 4.2 explains how the proposed measures are estimated. The asymptotic properties of the estimators are listed in Section 4.3, while Section 4.4 evaluates the performance of the asymptotic results in moderate size samples. We compare mean numbers of hospitalizations among kidney transplant recipients using our methods in Section 4.5. A discussion and some concluding remarks are contained in Section 4.6.

4.2 Proposed Methods

In this section, we describe the proposed methods, after establishing notation and explaining the data structure of interest.

4.2.1   Notation and Set-up

We begin by setting up necessary notation. For the ith subject $(i = 1, \cdots, n)$, Z_i and X_i denote the vector of covariates for the terminating event model and recurrent event model respectively. We set $Z_i = (Z_{i1}, Z'_{i2})'$, $X_i = (X_{i1}, X'_{i2})'$ where Z_{i1}, X_{i1} are (1/0) indicators for the treatment/placebo and Z_{i2} and X_{i2} represent adjustment covariates. Correspondingly, we let $\beta_0 = (\beta_{01}, \beta'_{02})'$ and $\theta_0 = (\theta_{01}, \theta'_{02})'$ represent the regression coefficients for Z_i and X_i. In addition, for convenience we devote the covariate vectors for a treated subject by $Z_i^1 = (1, Z'_{i2})'$ and $X_i^1 = (1, X'_{i2})'$. Similarly, for a placebo subject, $Z_i^0 = (0, Z'_{i2})'$ and $X_i^0 = (0, X'_{i2})'$. We use D_i for the time of the terminating event (e.g., death), and C_i for the censoring time. The counting process for the terminating event, is represented by $N_i^D(t) = I(D_i \leq t, D_i < C_i)$. We let $Y_i(t) = I(C_i \wedge D_i \geq t)$ be the at risk indicator for subject i at time t, and set $\widehat{\pi}(t) = n^{-1} \sum_{i=1}^n Y_i(t)$. The number of recurrent events as of time t is represented by $N_i^{R*}(t) = \int_0^t dN_i^{R*}(s)$, where $dN_i^{R*}(s) = N_i^{R*}(s) - N_i^{R*}(s^-)$. In the data structure under consideration, $N_i^{R*}(t) = N_i^{R*}(t \wedge D_i)$; that is, recurrent events can not occur after death. What we observe is a quantity subject to right censoring, $N_i^R(t) = N_i^R(t \wedge D_i \wedge C_i)$. The outcome being modeled is the mean number of recurrent events $E\{N_i^{R*}(t)|Z_i\}$, which is the integral of recurrent event rate function over time $E\{N_i^{R*}(t)|Z_i\} = \int_0^t E\{dN_i^{R*}(t)|Z_i\}$. It should be mentioned that we are not modeling the recurrent event intensity function, $E\{dN_i^{R*}(t)|Z_i, \mathcal{N}_i^{R*}(t)\}$, where $\mathcal{N}_i^{R*}(t) = \{N_i^{R*}(s); s \in (0, t)\}$, representing the event history up to time t, an approach considered for example by Andersen and Gill (1982).

As described in Section 4.1, in the presence of a terminating event, the recurrent event mean is a function of survival probability and the conditional recurrent event rate given survival. In estimating the mean function, it is natural to consider these two entities separately. We denote the conditional recurrent event rate given survival by

$$dR_i(t) \;\; = \;\; E\{dN_i^{R*}(t) \mid D_i > t\}.$$

The recurrent event mean function is then given by

$$\mu_i(t) = \int_0^t S_i(u^-)dR_i(u),$$

where $S_i(t^-) = Pr(D_i \geq t)$. We assume that the terminating event hazard function follows a proportional hazards model,

(4.1) $$d\Lambda_i(t) = d\Lambda(t \mid Z_i) = d\Lambda_0(t)e^{\beta_0' Z_i},$$

where $d\Lambda_0(t)$ is the unspecified baseline hazard function and the vector β_0 represents the true value of the regression coefficients. The parameter β_0 is estimated by $\widehat{\beta}$, the solution to the partial likelihood score function, $U^D(\beta) = 0$, where

$$U^D(\beta) = \sum_{i=1}^n \int_0^\tau \{Z_i - \overline{Z}(t; \beta)\}dN_i^D(t)$$

$$\overline{Z}(\beta; t) = \frac{S^{(1)}(t; \beta)}{S^{(0)}(t; \beta)}$$

$$S^{(k)}(t; \beta) = \sum_{i=1}^n Y_i(t)e^{\beta' Z_i} Z_i^{\otimes k},$$

with $Z_i^{\otimes 0} = 1$, $Z_i^{\otimes 1} = Z_i$ and $Z_i^{\otimes 2} = Z_i Z_i'$. The Breslow-Aalen baseline hazard estimator, $\widehat{\Lambda}_0(t; \widehat{\beta})$, is employed, where

$$\widehat{\Lambda}_0(t; \beta) = n^{-1} \sum_{i=1}^n \int_0^t S^{(0)}(s; \beta)^{-1}dN_i^D(s).$$

The subject-specific survival function is then estimated by

$$\widehat{S}_i(t) = \exp\{-\widehat{\Lambda}_i(t)\}.$$

Both multiplicative and additive models have been used in modeling the recurrent event mean. For certain outcomes such as costs or number of hospitalizations, investigators are typically more interested in the absolute (as opposed to relative) dollar amount or hospital admission numbers. Here we use an additive model to estimate the recurrent event rates among surviving subjects,

(4.2) $$dR_i(t) = dR(t \mid X_i) = dR_0(t) + \theta_0' X_i dt,$$

where $dR_0(t)$ is the baseline recurrent event rate function and the vector θ_0 represents the true additive effects of the corresponding covariate vector X_i. Usually, there are common covariates besides X_{i1} and Z_{i1} affecting both the terminating event hazard and recurrent event rates. As such, the covariate vectors Z_i and X_i usually overlap and will often be identical. Adapting the model of Schaubel et al. (2006) to the recurrent/terminal events setting, we estimate θ_0 as follows:

$$
\begin{aligned}
\widehat{\theta} &= \widehat{B}^{-1} U^R \\
\widehat{B} &= \sum_{i=1}^{n} \int_0^{\tau} Y_i(s)\{X_i - \overline{X}(s)\}^{\otimes 2} ds \\
\overline{X}(s) &= \pi(s)^{-1} n^{-1} \sum_{i=1}^{n} Y_i(s) X_i \\
U^R &= \sum_{i=1}^{n} \int_0^{\tau} \{X_i - \overline{X}(t)\} dN_i^R(t).
\end{aligned}
$$

Our notation is separately listed in Table 4.1.

4.2.2   Method I: Survival-Adjusted Difference in Means

Of interest in this method is the effect of treatment on the mean number of recurrent events, factoring out the differences in survival distributions for different treatment groups. We propose a semiparametric method to estimate the additive treatment effects on the cumulative number of recurrent events given a standard survival distribution to be applied to each treatment group. Specifically, the following process is proposed to represent the treatment effect

$$
\phi(t) = \int_0^t S(u^- | Z_{i1} = 0) d\{R(u | X_{i1} = 1) - R(u | X_{i1} = 0)\},
$$

In the presence of adjustment covariates, Z_{i2} and X_{i2}, it is useful to formulate the measure as

$$
\begin{aligned}
\phi(t) &= E[E[\phi(t) | Z_{i2}, X_{i2}]] \\
&= E\left[\int_0^t S(u^- | Z_i^0) d\{R(u | X_i^1) - R(u | X_i^0)\} \right],
\end{aligned}
$$

Function	Terminating event	Recurrent event
Event time	D_i	-
Censoring time	C_i	C_i
Risk set indicator	$Y_i(t) = I(C_i \wedge D_i \geq t)$	$Y_i(t) = I(C_i \wedge D_i \geq t)$
Counting process for event	$dN_i^D(t)$	$dN_i^R(t)$
Residual	$dM_i^D(t)$	$dM_i^R(t)$
Covariate vector	Z_i	X_i
Treatment indicator	Z_{i1}	X_{i1}
Adjusted covariate vector	Z_{i2}	X_{i2}
Regression coefficient vector	$\beta_0 = (\beta_{01}, \beta_{02}')'$	$\theta_0 = (\theta_{01}, \theta_{02}')'$
Regression coefficient for treatment	β_{01}	θ_{01}
Regression coefficient for adjusted covariates	β_{02}	θ_{02}
Covariate vector, treated subject	$Z_i^1 = (1, Z_{i2}')'$	$X_i^1 = (1, X_{i2}')'$
Covariate vector, placebo subject	$Z_i^0 = (0, Z_{i2}')'$	$X_i^0 = (0, X_{i2}')'$
Information matrix	$A(\beta)$	B
Baseline hazard/rate	$d\Lambda_0(t)$	$dR_0(t)$
Hazard/rate	$d\Lambda_i(t) = d\Lambda_0(t)\exp(\beta_0'Z_i)$	$dR_i(t) = dR_0(t) + \theta_0'X_i dt$

Table 4.1: Notation

which we propose estimating by

$$\widehat{\phi}(t) \;=\; \frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t}\widehat{S}(u^{-}|Z_{i}^{0})d\{\widehat{R}(u|X_{i}^{1}) - \widehat{R}(u|X_{i}^{0})\},$$

where $\widehat{R}(t \mid X_i) = \widehat{R}_0(t) + \widehat{\theta}'X_i t$ and $\widehat{R}_0(t) = n^{-1}\sum_{i=1}^{n}\int_{0}^{t}\widehat{\pi}(s)^{-1}Y_i(s)\{dN_i^R(s) - \widehat{\theta}'X_i ds\}$. After some reorganization, the proposed estimator reduces to

$$(4.3) \qquad\qquad \widehat{\phi}(t) \;=\; \widehat{\theta}\, n^{-1}\sum_{i=1}^{n}\int_{0}^{t}\widehat{S}(u^{-}|Z_{i}^{0})du.$$

This measure represents the average difference in the mean number of events between treatment categories while factoring out differences in treatment-specific survival.

The motivation for the proposed estimator is as follows. Typically, a cumulative effect measure, such as the mean number of events is more meaningful to investigators than the effect on an instantaneous quantity such as the rate. This is especially true if the treatment has a time-dependent effect, which is suspected in the motivating example, to be described in Section 4.5. If we compare the marginal means, we will not know if an estimated treatment effect is resulting from treatment-specific differences in the survival distribution, or from treatment-specific differences in the conditional event rate while subjects survive. For example, the marginal recurrent event mean could be greater for treated patients than placebo patients because either (i) conditional on survival the event rates are equal, but treated subjects live longer; or (ii) treated and placebo subjects have approximately equal survival, but $dR(t|X_i^1) > dR(t|X_i^0)$. If we conduct a fully conditional analysis, then integrate in an attempt to obtain a cumulative effect measure, $R(t) = \int_0^t dR_i(u)$ is only interpretable as the mean number of events in the unrealistic setting where death was impossible. The integrated conditional rate has no straightforward interpretation. We are attempting to obtain a meaningful cumulative effect estimator which appropriately acknowledges the occurrence of death, but is not affected by treatment-specific differences in survival. The estimator represents a form of standardization, and the quantity $S(t)$ is intended to represent the survival "average" or "overall" survival. In keeping with the standardization in objective, one could replace $S(t)$ with the sur-

vival for one of the treatment groups since, whatever is used to represent $Pr(D > t)$, the multiplier of $dR(t)$ will be equal across treatment groups.

The proposed procedure to estimate $\phi(t)$ consists of three steps. At the first step, $S(t)$ is estimated. The Kaplan-Meier estimator could be used, but other possibilities include $\widehat{S}(t|Z_i^j)$ for either $j = 0$ or $j = 1$, in which case a proportional hazards model would be fitted. The second step involves fitting the conditional rate model. We do not need to estimate $dR_0(t)$, only the regression coefficient, θ_0, since the baseline rate cancels out. In the third step, $\phi(t)$ is computed as in (4.3).

Note that neither the adjustment covariate effects nor the baseline rate are involved in $\phi(t)$. The quantity is the product of the rate model treatment effect and the placebo group truncated life expectancy (restricted survival time mean). This is the mean difference in the treatment-specific recurrent events, in the scenario where the treatment and placebo survival were equal. Ordinarily, the treatment effect is interpreted as "all other covariates being equal", while here, the interpretation is "all other covariates and survival being equal".

Models for the conditional event rate, $dR_i(t)$, and the terminating event hazard, $d\Lambda_i(t)$, assume covariate effects which are constant over time, t. Two points are important in this regard. First, unlike $\mu_i(t)$, covariate effects on both $dR_i(t)$ and $d\Lambda_i(t)$ will often be constant over time in practice. In a sense, it is preferable to model $dR_i(t)$ and $d\Lambda_i(t)$ separately, since the quantities are more mechanistic than a measure which integrates over their product. The model $d\Lambda_i(t)$ is a standard Cox model frequently employed in biomedical studies, while the quantity $dR_i(t)$ is analogous to the cause-specific hazard in competing risk studies. Second, models (4.1) and (4.2) could always be extended to allow for time-varying effects, and the procedures proposed in this Chapter would still be applicable.

4.2.3 Method II: Difference in Treatment-Specific Means

We now propose a treatment effect measure which is the difference in treatment-specific marginal recurrent event means. The means are given by

$$
\begin{aligned}
\mu_1(t) &= E\left[N_i^R(t)|Z_{i1}=1\right] \\
\mu_0(t) &= E\left[N_i^R(t)|Z_{i1}=0\right].
\end{aligned}
$$

As in method I, we model $dR_i(t)$ and $d\Lambda_i(t)$ separately and average over the observed adjustment covariates. From this perspective, it is useful to write the treatment-specific marginal means as

$$
\begin{aligned}
\mu_j(t) &= E\left[E[N_i^R(t)|Z_i^j, X_i^j]\right] \\
&= E\left[\int_0^t S(r^-|Z_i^j)dR(r|X_i^j)\right],
\end{aligned}
$$

for $j=0,1$, where the outer expectation is taken with respect to the distribution of other adjustment covariates besides the treatment, Z_{i2}. By substituting in survival and conditional rate function estimators, we obtain the proposed treatment-specific mean estimators:

$$
\begin{aligned}
\widehat{\mu}_1(t) &= n^{-1}\sum_{i=1}^n \int_0^t \widehat{S}(r^-|Z_i^1)\{d\widehat{R}_0(r) + (\widehat{\theta}_1 + X_{i2}'\widehat{\theta}_2)dr\} \\
\widehat{\mu}_0(t) &= n^{-1}\sum_{i=1}^n \int_0^t \widehat{S}(r^-|Z_i^0)\{d\widehat{R}_0(r) + X_{i2}'\widehat{\theta}_2 dr\}.
\end{aligned}
$$

The actual treatment effect (as opposed to standardized) on recurrent event numbers can be measured by

$$
\psi(t) = \mu_1(t) - \mu_0(t).
$$

which can be estimated as

$$
\widehat{\psi}(t) = \widehat{\mu}_1(t) - \widehat{\mu}_0(t)
$$

for $t \in (0, \tau]$ using $\widehat{S}(t|Z_i)$ from the proportional hazards model and $d\widehat{R}(t|X_i)$ from the additive conditional rates model.

Method I measures the treatment effects on the recurrent event number if the survival distribution was unaffected by the treatment. Under this hypothetical scenario, the treatment effect on recurrent events is isolated. Method II estimates the true difference in treatment-specific means, while incorporating (rather than factoring out) treatment-specific differences in survival. Thus, $\psi(t)$ combines the actual survival probabilities and recurrent event rates for the treatment and placebo respectively and reflects the difference in marginal recurrent event numbers which would be observed between treatment and placebo patients with the same adjustment covariates.

4.3 Asymptotic Properties

In this section, we describe the essential asymptotic properties of the proposed estimators. We begin by listing the assumed regularity conditions, for $i = 1, \ldots, n$.

(a) $\{N_i^{R*}(\cdot), D_i, C_i, Z_i, X_i\}$ are independent and identically distributed.

(b) $E[dN_i^{R*}(t)|D_i > t, C_i > t, X_i] = E[dN_i^{R*}(t)|D_i > t, X_i]$.

(c) $\lim_{\delta \to 0} \frac{1}{\delta} Pr\{t \le D_i < t + \delta | D_i > t, C_i > t, Z_i\} =$
$\lim_{\delta \to 0} \frac{1}{\delta} Pr\{t \le D_i < t + \delta | D_i > t, Z_i\}$.

(d) $Pr(Y_i(\tau) = 1) > 0$.

(e) $\int_0^\tau d\Lambda_0(t) < \infty$, $\int_0^\tau dR_0(t) < \infty$ and $N_i^R(\tau) < \infty$.

(f) Elements of Z_{i2} and X_{i2} are bounded almost surely.

(g) Positive-definiteness of the matrices, $A(\beta)$ and B, where

$$
\begin{aligned}
A(\beta) &= E\left[\int_0^\tau \{Z_i - \overline{z}(t;\beta)\}^{\otimes 2} Y_i(t) e^{\beta' Z_i} d\Lambda_0(t)\right] \\
\overline{Z}(t;\beta) &\xrightarrow{a.s.} \overline{z}(t;\beta) \\
B &= E\left[\int_0^\tau Y_i(s)\{X_i - \overline{x}(s)\}^{\otimes 2} ds\right] \\
\overline{X}(t) &\xrightarrow{a.s.} \overline{x}(t).
\end{aligned}
$$

Item (a) is the basis for Central Limit Theorem and is usually satisfied; an exception would be clustered data. Conditions (b) and (c) correspond to independent

censoring assumptions for the recurrent and terminating event processes, respectively. Item (d) is a standard identifiability condition.

We assume that there are no unmeasured factors which predict both $dR_i(t)$ and $d\Lambda_i(t)$. This assumption is stronger than the no-unmeasured confounders assumption typically applied in regression analysis. We evaluate the impact of violations of the no-unmeasured-predictors assumption in Section 4.5.

We now describe the essential asymptotic results for the proposed estimators. The proof of each theorem is outlined in the Appendix.

Theorem 1. Under conditions (a) to (g), $\widehat{\phi}(t)$ converges almost surely to $\phi(t)$ for $t \in (0, \tau]$. The proof of Theorem 1 proceeds through a Taylor Series expansion, followed by repeated application of the Uniform Strong Law of Large Numbers (USLLN).

Theorem 2. Under conditions (a) to (g), $n^{\frac{1}{2}}\{\widehat{\phi}(t) - \phi(t)\}$ is converges weakly to a zero-mean Gaussian process with covariance function $E[\gamma_i(s; \beta_0, \theta_0)\gamma_i(t; \beta_0, \theta_0)]$, where

$$
\begin{aligned}
\gamma_i(t; \beta, \theta) &= \gamma_{i1}(t; \beta, \theta) + \gamma_{i2}(t; \beta, \theta) + \gamma_{i3}(t; \beta, \theta) \\
\gamma_{i1}(t; \beta, \theta) &= -\theta_1 E\left[e^{\beta' Z_k^0}\int_0^t S(u^-|Z_k^0)\int_0^u \{Z_k^0 - \bar{z}(r; \beta)\}' d\Lambda_0(r) du\right] A(\beta)^{-1} U_i^D(\beta) \\
\gamma_{i2}(t; \beta, \theta) &= E[e(u^-|Z_k^0)]\{B^{-1}U_i^R(\theta)\}_{1,1} \\
\gamma_{i3}(t; \beta, \theta) &= -\theta_{01}\int_0^t E[e^{\beta' Z_k^0}\{e(t|Z_k^0) - e(r|Z_k^0)\}]\frac{dM_i^D(r; \beta)}{s^{(0)}(r; \beta)},
\end{aligned}
$$

with k indexing a hypothetical subject and with asymptotic score combinations given by

$$
\begin{aligned}
U_i^D(\beta) &= \int_0^\tau \{Z_i - \bar{z}(t; \beta)\} dM_i^D(t; \beta) \\
dM_i^D(t; \beta) &= dN_i^D(t) - Y_i(t)e^{\beta' Z_i} d\Lambda_0(t) \\
s^{(k)}(t; \beta) &= \lim_{n \to \infty} S^{(k)}(t; \beta),
\end{aligned}
$$

for $k = 0, 1, 2$, with the truncated mean lifetime denoted by $e(t|Z_k^0) = \int_0^t S(r|Z_k^0)dr$, and where $\{C\}_{1,1}$ denotes the $(1,1)th$ element of the matrix C.

The proof is done by decomposing $n^{\frac{1}{2}}\{\widehat{\phi}(t) - \phi(t)\}$ into three parts and dealing with each part separately. Taylor series expansions are applied, along with several

applications of the Strong Law of Large Numbers (Sen and Singer 1993). Results from each of the three parts are combined to show that $n^{\frac{1}{2}}\{\widehat{\phi}(t) - \phi(t)\}$ is asymptotically equivalent to $n^{-\frac{1}{2}}\sum_{i=1}^{n}\gamma_i(t;\beta_0,\theta_0)$, which can be shown to converge to a zero-mean Normal distribution for fixed t by the Central Limit Theorem. A demonstration of tightness completes the proof of weak convergence using results from empirical process theory (Pollard, 1990; Van derr Vaart and Wellner, 1996). The covariance function can be consistently estimated by replacing all limiting quantities with their empirical counterparts, then averaging across $Z_{12},\cdots,Z_{n2}$.

Theorem 3. Under conditions (a) to (g), $\widehat{\psi}$ is a consistent estimator of ψ; that is, $\widehat{\psi}(t) \xrightarrow{a.s.} \psi(t)$ for $t \in (0,\tau]$. The proof is similar to that of Theorem 1, at least with respect to the major tools, Taylor Series expansions and the USLLN.

Theorem 4. Under conditions (a) to (g), $n^{\frac{1}{2}}(\widehat{\psi} - \psi)$ converges weakly to a zero-mean Gaussian process with covariance function $E[\{\xi_{i1\cdot}(s;\beta_0,\theta_0) - \xi_{i0\cdot}(s;\beta_0,\theta_0)\}\{\xi_{i1\cdot}(t;\beta_0,\theta_0) - \xi_{i0\cdot}(t;\beta_0,\theta_0)\}]$, where

$$
\begin{aligned}
\xi_{ij\cdot}(t;\beta,\theta) &= \xi_{i11}(t;\beta,\theta) + \xi_{i12}(t;\beta,\theta) + \xi_{i13}(t;\beta,\theta) + \xi_{i14}(t;\beta,\theta)\\
\xi_{ij1}(t;\beta,\theta) &= -E\left[e^{\beta' Z_k^0}\int_0^t S(u^-|Z_k^0)\int_0^u \{Z_k^0 - \overline{z}(r;\beta)\}' d\Lambda_0(r)dR(u|X_k^j)\right]A(\beta)^{-1}U_i^D(\beta)\\
\xi_{ij2}(t;\beta,\theta) &= E\left[\int_0^t S(u^-|Z_i^j)\{X_i^j - \overline{x}(u)\}' du\right]B^{-1}U_i^R(\theta)\\
\xi_{ij3}(t;\beta,\theta) &= \int_0^t E\{S(u^-|Z_i^j)\}\pi(u)^{-1}dM_i^R(u;\theta)\\
\xi_{ij4}(t;\beta,\theta) &= -\int_0^t E\left[e^{\beta' Z_i^j}\{\mu(t|Z_i^j,X_i^j) - \mu(r|Z_i^j,X_i^j)\}\right]\frac{dM_i^D(r;\beta)}{s^{(0)}(r;\beta)}.
\end{aligned}
$$

Here

$$
\begin{aligned}
U_i^R(\theta) &= \int_0^t \{X_i - \overline{x}(u)\}dM_i^R(u;\theta)\\
M_i^R(t;\theta) &= N_i^R(t) - \int_0^t Y_i(u)\{dR_0(u) + \theta' X_i du\}.
\end{aligned}
$$

Since $n^{\frac{1}{2}}(\widehat{\psi}-\psi) = n^{\frac{1}{2}}(\widehat{\mu}_1-\mu_1) - n^{\frac{1}{2}}(\widehat{\mu}_0-\mu_0)$, we work on $n^{\frac{1}{2}}(\widehat{\mu}_1-\mu_1)$ and $n^{\frac{1}{2}}(\widehat{\mu}_0-\mu_0)$ separately as follows:

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\mu}_j(t;\widehat{\beta},\widehat{\theta}) - \mu_j(t)\} &= n^{\frac{1}{2}}\{\widehat{\mu}_j(t;\widehat{\beta},\widehat{\theta}) - \widehat{\mu}_j(t;\beta_0,\widehat{\theta})\} \\
&\quad + n^{\frac{1}{2}}\{\widehat{\mu}_j(t;\beta_0,\widehat{\theta}) - \widehat{\mu}_j(t;\beta_0,\theta_0)\} \\
&\quad + n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t \widehat{S}(r^-|Z_i^j)\{d\widehat{R}_0(r;\theta_0) - dR_0(r)\} \\
&\quad + n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t \{\widehat{S}(r^-;\beta_0|Z_i^j) - S(r^-|Z_i^j)\}\{dR_0(r) + \theta_0' X_i^j dr\}.
\end{aligned}
$$

for $j = 0, 1$. The major steps of the proof are then similar to those of the proof of Theorem 2, although there are several more steps. The $\xi_{ij\cdot}(s;\beta_0,\theta_0)$ quantities can be consistently estimated by replacing all limiting values by their empirical counterparts, then averaging across $i = 1, \cdots, n$.

4.4 Simulation Study

For each data configuration, we generated $n = 200$ independent and identically distributed subjects with both terminating and recurrent events. The terminating event hazard follows the following proportional hazards model,

$$
d\Lambda(t \mid Z_i) = d\Lambda_0 e^{\beta_1 Z_{i1} + \beta_2 Z_{i2}},
$$

where Z_{i1} (treatment) is distributed as Bernoulli (0.5), the adjustment covariate Z_{i2} follows a Uniform $(0, 10)$ distribution and $\lambda_0 = 0.04$. We set the coefficient β_1 at 0.5 and 1, to examine scenarios of low or high treatment effect on survival. Censoring is uniformly distributed on $(0, 20)$ which leads to an average censoring percentage of approximately 42%. The recurrent events follow a Poisson process, with rate function

$$
dR(t \mid X_i, Q_i) = dR_0 + Q_i dt + \theta' X_i dt,
$$

where Q_i follows a Gamma distribution with mean 0.25 and a variance of 0.5 and 1. The Q_i variate represents a frailty term shared by all the recurrent event times for the same subject and may be thought of as an unmeasured predictor. Note that the

assumptions of our proposed methods hold under this set-up. The above model was simulated by generating gap time between events as:

$$T_{i,j+1} = T_{i,j} - \log(U_{ij})\{dR_0 + Q_i + \theta'X_i\}^{-1},$$

for $j = 0, \cdots, 50$, where $U_{ij} \sim \text{Unif}(0, 1)$, $X_i = Z_{i1}$ and $T_{i,0} \equiv 0$. We varied the baseline recurrent event rate from 0.125 to 0.25. The covariate X_i is the same as Z_{i1} in the proportional hazards model for the terminating event, both representing the treatment or exposure of interest. The regression coefficient for X_i is set at $\theta = 0.5$.

Table 4.2 lists the performance of our proposed estimators $\widehat{\psi}(t)$ and $\widehat{\phi}(t)$ in eight scenarios of different β_1, $V(Q_i)$ and dR_0 combinations. The average observed number of recurrent events per subject ranges from 3.0 to 4.4. Three evenly spaced time points 5, 10 and 15 are picked to examine the performance of the estimators at early, middle and late follow-up times. In all the tested conditions, $\widehat{\psi}(t)$ and $\widehat{\phi}(t)$ are very close to their true values (obtained by numerical integration), and average asymptotic standard errors (ASE) agree well with empirical standard deviations (ESD). Correspondingly, empirical coverage probabilities (CP) are close to the nominal value of 0.95. Overall, the consistency, asymptotic normality and variance estimator appear to work very well in moderate size samples.

Next, we evaluated the sensitivity of our proposed methods to the no-unmeasured-predictors assumption. Specifically, we set up a model with the gamma frailty, Q_i, affecting both $d\Lambda_i(t)$ and $dR_i(t)$, in violation of our underlying assumptions. The simulated proportional hazards model (4.1) and the additive recurrent rate model (4.2) change to

$$
\begin{aligned}
d\Lambda(t \mid Z_i) &= Q_i e^{\beta'Z_i} d\Lambda_0 \\
dR(t \mid X_i) &= dR_0 + (Q_i + \theta'X_i)dt,
\end{aligned}
$$

respectively. In addition to the scenarios examined in the simulation study of the correct model, an extreme case where $V(Q_i) = 2$, which is 8 times $E(Q_i)$, was also examined to assess the impact of highly variable frailty terms on the performance of $\widehat{\psi}(t)$ and $\widehat{\phi}(t)$.

β_1	$V(Q_i)$	dR_0	t	Method I: $\widehat{\phi}(t)$					Method II: $\widehat{\psi}(t)$				
				$\phi(t)$	bias	ESD	ASE	CP	$\psi(t)$	bias	ESD	ASE	CP
1	0.5	0.25	5	2.12	0.03	0.38	0.38	0.96	1.16	0.02	0.34	0.34	0.95
			10	3.64	0.04	0.66	0.65	0.96	1.01	0.05	0.57	0.56	0.95
			15	4.73	0.07	0.87	0.85	0.96	0.53	0.06	0.74	0.74	0.95
		0.125	5	2.12	-0.00	0.38	0.37	0.95	1.28	0.01	0.34	0.33	0.95
			10	3.64	-0.01	0.66	0.64	0.95	1.34	0.02	0.54	0.54	0.95
			15	4.73	0.01	0.87	0.84	0.95	1.05	0.04	0.70	0.71	0.95
	1	0.25	5	2.12	0.01	0.50	0.48	0.95	1.16	-0.01	0.45	0.43	0.95
			10	3.64	0.02	0.87	0.83	0.95	1.01	0.00	0.74	0.710	0.94
			15	4.73	0.04	1.14	1.08	0.95	0.53	0.03	0.96	0.91	0.94
		0.125	5	2.12	0.04	0.47	0.47	0.96	1.28	0.00	0.44	0.43	0.96
			10	3.64	0.06	0.81	0.82	0.95	1.34	0.02	0.71	0.69	0.95
			15	4.73	0.09	1.07	1.07	0.96	1.05	0.05	0.89	0.89	0.95
0.5	0.5	0.25	5	2.12	0.00	0.36	0.36	0.96	1.71	-0.02	0.36	0.35	0.96
			10	3.64	-0.00	0.61	0.63	0.96	2.42	-0.03	0.62	0.61	0.96
			15	4.73	0.01	0.81	0.82	0.96	2.65	-0.03	0.83	0.84	0.97
		0.125	5	2.12	0.00	0.35	0.35	0.95	1.76	-0.02	0.33	0.34	0.96
			10	3.64	0.00	0.61	0.61	0.94	2.58	-0.04	0.58	0.59	0.96
			15	4.73	0.01	0.81	0.80	0.95	2.91	-0.03	0.77	0.80	0.97
	1	0.25	5	2.12	-0.00	0.47	0.46	0.96	1.71	0.02	0.47	0.45	0.95
			10	3.64	-0.00	0.80	0.79	0.96	2.42	0.03	0.80	0.77	0.95
			15	4.73	0.01	1.06	1.03	0.96	2.65	0.05	1.04	1.02	0.95
		0.125	5	2.12	-0.00	0.47	0.46	0.96	1.76	-0.03	0.47	0.44	0.95
			10	3.64	0.01	0.81	0.78	0.96	2.58	-0.05	0.80	0.75	0.95
			15	4.73	0.01	1.06	1.03	0.96	2.91	-0.04	1.05	0.99	0.95

ESD = empirical standard deviation

ASE = average asymptotic standard error

CP = coverage probability.

Table 4.2: Simulation Results: Performance of Proposed Estimators

$V(Q_i)$	β_1	dR_0	Method I: $\widehat{\phi}(t)$					Method II: $\widehat{\psi}(t)$				
			$\phi(t)$	bias	ESD	ASE	CP	$\psi(t)$	bias	ESD	ASE	CP
2	1	0.25	4.42	-0.04	0.35	0.35	0.93	4.18	0.01	0.39	0.38	0.95
		0.125	4.42	-0.06	0.31	0.31	0.95	4.22	0.01	0.33	0.33	0.94
	0.5	0.25	4.41	-0.01	0.35	0.35	0.94	4.30	0.00	0.37	0.38	0.94
		0.125	4.42	-0.00	0.32	0.31	0.95	4.31	0.01	0.33	0.32	0.94
1	1	0.25	4.09	-0.09	0.35	0.35	0.94	3.65	0.04	0.43	0.41	0.94
		0.125	4.08	-0.08	0.32	0.33	0.93	3.71	0.03	0.35	0.34	0.95
	0.5	0.25	4.09	-0.06	0.37	0.36	0.94	3.86	0.03	0.41	0.40	0.95
		0.125	4.08	-0.03	0.33	0.33	0.95	3.88	0.05	0.34	0.34	0.96
0.5	1	0.25	3.62	-0.11	0.37	0.36	0.93	2.91	0.01	0.44	0.43	0.94
		0.125	3.62	-0.12	0.33	0.34	0.93	3.01	0.04	0.37	0.36	0.94
	0.5	0.25	3.61	-0.05	0.37	0.36	0.94	3.25	0.01	0.43	0.42	0.94
		0.125	3.62	-0.08	0.35	0.34	0.93	3.30	-0.01	0.35	0.35	0.96

Estimators are evaluated at $t = 10$, the mean of the censoring distribution.

Table 4.3: Simulation Results:Robustness of Proposed Estimators under a Misspecified Model

Table 4.3 demonstrates the robustness of our estimators under this misspecified model. The estimators are slightly biased, while the proposed robust standard error estimators are close to the empirical standard deviations and hence coverage probabilities are still close to 0.95. In summary, the proposed methods appear fairly robust to the no-unmeasured-predictors assumption.

4.5 Application

We applied our two proposed estimators to the study of renal transplant patients. The goal of our study is to compare the mean number of hospitalizations for transplant recipients with an Expanded Criteria Donor (ECD) kidney and patients transplanted with a non-ECD kidney. An ECD is defined as a deceased donor either (i) age ≥ 60, or (ii) age $50 - 59$ and with at least two of the following three characteristics: hypertensive, serum creatinine concentration $> 1.5\text{mg/dl}$, or death due to stroke (Port et al., 2002). The ECD $(0, 1)$ classification is a well-accepted quality index for donated kidneys in the nephrology community. In fact, wait-listed patients willing to accept an ECD kidney are essentially listed separately and generally have

reduced waiting time until transplant. The relative frequency of ECD transplantation has increased over time, and several authors have examined the impact of an ECD on the post-transplant death hazard. Our analysis targets the impact of ECD transplantation on the mean number of post-transplant hospitalizations. Hospitalization is a composite index of patients' health status and resource utilization, and therefore serves as a concrete and objective measure of the post-transplant performance.

Patients began follow-up ($t = 0$) at the time of kidney transplant. Some received an ECD transplant, while others got a non-ECD kidney. Patients who receive a kidney transplant are subject to a greatly increased mortality hazard and hospitalization rate in the weeks immediately following the transplant, due to the risk of surgical complications, which are suspected to be more serious in ECD recipients. From a public health perspective, it is of interest to compare ECD and non-ECD transplanted patients with respect to mean number of hospitalizations. Due to the strong time-dependence in the effect of ECD kidney transplantation, the instantaneous effect on the hospitalization rate is generally of less interest than the cumulative effect. Moreover, survival probabilities are known to be reduced for ECD relative to non-ECD recipients (Port et al., 2002). Given these facts, proportionality or additivity is not expected to hold with respect to the mean function, meaning that estimating the ECD effect by fitting a multiplicative or additive model to the hospitalization mean directly would be inappropriate.

The goals of the investigators exemplify some of the limitations in the existing methods for analyzing recurrent/terminal event data. For example, in comparing the mean number of events (averaging across the survival experience), it is possible that non-ECD transplant recipients experience more hospitalizations than ECD patients simply because they survive much longer than ECD patients. If we were to compare the conditional hospitalization rate given survival, as did in Schaubel and Cai (2005), the transplant effect on the rate function, $dR_i(t)$, can be determined. However, $\int_0^t dR_i(s)$ can only be interpreted as the mean number of events in the setting where death were impossible. Using the proposed method, we can obtain the difference between the number of hospitalizations for ECD and non-ECD recipients

either considering the true different survival distributions in each type of transplant or factoring out differences in survival between the two transplant types.

We combine demographic, clinical and mortality data from the Scientific Registry of Transplant Recipients (SRTR) and hospitalization history information from Centers for Medicare and Medicaid Services (CMS). The study population is restricted to patients whose primary payer was Medicare. To increase homogeneity, we also exclude from our target population repeat and multi-organ transplant recipients. All Medicare patients age $\geq$ 18 who received a kidney transplant from a deceased donor in year 2000 were included in our study sample. In total, 3,816 recipients with follow-up and complete covariate information were included and tracked from the time of transplant until death, loss to follow-up, or at the end of the observation period (December 31, 2005). Among the 3,816 patients, 970 (25.42%) were observed to die during the 6-year follow-up period, with the remaining 2,846 (74.58%) recipients censored, either due to loss to follow-up or administratively at 12/31/2005. On average, each recipient experienced 0.85 hospitalizations during the follow-up period.

As stated, the treatment of interest is donor source: ECD ($Z_{i1} = 1$) or non-ECD ($Z_{i1} = 0$). We adjusted the same set of covariates in the proportional hazards and additive rates models: candidate demographics (gender, age, race), pre-transplant years on dialysis, various comorbid conditions (drug treated chronic obstructive pulmonary disease (COPD), angina, symptomatic cerebral vascular disease, symptomatic peripheral vascular disease, pretransplant malignancy), primary diagnosis (polycystic kidney disease, diabetes, hypertension), functional status (fully active to severely disabled), and stay in the intensive care unit (ICU).

Hazard ratios and recurrent event rate differences for each covariate are listed in Table 4.4, as well as their corresponding p-values. Recipients of an ECD kidney have a 1.3 times higher hazard of death compared to non-ECD recipients. At the same time, ECD kidney recipients experienced 84.4 more hospital admissions per 1,000 patients per year survived. Lower survival probabilities for ECD patients could lead to a reduced mean number of hospitalizations, since hospitalizations are terminated by death. However, ECD patients also experience a significantly elevated conditional

Covariate, $Z_{ik} = X_{ik}$	$e^{\widehat{\beta}_k}$	p	$\widehat{\theta}_k$	$\widehat{SE}(\widehat{\theta}_k)$	p
ECD	1.28	0.002	84.4	29.4	0.004
Female	1.00	0.951	-6.6	19.2	0.732
Age 18 - 24	0.34	0.003	-80.7	41.1	0.045
Age 25 - 34	0.55	0.0001	-87.6	29.0	0.003
Age 35 - 44	0.83	0.092	-56.2	26.4	0.033
Age 45 - 54	1	-	0	-	-
Age 55 - 64	1.62	< 0.0001	50.3	28.9	0.082
Age 65 - 70	2.12	< 0.0001	109.5	40.9	0.007
Age $\geq$ 70	2.88	< 0.0001	146.5	47.8	0.002
African American (vs. Caucasian)	0.99	0.908	5.7	24.3	0.813
COPD	1.25	0.457	7.1	136.5	0.960
Angina	1.40	0.0002	71.1	39.5	0.072
Pretransplant Malignancy	2.67	0.092	750.7	517.9	0.147
Cerebral vascular disease	1.19	0.323	95.4	71.5	0.182
Peripheral vascular disease	1.03	0.848	31.7	56.1	0.571
Polycystic kidneys	0.56	0.0006	-138.7	22.7	< 0.001
Diabetes	1.59	< 0.0001	123.3	32.2	< 0.001
Hypertension	1.19	0.067	16.5	24.3	0.500
Years on dialysis	1.04	0.002	-0.3	3.0	0.932
Functional status: minor disability	0.31	< 0.0001	-114.2	106.3	0.282
Intensive Care Unit	1.24	0.631	-220.1	136.2	0.106

θ: additional hospitalizations per 1,000 per year survived.

Table 4.4: Analysis of SRTR Data Estimated Regression Parameters from Proportional Hazards and Additive Rates Models

hospitalization rate given survival. The combination of these two effects results in the marginal effect of ECD on mean hospitalizations.

Based on Method II, ECD recipients on average experience 79 more hospital admissions per 1000 persons compared to a non-ECD recipients regardless the survival status at the end of one year. This difference increased with more post-transplant years and reaches 318 per 1,000 patients at the end of five years. At each time point, the difference between ECD transplant patients and non-ECD transplant patients is highly significant. The isolated ECD effect on hospitalization numbers are 81, 158, 233, 304, 372 per 1,000 patients at the end of one to five years. Take five years
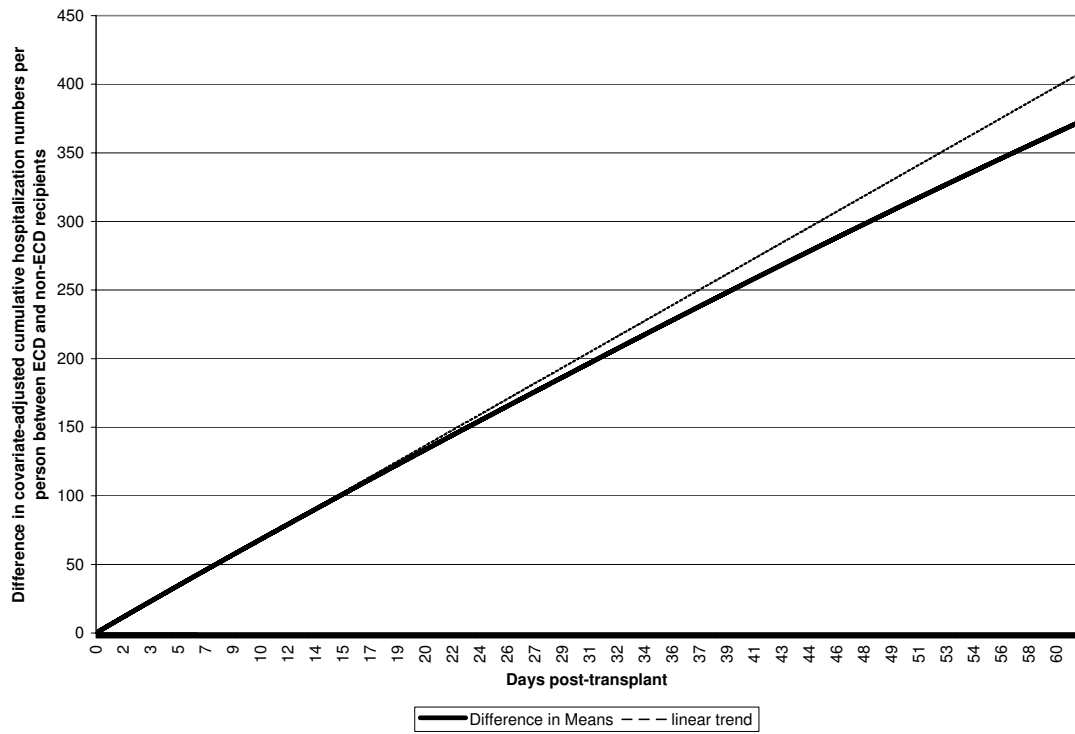
t	Method I			Method II		
	$\widehat{\phi}(t)$	$\widehat{SE}\{\widehat{\phi}(t)\}$	p	$\widehat{\psi}(t)$	$\widehat{SE}\{\widehat{\psi}(t)\}$	p
1 year	0.081	0.028	0.004	0.079	0.028	0.005
2 year	0.158	0.055	0.004	0.153	0.054	0.005
3 year	0.233	0.081	0.004	0.217	0.079	0.006
4 year	0.304	0.106	0.004	0.270	0.104	0.010
5 year	0.372	0.130	0.004	0.318	0.128	0.014

Table 4.5: Analysis of Kidney Transplant Data (ECD vs. non-ECD) Difference in Mean Numbers of Hospitalizations

as an example, factoring out the differences between ECD and non-ECD survival, an ECD recipient would experience 0.372 more hospitalizations in five years after transplant. All the differences at the selected time points are highly significant. In summary, receiving an ECD kidney leads to significantly higher hospitalizations post-transplant.

The difference in average hospitalization numbers for an ECD recipient compared to a non-ECD recipient with the same covariates pattern (i.e., $\widehat{\psi}(t)$) is plotted in Figure 4.1. We can see that ECD recipients have more hospitalizations immediately after transplant. The difference keeps increasing with time and reaches 363 hospital admissions per 1,000 patients at the end of the 5-year follow-up period. The increase in hospitalization rate due to ECD dominates the decrease in survival probability for ECD recipients and leads to positive ECD effects on mean hospitalizations. As Figures 4.1 and 4.2 illustrate, this positive difference accumulates over time.

It would be possible to set up a procedure to compute 95% confidence bands analogous to that proposed by Lin, Wei and Fleming (1994). However, due to the number of patients and event times, such a procedure is computationally inconvenient. As such, we propose a bootstrap method for calculating simultaneous confidence bands for the difference ψ over time. First, a bootstrap sample of the same size ($n = 3,816$) is selected by random sampling with replacement. Second, the difference $\widetilde{\psi}(t)$ is cal-

linear trend: The dashed lines in Figure 1-2 represents the difference in means which would result if the difference across the $(0, 400]$ day interval persisted across the $(400, 1825]$ day interval.

Figure 4.1: ECD Effects on Mean Number of Hospitalizations - Method I
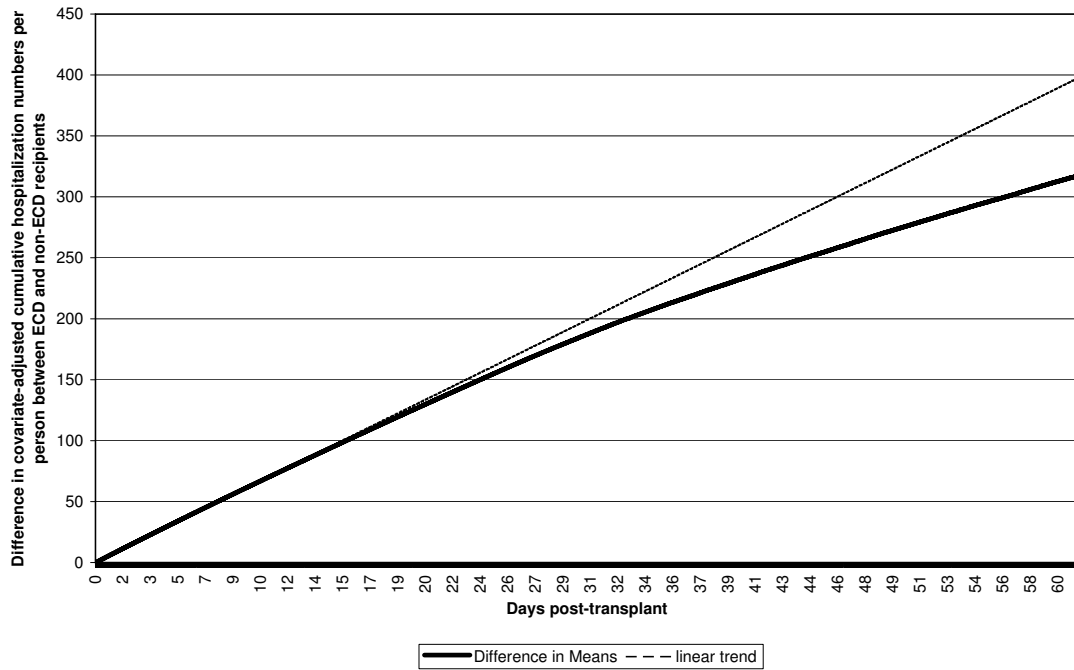
linear trend: The dashed lines in Figure 1-2 represents the difference in means which would result if the difference across the $(0, 400]$ day interval persisted across the $(400, 1825]$ day interval.

Figure 4.2: ECD Effects on Mean Number of Hospitalizations - Method II

culated for each day of the follow-up period because the cumulative hospitalization rate changes by day and survival probability drops at each death time. Third, find $\eta = \sup(|\widetilde{\psi}(t) - \psi(t)|)$. Here, $\psi(t)$ is estimated by $\widehat{\psi}(t)$ from the original data set. Then, the same procedure is repeated for 200 times and an empirical distribution of the η is obtained. Finally, the 95th quantile of η, $\widetilde{q}_{0.95}$, is computed through this empirical distribution, following which confidence bands are calculated as $\widehat{\psi}(t) \pm \widetilde{q}_{0.95}$.

¿From Figure 4.3 and 4.4, confidence bands are wider than the pointwise confidence interval since confidence bands represent the variation of the supremum over five years of follow up. Also by comparing Figure 4.3 and 4.4, $\widehat{\phi}(t)$ from Method I has better precision and narrower confidence bands possibly because estimator $\widehat{\phi}(t)$ employs a standard survival distribution, $d\widehat{R}_0(u)$ and $\widehat{\theta}'_{i2}Z_{i2}$ cancel out in taking the difference of $\int_0^t \widehat{S}(u; \widehat{\beta}|Z_i^0)d\widehat{R}_i(u; \widehat{\theta}|X_i^1)$ and $\int_0^t \widehat{S}(u; \widehat{\beta}|Z_i^0)d\widehat{R}_i(u; \widehat{\theta}|X_i^0)$. The isolated treatment effect on hospitalization means is highly significant at the end of $1-5$ years. But the process $\widehat{\phi}(t)$ is not significantly different from zero in the first 39 months post transplant, and the difference becomes significant after the 39th month. In Method II, although post-transplant hospitalization means in ECD and non-ECD transplant patients, $\widehat{\mu}_1(t)$ and $\widehat{\mu}_0(t)$, differ significantly pointwise, the process $\widehat{\psi}(t)$ is not significantly different from zero over five years post transplant.

4.6 Discussion

We propose semiparametric methods to compare marginal treatment-specific mean number of recurrent events in the presence of a terminating event. The proposed methods involve modeling terminating event survival probability and the conditional recurrent event rate given survival separately, then integrating to estimate treatment-specific marginal recurrent event means. Two different effect measures are proposed, with their asymptotic properties derived and evaluated in moderate size samples. We demonstrate that these two estimators work reasonably well under misspecified models; that is, models failing to incorporate unmeasured predictors of both the terminating event hazard function and conditional recurrent event rate. Both estimators are applied to national kidney transplant data to study the effect of ECD
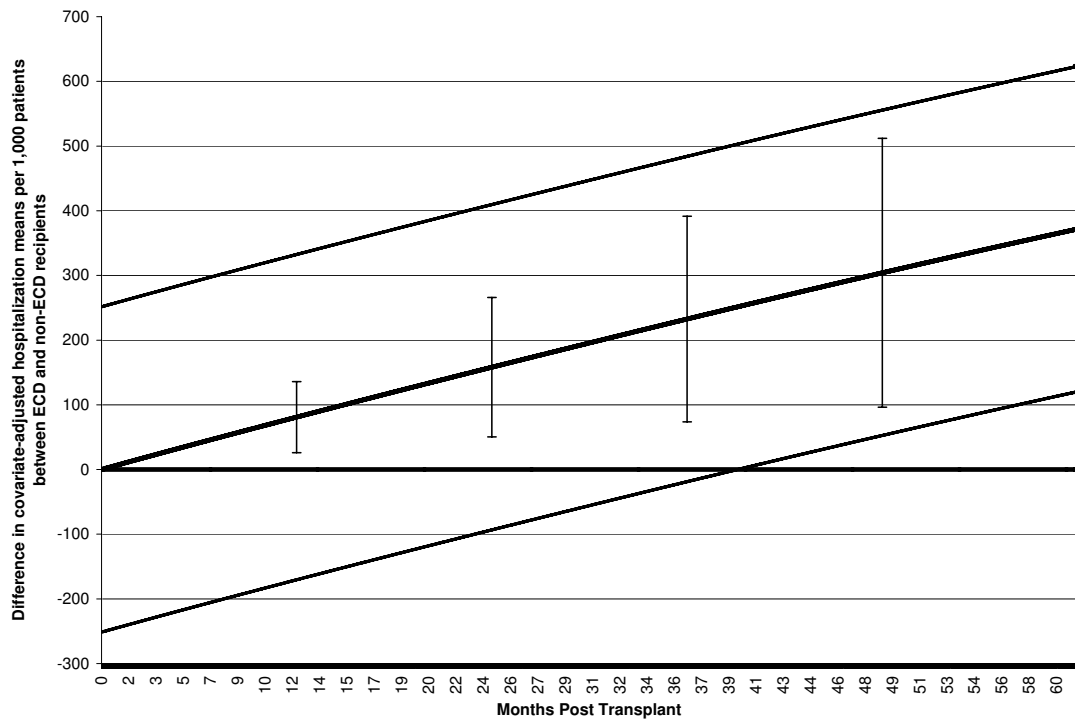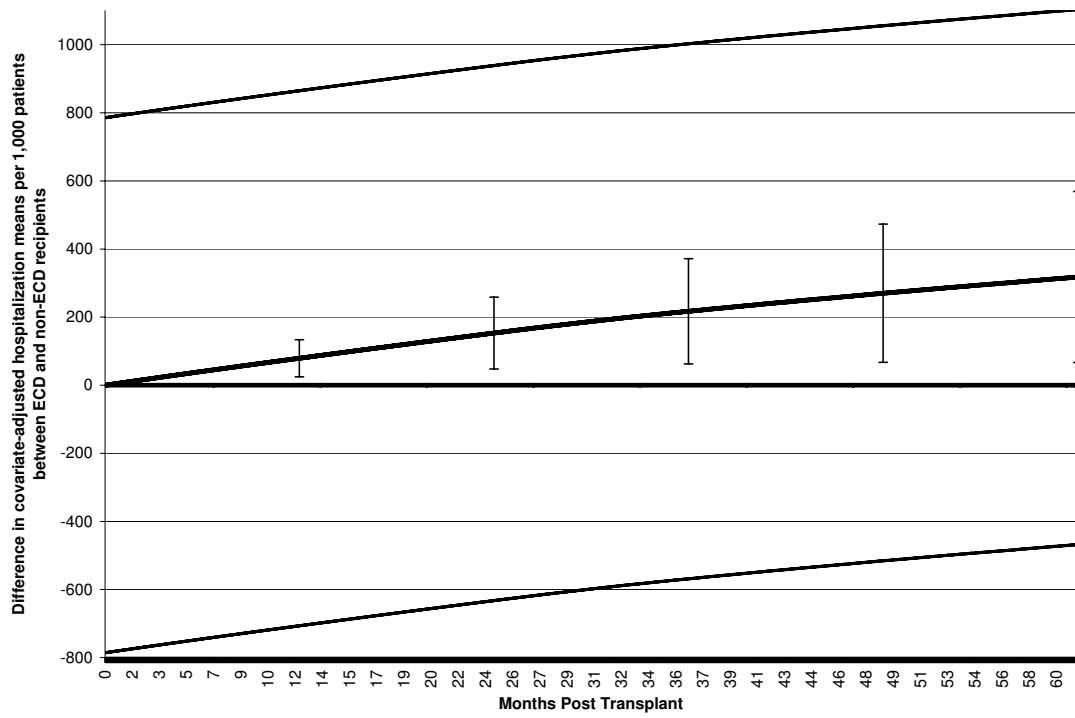
The two outside lines are 95% confidence bands for the year 0-5 interval.

The five vertical bars represent 95% pointwise confidence intervals at the end of year $1, \cdots, 5$.

Figure 4.3: 95% Confidence Bands and Pointwise Confidence Intervals for the Difference between ECD and non-ECD recipients in Hospitalization Means by Method I

The two outside lines are 95% confidence bands for the year 0-5 interval.

The five vertical bars represent 95% pointwise confidence intervals at the end of year $1, \cdots, 5$.

Figure 4.4: 95% Confidence Bands and Pointwise Confidence Intervals for the Difference between ECD and non-ECD recipients in Hospitalization Means by Method II

transplantation on post-transplant hospitalization admission numbers. ECD recipients are found to have consistently more hospitalizations during the whole follow-up period, with the difference between ECD and non-ECD recipients increasing with time.

There are several possible variations to the proposed methods; for example, a multiplicative recurrent event rate model, or nonparametric estimators for the terminating event survival probabilities. In practice, the appropriateness of the chosen models depends on the degree to which its assumptions suit the data and its interpretation answers the question of interest. Our choice of the proportional hazards and additive rates models is a combination of appropriate model and straightforward interpretation.

Both proposed estimators, $\widehat{\phi}(t)$ and $\widehat{\psi}(t)$, are asymptotically normal, with explicit variance estimators, and one can construct point-wise confidence intervals and simultaneous confidence bands for each. However, $\widehat{\phi}(t)$ and $\widehat{\psi}(t)$ have different interpretations and are answering different questions. For investigators interested in the actual difference between group-specific recurrent event means, $\widehat{\psi}(t)$ would be preferred. On occasions where investigators wish to isolate the treatment effect on the recurrent events, $\widehat{\phi}(t)$ provides a standardized difference between treatment-specific recurrent event means, since it factors out the treatment-specific differences in survival probabilities. Although such a comparison is hypothetical, standardized comparisons are widely used, with the common objective of describing a single effect in a hypothetical setting wherein all other factors are equal.

Our strongest assumption is that there are no unmeasured predictors of both the terminating event hazard and conditional event rate. In observational settings, this assumption will fail frequently. Several issues are important in this regard. First, the unmeasured predictor must be a risk factor for both the terminating and recurrent event conditional on all measured covariates. If the most important predictors are included in the adjustment covariate vector, bias may be minimized and the estimated treatment effect may be a good approximation to the reality. Second, the most popular alternative to a no-unmeasured-predictor assumption is to incorporate a frailty.

However, most frailty approaches for recurrent event data employed assume that the events follow a Poisson process, which is restrictive in its own right. To the best of our knowledge, there is only one method (Ye, Kalbfleisch and Schaubel; unpublished manuscript) which employs a frailty in the absence of the Poisson-process assumption, and that method does not propose cumulative effect measures. Third, our sensitivity analysis reveals that if an unmeasured frailty affects both the terminating event hazard and conditional rate, bias is quite low.

4.7   Appendix

Proof of Theorem 1: Under conditions (a) to (g), we have essentially the same framework as the additive rates model of Schaubel et al. (2006), with the at risk indicators redefined. As such, the almost sure convergence of $\widehat{\theta}$ to θ_0 holds from Theorem 2 of Schaubel et al. (2006). Therefore, by the Cramer-Wold Theorem (Sen and Singer, 1993), the first element of $\widehat{\theta}$, denoted by $\widehat{\theta}_1$, is a consistent estimator for θ_{01}.

Since $\widehat{\beta} \xrightarrow{a.s.} \beta_0$ and $\widehat{\Lambda}_0(t) \xrightarrow{a.s.} \Lambda_0(t)$ for all $t \in (0, \tau]$ (Andersen and Gill, 1982), by the continuous mapping theorem, $\widehat{S}(u^-|Z_i) \xrightarrow{a.s.} S(u^-|Z_i)$ and hence $\widehat{\phi}(t; \widehat{\theta}, \widehat{\beta}) \xrightarrow{a.s.} \widehat{\phi}(t; \theta_0, \beta_0)$. Finally, the empirical average taken in $\widehat{\phi}(t)$ converges to its expectation for $t \in (0, \tau]$ by the Uniform Strong Law of Large Numbers (USLLN; Pollard, 1990).

Proof of Theorem 2: We can decompose $n^{\frac{1}{2}}\{\widehat{\phi}(t; \widehat{\beta}, \widehat{\theta}) - \phi(t)\}$ into three parts as follows

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\phi}(t; \widehat{\beta}, \widehat{\theta}) - \phi(t)\} &= n^{\frac{1}{2}}\{\widehat{\phi}(t; \widehat{\beta}, \widehat{\theta}) - \widehat{\phi}(t; \beta_0, \widehat{\theta})\} \\
&+ n^{\frac{1}{2}}\{\widehat{\phi}(t; \beta_0, \widehat{\theta}) - \widehat{\phi}(t; \beta_0, \theta_0)\} \\
&+ n^{\frac{1}{2}}\{\widehat{\phi}(t; \beta_0, \theta_0) - \phi_0(t)\}.
\end{aligned}
$$

(4.4)

For the first part of (4.4), applying a Taylor expansion around β_0, we get

$$n^{\frac{1}{2}}\{\widehat{\phi}(t;\widehat{\beta},\widehat{\theta}) - \widehat{\phi}(t;\beta_0,\widehat{\theta})\} = \widehat{\theta}_1 n^{-1}\sum_{i=1}^{n}\int_{0}^{t}\left.\frac{\partial S(u^-;\beta,\widehat{\theta}|Z_i^0)}{\partial \beta'}\right|_{\beta=\beta_*}n^{\frac{1}{2}}(\widehat{\beta}-\beta_0)du$$

$$(4.5) \qquad = -\widehat{\theta}_1 E\left[\int_{0}^{t}S(u^-|Z_i^0)e^{\widehat{\beta}'Z_i^0}\int_{o}^{u}\{Z_i^0-\bar{z}(s)\}d\Lambda_0(s)\right]n^{\frac{1}{2}}(\widehat{\beta}-\beta_0)du$$

as $n \to \infty$, where $N^D(u) = \sum_{1}^{n}N_i^D(u)$. Furthermore, through another Taylor expansion,

$$(4.6) \qquad n^{\frac{1}{2}}(\widehat{\beta}-\beta_0) = A(\beta_0)^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}U_i^D(\beta_0).$$

as $n \to \infty$. Combining (4.5) and (4.6),

$$(4.7) \qquad n^{\frac{1}{2}}\{\widehat{\phi}(t;\widehat{\beta},\widehat{\theta}) - \widehat{\phi}(t;\beta_0,\widehat{\theta})\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\gamma_{i1}(t;\beta_0,\theta_0)$$

as $n \to \infty$.

For the second term of (4.4),

$$n^{\frac{1}{2}}\{\widehat{\phi}(t;\beta_0,\widehat{\theta}) - \widehat{\phi}(t;\beta_0,\theta_0)\} = n^{-1}\sum_{i=1}^{n}\int_{0}^{t}\widehat{S}(u^-|Z_i^0)du\,n^{\frac{1}{2}}(\widehat{\theta}_1-\theta_{01}).$$

The difference between $\widehat{\theta}$ and θ_0 can be expressed as sum of independent identical contributions from each subject (Schaubel et al. 2006). When $n \to \infty$,

$$n^{\frac{1}{2}}(\widehat{\theta}-\theta_0) = B^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}U_i^R(\theta_0),$$

while $n^{\frac{1}{2}}(\widehat{\theta}_1-\theta_{01})$ corresponds to the first element of $n^{\frac{1}{2}}(\widehat{\theta}-\theta_0)$; that is,

$$(4.8) \qquad n^{\frac{1}{2}}(\widehat{\theta}_1-\theta_{01}) = \{B^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}U_i^R(\theta_0)\}_{1,1}$$

as $n \to \infty$. By continuous mapping, $\widehat{S}(u^-|Z_i^0) \xrightarrow{a.s.} S(u^-|Z_i^0)$. Through the USLLN,

$$(4.9) \qquad n^{-1}\sum_{i=1}^{n}\int_{0}^{t}\widehat{S}(u^-|Z_i^0)du \xrightarrow{a.s.} E[\int_{0}^{t}S(u^-|Z_i^0)du].$$

Combining (4.8) and (4.9), we get

$$(4.10) \qquad n^{\frac{1}{2}}\{\widehat{\phi}(t;\beta_0,\widehat{\theta}) - \widehat{\phi}(t;\beta_0,\theta_0)\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\gamma_{i2}(t;\beta_0,\theta_0)$$

as $n \to \infty$.

For the third term of (4.4), as $n \to \infty$,

$$
\begin{aligned}
n^{\frac{1}{2}}\widehat{\phi}(t;\beta_0,\theta_0) - \phi_0(t) &= \theta_{01}\, n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t \{\widehat{S}(u^-;\beta_0|Z_i^0) - S(u^-|Z_i^0)\}du \\
&= \theta_{01}\, n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t [\exp\{-\widehat{\Lambda}_0(u;\beta_0)e^{\beta_0'Z_i^0}\} - \exp\{-\Lambda_0(u)e^{\beta_0'Z_i^0}\}]du \\
&= \theta_{01}\, n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t -S(u^-|Z_i^0)e^{\beta_0'Z_i^0}\{\widehat{\Lambda}_0(u;\beta_0) - \Lambda_0(u)\}du \\
&= -\theta_{01}n^{-1}\sum_{i=1}^{n}\int_0^t \{S(u|Z_i^0)e^{\beta_0'Z_i^0}\}n^{\frac{1}{2}}\{\widehat{\Lambda}_0(u;\beta_0) - \Lambda_0(u)\}du.
\end{aligned}
$$

while

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\Lambda}_0(u;\beta_0) - \Lambda_0(u)\} &= n^{\frac{1}{2}}\left\{\int_0^u \frac{\sum_{i=1}^{n}dN_i^D(r;\beta_0)}{\sum_{i=1}^{n}Y_i(r)e^{\beta_0'Z_i}} - \Lambda_0(u)\right\} \\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^u \frac{dM_i^D(r;\beta_0)}{S^{(0)}(r;\beta_0)} \\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t \frac{dM_i^D(r;\beta_0)}{s^{(0)}(r;\beta_0)} \\
&\quad +n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t [S^{(0)}(r;\beta_0)^{-1} - s^{(0)}(r;\beta_0)^{-1}]dM_i^D(r;\beta_0).
\end{aligned}
$$

The second term in the summation converges to zero, which can be shown by the strong convergence of $S^{(0)}(r;\beta_0)$ to $s^{(0)}(r;\beta_0)$, the continuous mapping theorem and the USLLN. Thus,

$$
(4.11) \qquad n^{\frac{1}{2}}\{\widehat{\Lambda}(u;\beta_0) - \Lambda(u)\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^u \frac{dM_i^D(r;\beta_0)}{s^{(0)}(r;\beta_0)}
$$

as $n \to \infty$.

In summary, combining (4.7), (4.10) and (4.11)

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\phi}(t;\widehat{\beta},\widehat{\theta}) - \widehat{\phi}(t;\beta_0,\widehat{\theta})\} &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\gamma_{i1}(t;\beta_0,\theta_0) \\
n^{\frac{1}{2}}\{\widehat{\phi}(t;\beta_0,\widehat{\theta}) - \widehat{\phi}(t;\beta_0,\theta_0)\} &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\gamma_{i2}(t;\beta_0,\theta_0) \\
n^{\frac{1}{2}}\{\widehat{\phi}(t;\beta_0,\theta_0) - \phi_0(t)\} &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\gamma_{i3}(t;\beta_0,\theta_0),
\end{aligned}
$$

as $n \to \infty$, such that

$$n^{\frac{1}{2}}\{\widehat{\phi}(t;\widehat{\beta},\widehat{\theta}) - \phi_0(t)\} \;\;=\;\; n^{-\frac{1}{2}}\sum_{i=1}^{n}\gamma_i(t;\beta_0,\theta_0),$$

as $n \to \infty$, and hence $n^{\frac{1}{2}}\{\widehat{\phi}(t;\widehat{\beta},\widehat{\theta}) - \phi_0(t)\}$ behaves asymptotically as a scaled sum of independent and identically distributed zero mean variables.

The quantity $\gamma_i(t;\beta_0,\theta_0)$ can shown asymptotically to be martingale (Fleming and Harrington, 1990), meaning that $n^{-\frac{1}{2}}\sum_{i=1}^{n}\gamma_i(t;\beta_0,\theta_0)$ is tight, which combined with the normality results completes the weak convergence proof. The covariance function for $n^{\frac{1}{2}}\{\widehat{\phi}(t;\widehat{\beta},\widehat{\theta}) - \phi(t)\}$, $\Phi(s,t) = E[\gamma_i(s;\beta_0,\theta_0)\gamma_i(t;\beta_0,\theta_0)]$, can be obtained through the Multivariate Central Limit Theorem (Sen and Singer 1993), and can be consistently estimated by

$$\widehat{\Phi}(s,t) = n^{-1}\sum_{i=1}^{n}\widehat{\gamma}_i(s;\widehat{\beta},\widehat{\theta})\widehat{\gamma}_i(t;\widehat{\beta},\widehat{\theta}).$$

Proof of Theorem 3:

We prove the consistency of $\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta})$ and $\widehat{\mu}_0(t;\widehat{\beta},\widehat{\theta})$ separately, following which the consistency of $\widehat{\psi}(t;\widehat{\beta},\widehat{\theta})$ is directly obtained.

We estimate $\psi(t)$ by plugging in the corresponding empirical estimates for all the quantities,

$$\begin{aligned}\widehat{\psi}(t) \;\;=\;\; & n^{-1}\sum_{i=1}^{n}\int_{0}^{t}[\widehat{S}(r^{-}|Z_i^1)\{d\widehat{R}_0(r) + (\widehat{\theta}_1 + X_{i2}'\widehat{\theta}_2)dr\} \\ & -\widehat{S}(r^{-}|Z_i^0)\{d\widehat{R}_0(r) + X_{i2}'\widehat{\theta}_2 dr\}].\end{aligned}$$

As stated in the proof of Theorem 1, $\widehat{\theta} \xrightarrow{a.s.} \theta_0$, $\widehat{\beta} \xrightarrow{a.s.} \beta_0$, $\widehat{\Lambda}_0(t) \xrightarrow{a.s.} \Lambda_0(t)$, and $\widehat{S}(u^{-}|Z_i) \xrightarrow{a.s.} S(u^{-}|Z_i)$, for all $t \in (0,\tau]$. Therefore, by the Continuous Mapping Theorem,

$$\widehat{\mu}_1(t;\widehat{\theta},\widehat{\beta}) \;\;\xrightarrow{a.s.}\;\; \widehat{\mu}_1(t;\theta_0,\beta_0),$$

for all $t \in (0,\tau]$.

Finally, applying the USLLN (Pollard, 1990) to $\widehat{\mu}_1(t; \theta_0, \beta_0) = n^{-1} \sum_{i=1}^{n} \mu_1(t; \theta_0, \beta_0 | Z_i, X_i)$, we obtain $n^{-1} \sum_{i=1}^{n} \mu_1(t; \theta_0, \beta_0 | Z_i, X_i) \xrightarrow{a.s.} E\{\mu_1(t; \theta_0, \beta_0 | Z_i, X_i)\}$, which is essentially $\mu_1(t)$. Similar arguments can be applied to $\widehat{\mu}_0(t; \widehat{\theta}, \widehat{\beta})$.

Proof of Theorem 4:

We describe here the distribution of $n^{\frac{1}{2}}\{\widehat{\mu}_1(t; \widehat{\beta}, \widehat{\theta}) - \mu_1(t)\}$. A similar development applies to $n^{\frac{1}{2}}\{\widehat{\mu}_0(t; \widehat{\beta}, \widehat{\theta}) - \mu_0(t)\}$. First, $n^{\frac{1}{2}}\{\widehat{\mu}_1(t; \widehat{\beta}, \widehat{\theta}) - \mu_1(t)\}$ is decomposed into four parts as follows:

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\mu}_1(t; \widehat{\beta}, \widehat{\theta}) - \mu_1(t)\} = \ & n^{\frac{1}{2}}\{\widehat{\mu}_1(t; \widehat{\beta}, \widehat{\theta}) - \widehat{\mu}_1(t; \beta_0, \widehat{\theta})\} \\
& + n^{\frac{1}{2}}\{\widehat{\mu}_1(t; \beta_0, \widehat{\theta}) - \widehat{\mu}_1(t; \beta_0, \theta_0)\} \\
& + n^{\frac{1}{2}} \int_0^t n^{-1} \sum_{i=1}^{n} \widehat{S}(r^-|Z_i^1)\{d\widehat{R}_0(r; \theta_0) - dR_0(r)\} \\
& + n^{\frac{1}{2}} \int_0^t n^{-1} \sum_{i=1}^{n} \{\widehat{S}(r^-; \beta_0|Z_i^1) - S(r^-|Z_i^1)\}\{dR_0(r) + \theta_0' X_i^j dr\}.
\end{aligned}
$$

(4.12)

For the first part of (4.12) , we apply a Taylor expansion around β_0 to obtain

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\mu}_1(t; \widehat{\beta}, \widehat{\theta}) - \widehat{\mu}_1(t; \beta_0, \widehat{\theta})\} = \ & n^{-1} \sum_{i=1}^{n} \int_0^t \left. \frac{\partial S(r^-; \beta, \widehat{\theta}|Z_i^1)}{\partial \beta'} \right|_{\beta=\beta_*} n^{\frac{1}{2}}(\widehat{\beta} - \beta_0) \\
& \{d\widehat{R}_0(r) + (\widehat{\theta}_1 + X_{i2}'\widehat{\theta}_2)dr\} \\
= \ & -n^{-1} \sum_{i=1}^{n} \int_0^t \Big[\{\exp\{-e^{\beta_*' Z_i^1}\widehat{\Lambda}_0(r) + \beta_*' Z_i^1\} Z_i^1 \widehat{\Lambda}_0(r) \\
& - \exp\{-e^{\beta_*' Z_i^1}\widehat{\Lambda}_0(r) + \beta_*' Z_i^1\} \int_0^r \frac{S^{(1)}(u; \beta_*)}{n\{S^{(0)}(u; \beta_*)\}^2} dN(u) \Big\} \\
& \{d\widehat{R}_0(r) + (\widehat{\theta}_1 + X_{i2}'\widehat{\theta}_2)dr\} \Big] n^{\frac{1}{2}}(\widehat{\beta} - \beta_0) \\
= \ & -E\left[e^{\beta' Z_k^0} \int_0^t S(u^-|Z_k^0) \int_0^u \{Z_k^0 - \overline{z}(r; \beta)\}' d\Lambda_0(r) dR(u|X_k^j) \right] \\
& n^{\frac{1}{2}}(\widehat{\beta} - \beta_0),
\end{aligned}
$$

as $n \rightarrow \infty$, where β_* lies between $\widehat{\beta}$ and β_0 in $\mathcal{R}^p$. Furthermore, given (4.7), when $n \rightarrow \infty$,

(4.13)
$$
n^{\frac{1}{2}}\{\widehat{\mu}_1(t; \widehat{\beta}, \widehat{\theta}) - \widehat{\mu}_1(t; \beta_0, \widehat{\theta})\} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_{i11}(t; \beta, \theta).
$$

.

For the second part of (4.12), when $n \to \infty$,

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\beta_0,\widehat{\theta}) - \widehat{\mu}_1(t;\beta_0,\theta_0)\} &= n^{-1}\sum_{i=1}^{n}\int_0^t \widehat{S}(r^-|Z_i^1)\{\frac{\partial d\widehat{R}_0(r)}{\partial\theta'} + X_i'dr\}n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \\
&= n^{-1}\sum_{i=1}^{n}\int_0^t \widehat{S}(r^-|Z_i^1)(-\overline{X}'(r) + X_i^{1'})dr\{B^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}U_i^R\} \\
(4.14)\qquad &= E\left[\int_0^t S(u^-|Z_i^j)\{(X_i^j)' - \overline{x}'(u)\}du\right]\{B^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}U_i^R\},
\end{aligned}
$$

by the USLLN and the results of Schaubel et al. (2006) for $\widehat{\theta}$.

For the third part of (4.12), when $n \to \infty$,

$$
\begin{aligned}
n^{\frac{1}{2}}\int_0^t n^{-1}\sum_{i=1}^{n}\{\widehat{S}(r^-|Z_i^1)\}\{d\widehat{R}_0(r;\theta_0) - dR_0(r)\} &= \int_0^t E\{S(r^-|Z_i^1)\}n^{\frac{1}{2}}\{d\widehat{R}_0(r;\theta_0) - dR_0(r)\} \\
(4.15)\qquad &= \int_0^t E\{S(r^-|Z_i^1)\}n^{-\frac{1}{2}}\sum_{i=1}^{n}\pi(r)^{-1}dM_i^R(r).
\end{aligned}
$$

For the fourth part of (4.12), when $n \to \infty$,

$$
\begin{aligned}
n^{\frac{1}{2}}\{\widehat{\psi}(t;\beta_0,\theta_0) - \psi_0(t)\} &= n^{-1}\sum_{i=1}^{n}\int_0^t n^{\frac{1}{2}}\{\widehat{S}(u^-;\beta_0|Z_i^1) - S_i(u^-|Z_i^1)\} \\
&\qquad \{dR_0(u) + (\theta_1 + X_{i2}'\theta_2)du\} \\
&= n^{-1}\sum_{i=1}^{n}\int_0^t n^{\frac{1}{2}}[\exp\{-\widehat{\Lambda}_0(u;\beta_0)e^{\beta_0'Z_i^0}\} - \exp\{-\Lambda_0(u)e^{\beta_0'Z_i^0}\}] \\
&\qquad \{dR_0(u) + (\theta_1 + X_{i2}'\theta_2)du\} \\
&= -n^{-1}\sum_{i=1}^{n}\int_0^t \exp\{-\Lambda_0(u)e^{\beta_0'Z_i^0}\}(e^{\beta_0'Z_i^0})n^{\frac{1}{2}}\{\widehat{\Lambda}_0(u;\beta_0) - \Lambda_0(u)\} \\
&\qquad \{dR_0(u) + (\theta_1 + X_{i2}'\theta_2)du\} \\
&= -\int_0^t n^{-1}\sum_{i=1}^{n}\{e^{-\Lambda_0(u)e^{\beta_0'Z_i^0}}e^{\beta_0'Z_i^0}\}\{dR_0(u) + (\theta_1 + X_{i2}'\theta_2)du\} \\
&\qquad n^{\frac{1}{2}}\{\widehat{\Lambda}_0(u;\beta_0) - \Lambda_0(u)\} \\
(4.16)\qquad &= -\int_0^t E[e^{\beta_0'Z_i^0}S(u^-|Z_i^1)dR(u|X_i^1)]n^{\frac{1}{2}}\{\widehat{\Lambda}_0(u;\beta_0) - \Lambda_0(u)\},
\end{aligned}
$$

while $n^{\frac{1}{2}}\{\widehat{\Lambda}_0(u;\beta_0) - \Lambda_0(u)\}$ is derived in (4.11).

In summary, combining (4.13), (4.14), (4.15) and (4.16), when $n \to \infty$,

$$n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta}) - \mu_1(t)\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\xi_{i1.}(t;\beta,\theta)$$

$$n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta}) - \widehat{\mu}_1(t;\beta_0,\widehat{\theta})\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\xi_{i11}(t;\beta,\theta)$$

$$n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\beta_0,\widehat{\theta}) - \widehat{\mu}_1(t;\beta_0,\theta_0)\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\xi_{i12}(t;\beta,\theta)$$

$$n^{\frac{1}{2}}\int_0^t n^{-1}\sum_{i=1}^{n}\widehat{S}(r^-|Z_i^1)\{d\widehat{R}_0(r;\theta_0) - dR_0(r)\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\xi_{i13}(t;\beta,\theta)$$

$$n^{\frac{1}{2}}\int_0^t n^{-1}\sum_{i=1}^{n}\{\widehat{S}(r^-;\beta_0|Z_i^1) - S(r^-|Z_i^1)\}\{dR_0(r) + \theta_0'X_i^j dr\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\xi_{i14}(t;\beta,\theta).$$

Therefore $n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta}) - \mu_1(t)\}$ is asymptotically $n^{-\frac{1}{2}}$ times sum of independent and identically distributed terms. The tightness of $n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta}) - \mu_1(t)\}$ can be proved by examining each component of $\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta})$ so that $n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta}) - \mu_1(t)\}$ can be written as sum of monotone bounded functions. Through similar proofs, $n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta}) - \mu_1(t)\}$ can be written as scaled sum of contributions from each of the independent and identically distributed subject. After expressing the difference, $n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta}) - \mu_1(t)\} - n^{\frac{1}{2}}\{\widehat{\mu}_0(t;\widehat{\beta},\widehat{\theta}) - \mu_0(t)\}$, at the subject level, weak convergence to a Gaussian process follows from the application of various tools from empirical process theory (Pollard, 1990; Van der Vaart and Wellner, 1996). The covariance function for $n^{\frac{1}{2}}\{\widehat{\mu}_1(t;\widehat{\beta},\widehat{\theta}) - \widehat{\mu}_0(t;\widehat{\beta},\widehat{\theta})\}$ is given by

$$\Phi(s,t) = E[\{\xi_{i1.}(s;\beta_0,\theta_0) - \xi_{i0.}(s;\beta_0,\theta_0)\}\{\xi_{i1.}(t;\beta_0,\theta_0) - \xi_{i0.}(t;\beta_0,\theta_0)\}],$$

which can be estimated by replacing limiting values with their empirical counterparts, then averaging over the sample.

4.8   References

ANDERSEN, P. K., & GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-20.

BILIAS, Y., GU, M. & YING, Z. (1997). Towards a general asymptotic theory for the Cox model with staggered entry. *Ann. Statist.* **25**, 662-82.

COOK, R. J. & LAWLESS, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Statist. Med.* **16**, 911-24.

COX, D. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B **34**, 187-220.

COX, D. (1975). Partial likelihood. *Biometrika* **62**, 262-76.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis.* New York: Wiley.

GHOSH, D. & LIN, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554-62.

GHOSH, D. & LIN, D. Y. (2002). Marginal regression models for recurrent and terminal event. *Statist. Sinica* **12**, 663-88.

HUANG, C. Y. & WANG, M. C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Am. Statist. Assoc.* **99**, 1153-65.

KALBFLEISCH, J. D & PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data.* Hoboken, NJ: Wiley.

LAWLESS, J. F. & NADEAU, C. (1995). Some simple robust methods for the analysis of recurrent event. *Technometrics* **37**, 158-68.

LI, Q. & LAGAKOS, S. (1997). Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. *Statist. Med.* **16**, 925-40.

LIN, D. Y., FLEMING, T. R. & 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., & YING, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Statist. Soc.* B **62**, 711-30.

LIU, L., WOLFE, R. A., & HUANG, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 747-56.

PEPE, M. S. & CAI, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *J. Am. Statist. Assoc.* **88**, 811-20.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* Hayward, CA: Institute of Mathematical Statistics.

PRENTICE, R. L., WILLIAMS, B., & PETERSON, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373-9.

SCHAUBEL, D. E., ZENG, D. & CAI, J. (2006). A semiparametric additive rates model for recurrent event data. *Lifetime Data Analysis* **12**, 389-406.

SEN, P. K. & SINGER, J. M. (1993). *Large Sample Methods in Statistics..* New York: Chapman and Hall.

VAN DER VAART, A. W. (2000). *Asymptotic Statistics.* Melbourne: Cambridge.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* New York: Springer.

WEI, L. J., LIN, D. Y. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065-73.

YE, Y., KALBFLEISCH, J. D. & SCHAUBEL, D.E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* , **63**, 78-87.

YE, Y., KALBFLEISCH, J. D. & SCHAUBEL, D.E. (2007). An estimating function

approach to the analysis of recurrent and terminal events. (Submitted manuscript).

CHAPTER V

Conclusion

This dissertation proposes three novel survival analysis methods for use in observational studies, each motivated by real problems in applying the classic survival analysis techniques. A common aspect of the three papers is that they were motivated by specific needs in applying statistical models to real research questions. Hence, the contribution of this research are of clinical, in addition to biostatistical, significance. Specifically, Chapter 2 addresses the problem of fitting a proportional hazards model when subjects have different probabilities of being sampled/observed. Chapter 3 proposes testing procedures with considerable statistical power to formally evaluate the degree of selection bias. Chapter 4 considers estimating the treatment effect on the recurrent event mean, in the presence of a terminating event.

Chapters 2 and 3 feature the idea of "inverse-selection-probability-weighting", in which each subject is weighted by the inverse of their selection probability. Under this approach, inference is valid for the target population before selection. Chapter 3 proposes, as a test statistic, the difference between parameters before and after selection. Chapter 4 starts from the idea of separately estimating the terminating event survival probability and recurrent event conditional rate given survival. A proportional hazards model and an additive rates model are fitted. When we integrate the estimated survival probability and conditional recurrent event rate, we can choose either a "standard" survival distribution for both groups, or the treatment-specific survival distribution for the corresponding group. The choice between the two measures (which have different interpretations) will depend on the goals of the

investigator.

As stated in the beginning of this chapter, each method proposed in this dissertation was applied to address an existing medical research question. In Chapter 2, we determined that Expanded Criteria Donor (ECD) kidneys are associated with a much higher increase in transplant failure when inference pertains to all procured kidneys instead of all transplanted kidneys. Inference applicable to all procured kidneys has more relevance since wait-listed patients are offered procured kidneys, not kidneys already transplanted. Previous results based on only transplanted kidneys underestimate the ECD effect. In Chapter 3, we are interested in the effect of wait-listed patients' characteristics on post transplant survival for the purpose of ranking candidates by predicted "survival benefit". Selection bias occurs since transplanted patients are not a representative sample of all wait listed patients, for whom we want to make inference. The potential biasing factor, morbidity history, could not be adjusted at time of ranking. Through the proposed test statistic, chronic obstructive pulmonary disease (COPD), age 35 to 44 and years on dialysis have significantly different impact on post transplant survival in the transplant candidate and recipient populations. Chapter 4 continues our interest in ECD transplantation. It is discovered that ECD transplant recipients have higher post-transplant death hazard and that patients have a higher hospitalization rate while they survive. Overall, ECD patients have more hospitalizations, despite the fact that they also have significantly lower survival than non-ECD transplant recipients over the whole post-transplant follow-up period. Each of these findings is important information for kidney transplantation surgeons and patients, as well as policy makers, in addressing questions such as how restrictive patients should be with respect to accepting lower quality organs.

CHAPTER VI

Supplementary Materials for "Evaluating Bias Correction in Weighted Proportional Hazards Regression"

6.1   Appendix A

We assume the following regularity conditions:

(a) $\{T_i, C_i, Z_i, I_i, X_i\}$ are independent and identically distributed for $i = 1, \ldots, N$.

(b) $P(\widetilde{T}_i > \tau) > 0$.

(c) $\int_0^\tau \lambda_{0S}(t)dt < \infty$ and $\int_0^\tau \lambda_{0T}(t)dt < \infty$.

(d) $|X_{ik}| < \infty$; $|Z_{ik}(0)| + \int_0^\tau |Z_{ik}(s)|ds < \infty$ almost surely, where the second subscript refers to vector element.

(e) Positive-definiteness of the matrices, $A(\beta_0)$, $A^w(\beta_0, \theta_0)$ and $B(\theta_0)$, where, for the quantities applicable to the possibly biased sample,

$$A(\beta) = E\left[\int_0^\tau \{Z_i(t) - \overline{z}(t; \beta)\}^{\otimes 2} Y_i(t) e^{\beta' Z_i(t)} \lambda_{0S}(t)dt\right]$$

$$S^{(k)}(t; \beta) = n^{-1} \sum_{i=1}^N I_i Y_i(t) e^{\beta' Z_i(t)} Z_i(t)^{\otimes k}, k = 0, 1, 2$$

$$\overline{Z}(t; \beta) = \frac{S^{(1)}(t; \beta)}{S^{(0)}(t; \beta)}$$

$$s^{(k)}(t; \beta) = \lim_{N \to \infty} S^{(k)}(t; \beta), k = 0, 1, 2$$

$$\overline{z}(t; \beta) = \frac{s^{(1)}(t; \beta)}{s^{(0)}(t; \beta)}$$

and for the quantities applicable to the target population,

$$
\begin{aligned}
A^w(\beta, \theta) &= E\left[\int_0^\tau \{Z_i(t) - \overline{z}_w(\beta, \theta; t)\}^{\otimes 2} w_i(\theta) Y_i(t) e^{\beta' Z_i(t)} \lambda_{0T}(t) dt\right] \\
S_w^{(k)}(t; \beta, \theta) &= N^{-1} \sum_{i=1}^N w_i(\theta) Y_i(t) e^{\beta' Z_i(t)} Z_i(t)^{\otimes k}, k = 0, 1, 2 \\
\overline{Z}_w(t; \beta, \theta) &= \frac{S_w^{(1)}(t; \beta, \theta)}{S_w^{(0)}(t; \beta, \theta)} \\
s_w^{(k)}(t; \beta, \theta) &= \lim_{N \to \infty} S_w^{(k)}(t; \beta, \theta), k = 0, 1, 2 \\
\overline{z}_w(t; \beta, \theta) &= \frac{s_w^{(1)}(t; \beta, \theta)}{s_w^{(0)}(t; \beta, \theta)} \\
B(\theta) &= E[X_i p_i(\theta)\{1 - p_i(\theta)\} X_i']
\end{aligned}
$$

with $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa'$.

(f) There exists a δ such that $p_i(\theta) > \delta > 0$ almost surely.

6.2 Appendix B

Proof of Theorem 1: The strong consistency of $\widehat{\beta}_S$ was demonstrated by Andersen and Gill (1982).

The consistency of the weighted parameter estimate, $\widehat{\beta}_T$, for the true parameter values applicable to the target population, β_T, is proved in Theorem 1 of Pan and Schaubel (2007).

Combining these two results, as a linear combination of the difference of $\widehat{\beta}_S$ and $\widehat{\beta}_T$, $\widehat{D}_j \xrightarrow{a.s.} D_j$ by the continuous mapping theorem.

Through a Taylor expansion around $\beta = \beta_S$,

$$
U(\widehat{\beta}_S) - U(\beta_S) = \left.\frac{\partial U(\beta)}{\partial \beta'}\right|_{\beta = \beta_S^*} (\widehat{\beta}_S - \beta_S),
$$

where β_S^* lies between $\widehat{\beta}_S$ and β_S in $\mathcal{R}^p$. Define

$$
\widehat{A}(\beta) = -n^{-1} \frac{\partial U(\beta)}{\partial \beta'} = -n^{-1} \sum_{i=1}^N \int_0^\tau I_i \left[\frac{S^{(2)}(t; \beta)}{S^{(0)}(t; \beta)} - \overline{Z}(t; \beta)^{\otimes 2}\right] dN_i(t).
$$

Then, we have

$$
n^{\frac{1}{2}}(\widehat{\beta}_S - \beta_S) = \widehat{A}(\beta_S^*)^{-1} n^{-\frac{1}{2}} U(\beta_S)
$$

since $U(\widehat{\beta}_S) = 0$. Under the assumed conditions,

$$\sup_{t \in [0,\tau]} ||S^{(d)}(t; \beta) - s^{(d)}(t; \beta)|| \xrightarrow{a.s.} 0,$$

for $d = 0, 1$, or 2 and any β in a compact set. Since $\widehat{\beta}_S \xrightarrow{a.s.} \beta_S$ and $\| \beta_* - \beta_S \| \leq \| \widehat{\beta}_S - \beta_S \|$, $\beta_S^* \xrightarrow{a.s.} \beta_S$. Using the continuous mapping theorem, $S^{(d)}(t; \beta_S^*) \xrightarrow{a.s.} s^{(d)}(t; \beta_S)$, with $s^{(0)}(t; \beta)$ bounded away from zero for all β and $t \in [0, \tau]$. Using the almost sure convergence of $S^{(d)}(t; \beta_S^*)$ to $s^{(d)}(t; \beta_S)$ and the continuous mapping theorem, we get

$$\frac{S^{(2)}(t; \beta_S^*)}{S_w^{(0)}(t; \beta_S^*)} - \overline{Z}(t; \beta_S^*)^{\otimes 2} \longrightarrow \frac{s^{(2)}(t; \beta_S)}{s^{(0)}(t; \beta_S)} - \overline{z}(t; \beta_S)^{\otimes 2} \equiv v(t; \beta_S)$$

uniformly in $t \in (0, \tau]$. Using the Strong Law of Large Numbers (SLLN) and the continuous mapping theorem,

$$\widehat{A}(\beta_S^*) \xrightarrow{a.s.} A(\beta_S),$$

where $A(\beta) \equiv E\left[\int_0^\tau v(t; \beta)dN_i(t)\right]$, as defined in condition (e). Hence

(6.1) $$n^{\frac{1}{2}}(\widehat{\beta}_S - \beta_S) = A(\beta_S)^{-1} n^{-\frac{1}{2}} U_N(\beta_S) + o_p(1).$$

Through the definition of $S^{(d)}(t; \beta)$ and with some basic algebra,

$$U(\beta) = \sum_{i=1}^N \int_0^\tau I_i[Z_i(t) - \overline{Z}(t; \beta)]dM_i(t; \beta).$$

Furthermore, we can write

(6.2) $$n^{-\frac{1}{2}} U(\beta_S) = n^{-\frac{1}{2}} \sum_{i=1}^N \int_0^\tau I_i \psi_i(\beta_S) + o_p(1),$$

where $\psi_i(\beta) = [Z_i(t) - \overline{Z}(t; \beta)]dM_i(t; \beta) = [Z_i(t) - \overline{z}(t; \beta)]dM_i(t; \beta)$, using the fact that

$$\left\| n^{-\frac{1}{2}} \sum_{i=1}^N \int_0^\tau I_i[\overline{Z}_w(t; \beta_S) - \overline{z}(t; \beta_S)]dM_i(t; \beta_S) \right\| \xrightarrow{P} 0_{p \times 1}.$$

which can be demonstrated by employing various empirical process results (van der Vaart 2000) and results from empirical process theory (Pollard 1990, van der Vaart

1996, Bilias, Gu and Ying 1997). Combining (6.1) and (6.2), we get

$$
(6.3) \qquad n^{\frac{1}{2}}(\widehat{\beta}_S - \beta_S) = A(\beta_S)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{N} \int_0^{\tau} I_i \psi_i(\beta_S) + o_p(1).
$$

Pan and Schaubel (2007) demonstrated that

$$
(6.4) \qquad N^{\frac{1}{2}}(\widehat{\beta}_T - \beta_T) = A^w(\beta_T, \theta_0)^{-1} N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi_i^w(\beta_T, \theta_0) + o_p(1),
$$

where

$$
\begin{aligned}
\psi_i^w(\beta, \theta) &= \int_0^{\tau} w_i(\theta)\{Z_i(t) - \bar{z}_w(t; \beta, \theta)\}dM_i(t) - G(\beta, \theta)B(\theta)^{-1}U_{Li} \\
G(\beta, \theta) &= E\left[\int_0^{\tau} \{Z_i(t) - \bar{z}_w(t; \beta, \theta)\}X_i' e^{-\theta' X_i} dM_i(t; \beta)\right], \\
U_{Li}(\theta) &= X_i\{I_i - p_i(\theta)\}.
\end{aligned}
$$

Combining (6.3) and (6.4),

$$
n^{\frac{1}{2}}(\widehat{D} - D) = \bar{p}^{\frac{1}{2}} c_j' A^w(\beta_T, \theta_0)^{-1} N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi_i^w(\beta_T, \theta_0) - c_j' A(\beta_S)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{N} \int_0^{\tau} I_i \psi_i(\beta_S).
$$

We now define

$$
\begin{aligned}
D_1 &= \bar{p}^{\frac{1}{2}} c_j' A^w(\beta_T, \theta_0)^{-1} N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi_i^w(\beta_T, \theta_0) \\
D_2 &= c_j' A(\beta_S)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{N} \int_0^{\tau} I_i \psi_i(\beta_S),
\end{aligned}
$$

Such that

$$
(6.5) \qquad \mathrm{Var}(n^{\frac{1}{2}}(\widehat{D} - D)) = \mathrm{Var}(D_1) + \mathrm{Var}(D_2) - 2\mathrm{Cov}(D_1, D_2).
$$

For D_1, $\bar{p}^{\frac{1}{2}}$ converges to a constant between 0 and 1, indicator vector c_j is pre-specified and the asymptotic information matrix for the weighted Cox regression $A^w(\beta_T, \theta_0)^{-1}$ is also fixed. Hence the variance of D_1 is the product of several constants and the variance of $N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi_i^w(\beta_T, \theta_0)$, namely $\Sigma_w(\beta_T, \theta_0) = E\{\psi_i^w(\beta_T, \theta_0)^{\otimes 2}\}$, which was derived by Pan and Schaubel through the Multivariate Central Limit Theorem. Therefore

$$
(6.6) \qquad \mathrm{Var}(D_1) = \bar{p} c_j' A^w(\beta_T, \theta_0)^{-1} \Sigma_w(\beta_T, \theta_0) A^w(\beta_T, \theta_0)^{-1} c_j.
$$

Similar arguments can be applied to $\text{Var}(D_2)$. The matrix $A(\beta_S)$ is the limiting value for $\widehat{A}(\widehat{\beta}_S)$ and hence can be treated as constant matrix. The variance of $n^{-\frac{1}{2}} \sum_{i=1}^{N} \int_0^{\tau} I_i \psi_i(\beta_S)$ can be derived through the Martingale Central Limit Theorem (Andersen and Gill 1982). That is, $\Sigma(\beta_S) = E\{\psi_i(\beta_S)^{\otimes 2}\}$, such that

$$(6.7) \qquad \text{Var}(D_2) = c_j' A(\beta_S)^{-1} \Sigma(\beta_S) A(\beta_S)^{-1} c_j.$$

In terms of $\text{Cov}(D_1, D_2)$, the quantities c_j, $A^w(\beta_T, \theta_0)^{-1}$ and $A(\beta_S)$ are fixed. For the covariance between $N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi_i^w(\beta_T, \theta_0)$ and $n^{-\frac{1}{2}} \sum_{i=1}^{N} \int_0^{\tau} I_i \psi_i(\beta_S)$, we consider the three possible scenarios for the combination of subject i and j. If $i \neq j$, $\text{Cov}\{\psi_i^w(\beta_T, \theta_0), I_j \psi_j(\beta_S)\} = \text{Cov}(\psi_j^w(\beta_T, \theta_0), I_i \psi_i(\beta_S)) = 0$ since subjects are all independent from each other. If $i = j$ and $I_i = 0$, $\text{Cov}(\psi_i^w(\beta_T, \theta_0), I_j \psi_j(\beta_S)) = \text{Cov}(\psi_i^w(\beta_T, \theta_0), 0) = 0$ because subject i does not contribute to the estimation of β_S. If $i = j$ and $I_i \neq 0$, $\text{Cov}\{\psi_i^w(\beta_T, \theta_0), I_j \psi_j(\beta_S)\} = E\{\psi_i(\beta_S)\psi_i^w(\beta_T, \theta_0)'\}$ because both $\psi_i(\beta_S)$ and $\psi_i^w(\beta_T, \theta_0)$ have expectation zero. Summing up the three types of covariances,

$$\text{Cov}\left\{N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi_i^w(\beta_T, \theta_0), n^{-\frac{1}{2}} \sum_{i=1}^{N} \int_0^{\tau} I_i \psi_i(\beta_S)\right\} = (Nn)^{-\frac{1}{2}} n \Sigma_{\beta}(\beta_S, \beta_T, \theta_0)$$

where

$$\Sigma_{\beta}(\beta_1, \beta_2, \theta) = E\{\psi_i(\beta_1)\psi_i^w(\beta_2, \theta)'\}.$$

Combining the above results, we obtain

$$(6.8) \qquad \text{Cov}(D_1, D_2) = \bar{p}^{\frac{1}{2}} c_j' A(\beta_S)^{-1} \Sigma_{\beta}(\beta_S, \beta_T, \theta_0) A^w(\beta_T, \theta_0)^{-1} c_j.$$

Combining (6.5), (6.6), (6.7) and (6.8), the variance of $n^{\frac{1}{2}}(\widehat{D} - D)$ in Theorem 1 is proved.

6.3 Appendix C

Proof of Theorem 2: The proof of uniform consistency of of $\widehat{\Lambda}_0(t; \widehat{\beta}_S)$ to $\Lambda_{0S}(t)$ for $t \in (0, \tau]$ begins by decomposing $\alpha(t) = \widehat{\Lambda}_0(t; \widehat{\beta}_S) - \Lambda_{0S}(t)$ into two parts, $\alpha(t) =$

$\alpha_1(t) + \alpha_2(t)$ with

$$\alpha_1(t) = \widehat{\Lambda}_0(t; \widehat{\beta}_S) - \widehat{\Lambda}_0(t; \beta_S)$$

$$\alpha_2(t) = \widehat{\Lambda}_0(t; \beta_S) - \Lambda_{0S}(t)$$

Applying a Taylor expansion about β_S,

$$\alpha_1(t) = N^{-1} \sum_{i=1}^{N} \int_0^t \frac{\partial}{\partial \beta'} [S^{(0)}(s; \beta)] \Big|_{\beta=\beta_*} dN_i(s)(\widehat{\beta}_S - \beta_S)$$

$$= -N^{-1} \sum_{i=1}^{N} \int_0^t \frac{S^{(1)}(s; \beta)}{S^{(0)}(s; \beta)^2} \Big|_{\beta=\beta_*} dN_i(s)(\widehat{\beta}_S - \beta_S)$$

$$= -\int_0^t \overline{Z}'(s; \beta_*) d\widehat{\Lambda}_0(s; \beta_*)(\widehat{\beta}_S - \beta_S),$$

where β_* lies between $\widehat{\beta}_S$ and β_S in $\mathcal{R}^p$. Since the quantities $\overline{Z}(s; \beta_*)$ and $d\widehat{\Lambda}_0(s; \beta_*)$ are bounded, along with $\widehat{\beta}_S \xrightarrow{a.s.} \beta_S$, $\mid \alpha_1(t) \mid \xrightarrow{a.s.} 0$.

The second component, $\alpha_2(t)$, can be rewritten as $n^{-1} \sum_{i=1}^{N} \int_0^t I_i S^{(0)}(s; \beta_S)^{-1} dM_i(s; \beta_S)$. By the Uniform Strong Law of Large Numbers (USLLN; Pollard 1990), $n^{-1} \sum_{i=1}^{N} \int_0^t I_i dM_i(s; \beta_S) \xrightarrow{a.s.}$ 0 for $t \in [0, \tau]$. As $n \longrightarrow \infty$, $S^{(0)}(s; \beta_S) \longrightarrow s^{(0)}(s; \beta_S)$ which is bounded away from 0. Therefore, $\mid \alpha_2(t) \mid \xrightarrow{a.s.} 0$.

Combining results for $\alpha_1(t)$ and $\alpha_2(t)$ and the triangle inequality,

$$\sup_{t \in [0,\tau]} \mid \alpha(t) \mid \leq \sup_{t \in [0,\tau]} \mid \alpha_1(t) \mid + \sup_{t \in [0,\tau]} \mid \alpha_2(t) \mid,$$

yields the required result,

$$(6.9) \qquad \sup_{t \in [0,\tau]} \mid \widehat{\Lambda}_0(t; \widehat{\beta}_S) - \Lambda_{0S}(t) \mid \xrightarrow{a.s.} 0.$$

Combining (6.9) and the convergence of $\widehat{\Lambda}_0^w(t; \widehat{\beta}_T, \widehat{\theta})$ for $\Lambda_{0T}(t)$ (Pan and Schaubel 2007), $\widehat{\Delta}(t; \widehat{\beta}_S, \widehat{\beta}_T, \widehat{\theta})$ converges uniformly to $\Delta(t)$ for $t \in [0, \tau]$ by continuous mapping theorem.

6.4 Appendix D

Proof of Theorem 3: First, we decompose

$$n^{\frac{1}{2}} \{ \widehat{\Delta}(t; \widehat{\beta}_S, \widehat{\beta}_T, \widehat{\theta}) - \Delta(t) \} = \overline{p}^{\frac{1}{2}} N^{\frac{1}{2}} \{ \widehat{\Lambda}_0^w(t; \widehat{\beta}_T, \widehat{\theta}) - \Lambda_{0T}(t) \} - n^{\frac{1}{2}} \{ \widehat{\Lambda}_0(t; \widehat{\beta}_S) - \Lambda_{0S}(t) \}$$

$$= \Delta_1(t) - \Delta_2(t).$$

As shown by Pan and Schaubel,

$$\Delta_1(t) = \overline{p}^{\frac{1}{2}} N^{-\frac{1}{2}} \sum_{i=1}^{N} \phi_i^w(t; \beta_T, \theta_0) + o_p(1),$$

where

$$
\begin{aligned}
b(t; \beta, \theta) &= E\{X_i e^{-\theta' X_i} Y_i(t) e^{\beta' Z_i(t)}\} \\
k(t; \beta, \theta) &= -E\left\{\int_0^t \frac{X_i e^{-\theta' X_i}}{s_w^{(0)}(s; \beta, \theta)} dN_i(s) + \int_0^t \frac{w_i(\theta) b(s; \beta, \theta)}{s_w^{(0)}(s; \beta, \theta)^2} dN_i(s)\right\} \\
h(t; \beta, \theta) &= -\int_0^t \overline{z}_w(s; \beta, \theta) d\Lambda_0(s) \\
\phi_i^w(t; \beta, \theta) &= k'(t; \beta, \theta) B(\theta)^{-1} U_{Li} + h'(t; \beta, \theta) A^w(\beta, \theta)^{-1} \psi_i^w(\beta, \theta) + \int_0^t \frac{w_i(\theta) dM_i(s; \beta, \theta)}{s_w^{(0)}(s; \beta, \theta)}.
\end{aligned}
$$

Through a similar derivation,

$$\Delta_2(t) = n^{-\frac{1}{2}} \sum_{i=1}^{N} I_i \phi_i(t; \beta_S) + o_p(1),$$

where

$$\phi_i(t; \beta) = -\int_0^t \overline{z}'(s; \beta) d\Lambda_{0S}(s)(t; \beta) A(\beta)^{-1} \psi_i(\beta) + \int_0^t \frac{dM_i(s; \beta, \theta)}{s_w^{(0)}(s; \beta, \theta)}.$$

Asymptotic normality extends to any finite set of time points. Finally, weak convergence to a Gaussian process follows from the tightness of $n^{\frac{1}{2}}\{\widehat{\Delta}(t) - \Delta(t)\}$, which follows from the manageability of both $\phi_i^w(t; \beta_T, \theta_0)$ and $\phi_i(t; \beta_S)$. See Pan and Schaubel (2007) for pertinent details. For two time points $(s, t) \in (0, \tau] \times (0, \tau]$,

$$
\begin{aligned}
&\text{Cov}(n^{\frac{1}{2}}\{\widehat{\Delta}(s; \widehat{\beta}_S, \widehat{\beta}_T, \widehat{\theta}) - \Delta(t)\}, n^{\frac{1}{2}}\{\widehat{\Delta}(t; \widehat{\beta}_S, \widehat{\beta}_T, \widehat{\theta}) - \Delta(t)\}) \\
&= \text{Cov}(\Delta_1(s), \Delta_1(t)) + \text{Cov}(\Delta_2(s), \Delta_2(t)) - \text{Cov}(\Delta_1(s), \Delta_2(t)) - \text{Cov}(\Delta_2(s), \Delta_1(t))
\end{aligned}
$$

where

$$
\begin{aligned}
\text{Cov}(\Delta_1(s), \Delta_1(t)) &= \overline{p} E\{\phi_i^w(s; \beta_T, \theta_0) \phi_i^w(t; \beta_T, \theta_0))\} \\
\text{Cov}(\Delta_2(s), \Delta_2(t)) &= E\{\phi_i(s; \beta_S) \phi_i(t; \beta_S)\} \\
\text{Cov}(\Delta_1(s), \Delta_2(t)) &= \overline{p} E\{\phi_i^w(s; \beta_T, \theta_0) \phi_i(t; \beta_S)\} \\
\text{Cov}(\Delta_2(s), \Delta_1(t)) &= \overline{p} E\{\phi_i(s; \beta_S) \phi_i^w(t; \beta_T, \theta_0)\}.
\end{aligned}
$$

through arguments similar to the derivation for $\text{Cov}(D_1, D_2)$ and the Functional CLT (Pollard 1990).
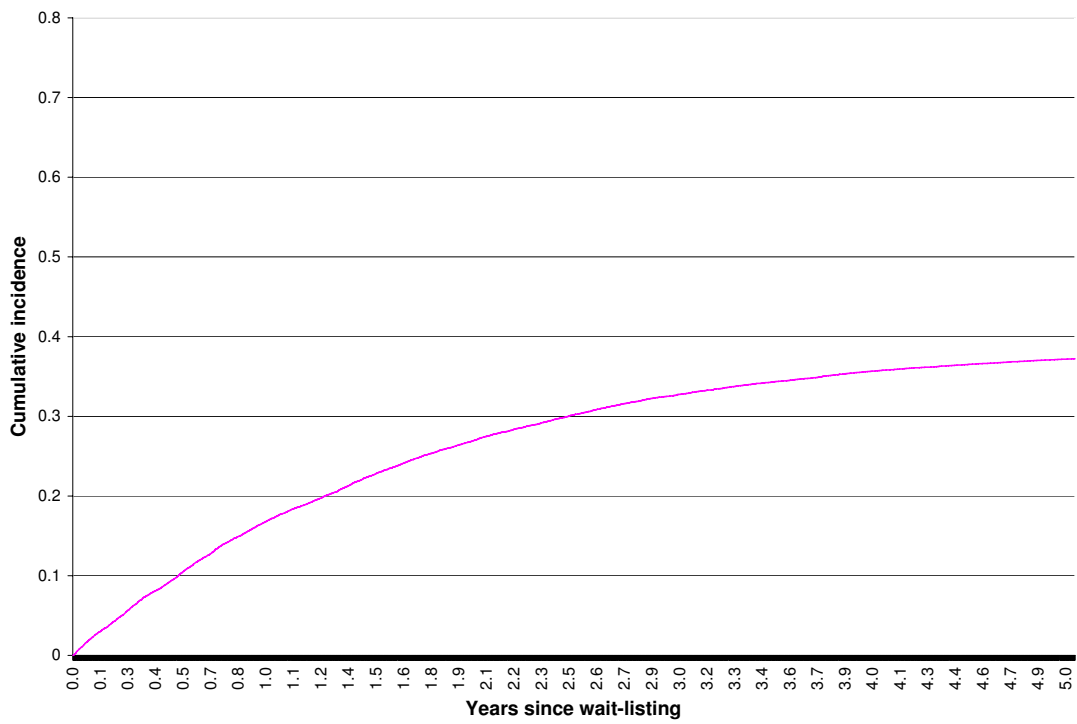
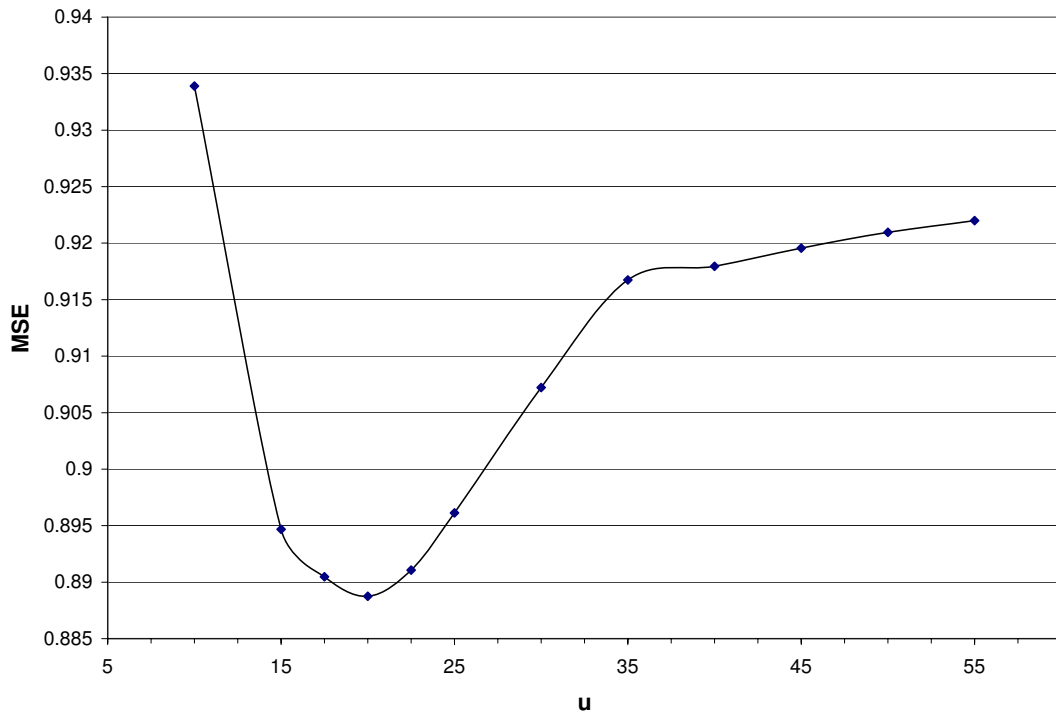Figure 6.1: Analysis of SRTR data: Cumulative incidence of kidney transplantation

Figure 6.2: Analysis of SRTR data: Effect of bounding $w_i(\widehat{\theta})$ on $MSE\{\widehat{\beta}_T(u)\}$, where MSE=mean square error and $\widehat{\beta}_T(u)$ is the estimate of β_T with upper bound, u.
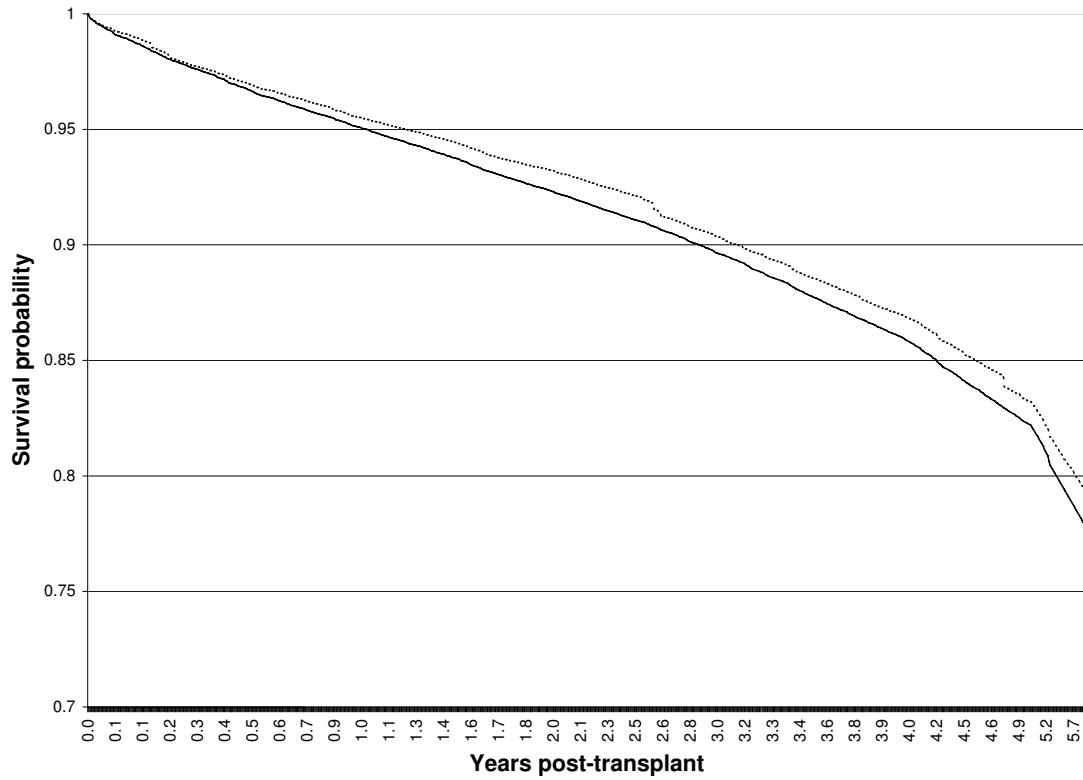
Figure 6.3: Analysis of SRTR data: Plot of estimated survival functions, $\exp\{-\widehat{\Lambda}_{0T}(t; \widehat{\beta}_T, \widehat{\theta})\}$ (solid line) and $\exp\{-\widehat{\Lambda}_{0S}(t; \widehat{\beta}_S)\}$ (dotted line).