Matching Methods for Estimating Effects of Time-dependent Treatment on Survival Outcomes

by

Danting Zhu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2019

Doctoral Committee:

Professor Douglas Schaubel, Chair
Associate Professor Ben Hansen
Research Assistant Professor Zhi He
Research Associate Professor Yun Li
Professor Randall Sukhyun Sung

Danting Zhu
dantzhu@umich.edu
ORCID iD: 0000-0002-2452-1703

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

In observational studies, treatment is often evaluated through its impact on survival time. However, when treatment initiation is time-dependent, existing methods are either inapplicable or yield treatment effect parameters with unsatisfactory interpretation. In this dissertation we propose methodology that evaluates the effect of time-dependent treatments in the context of survival functions.

In Chapter II, we estimate the average treatment effect among the treated (ATT) in the setting where the covariates remain constant over time. Since the counterfactual absence-of-treatment experience is not observable for treated patients, we match (using prognostic scores) to similar yet-untreated subjects to mimic this counterfactual experience. Novel components of the work include the emphasis on big data sets; use of personalized nonparametric survival function estimators; and the fact that, through grouping, survival curves (as opposed to patients) are ultimately matched.

In Chapter III, we propose alternative methods for the same general data structure as Chapter II. It is assumed that the data set is much smaller, implying different techniques to leverage the matching. Like Chapter II, methods proposed in Chapter III are applied to kidney transplant data.

In Chapter IV, we extend our method to the setting where adjustment covariates are time-dependent. As a generalization of Chapter II, methods in Chapter IV use matching and in addition, they incorporate the partly conditional model for the pretreatment death hazard to adjust for the time-dependent variables. Patients were matched on their residual survival time. Methods were then applied to the deceased donor liver transplant data to quantify the transplant effect.

CHAPTER I

Introduction

In observational studies, treatments are often evaluated by their perceived impact on survival time. However, when treatment initiation is time-dependent, existing methods are either inapplicable or yield treatment effect parameters with unsatisfactory interpretations. In medical studies with a time-to-event response, the treatment effect is usually measured by the hazard ratio (HR). We consider the setting wherein treatment can be represented as a non-reversible binary (0/1) indicator function. In this dissertation, we propose methodology that evaluates the effect of time-dependent treatments in the context of survival functions.

In Chapter II, we consider a time-dependent treatment scenario where the adjustment covariates remain constant over time; the effect of the treatment is quantified by the average treatment effect among the treated (ATT). Since treatment is initiated after the start of follow-up, the ideal comparison is between a subject that receives treatment at time s and the same subject under the scenario where the treatment does not exist. Since the counterfactual absence-of-treatment experience is not observable in practice for patients who receive treatment, our goal is to use other 'similar' subjects to mimic the treated subject's counterfactual absence-of-treatment experience. We use a conditional prognostic (risk) score to identify subjects with a

similar death rate in the absence of treatment. Subjects that remain untreated at time s are then matched to the subject treated at time s if their prognostic score is within a given distance from the score of the treated subject. We then obtain the post-treatment survival function for the treated patient by using the matched patients. Finally, we estimated the post-treatment survival function for treated patients through properly weighted survival curves. The methods in Chapter II are applied to data from Scientific Registry of Transplant Recipients to estimate the effect of deceased-donor kidney transplantation (KT) on survival among waitlisted end-stage renal disease (ESRD) patients.

In Chapter III, we also estimate the ATT of a time-dependent treatment, but under somewhat different conditions. The proposed method in Chapter II essentially estimates a survival curve for each matched set of yet-untreated patients, as such, this method requires a very large sample size. When the sample size is not big enough, the number of matched patients may not be sufficient to obtain sufficiently-precise individualized survival curves for each treated patient. In Chapter III, we proposed to use the conditional survival based on shared absence-of-treatment survival curves to obtain the post-treatment survival for treated patients, instead of matching . Similar to the matching method, we need to first obtain the pre-treatment prognostic score; but, instead of matching, we group patients based on the prognostic score. If we group patients finely enough, within each group patients would have very close pretreatment hazard, such that patients in the same group should have very similar marginal treatment-free survival probability. Finally, we estimate the post-treatment survival by the marginal survival probability of the corresponding group of patients, using the connection between marginal and conditional survival probability. Unlike Chapter II, this method does not require big data and is computationally efficient.

For the real- data application, we are intending to apply the proposed method on the same SRTR data, as in Chapter II, but within Region, on the center-level data. The purpose is to compare ATT for kidney transplantation between different centers in Region 10. The motivation of using the center-level data is to demonstrate the advantage of the methods in data sets that are not large.

In Chapter IV, we extend the method from Chapter II to a more complex setting. In the previous setting, there were no time-dependent adjustment covariates. However, many medical studies feature important time-dependent covariates that affect both the survival time and treatment assignment. For example, in liver transplantation, there are several time-dependent variables related to health status that determines eligibility of receiving a transplant (the 'treatment' of interest). In such scenarios, we should take these variables into account. Here, we propose to modify the matching method proposed in Chapter II. In particular, matching is based on a prognostic score derived from a partly conditional model; the score reflects residual survival probability (i.e., after the index patient's treated time). The partly conditional model for pretreatment survival, explicitly accounts for the information from the time-dependent adjustment covariates. The methods in Chapter IV are then applied to the SRTR data to evaluate the effect for deceased donor liver transplantation.

In summary, the proposed methods all deal with time-dependent treatment problems. Methods in Chapter II and Chapter III do not account for time-dependent adjustment covariates; each has various advantage over the other. Methods in Chapter IV was generalized from Chapter II and intended for more complex data with time-dependent adjusted covariates.

CHAPTER II

Matching Methods for Evaluating the Effect of A Time-dependent Treatment on the Survival Function

2.1 Introduction

Treatments are often evaluated through their perceived impact on survival time. In medical studies with a time-to-event response, the treatment effect is usually measured by the hazard ratio. We consider the setting wherein treatment can be represented as a non-reversible binary (0/1) indicator function. This set-up has a long history in survival analysis, dating back to the seminal analysis of the Stanford heart transplant data [6]. When treatment is assigned at baseline, estimation of the treated and untreated survival functions is straightforward. In the setting of our interest, treatment is time-dependent and not randomized. Specifically, each patient begins follow-up untreated with some patients eventually receiving treatment at some time point after baseline. We will use the average treatment effect among the treated (ATT) to quantify the treatment effect.

Our motivating example involves kidney transplantation. End-stage renal disease patients typically begin renal replacement therapy on dialysis, with some later receiving a kidney transplant. Usually, the referral for transplantation is not random, such that only medically suitable patients will be waitlisted for transplantation. Our goal is to estimate the effect of deceased-donor kidney transplantation compared

to dialysis ('untreated') with respect to the survival function and corresponding restricted mean survival time (RMST; i.e, area under the survival curve out to a fixed point). In our case, the ATT will be expressed as the difference between the average post-treatment survival function and the average survival function that would have been observed (among the transplanted patients) had, contrary to fact, kidney transplantation been unavailable. This counterfactual experience is unobservable in practice, however.

A number of methods have been proposed for estimating the effect of a time-dependent treatment. The most frequently used method is Cox regression with a time-dependent treatment indicator. The output of such a model is typically the hazard ratio (HR) comparing the treated versus untreated mortality hazards. However, investigators are often more interested in a contrast between treated and untreated survival functions, as opposed to the HR. The difference in survival functions is generally more interpretable than a HR, which is an instantaneous treatment effect, which requires proportional hazards to be meaningful. A non-parametric contrast between survival functions does not have the restriction of assuming proportionality between the pre- and post-treatment hazards. In addition to standard Cox regression, various models have been proposed in the arena. However, most existing methods do not express the treatment effect specifically in terms of survival functions. Marginal structural models [27, 12, 13] and their history-adjusted versions [23] estimate average causal effect (ACE) of treatment through a so-called causal HR. Structural nested failure time models [28, 21, 14] often use the accelerated failure time model to measure the treatment effect, such that the causal effect is estimated in terms of a ratio of mean survival times.

In order to compare each treated patient with their unobserved treatment-free

experience, in this report we will use prognostic score matching. Specifically, for a patient receiving treatment at time T, we select matches from patients currently alive, uncensored and untreated as of time T. The selected matches are intended to be very similar to the treated patients in terms of pre-treatment prognosis, such that their resulting untreated follow-up reflects what would have been observed for the treated patient in the absence of treatment. Here we will consider one-to-many matching since we want to construct a survival function for each set of matched patients. After obtaining patient-specific treatment-absent survival curves, the final treatment-absent survival curve is estimated through appropriate reweighting and averaging patient-level survival curves. To obtain the ATT, we need to estimate the restricted mean survival time for both the treatment and treatment-absent groups. For the treated group, we will use the analog of the treament-absent side.

The remainder of this chapter proceeds as follows. In Section 2, we introduce the notation and proposed methods. Section 3 presents simulation studies to demonstrate the performance of the treatment effect estimator in finite sample sizes and in various settings. An application to kidney transplantation is described in Section 4 using data from Scientific Registry of Transplant Recipients (SRTR). Some concluding remarks are offered in Section 5.

2.2 Proposed Methods

2.2.1 Notation and Set-up

We define the parameter of interest in the causal inference framework. Typically, this framework hypothesizes the setting wherein each individual has two potential outcomes, corresponding to the two possible treatment regimes (e.g., treated and untreated). In the counterfactual world, let $D_i^1(T_i)$ denote the potential death time (measured from 0) if patient i is treated at T_i. The counterfactual quantity $D_i^0(T_i)$ de-

notes the potential death time if, contrary to fact, patient i never received treatment. By definition, both $D_i^1(T_i)$ and $D_i^0(T_i)$ are greater than T_i and the counterfactuals are meaningfully defined only for individuals that receive treatment. Let $\boldsymbol{Z}_i$ be the covariate vector, which is assumed not dependent on time. We assume that $D_i^0(T_i)$ and $D_i^1(T_i)$ are conditionally independent given T_i and the observed covariates $\boldsymbol{Z}_i$, known as the strong ignorability assumption [31].

Next, we define notation for the observed data. Let D_i denote death time for subject i. The obsevation time for subject i is denoted by $U_i = D_i \wedge C_i$, with $a \wedge b = \min\{a, b\}$. The death indicator is given by $\Delta_i = I(D_i < C_i)$. The at-risk indicator is defined as $Y_i(t) = I(U_i \geq t)$ and the treatment indicator is defined by $\Delta_i^T = I(T_i < U_i)$. We also define $Y_i^1(t) = I(U_i \geq T_i + t)$, which equals 1 when subject i is at risk at time t and has already initiated treatment. Correspondingly, we define the post-treatment counting process increment, $dN_i^1(t) = Y_i^1(t)\, dN_i(T_i + t)$.

Since the treatment decision depends on $\boldsymbol{Z}_i$, and since some untreated patients may never be eligible for treatment, the ATT may be a more desirable treatment effect than the ACE in our setting. Our objective is hence to estimate the average treatment effect among the treated. For patient i, let $\widetilde{D}_i^1(T_i)$ denote the potential remaining survival time following treatment assignment at T_i, such that $\tilde{D}_i^1(T_i) = [D_i^1(T_i) - T_i]_+$. Conversley, let $\widetilde{D}_i^0(T_i)$ denote the potential remaining survival time if the patient never receives treatment such that $\tilde{D}_i^0(T_i) = [D_i^0(T_i) - T_i]_+$. The post-treatment survival function of our interest can then be defined as,

$$S_i^j(t) = P\left\{\widetilde{D}_i^j(T_i) > t | T_i, \boldsymbol{Z}_i, \right\}, \qquad j = 0, 1$$

and the subject-specific treatment effect can be defined as

$$\delta_i(t) = S_i^1(t) - S_i^0(t).$$

Hence, the average causal treatment effect among treated is given by

$$\delta(t) = S_1(t) - S_0(t),$$

where $S_1(t)$ and $S_0(t)$ are average survival functions,

$$S_j(t) = E\left\{S_i^j(t)\right\},$$

with the expectation being with respect to the distribution of $\{T, \boldsymbol{Z}|T < D\}$; i.e., the joint distribution of $(T, \boldsymbol{Z})$ among patients with $T < D$. To avoid identifiability issues, we need to have some restrictions pertaining to follow-up time. Specifically, if we let τ_C be the maximum censoring time, then our inference is restricted to $T \in [0, \tau_T]$ with $S_1(t)$ estimable on $t \in [0, \tau_1]$ for $\tau_T + \tau_1 \leq \tau_C$.

We also define the restricted mean survival time on $[0, L]$ with $L < \tau_1$ for both groups as $\mu_0(L) = \int_0^L S_0(u)du$ and $\mu_1(L) = \int_0^L S_1(u)du$, so that the difference in restricted mean life is denoted as $\Delta(L) = \mu_1(L) - \mu_0(L)$. Note that $\Delta(L) = \int_0^L \delta(t)dt$.

Our proposed method will use the risk class of each individual for each of the post-treatment and treatment-absent period, rather than use $(T_i, \boldsymbol{Z}_i)$ explicitly. We therefore define:

$$S^1(t|T_i, Z_i) = S^1(t|G_i^1, T_i)$$

$$S^0(t|T_i, Z_i) = S^0(t|G_i^0, T_i),$$

where G_i^1 and G_i^0 are the post-treatment and treatment-absent risk classes for treated individual i. Hence, instead of estimating $\delta(t) = E\left\{\delta_i(t|T_i, Z_i)\right\}$, we are instead estimating the very closely related quantity $\delta(t) = E\left\{\delta_i(t|G_i^1, G_i^0, T_i)\right\}$.

2.2.2 Estimation of $S^1(t)$

Since T_k is subject to right censoring by C_k, the uncensored T_k represent a biased sample of shorter values of time-to-treatment. A method that explicitly accounts

for censoring is required here so that the resulting nonparametric estimator of $S^1(t)$ represents an appropriate average over the $\{T, \boldsymbol{Z}|T < D\}$. Such an average should, naturally, not depend on the C distribution.

We use the Inverse Probability of Censoring Weighting (IPCW; [28]) to remedy the issue of dependent censoring. Specifically, for patient i, the weight is given by

$$w_i = \frac{\Delta_i^T}{P(C > T_i|T_i, \boldsymbol{Z})}.$$

For untreated patients $\Delta_i^T = 0$ such that $w_i = 0$. To estimate $P(C > T_i|T_i, \boldsymbol{Z})$, we assume the following Cox model for censoring,

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

which can be fiited using standard partial likelihood ([5]). To estimate $S_1(t)$, we focus on the prognostic score which is based on the hazard of death at time t given treated at time T_i,

$$\lambda_i^1\left(t|T_i, \boldsymbol{Z}_i\right) = \lim_{dt \to 0} \frac{1}{dt} P(t \leq D_i^1(T_i) < t + dt|\boldsymbol{Z}_i, T_i, T_i < D_i),$$

for which we assume the following post-treatment hazard model,

$$\lambda_i^1\left(t|T_i, \boldsymbol{Z}_i\right) = h\left\{\lambda_0^1(t),\ \boldsymbol{\beta}_1' \boldsymbol{Z}_i + \boldsymbol{\beta}_T{}' \boldsymbol{g}(T_i)\right\}, \tag{2.1}$$

such that $\lambda_i^1\left(t|T_i, \boldsymbol{Z}_i\right)$ presents a semi-parametric function of h; e.g., Cox model ([5]), additive hazards model ([1]), etc. Note that $\boldsymbol{\beta}_1$ is a vector of unknown parameters and $g(\bullet)$ is a vector of functions such that the effect of T is parametrized; and $\lambda_0^1(t)$ is the baseline hazard for post-treatment death.

For every treated patient we obtain the prognostic score $\boldsymbol{\beta}_1' \boldsymbol{Z}_i + \boldsymbol{\beta}_T \boldsymbol{g}(T_i)$ for post-treatment death. Then we group patients based on (2.1), which can be done by simply building grids or using empirical quantiles. Suppose eventually we have K

groups of treated patients. Patients in the same group have similar prognostic scores, such that we have approximately homogeneity with respect to post-treatment death risk within each group k. In order to estimate $S_1(t)$, we employ the weighted survival function

$$\hat{S}^1(t) = \left[\sum_{k=1}^{K}\sum_{i=1}^{n_k} w_{ki}\right]^{-1} \times \sum_{k=1}^{K}\sum_{i=1}^{n_k} w_{ki}\hat{S}^1_{ki}(t),$$

where $w_{ki} = w_i G^1_{ik}$ and $G^1_{ik} = I(\text{patient } i \text{ is in group } k)$, with $\hat{S}^1_{ki}(t)$ being the estimated survival probability for the ith patient in group k. Here, the $\widehat{S}^1_{ki}(t)$ can be based on Kaplan-Meier or Nelson-Aalen methods. Since patients in each group l have approximate homogeneity with respect to post-treatment death risk, $\hat{S}^1_{ki}(t)$ is the same across all patients in group k. Therefore we, have

$$\hat{S}^1(t) = \left[\sum_{k=1}^{K}\sum_{i=1}^{n_k} w_{ki}\right]^{-1} \times \sum_{k=1}^{K}\left(\sum_{i=1}^{n_k} w_{ki}\right)\hat{S}^1_k(t),$$

where $\hat{S}^1_k(t)$ is the estimated survival probability for group k. After rearranging the terms we have:

$$\hat{S}^1(t) = \left[\sum_{i=1}^{n} w_i\right]^{-1} \times \sum_{i=1}^{n} w_i \sum_{k=1}^{K} G^1_{ik}\hat{S}^1_k(t).$$

2.2.3 Estimation of $S^0(t)$

In this section, we will introduce a nonparametric estimator for $S_0(t)$. We begin by defining some additional notation. Specifically, let $Y^0_i(t) = I(U_i \wedge T_i \geq t)$, an indicator for being at risk and untreated as of time t, and define the following counting process increment, $dN^0_i(t) = Y^0_i(t)dN_i(t)$.

Since in practice we do not observe data to estimate $P\{D^0(T) > t|\boldsymbol{Z}, T, T < D\}$, the basic idea is to first obtain a pertinent estimator $\widehat{S}^0_i(t)$ for each treated patient with $\widehat{S}^0(t)$ then defined as an appropriately weighted average of $\widehat{S}^0_i(t)$ across all

treated patients. As an analog to $S^1(t)$, we propose to use the following estimator,

$$\widehat{S}^0(t) = \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} w_i \widehat{S}_i^0(t),$$

where w_i is inherited from each corresponding treated patient and hence has the same definition as in Section 2.2.2.

Next, we will describe a method for estimating $S_i^0(t)$, which involves matching methods to choose proper substitutions from the alive, uncensored and not-yet-treated patients.

2.2.4 Matching Method

The basic idea of the matching method is to match not-yet-treated and at-risk patients with similar pre-treatment hazard to each treated patient. Before matching, we first calculate the pre-treatment prognostic score, which reflects the treatment-free death hazard. The prognostic score is obtained through the working model,

$$\lambda_i^0(t|\boldsymbol{Z}_i) = \lambda_{00}(t) \exp\left\{\boldsymbol{\beta}_0' \boldsymbol{Z}_i\right\}.$$

We compare treated patient, k, and a potential control, ℓ, with respect to treatment-free prognostic score through the difference in linear predictor

$$\psi_{\ell:k} \equiv \boldsymbol{\beta}_0'(\boldsymbol{Z}_\ell - Z_k).$$

Patient ℓ is a suitable match to treated patient k to the extent that $\psi_{\ell:k}$ is close to 0. To avoid inappropriate matches, we add the restriction that $\psi_{\ell:k}$ needs to be within a caliper, $|\psi_{\ell:k}| \leq \varepsilon$,where ε is a predefined small number. Note that if we use different models (e.g., additive hazards model) to obtain the prognostic score, then the criteria to select matched patients is identical as the one for proportional hazards model; i.e., the difference in linear predictors.

By matching 'qualified' patients to each treated patients, we will obtain matched sets for each of treated patient. Within each matched set, patients have approximate homogeneity with respect to pre-treatment death hazard and can be viewed as the counterfactual cases corresponding to the treated patient to which they were matched. Using the matched patients, we can estimate the survival probability $S_i^0(t)$ for each treated patient i by using non-parametric estimators (Kaplan-Meier or Nelson-Aalen).

2.2.5 Variance Estimator for $\widehat{S}^1(t)$

In this subsection we first present heuristic argument regarding the asymptotic behavior of $\widehat{S}^1(t)$. Asymptotic properties for matching-derived estimators are notoriously complex and difficult to establish. In order to maintain the original focus of the dissertation, we therefore provide arguments that lead to a reasonably (if not tightly) defensible variance estimator. As n goes to ∞, we let the number of groups K go to ∞ as well, but at a slower rate than n such that the the number of individuals in each group will also go to ∞. This being the case, individuals in the same group can be viewed as identical with respect to pre-treatment death hazard. If there is no censoring, by the definition of $S^1(t)$, $\widehat{S}^1(t) = n^{-1}\sum_{k=1}^{K} n_k \widehat{S}_k^1(t)/n$. When $K \to \infty$, $\widehat{S}^1(t) = n^{-1}\sum_{i=1}^{n} S_i^1(t)$, by Uniform Weak Law of Large Numbers (UWLLN), $\widehat{S}^1(t) \to S^1(t)$ in probability for all $S^1(t)$.

When there is no censoring , $w_{k\cdot}\left\{\sum_{k=1}^{K} w_{k\cdot}\right\}^{-1}$ can be viewed as a density function corresponding to a certain function of $(T, \mathbf{Z})$ given prognostic score $\boldsymbol{\beta}_P^{'}\boldsymbol{Z}_i + \beta_T T_i$ falling in kth interval. We denote this density as $f_\phi(k) = w_{k\cdot}\left\{\sum_{k=1}^{K} w_{k\cdot}\right\}^{-1}$. As $n \to \infty$, $K \to \infty$, such that each interval of the prognostic score will be close to a value on the domain. Thus, we can re-write $f_\phi(k)$ as $f_\phi(x)$ with $x \in (-\infty, \infty)$.

Therefore we have

$$\widehat{S}^1(t) = \int_{-\infty}^{\infty} f_\phi(x)\widehat{S}^1_x(t)dx = \int_{-\infty}^{\infty} \widehat{S}^1_x(t)dF_\phi(x), \tag{2.2}$$

where F_ϕ is the CDF of ϕ.

In order to derive the asymptotic distribution of $\widehat{S}^1(t)$ we need to define an additional set of notation. Let $Y_{ik}(t) = G^1_{ik}Y_i(t)$, $dM^1_{ik}(t) = G^1_{ik}\left\{dN_i(t) - Y_i(t)d\Lambda^1_k(t)\right\}$. We also define $\rho = E(\Delta^T_i)$ which can be estimated by $n^{-1}\sum_{i=1}^n \Delta^T_i$. We then have

$$n^{1/2}\rho^{1/2}\left\{\widehat{\Lambda}^1_k(t) - \Lambda^1_k(t)\right\} = n^{1/2}\rho^{1/2}\sum_{i=1}^n \varphi^1_{ik}(t),$$

where $\widehat{\Lambda}^1_j(t)$ is the estimator of post-treatment cumulated hazard function for group k, and $\varphi^1_{ik}(t) = \int_0^t \pi^{-1}(u)dM^1_{ik}(u)$, where $\pi(u) = P(U \geq u)$. Under mild regularity conditions, $\{\varphi^1_{1k}(t), \ldots, \varphi^1_{nk}(t)\}$ are independent and identically distributed mean 0 variates. As a result, $n^{1/2}\left\{\widehat{\Lambda}^1_k(t) - \Lambda^1_k(t)\right\}$ would be expected to converge to asymptotically to a mean-zero Normal distribution with variance $E\left[\varphi^1_{1k}(t)^2\right]$ by the Multivariate Central Limit Theorem. By applying the Functional Delta Method, we obtain that $n^{1/2}\left\{\widehat{S}^1_k(t) - S^1_k(t)\right\}$ is also asymptotically mean-zero Normal with variance estimator,

$$\widehat{\sigma}^2_k(t) = n^{-1}\sum_{i=1}^n \left\{\widehat{S}^1(t)\widehat{\varphi}^1_{ik}(t)\right\}^2,$$

where $\widehat{\varphi}^1_{ik}(t) = \int_0^t \widehat{\pi}^{-1}(u)d\widehat{M}^1_{ik}(u)$ and $d\widehat{M}^1_{ik}(u) = G_{ik}\left\{dN^1_i(u) - Y^1_{ik}(u)d\Lambda^1(u)\right\}$.

As discussed above, $\widehat{S}^1(t)$ will converge in probability to its limiting value $S^1(t) = \int_{-\infty}^{\infty} S^1_x(t)dF_\phi(x)$, where we define $S^1(t) = \sum_{k=1}^K f_\phi(k)S^1_k(t)$.

Next, we consider the asymptotic distribution of $\widehat{S}^1(t)$. We can write,

$$\begin{aligned}
\widehat{S}^1(t) - S^1(t) &= \sum_{k=1}^{K} f_\phi(k) \left\{ \widehat{S}^1_k(t) - S^1_k(t) \right\} \\
&= \sum_{k=1}^{K} f_\phi(k) \left\{ - \sum_{i=1}^{n_T} S_k(t) \varphi^1_{ik}(t) \right\} \\
&= - \sum_{k=1}^{K} f_\phi(k) S^1_k(t) \sum_{i=1}^{n} \varphi^1_{ik}(t) \\
&= - \sum_{i=1}^{n} \sum_{k=1}^{K} f_\phi(k) S^1_k(t) \varphi^1_{ik}(t) \\
&= - \sum_{i=1}^{n} \varphi^1_{i\bullet}(t),
\end{aligned}$$

where $\varphi^1_{i\bullet}(t) = \sum_{k=1}^{K} f_\phi(k) S^1_k(t) \varphi^1_{ik}(t)$. Since $\varphi^1_{i\bullet}(t)$ are also independent and mean-zero, $n^{1/2} \left\{ \widehat{S}^1(t) - S^1(t) \right\}$ converges in distribution to a zero-mean Normal with a variance that can be consistently estimated by

$$\widehat{\sigma}^2_1 = n^{-1} \sum_{i=1}^{n} (\widehat{\varphi}^1_{i\bullet})^2,$$

where we define

$$\widehat{\varphi}_{i\bullet}(t) = \sum_{k=1}^{K} f_\phi(k) \widehat{S}^1_k(t) \widehat{\varphi}^1_{ik}(t).$$

2.2.6 Variance Estimator for $\widehat{S}^0(t)$

As long as $\widehat{S}^0_i(t)$ is an consistent estimator (e.g. Kaplan-Meier or Nelson-Aalen estimator) for $S^0_i(t)$, $\widehat{S}^0(t)$ be a weighted average of constant estimators and, therefore, should converge in probability to $S^0(t)$, where $S^0(t)$ is defined as

$$S^0(t) = \int_{-\infty}^{\infty} f_\phi(x) S^0_x(t) dx = \int_{-\infty}^{\infty} S^0_x(t) dF_\phi(x), \tag{2.3}$$

with $f_\phi(x)$ defined as in the previous section. A discretized version of (4.10) can be expressed as $S^0(t) = \sum_{k=1}^{K} f_\phi(k) S^0_k(t)$, where $f_\phi(k) = w_{k\bullet}$.

Next we derive the limiting distribution for $\widehat{S}^0(t)$, we start with the limiting distribution for $\widehat{S}^0_i(t)$.

For the matching method, we need to first define several new quantities. Let $G^0_{ik} = I(G^0_i = k)$. Consistently we define $Y^0_{ik} = G^0_{ik}Y^0_i(t)$ and $dM^0_{ik}(t) = G^0_{ik}\{dN^0_i(t) - Y^0_i(t)d\Lambda^0(t)\}$. Similar to the process for $\widehat{S}^1_k(t)$, analogous arguments leads to

$$n^{1/2}\left\{\widehat{S}^0_k(t) - S^0_k(t)\right\} = -n^{1/2}\sum_{i=1}^{n} S^0_k(t)\varphi^0_{ik}(t)$$

asymptotically. Different from the treatment side where each subject k can appear only once, a given subject i in the treatment-free side can be matched to sevreal treated patients. As such, the asymptotically independent terms with respect to the treatment-free side are given by

$$\varphi^0_{ik}(t) = \int_0^t \pi_0^{-1}(u)dM^0_{ik}(u),$$

such that $n^{1/2}\left\{\widehat{S}^0_k(t) - S^0_k(t)\right\}$ converges in distribution to a zero-mean Normal with a variance that can be consistently estimated by

$$(\widehat{\sigma}^0_k)^2(t) = n^{-1}\sum_{i=1}^{n}\left\{\widehat{S}^0_k(t)\widehat{\varphi}^0_{ik}\right\}^2,$$

where we define $\widehat{\varphi}^0_{ik} = \int_0^t \widehat{\pi}^{-1}(u)d\widehat{M}^0_{ik}(u)$ with $d\widehat{M}^0_{ik}(u) = G^0_{ik}\left\{dN^0_i(u) - Y^0_i(u)d\widehat{\Lambda}^0(u)\right\}$.

Similar to $\widehat{S}^1(t)$, $\widehat{S}^0(t) - S^0(t) = -\sum_{i=1}^n \varphi^0_{i\bullet}$ with $\varphi^0_{i\bullet}(t) = \sum_{k=1}^{n_T} f_\phi(k)S^0_k(t)\varphi^0_{ik}(t)$. Therefore $n^{1/2}\left\{\widehat{S}^0(t) - S^0(t)\right\}$ should then be asymptotically mean-zero Normal with variance estimator,

$$\widehat{\sigma}^2_0(t) = n^{-1}\sum_{i=1}^{n}(\widehat{\varphi}^0_{i\bullet})^2$$

where we define $\widehat{\varphi}^0_{i\bullet}(t) = \sum_{k=1}^{n_T} f_\phi(k)\widehat{S}^0_k(t)\widehat{\varphi}^0_{ik}(t)$.

For the matching method, combining the results above, we can represent $n^{1/2}\left\{\widehat{\delta}(t) - \delta(t)\right\} = n^{1/2}\sum_{i=1}^n\{\varphi^1_{i\bullet}(t) - \varphi^0_{i\bullet}(t)\}$, where $\{\varphi^1_{i\bullet}(t) - \varphi^0_{i\bullet}(t)\}$ components are independent and

identically distributed with mean 0. Note that the representation above accounts for the possibility that patients may contribute follow-up on both the $S^0(t)$ and $S^1(t)$ sides. The quantity $n^{1/2}\left\{\widehat{\delta}(t)-\delta(t)\right\}$ should converge asymptotically to a Normal variate with mean 0 and a variance that can be consistently estimated by

$$\widehat{\sigma}^2_\delta(t) = n^{-1}\sum_{i=1}^{n}\left\{\varphi^1_{i\bullet}(t)-\varphi^0_{i\bullet}(t)\right\}^2.$$

For computational convenience, we propose to use the bootstrap method to compute the asymptotic variance, as the point estimators are fast to compute. We evaluate the performance of bootstrap in the next section.

2.3 Simulation Studies

We conducted simulations to demonstrate the properties of the proposed method in finite samples. The treatment time T was generated from an exponential distribution with hazard $\lambda_0^T \exp\{\beta_{T1}Z_1+\beta_{T2}Z_2\}$ while treatment-free death times D^0 were generated as exponential with hazard $\lambda_0^D \exp\{\beta_{D1}Z_1+\beta_{D2}Z_2\}$. Here both Z_1 and Z_2 are confounders, affecting both T and D^0. Censoring times C were generated from an exponential distribution with hazard $\lambda_0^C \exp\{\beta_{C1}Z_1+\beta_{C2}Z_2\}$. Times between treatment and death $(D^1-T)_+$ were generated from exponential distribution with rate $\lambda_0^1 \exp\{\beta_{10}Z_1+\beta_{11}Z_2+\beta_{12}T\}$, where we set $\lambda_0^1 = a\lambda_0^D$, $\beta_{10}=\beta_{D1}$ and $\beta_{11}=\beta_{D2}$, for various values as given below. Baseline covariates Z_1 and Z_2 were each generated from a Uniform(-1, 1) distributions. if we denote the actual realized death time as D, then for treated patients $D = D^1 = T + (D^1-T)_+$ and for untreated patients $D = D^0$. There were $n = 2500$ subjects in all simulation configurations, with each data configuration replicated 500 times. To obtain the standard deviation we generate 25 bootstrap data sets for each replicate.

In practice, we observe the minimum of T, D^0 and C. In simulations, however, we

always observe T, D^0, $(D^1 - T)_+$ and C for all patients. True values of $S_1(t)$, $S_0(t)$, $\delta(t)$ and $\Delta(L)$ were obtained using Monte-Carlo on these counterfactuals. Naturally, for the purpose of computing $\widehat{S}_0(t)$, only $[(D \wedge C \wedge T), N^D(D \wedge C \wedge T)]$ were used; similarly, only $[(D \wedge C - T)_+, N^D(D \wedge C)]$ were used for subjects with $(D \wedge C > T)$ for the purpose of computing $\widehat{S}_1(t)$. Hence, $\widehat{\delta}(t)$ and $\widehat{\Delta}(L)$ were, for each replicate, only based on data that would in reality be observed.

After generating the data, prognostic scores representing pre-treatment history were obtained from model $\lambda_i^0(t|\boldsymbol{Z}_i) = \lambda_{00} \exp\{\beta_{00} Z_1 + \exp \beta_{01} Z_2\}$. Subjects are matched if $|\log(\psi_{\ell:k})| \leq 0.05$. For all simulations configurations we set $\tau_T = 10$.

In the first set of simulations, we examine the bias, empirical standard deviation and bootstrap standard error (BSE) of the proposed estimators with various treatment effect under light censoring with $\lambda_0^C = 0.015$, where approximately 15% of individuals get censored. We also compute 95% empirical coverage probability (CP). We vary a from 0.8, 1 and 1.2 to change the treatment effect from moderate, null to negative. The remaining parameters are set equal across the three scenarios: $\lambda_0^D = 0.05$, $\lambda_0^T = 0.03$, $\beta_{D1} = \beta_{T1} = \beta_{C1} = \log 2$, $\beta_{12} = \log 3/500$ and $\beta_{D2} = \beta_{T2} = \beta_{C2} = \log 3$.

In the second set of simulations, we examine the properties of the proposed estimators under moderate censoring. The parameter setting is same as the first set of simulations except the censoring parameter was changed to $\lambda_0^C = 0.02$, which results in approximately 30% censoring.

Results for the first and second set of simulations are shown in Table 2.1 and Table 2.2, respectively. In both tables, the absolute bias of $\widehat{S}^1(t)$ $\widehat{S}^0(t)$ and $\widehat{\delta}(t)$ range from 0.001 to 0.015. The biases of $\widehat{\mu}_0(15)$, $\widehat{\mu}_1(15)$ and $\widehat{\Delta}(15)$ were somewhat bit larger but, considering the scale of these quantities, was still negligible. The

BSE is generally close to the ESD, such that the empirical coverage probability was around 0.95, except for $S_0(15)$. The estimation of $S_0(t)$ is more sensitive to the censoring percentage the $S^1(t)$. The censoring percentage did not appear to effect the results of $\widehat{S}_1(t)$, but the bias of $\widehat{S}_0(t)$ becomes more pronounced with larger censoring percentages.

2.4 Application

We applied our proposed methods to estimate the effect of deceased-donor kidney transplantation (KT) ($j = 1$) on survival in the absence of KT ($j = 0$) among waitlisted end-stage renal disease patients. Data were obtained from Scientific Registry of Transplant Recipients. The SRTR data system includes data on all donors, waitlisted candidates, and transplant recipients in the United States, as submitted by the members of the Organ Procurement and Transplantation Network (OPTN), and has been described elsewhere. The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors.

The study population included $n = 112,901$ patients aged ≥ 18 and listed between 01/01/2003 and 12/31/2013. Follow-up time begins at the date when patients got listed and ends at earliest of death, loss to follow-up, or the end of the observation period (12/31/2013). Adjustment covariates for the $\lambda_i^0(t)$ model included height, weight, years on dialysis (prior to waitlisting), calender year of listing, albumin, diabetes, hypertension, panel reactive antibiotics (PRA), age, angina, blood type, symptomatic peripheral vascular disease (PVD), race, gender, calendar year of transplant and Kidney Donor Risk Index (KDRI) [26].

In this application, we set τ_T and τ_1 to 3 years and 5 years, respectively. A total

Table 2.1: Simulation results for light censoring. ESD=empirical standard deviation; BSE=bootstrapped standard error (based on 25 bootstrap samples); CP=empirical coverage probability

Setting	t	Quantity	True	BIAS	ESD	BSE	CP
Moderate	5	$S^1(t)$	0.731	-0.002	0.022	0.021	0.93
treatment	10		0.554	-0.002	0.026	0.025	0.94
effect	15		0.431	0.001	0.025	0.025	0.95
	5	$S^0(t)$	0.678	0.004	0.013	0.014	0.98
	10		0.486	0.008	0.018	0.018	0.93
	15		0.364	0.012	0.020	0.020	0.88
	5	$\delta(t)$	0.054	-0.006	0.024	0.024	0.97
	10		0.068	-0.010	0.029	0.028	0.92
	15		0.067	-0.011	0.029	0.029	0.93
	15	$\mu_0(t)$	9.096	0.086	0.198	0.200	0.92
	15	$\mu_1(t)$	9.911	-0.019	0.271	0.263	0.94
	15	$\Delta(t)$	0.815	-0.105	0.300	0300	0.93
Null	5	$S^1(t)$	0.680	-0.002	0.023	0.022	0.96
treatment	10		0.487	-0.001	0.026	0.025	0.93
effect	15		0.365	0.002	0.024	0.024	0.95
	5	$S^0(t)$	0.678	0.004	0.013	0.014	0.98
	10		0.486	0.008	0.018	0.018	0.93
	15		0.364	0.013	0.020	0.020	0.88
	5	$\delta(t)$	0.003	-0.006	0.024	0.024	0.96
	10		0.001	-0.009	0.028	0.028	0.95
	15		0.001	-0.010	0.028	0.029	0.93
	15	$\mu_0(t)$	9.112	0.086	0.198	0.200	0.92
	15	$\mu_1(t)$	9.123	-0.014	0.272	0.269	0.94
	15	$\Delta(t)$	0.011	-0.100	0.295	0.301	0.94
Negative	5	$S^1(t)$	0.635	-0.003	0.024	0.023	0.96
treatment	10		0.431	-0.000	0.024	0.024	0.97
effect	15		0.310	0.004	0.023	0.023	0.94
	5	$S^0(t)$	0.678	0.004	0.013	0.014	0.99
	10		0.486	0.008	0.018	0.018	0.93
	15		0.364	0.012	0.020	0.020	0.88
	5	$\delta(t)$	-0.043	-0.007	0.024	0.025	0.95
	10		-0.055	-0.008	0.027	0.028	0.94
	15		-0.054	-0.007	0.028	0.028	0.94
	15	$\mu_0(t)$	9.108	0.086	0.198	0.200	0.92
	15	$\mu_1(t)$	8.442	-0.008	0.269	0.266	0.94
	15	$\Delta(t)$	-0.666	-0.094	0.291	0.300	0.95

Table 2.2: Simulation results for moderate censoring. ESD=empirical standard deviation; BSE=bootstrapped standard error (based on 25 bootstrap samples); CP=empirical coverage probability

Setting	t	Quantity	True	BIAS	ESD	BSE	CP
Moderate	5	$S^1(t)$	0.731	-0.002	0.022	0.022	0.92
treatment	10		0.554	-0.001	0.024	0.025	0.95
effect	15		0.431	0.006	0.025	0.025	0.94
	5	$S^0(t)$	0.678	0.005	0.013	0.015	0.99
	10		0.486	0.010	0.019	0.019	0.95
	15		0.364	0.016	0.021	0.021	0.89
	5	$\delta(t)$	0.054	-0.007	0.025	0.025	0.97
	10		0.068	-0.010	0.030	0.030	0.95
	15		0.067	-0.011	0.030	0.031	0.93
	15	$\mu_0(t)$	9.101	0.112	0.202	0.213	0.92
	15	$\mu_1(t)$	9.908	0.004	0.262	0.269	0.92
	15	$\Delta(t)$	0.807	-0.108	0.313	0319	0.93
Null	5	$S^1(t)$	0.680	-0.002	0.024	0.023	0.95
treatment	10		0.487	0.000	0.027	0.025	0.93
effect	15		0.365	0.006	0.025	0.025	0.94
	5	$S^0(t)$	0.678	0.005	0.013	0.014	0.99
	10		0.486	0.001	0.019	0.019	0.95
	15		0.364	0.015	0.021	0.021	0.86
	5	$\delta(t)$	0.003	-0.007	0.025	0.025	0.96
	10		0.001	-0.010	0.030	0.030	0.93
	15		0.001	-0.010	0.029	0.031	0.94
	15	$\mu_0(t)$	9.107	0.106	0.204	0.211	0.91
	15	$\mu_1(t)$	9.120	0.000	0.282	0.274	0.93
	15	$\Delta(t)$	0.013	-0.106	0.310	0.315	0.93
Negative	5	$S^1(t)$	0.635	-0.004	0.024	0.024	0.96
treatment	10		0.431	0.001	0.025	0.025	0.96
effect	15		0.310	0.008	0.024	0.024	0.93
	5	$S^0(t)$	0.678	0.005	0.013	0.014	0.99
	10		0.486	0.009	0.019	0.019	0.95
	15		0.364	0.015	0.021	0.021	0.86
	5	$\delta(t)$	-0.043	-0.008	0.026	0.026	0.96
	10		-0.055	-0.008	0.028	0.029	0.94
	15		-0.054	-0.007	0.029	0.030	0.95
	15	$\mu_0(t)$	9.111	0.106	0.204	0.211	0.92
	15	$\mu_1(t)$	8.437	0.008	0.276	0.275	0.95
	15	$\Delta(t)$	-0.664	-0.098	0.304	0.315	0.94

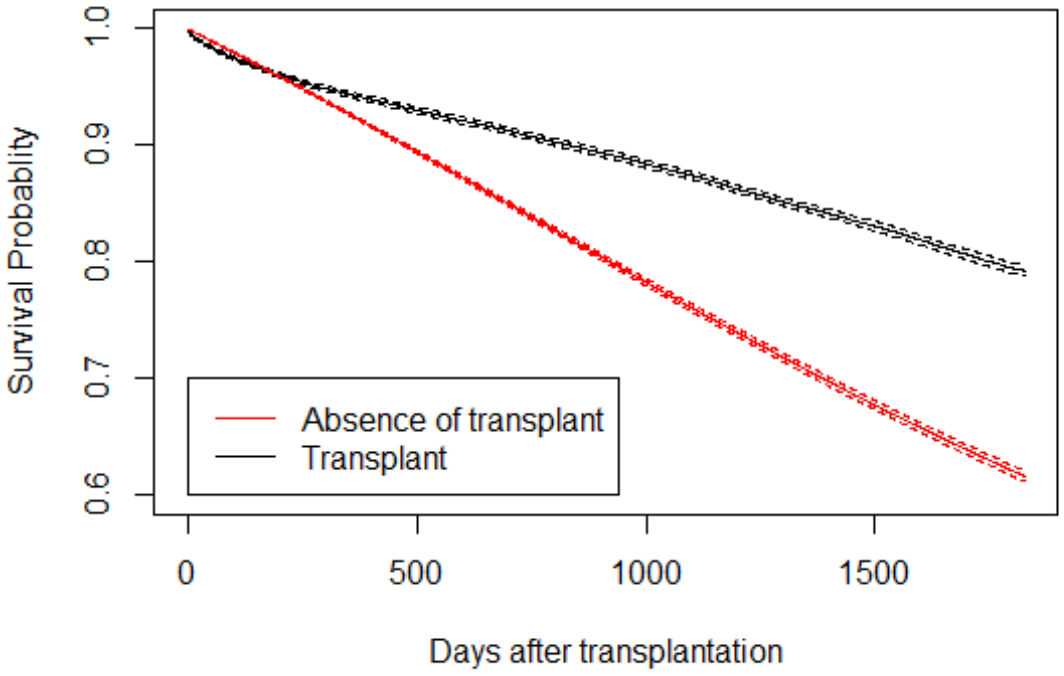

Figure 2.1: Analysis of SRTR Data (n=112,901) for 5-year post-transplant survival.

of 37,724 patients received KT (33.4%) and 41,453 deaths were observed. We set the caliper width of prognostic score matching $\epsilon = 0.02$.

The estimated survival curves on [0,5] year interval of the two groups are presented in Figure 1, as well as the corresponding confidence band. We also magnify $S_1(t)$ and $S_0(t)$ on the [0,1] year interval in Figure 2 such that the crossing of the survival curves become more apparent.

Table 3 shows the estimates of $S_1(t)$, the average survival probability from the time of transplant among patients received transplantation, and $S_0(t)$ intended to represent the survival probability to which the transplanted patients would have been observed in the absence of transplantation, and $\widehat{\delta}(t) = \widehat{S}_1(t) - \widehat{S}_0(t)$ and $\widehat{\Delta}(5) =$

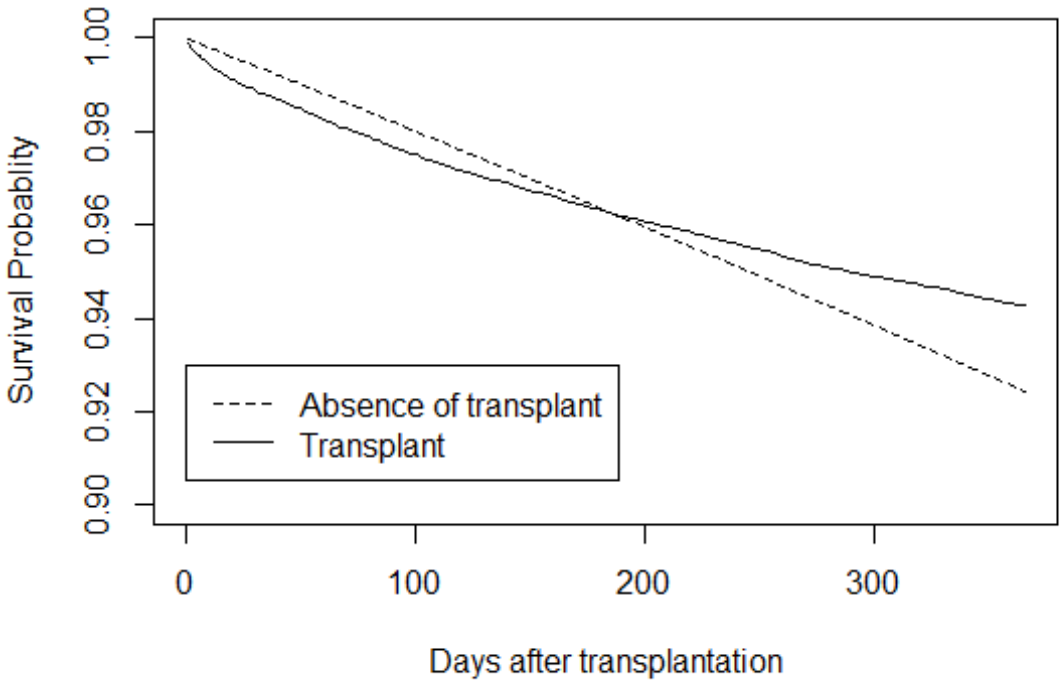

Figure 2.2: Analysis of SRTR Data (n=112,901) for 1-year post-transplant survival.

Table 2.3: Analysis of SRTR data (n=112,901): Evaluation of the effect on survival (and RMST) of kidney transplantation.

t (year)	Quantity	Estimate	BSE
1	$S_1(t)$	0.943	0.001
3		0.875	0.002
5		0.793	0.002
1	$S_0(t)$	0.924	0.001
3		0.762	0.002
5		0.616	0.002
1	$\delta(t)$	0.018	0.001
3		0.113	0.002
5		0.176	0.003
5	$\mu_1(t)$	1625 days	2.387
5	$\mu_0(t)$	1468 days	2.280
5	$\Delta(t)$	157 days	3.277

$\int_0^5 \widehat{\delta}(t)dt$. All quantities are estimated at 1, 3, 5 years with their corresponding bootstrap standard error. We also presented the restricted mean lifetime at 5 years for both groups $\mu_0(t)$, $\mu_1(t)$ and their contrast $\Delta(t)$. Based on the test of $\delta(t)$, there are significant difference in survival probability between the two groups at all three time points. Deceased donor kidney transplantation is significantly beneficial for survival as the restricted mean survival time for transplant patients is approximately 5 months longer than that of matched waitlisted patients.

2.5 Discussion

In this report, we proposed methods to estimate the average causal effect among the treated of a time-dependent treatment. In particular, the proposed treatment effect contrasts post-treatment survival with the survival function that would apply to treated patients had they, contrary to fact, not received treatment. To estimate the survival of treated patients in the absence of treatment, we proposed a matching method to create a group of patients that is considered as counterfactual version for each treated patient and then properly average over survival functions. Heuristic argument towards establishing asymptotic variances were provided. For computa-

tional conveniences, a bootstrap method is employed to estimate the standard error. The proposed methods were shown through simulations to work well in finite sample sizes.

The proposed method is non-parametric in the sense that the assumed models only contribute to the prognostic score used for matching. For the treatment effect, we target directly at survival function and restricted mean lifetime, which is more flexible compared to a hazard ratio. Due to the nature of non-parametric estimation, the proposed methods do not require that pre- and post-treatment hazards are proportional or have any particular relationship to each other. In addition, our methods can handle big data sets, since the number of survival function estimators to average over increases much slower than total sample size.

Another similar matching method is proposed by [18] where they first match patients and then pool the strata together. This method relies heavily on IPCW, which, in conjunction with the required data augmentation, makes the method computationally burdensome in large data sets. Compared to [18], the proposed methods only needs inverse weighting probability on each treatment time, (i.e. not a time-dependent weight) which speeds up computing time considerably in big-data settings.

There are several existing methods related to those proposed. However, these methods either do not target at the survival function, or do not estimate average treatment effect among the treated. Structural nested failure time model assumed g-estimation [29] measures the ratios of mean survival time and marginal structural models [27] usually targets the hazard ratio. The time-dependent propensity score matching proposed by Lu et al. [22]. also targets the hazard ratio. Obtaining survival function from these methods is either difficult or impossible. In contrasting the proposed method with time-dependent propensity score matching method or other

sequential stratification method [32, 33], the proposed method also has advantage in big datasets as rather than combining matched subjects and computing suvrival function, we are estimating tons of survival functions with small sample size and then combining them.

The proposed method makes use of the prognostic score to match yet-untreated patients to each treated patient. As mentioned above, another viable alternative would be propensity socre matching, i.e., matching on the probability of receiving treatment [22]. A propensity score measures the patient-specific rate of treatment assignment, given the covariates. Our goal, however,was to create a comparison group that mirrored the treatment-free experience of a subject treated at time s. It was therefore necessary to ensure that the event trajectories up until s were the same between treated and control subjects, a property that the propensity score does not preserve.

Due to the nature of matching method, proposed methods have the advantages of handling covariates higher dimensions with greater robustness towards model misspecification. On the other hand, matching also relies on several assumptions such as no unmeasured confounding and overlapping support between treated and untreated groups. In the process of matching, caliper width is subject to change from data to data. We need to ensure every treated subject has a sufficient number of matched subjects and that the matched sets contain only subjects that are sufficiently similar.

One limitation of the proposed method is we only considered time-constant covariates, but in practice it is possible that time-dependent covariates exist which may lead to violations of the ignorability assumption and bias the treatment effect. There are very few existing methods [9] can estimate time-dependent treatment effect with time-dependent covariates, but these methods are usually hard to obtain

survival functions and rely more on model assumptions. Although in the context of kidney transplantation the issue of unmeasured time-dependent confounding may not be so severe, time-dependent variables are common in other data applications, for example, liver transplantation data where the rank on waiting list in the U.S. can change dynamically depending on a patient's health conditions. Moreover, as health care systems become increasingly digitalized, longitudinal information will be more available in increasing number of registry databases. Therefore a meaningful further step will be to investigating estimating the same treatment effect in the presence of time-dependent covariates. This challenge is addressed in Chapter IV.

CHAPTER III

Semiparametric Survival Methods for Evaluating the Effect of A Time-dependent Treatment on the Survival Function

3.1 Introduction

We will present this method under the same framework of Method 1, as the two methods target similar settings. We still consider kidney transplantation as a motivating example. Our goal is to estimate the effect of kidney transplantation compared to dialysis ("untreated") with respect to the survival function and corresponding restricted mean survival time (RMST; i.e, area under the survival curve out to a fixed point). In our case, the ATT will represent the difference between the average post-treatment survival function and the average survival function that would have been observed (among the transplanted patients) had, contrary to fact, kidney transplantation been unavailable. This counterfactual experience is unobservable in practice.

Methods proposed in Chapter II used matching to estimate the survival probability in the absence of treatment, among treated patients. Since treatment-absent survival curves are essentially estimated, individually for each treated patient, the methods will work best on big data sets. For smaller data sets, the estimation of the matched-set-specific survival curves may be imprecise due to a smaller number of available matches. In this chapter, we consider a conditional survival method that

targets the same causal estimand as in Chapter II, but is applicable to much smaller sample sizes.

There have been a number of methods proposed for estimating the effect of a time-dependent treatment effect. The most widely used quantity to measure the effect size is the hazard ratio (HR). However, investigators are often more interested in a contrast between treated and untreated survival functions as opposed to HR. Moreover, a non-parametric contrast between survival functions does not have the restriction of assuming proportionality between the pre- and post-treatment hazard functions. In addition to standard Cox regression, various pertinent methods have been proposed. However, most existing method do not express the treatment effect specifically in terms of the survival function. Marginal Structural Models [27, 12, 13] and their history-adjusted versions [23] estimate the average causal effect (ACE) of treatment through the HR. Structural Nested Failure Time models [28, 21, 14] often use the accelerated failure time model as a basis for estimating the treatment effect, such that the causal effect is estimated in terms of a ratio of mean survival times.

In order to compare each treated patient with their unobserved treatment-free experience, in this chapter we will use a conditional survival method. By grouping the pre-treatment prognostic score for every patient, we obtain the corresponding 'similar' group of patients for each treated patient. Therefore the treatment-absent survival curves can be estimated by conditional method on the group of qualified patients. For the treated side, we will continue to use the method proposed in Chapter II. The ATT will be defined in terms of the integral of the difference in two survival curves, as in Chapter II.

3.2 Methods

3.2.1 Notation

The notation for the set up om Chapter III will be the same as that of Chapter II. We define the parameter of interest in the causal inference framework. In the counterfactual world, let $D_i^1(T_i)$ denote the potential death time (measured from 0) if patient i is treated at T_i. The counterfactual quantity $D_i^0(T_i)$ denotes the potential death time if, contrary to fact, patient i never received treatment. By definition, both $D_i^1(T_i)$ and $D_i^0(T_i)$ are greater than T_i and the counterfactuals are meaningfully defined only for individuals that receive treatment. Let $\boldsymbol{Z}_i$ be the covariate vector, which is assumed to not be dependent on time. We assume that $D_i^0(T_i)$ and $D_i^1(T_i)$ are conditionally independent given T_i and the observed covariates $\boldsymbol{Z}_i$, known as the strong ignorability assumption [31].

Next, we define notation for the observed data. Let D_i denote death time for subject i. The obsevation time is denoted by $U_i = D_i \wedge C_i$, with $a \wedge b = \min\{a, b\}$. The death indicator is given by $\Delta_i = I(D_i < C_i)$. The at-risk indicator is defined as $Y_i(t) = I(U_i \geq t)$ and the treatment indicator is defined by $\Delta_i^T = I(T_i < U_i)$. We also define $Y_i^1(t) = I(U_i \geq t, T < t)$, which equals 1 when subject i is at risk at time t and has already initiated treatment. Correspondingly, we define the post-treatment counting process increment, $dN_i^1(t) = Y_i^1(t)\, dN_i(t)$.

Our target is to estimate the average treatment effect among the treated (ATT). For patient i, let $\widetilde{D}_i^1(T_i)$ denote the potential remaining survival time following treatment assignment at T_i, such that $\tilde{D}_i^1(T_i) = [D_i^1(T_i) - T_i]_+$. Conversely, let $\tilde{D}_i^0(T_i)$ denote the potential remaining survival time if the patient never receives treatment such that $\tilde{D}_i^0(T_i) = [D_i^0(T_i) - T_i]_+$. The post-treatment survival functions of our

interest can then be defined as,

$$S_i^j(t) = P\left\{\tilde{D}_i^j(T_i) > t | T_i, \boldsymbol{Z}_i, \right\}, \qquad j = 0, 1$$

and the subject-specific treatment effect can be defined as

$$\delta_i(t) = S_i^1(t) - S_i^0(t).$$

Hence, the average causal treatment effect among treated is given by

$$\delta(t) = S_1(t) - S_0(t),$$

where $S_1(t)$ and $S_0(t)$ are average survival functions,

$$S_j(t) = E\left\{S_i^j(t)\right\},$$

with the expectation being with respect to the distribution of $\{T, \boldsymbol{Z} | T < D\}$; i.e., the joint distribution of $(T, \boldsymbol{Z})$ among patients with $T < D$. To avoid identifiability issues, we need to have some restrictions pertaining to follow-up time. Specifically, if we let τ_C be the maximum censoring time, then our inference is restricted to $T \in [0, \tau_T]$ with $S_1(t)$ estimable on $t \in [0, \tau_1]$ for $\tau_T + \tau_1 \leq \tau_C$.

We also define the restricted mean survival time on $[0, L]$ with $L < \tau_1$ as $\mu_j(L) = \int_0^L S_j(u)du$ for $j = 0, 1$, so that the difference in restricted mean life is denoted as $\Delta(L) = \mu_1(L) - \mu_0(L)$; note that $\Delta(L) = \int_0^L \delta(t)dt$.

Since our proposed method is using the risk class of each individual in both post-treatment and treatment-absent period, rather than use $(T_i, \boldsymbol{Z}_i)$ explicitly, we define:

$$S^1(t|T_i, Z_i) = S^1(t|G_i^1, T_i)$$

$$S^0(t|T_i, Z_i) = S^0(t|G_i^0, T_i),$$

where G_i^1 and G_i^0 are the post-treatment and treatment-absent risk classes for treated individual i. Hence, instead of estimating $\delta(t) = E\left\{\delta_i(t|T_i, Z_i)\right\}$, we are instead estimating the very closely related quantity $\delta(t) = E\left\{\delta_i(t|G_i^1, G_i^0, T_i)\right\}$.

3.2.2 Estimation of $S^1(t)$

We will use the same estimation process for $S^1(t)$ as shown in Chapter II. Since T_k is subject to right censoring by C_k, the uncensored T_k represent a biased sample of shorter values of time-to-treatment. A method that explicitly accounts for censoring is required here so that the resulting nonparametric estimator of $S^1(t)$ represents an appropriate average over the $\{T, \boldsymbol{Z}|T < D\}$. Such an average should, naturally, not depend on the C distribution.

We use the Inverse Probability of Censoring Weighting (IPCW; Robins and Rotnitzky, 1992) to remedy the issue of dependent censoring. Specifically, the weight of patient i is given by

$$w_i = \frac{\Delta_i^T}{P(C > T_i|T_i, \boldsymbol{Z})}.$$

For a untreated patient, such that $w_i = 0$. To estimate $P(C > T_i|T_i, \boldsymbol{Z})$, we assume the following Cox model for censoring,

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

which can be fiited using standard partial likelihood [5]. To estimate $S_1(t)$, we focus on the prognostic score which is based on the hazard of death at time t given treated at time T_i,

$$\lambda_i^1\left(t|T_i, \boldsymbol{Z}_i\right) = \lim_{dt \to 0} \frac{1}{dt} P(t \leq D_i^1(T_i) < t + dt|\boldsymbol{Z}_i, T_i, T_i < D_i),$$

for which we assume the following post-treatment hazard model,

$$\lambda_i^1\left(t|T_i, \boldsymbol{Z}_i\right) = h\left\{\lambda_0^1(t),\ \boldsymbol{\beta}_1' \boldsymbol{Z}_i + \boldsymbol{\beta}_T{}' \boldsymbol{g}(T_i)\right\}, \tag{3.1}$$

such that $\lambda_i^1\left(t|T_i, \boldsymbol{Z}_i\right)$ presents a semi-parametric function of h; e.g., Cox model [5], additive hazards model [1], etc. Note that $\boldsymbol{\beta}_1$ is a vector of unknown parameters and

$g(\bullet)$ is a vector of functions such that the effect of T is parametrized; and $\lambda_0^1(t)$ is the baseline hazard for post-treatment death. For each treated patient, we obtain the prognostic score $\boldsymbol{\beta}_1^{'}\boldsymbol{Z}_i + \boldsymbol{\beta}_T\boldsymbol{g}(T_i)$ for post-treatment death. Then we group patients based on (3.1), which can be done by simply building grids or using empirical quantiles. Suppose eventually we have K groups of treated patients. Patients in the same group have similar prognostic scores, such that we have approximate homogeneity with respect to post-treatment death risk within each k grouping. WE propose estimating $S_1(t)$ through the following weighted survival function,

$$\hat{S}^1(t) = \left[\sum_{k=1}^{K}\sum_{i=1}^{n_k} w_{ki}\right]^{-1} \times \sum_{k=1}^{K}\sum_{i=1}^{n_k} w_{ki}\hat{S}_{ki}^1(t),$$

where $w_{ki} = w_i G_{ik}^1$ and $G_{ik}^1 = I(\text{patient } i \text{ is in group } k)$, with $\hat{S}_{ki}^1(t)$ being the estimated survival probability for the ith patient in group k. Here, the $\widehat{S}_{ki}^1(t)$ can be based on Kaplan-Meier or Nelson-Aalen methods. Since patients in each group l have homogeneity on death risk, $\hat{S}_{ki}^1(t)$ is the same across all patients in group k. Therefore we, have

$$\hat{S}^1(t) = \left[\sum_{k=1}^{K}\sum_{i=1}^{n_k} w_{ki}\right]^{-1} \times \sum_{k=1}^{K}\left(\sum_{i=1}^{n_k} w_{ki}\right)\hat{S}_k^1(t),$$

where $\hat{S}_k^1(t)$ is the estimated survival probability for group k. After rearranging the terms we have:

$$\hat{S}^1(t) = \left[\sum_{i=1}^{n} w_i\right]^{-1} \times \sum_{i=1}^{n} w_i \sum_{k=1}^{K} G_{ik}^1\hat{S}_k^1(t).$$

3.2.3 Estimation of $S^0(t)$

In this section, we will introduce a nonparametric estimator for $S_0(t)$. We begin by defining some additional notation. Specifically, let $Y_i^0(t) = I(U_i \wedge T_i \geq t)$, an indicator for being at risk and untreated as of time t, and define the following counting process increment, $dN_i^0(t) = Y_i^0(t)dN_i(t)$.

Since in practice we do not observe data to estimate $P\{D^0(T) > t|\boldsymbol{Z}, T, T < D\}$, the basic idea is to first obtain a pertinent estimator $\widehat{S}_i^0(t)$ for each treated patient, $\widehat{S}^0(t)$ then being an appropriately weighted average of $\widehat{S}_i^0(t)$ across $i = 1, \ldots n_T$. As an analog to $S^1(t)$, we propose to use the following estimator,

$$\widehat{S}^0(t) = \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} w_i \widehat{S}_i^0(t), \tag{3.2}$$

where w_i is inherited from each corresponding treated patient.

Similar to the matching method described in Chapter II, we need to first obtain the pre-treatment prognostic score as in Section 2.2.4 but instead of matching we apply conditional survival function. We first calculate the pre-treatment prognostic score, which reflects the treatment-free death hazard. The prognostic score is obtained through the model,

$$\lambda_i^0(t|\boldsymbol{Z}_i) = h\left\{\lambda_0^0(t),\ \boldsymbol{\beta}_0'\boldsymbol{Z}_i\right\}.$$

As covariates are time-constant, we will define prognostic score classes which is also time-constant. If we group patients finely enough, within each group patients would have very close pre-treatment hazard. Therefore for patients in the same group, they should have similar treatment-free survival probability $P(D^0 > t)$. Suppose we have K groups of prognostic score classes and treated patient i is included in group k, then

$$S_{ki}^0(t; s) = \frac{S_{ki}^0(s+t)}{S_{ki}^0(s)},$$

where $S(t; s)$ is the survival probability of living t time units more after living s units already. As discussed above, in this case we can use $S_k^0(t)$ as an estimator for $S_{ki}^0(t)$. Here we choose Nelson-Aalen estimator for $S_k^0(t)$, then $S_{ki}^0(t; s)$ can be estimated by

$$\widehat{S}_{ki}^0(t; s) = \exp\left\{-\widehat{\Lambda}_k^0(t; s)\right\}, \tag{3.3}$$

where we define $\widehat{\Lambda}_k^0(t;s) = \widehat{\Lambda}_k^0(t+s) - \widehat{\Lambda}_k^0(s)$ and $\widehat{\Lambda}_k^0(t)$ is the Nelson-Aalen estimator for the cumulative hazard function for group k.

As we are interested in estimating the post-treatment survival probability in the absence of treatment, naturally we only need to substitute s by T_i in formula (3.3). To be consistent with equation (3.2), let

$$\widehat{S}_i^0(t) = \sum_{k=1}^{K} G_{ki}^0 \widehat{S}_{ki}^0(t;T_i),$$

where $G_{ki}^0 = 1$ if patient i is group k, otherwise 0. We have the final version of equation (3.2),

$$\widehat{S}^0(t) = \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} w_i \sum_{k=1}^{K} G_{ki}^0 \widehat{S}_{ki}^0(t;T_i).$$

3.2.4 Vaiance Estimator for $\widehat{S}^1(t)$

In this subsection we aim to heuristically derive a variance estimator for $\widehat{S}^1(t)$. Technical details of the arguments are omitted, in keeping withe the emphasis of the work.

As n goes to infinity, we let the number of groups K goes to infinity as well but at a slower rate than n such that the the number of individuals in each group will also go to infinity. This being the case, individuals in the same group can be viewed as identical with respect to pre-treatment death hazard. If there is no censoring, by the definition of $S^1(t)$, $\widehat{S}^1(t) = n^{-1}\sum_{k=1}^{K} n_k \widehat{S}_k^1(t)/n$. When $K \to \infty$, $\widehat{S}^1(t) = n^{-1}\sum_{i=1}^{n} S_i^1(t)$, by UWLLN, $\widehat{S}^1(t) \to S^1(t)$ in probability for all $S^1(t)$.

When there is no censoring , $w_{k\cdot}\left\{\sum_{k=1}^{K} w_{k\cdot}\right\}^{-1}$ can be viewed as a density function of certain function of $(T, \mathbf{Z})$ given prognostic score $\boldsymbol{\beta}_P^{'}\boldsymbol{Z}_i + \beta_T T_i$ falling in kth interval. We denote this density as $f_\phi(k) = w_{k\cdot}\left\{\sum_{k=1}^{K} w_{k\cdot}\right\}^{-1}$. As $n \to \infty$, $K \to \infty$, such that each interval of the prognostic score will be close to a value on the domain.

Thus we can re-write $f_\phi(k)$ as $f_\phi(x)$ with $x \in (-\infty, \infty)$. Therefore we have

$$\widehat{S}^1(t) = \int_{-\infty}^{\infty} f_\phi(x)\widehat{S}^1_x(t)dx = \int_{-\infty}^{\infty} \widehat{S}^1_x(t)dF_\phi(x), \tag{3.4}$$

where F_ϕ is the CDF of ϕ.

In order to derive the asymptotic distribution of $\widehat{S}^1(t)$ we need to define an additional set of notation. Let $Y_{ik}(t) = G^1_{ik}Y_i(t)$, $dM^1_{ik}(t) = G^1_{ik}\left\{dN_i(t) - Y_i(t)d\Lambda^1_k(t)\right\}$. We also define $\rho = \Delta_i^T$ which can be estimated by $n^{-1}\sum_{i=1}^n \Delta_i^T$. We then have

$$n^{1/2}\rho^{1/2}\left\{\widehat{\Lambda}^1_k(t) - \Lambda^1_k(t)\right\} = n^{1/2}\rho^{1/2}\sum_{i=1}^n \varphi^1_{ik}(t),$$

where $\widehat{\Lambda}^1_j(t)$ is the estimator of post-treatment cumulated hazard function for group k, and $\varphi^1_{ik}(t) = \int_0^t \pi^{-1}(u)dM^1_{ik}(u)$, where $\pi(u) = P(U \geq u)$. Under mild regularity conditions, $\{\varphi^1_{1k}(t), \ldots, \varphi^1_{nk}(t)\}$ are independent and identically distributed mean 0 variates. As a result, $n^{1/2}\left\{\widehat{\Lambda}^1_k(t) - \Lambda^1_k(t)\right\}$ converges to asymptotically to a mean-zero normal distribution with variance $E\left[\varphi^1_{1k}(t)^2\right]$ by the Multivariate Central Limit Theorem. By applying the Functional Delta Method, we obtain that $n^{1/2}\left\{\widehat{S}^1_k(t) - S^1_k(t)\right\}$ is also asymptotically mean-zero Normal with variance estimator,

$$\widehat{\sigma}^2_k(t) = n^{-1}\sum_{i=1}^n \left\{\widehat{S}^1(t)\widehat{\varphi}^1_{ik}(t)\right\}^2,$$

where $\widehat{\varphi}^1_{ik}(t) = \int_0^t \widehat{\pi}^{-1}(u)d\widehat{M}^1_{ik}(u)$ and $d\widehat{M}^1_{ik}(u) = G_{ik}\left\{dN^1_i(u) - Y^1_{ik}(u)d\Lambda^1(u)\right\}$

As we discussed above, $\widehat{S}^1(t)$ will converge in probability to its limiting value $S^1(t) = \int_{-\infty}^{\infty} S^1_x(t)dF_\phi(x)$, in discretized case we define $S(t) = \sum_{k=1}^K f_\phi(k)S^1_k(t)$.

Next, we consider the asymptotic distribution of $\widehat{S}^1(t)$. We can write,

$$\begin{aligned}
\widehat{S}^1(t) - S^1(t) &= \sum_{k=1}^{K} f_\phi(k) \left\{ \widehat{S}^1_k(t) - S^1_k(t) \right\} \\
&= \sum_{k=1}^{K} f_\phi(k) \left\{ - \sum_{i=1}^{n_T} S_k(t) \varphi^1_{ik}(t) \right\} \\
&= - \sum_{k=1}^{K} f_\phi(k) S^1_k(t) \sum_{i=1}^{n} \varphi^1_{ik}(t) \\
&= - \sum_{i=1}^{n} \sum_{k=1}^{K} f_\phi(k) S^1_k(t) \varphi^1_{ik}(t) \\
&= - \sum_{i=1}^{n} \varphi^1_{i\bullet}(t),
\end{aligned}$$

where $\varphi^1_{i\bullet}(t) = \sum_{k=1}^{K} f_\phi(k) S^1_k(t) \varphi^1_{ik}(t)$. Since $\varphi^1_{i\bullet}(t)$ are also independent and mean-zero, $n^{1/2} \left\{ \widehat{S}^1(t) - S^1(t) \right\}$ converges in distribution to a zero-mean Normal with a variance that can be consistently estimated by

$$\widehat{\sigma}^2_1 = n^{-1} \sum_{i=1}^{n} (\widehat{\varphi}^1_{i\bullet})^2,$$

where we define

$$\widehat{\varphi}_{i\bullet}(t) = \sum_{k=1}^{K} f_\phi(k) \widehat{S}^1_k(t) \widehat{\varphi}^1_{ik}(t).$$

3.2.5 Variance Estimator for $\widehat{S}^0(t)$

Since $\widehat{S}^0(t)$ is an analog to $\widehat{S}^1(t)$, as long as $\widehat{S}^0_i(t)$ is an consistent estimator for $S^0_i(t)$, $\widehat{S}^0(t)$ will converge in probability to $S^0(t)$, where $S^0(t)$ is defined as

$$S^0(t) = \int_{-\infty}^{\infty} f_\phi(x) S^0_x(t) dx = \int_{-\infty}^{\infty} S^0_x(t) dF_\phi(x), \tag{3.5}$$

where $f_\phi(x)$ has the same definition as in previous section. Note that when (3.5) is discretized, we have0 $S^0(t) = \sum_{k=1}^{n_T} f_\phi(k) S^0_k(t)$, where $f_\phi(k) = w_{k\bullet}$.

Next we derive the limiting distribution for $\widehat{S}^0(t)$. We start with the limiting distribution for $\widehat{S}^0_i(t)$. We first define $Y^0_{ki}(t) = G^0_{ki} Y^0_i(t)$, $dM^0_{ki}(t) = G^0_{ki} \{ dN^0_i(t) - Y^0_i(t) d\Lambda^0(t) \}$.

Suppose in total we have grouped patients into K risk groups. Then, for the jth treated patient:

$$n^{1/2}\left\{\widehat{\Lambda}_j^0(t;T_j)-\Lambda_j^0(t;T_j)\right\}=n^{1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}G_{kj}^0\varphi_{ki}^0(t;T_j)$$

where $\varphi_{ki}^0(t;T_j)=\int_{T_j}^{t+T_j}\pi^{-1}(u)dM_{ki}^0(u)$.

Analogous to $\widehat{S}_k^1(t)$, similar arguments lead to

$$n^{1/2}\left\{\widehat{S}_j^0(t)-S_j^0(t)\right\}=n^{1/2}\sum_{i=1}^{n}S_j^0(t)\sum_{k=1}^{K}G_{kj}^0\varphi_{ki}^0(t;T_j)$$

and $n^{1/2}\left\{\widehat{S}_j^0(t)-S_j^0(t)\right\}$ converges in distribution to a zero-mean normal with a variance estimator given by

$$(\sigma_j^0)^2(t)=n^{-1}\sum_{i=1}^{n}\left\{\widehat{S}_j^0(t)\sum_{k=1}^{K}G_{kj}^0\widehat{\varphi}_{ki}^0(t;T_j)\right\}^2$$

where $\widehat{\varphi}_{ki}^0(t;T_j)=\int_{T_j}^{t+T_j}\widehat{\pi}^{-1}(u)d\widehat{M}_{ki}^0(u)$ and $d\widehat{M}_{ki}^0(u)=G_{ki}^0\left\{dN_i^0(u)-Y_i^0(u)d\widehat{\Lambda}^0(u)\right\}$. Following arguments similar to Section 3.2.4, for $\widehat{S}^1(t)$, $\widehat{S}^0(t)-S^0(t)=-\sum_{i=1}^{n}\varphi_{i\bullet}^0(t)$, where

$$\varphi_{i\bullet}^0=\sum_{j=1}^{n_T}f_\phi(j)S_j^0(t)\sum_{k=1}^{K}G_{kj}^0\varphi_{ki}^0(t;T_j).$$

Thus, by the independence across the $\varphi_{i\bullet}$ and using central limit theorem, we have $n^{1/2}(\widehat{S}^0(t)-S^0(t))$, should converge to a mean zero normal with variance estimator

$$\widehat{\sigma}_0^2(t)=n^{-1}\sum_{i=1}^{n}(\widehat{\varphi}_{i\bullet})^2,$$

where $\widehat{\varphi}_{i\bullet}=\sum_{j=1}^{n_T}f_\phi(j)\widehat{S}_j^0(t)\sum_{k=1}^{K}G_{kj}^0\widehat{\varphi}_{ki}^0(t;T_j)$.

Combining the results above, we can represent

$$n^{1/2}\left\{\widehat{\delta}(t)-\delta(t)\right\}=n^{1/2}\left\{\varphi_{i\bullet}^1(t)-\varphi_{i\bullet}^0(t)\right\},$$

where $\varphi_{i\bullet}^1(t)-\varphi_{i\bullet}^0(t)$ components are independent and identically distributed with mean 0. Note that the development above accounts for the possibility that patients may contribute follow-up to both the $\widehat{S}^0(t)$ and $\widehat{S}^1(t)$ sides. The quantity

$n^{1/2}\left\{\widehat{\delta}(t)-\delta(t)\right\}$ should converge asymptotically to a Normal variate with mean 0 and a variance that can be consistently estimated by

$$\widehat{\sigma}^2_\delta(t) = n^{-1}\left\{\varphi^1_{i\bullet}(t)-\varphi^0_{i\bullet}(t)\right\}^2.$$

3.3 Simulation Study

Simulation was used to assess the performance of the proposed methods in moderate sized samples. The treatment time T was generated from an exponential distribution with hazard $\lambda_{T0}\exp\{\beta_{T1}Z_1+\beta_{T2}Z_2\}$ while treatment-free death times D^0 were generated as exonential with hazard $\lambda_{D0}\exp\{\beta_{D1}Z_1+\beta_{D2}Z_2\}$. Here both Z_1 and Z_2 are confounders that affects both T and D^0. Censoring times C were generated from an exponential distribution with hazard $\lambda_{C0}\exp\{\beta_{C1}Z_1+\beta_{C2}Z_2\}$. Times between treatment and death $(D^1-T)_+$ were generated from exponential distribution with rate $\lambda_{10}\exp\{\beta_{10}Z_1+\beta_{11}Z_2+\beta_{12}T\}$, where we set $\lambda_{10}=a\lambda_{D0}$, $\beta_{10}=\beta_{D1}$ and $\beta_{11}=\beta_{D2}$. Baseline covariates Z_1 and Z_2 were generated from a Uniform(-1,1). We denote the actual death times as D, for treated patients $D=D^1=T+(D^1-T)_+$ and for untreated patients $D=D^0$. There were $n=2500$ subjects in all simulation configurations, with each data configuration replicated 500 times. To estimate the standard errors, we bootstrap 25 datasets for per replicate.

In practice, we observe the minimum of T, D^0 and C. In simulations, however, we always observe T, D^0, $(D^1-T)_+$ and C for all patients. True values of $S_1(t)$, $S_0(t)$, $\delta(t)$ and $\Delta(L)$ were obtained using monte-carlo on these counterfactuals. Naturally, for the purpose of computing $\widehat{S}_0(t)$, only $[(D\wedge C\wedge T), N^D(D\wedge C\wedge T)]$ were used; similarly, only $[(D\wedge C-T)_+, N^D(D\wedge C)]$ were used for subjects with $(D\wedge C>T)$ for the purpose of computing $\widehat{S}_1(t)$. Hence, $\widehat{\delta}(t)$ and $\widehat{\Delta}(L)$ were, for each replicate, only based on data that would in reality be observed.

After generating the data, prognostic scores representing pre-treatment history were obtained from model the $\lambda_i^0(t|\boldsymbol{Z}_i) = \lambda_{00}(t)\exp\{\beta_{00}Z_1 + \exp\beta_{01}Z_2\}$. Patients are grouped by half deciles of prognostic score. For all simulations configurations we set $\tau_T = 10$.

In the first set of simulations, we examine the bias, empirical standard deviation (ESD) and bootstrap standard error (BSE) of the proposed estimators under various treatment effects and in the presence of light censoring with $\lambda_{C0} = 0.015$, where around 15% of individuals get censored. We vary a from 0.8, 1 and 1.2 to change the treatment effect from moderate, null to negative. The remaining parameters are set equal across the three scenarios: $\lambda_{D0} = 0.05$, $\lambda_{T0} = 0.03$, $\beta_{D1} = \beta_{T1} = \beta_{C1} = \log 2$, $\beta_{12} = \log 3/500$ and $\beta_{D2} = \beta_{T2} = \beta_{C2} = \log 3$.

In the second set of simulations, we examine the properties of the proposed estimators under moderate censoring. The parameter setting is same as the first set of simulations except for me change the censoring parameter to $\lambda_{C0} = 0.02$, which results in approximately 30% censoring.

Results for the first and second set of simulations are shown in Table 3.1 and Table 3.2, respectively. In both tables, the absolute bias of $\widehat{S}_1(t)$ $\widehat{S}_0(t)$ and $\widehat{\delta}(t)$ range from 0.001 to 0.015. The bias of $\widehat{\mu}_0$, $\widehat{\mu}_1$ and Δ is little bit larger, but is still negligible considering the scale of these quantities the bias. The BSEs are generally close to ESDs. The empirical coverage probability (CP) is around 0.95, except for $S_0(15)$. The estimation of $S_0(t)$ is more sensitive to censoring percentage, as the censoring percentage does not effect the results of $S_1(t)$ but the bias of $S_0(t)$ becomes more pronounced with larger censoring percentage.

Table 3.1: Simulation: Light censoring

Setting	t	Quantity	True	BIAS	ESD	BSE	CP
Moderate effect	5	$S^1(t)$	0.731	-0.001	0.022	0.021	0.93
	10		0.554	-0.001	0.025	0.025	0.94
	15		0.431	0.002	0.025	0.025	0.95
	5	$S^0(t)$	0.678	0.002	0.013	0.014	0.99
	10		0.486	0.004	0.019	0.019	0.97
	15		0.364	0.007	0.020	0.021	0.93
	5	$\delta(t)$	0.054	-0.003	0.024	0.024	0.97
	10		0.068	-0.005	0.030	0.029	0.95
	15		0.067	-0.005	0.030	0.030	0.94
	15	$\mu_0(t)$	9.10	0.044	0.205	0.210	0.94
	15	$\mu_1(t)$	9.91	-0.013	0.271	0.263	0.94
	15	$\Delta(t)$	0.81	-0.057	0.310	0305	0.94
Null effect	5	$S^1(t)$	0.680	-0.002	0.023	0.022	0.94
	10		0.487	-0.001	0.025	0.024	0.94
	15		0.365	0.002	0.024	0.024	0.95
	5	$S^0(t)$	0.678	0.002	0.013	0.014	0.99
	10		0.486	0.004	0.019	0.019	0.97
	15		0.364	0.007	0.021	0.021	0.93
	5	$\delta(t)$	0.003	-0.004	0.025	0.025	0.96
	10		0.001	-0.004	0.029	0.029	0.94
	15		0.001	-0.004	0.029	0.029	0.93
	15	$\mu_0(t)$	9.11	0.044	0.205	0.210	0.94
	15	$\mu_1(t)$	9.12	-0.008	0.272	0.265	0.94
	15	$\Delta(t)$	0.02	-0.052	0.305	0.305	0.93
Negative effect	5	$S^1(t)$	0.635	-0.003	0.024	0.023	0.95
	10		0.431	-0.000	0.024	0.024	0.96
	15		0.310	0.005	0.023	0.023	0.93
	5	$S^0(t)$	0.678	0.002	0.013	0.014	0.99
	10		0.486	0.004	0.019	0.020	0.97
	15		0.364	0.007	0.021	0.021	0.93
	5	$\delta(t)$	-0.043	-0.005	0.025	0.025	0.96
	10		-0.055	-0.004	0.028	0.028	0.95
	15		-0.054	-0.002	0.029	0.029	0.93
	15	$\mu_0(t)$	9.11	0.044	0.205	0.210	0.94
	15	$\mu_1(t)$	8.44	-0.002	0.272	0.266	0.94
	15	$\Delta(t)$	-0.661	-0.046	0.303	0.303	0.93

Table 3.2: Simulation: Moderate censoring

Setting	t	Quantity	True	BIAS	ESD	BSE	CP
Moderate effect	5	$S^1(t)$	0.731	-0.001	0.023	0.022	0.93
	10		0.554	-0.001	0.027	0.026	0.94
	15		0.431	0.005	0.026	0.026	0.94
	5	$S^0(t)$	0.678	0.003	0.014	0.015	0.99
	10		0.486	0.005	0.020	0.021	0.97
	15		0.364	0.009	0.022	0.023	0.93
	5	$\delta(t)$	0.054	-0.004	0.026	0.025	0.98
	10		0.068	-0.005	0.031	0.031	0.95
	15		0.067	-0.005	0.031	0.033	0.95
	15	$\mu_0(t)$	9.10	0.056	0.213	0.225	0.94
	15	$\mu_1(t)$	9.91	-0.004	0.282	0.272	0.93
	15	$\Delta(t)$	0.81	-0.059	0.323	0.326	0.94
Null effect	5	$S^1(t)$	0.680	-0.002	0.024	0.023	0.95
	10		0.487	0.001	0.027	0.025	0.94
	15		0.365	0.006	0.026	0.025	0.94
	5	$S^0(t)$	0.678	0.003	0.014	0.015	0.99
	10		0.486	0.005	0.019	0.021	0.97
	15		0.364	0.009	0.022	0.023	0.93
	5	$\delta(t)$	0.003	-0.005	0.026	0.026	0.97
	10		0.001	-0.004	0.031	0.030	0.95
	15		0.001	-0.003	0.031	0.032	0.94
	15	$\mu_0(t)$	9.11	0.056	0.213	0.224	0.94
	15	$\mu_1(t)$	9.12	0.006	0.286	0.274	0.93
	15	$\Delta(t)$	0.01	-0.049	0.324	0.325	0.94
Negative effect	5	$S^1(t)$	0.635	-0.003	0.025	0.024	0.95
	10		0.431	0.002	0.026	0.025	0.96
	15		0.310	0.008	0.024	0.024	0.93
	5	$S^0(t)$	0.678	0.003	0.014	0.015	0.99
	10		0.486	0.005	0.020	0.021	0.97
	15		0.364	0.009	0.022	0.023	0.93
	5	$\delta(t)$	-0.043	-0.006	0.026	0.026	0.97
	10		-0.055	-0.003	0.029	0.030	0.95
	15		-0.054	-0.001	0.030	0.031	0.95
	15	$\mu_0(t)$	9.11	0.056	0.213	0.215	0.93
	15	$\mu_1(t)$	8.44	0.013	0.276	0.275	0.94
	15	$\Delta(t)$	-0.661	-0.043	0.318	0.323	0.95

3.4 Application

We applied our proposed methods in order to estimate the effect of deceased-donor kidney transplantation (KT) ($j = 1$) on survival in the absence of KT ($j = 0$) among waitlisted end-stage renal disease patients. Data were obtained from Scientific Registry of Transplant Recipients. The SRTR data system includes data on all donors, wait-listed candidates, and transplant recipients in the United States, as submitted by the members of the Organ Procurement and Transplantation Network (OPTN), and has been described elsewhere. The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors. Different from the application in Chapter II, here we focus on emphasizing the power the proposed method on smaller data sets. Instead of using the entire population, we only apply our methods on patients within Region 10. In Region 10, there are 6 centers in total, our goal is to compare the treatment effect of KT between different centers within Region 10.

The study population included $n = 7,209$ patients aged ≥ 18 and listed between 01/01/2003 and 12/31/2013 in Region 10. The sample size varies from 274 to 2059 from center to center. Follow-up time begins at the date when patients got listed and ends at earliest of death, loss to follow-up, or the end of the observation period (12/31/2013). Adjustment covariates for the $\lambda_i^0(t)$ model included height, weight, years on dialysis (prior to waitlisting), calender year of listing, albumin, diabetes, hypertension, panel reactive antibiotics (PRA), age, angina, blood type, symptomatic peripheral vascular disease (PVD), race, gender, calendar year of transplant and Kidney Donor Risk Index (KDRI) [26].

In this application, we set τ_T and τ_1 to 3 years and 5 years, respectively. A total of 3,267 patients received KT (45%) and 2,835 deaths were observed. We set the caliper width of prognostic score matching $\epsilon = 0.02$.

The estimated post-treatment and treatment-free survival curves on [0,5] year interval of each center are presented in Figure 1 and Figure 2, respectively. The thicker curve in figures represent the survival of the overall population.

Table 3 shows the estimates of $S_1(t)$, the average survival probability from the time of transplant among patients received transplantation, and $S_0(t)$ intended to represent the survival probability to which the transplanted patients would have been observed in the absence of transplantation, and $\widehat{\delta}(t) = \widehat{S}_1(t) - \widehat{S}_0(t)$ and $\widehat{\Delta}(5) = \int_0^5 \widehat{\delta}(t)dt$. All quantities are estimated at 1, 3, 5 years for each center. We also presented the restricted mean lifetime at 5 years for both groups ($\mu_0(t)$, $\mu_1(t)$) and their contrast ($\Delta(t)$). The benefit of kidney transplantation varies from 133 days to 227 days in 5-year post-treatment survival restricted mean survival time.

Table 4 shows the tests results for the RMST of treated and treatment-free group of each center, where $\bar{\mu}_1$ and $\bar{\mu}_0$ are the restricted mean survival time in the overall population. Based on the test of $\Delta(t)$, there are significant difference between the all the centers and the overall population except for center with ID=472. Centers 220, 470, 471 and 480 each have a significant better treatment effect than the average while center 330 has worse treatment effect than the average.

3.5 Discussion

In this chapter, we proposed methods to estimate the average causal effect among the treated of a time-dependent treatment. In particular, the proposed treatment effect contrasts the post-treatment survival function that applied to treated patients

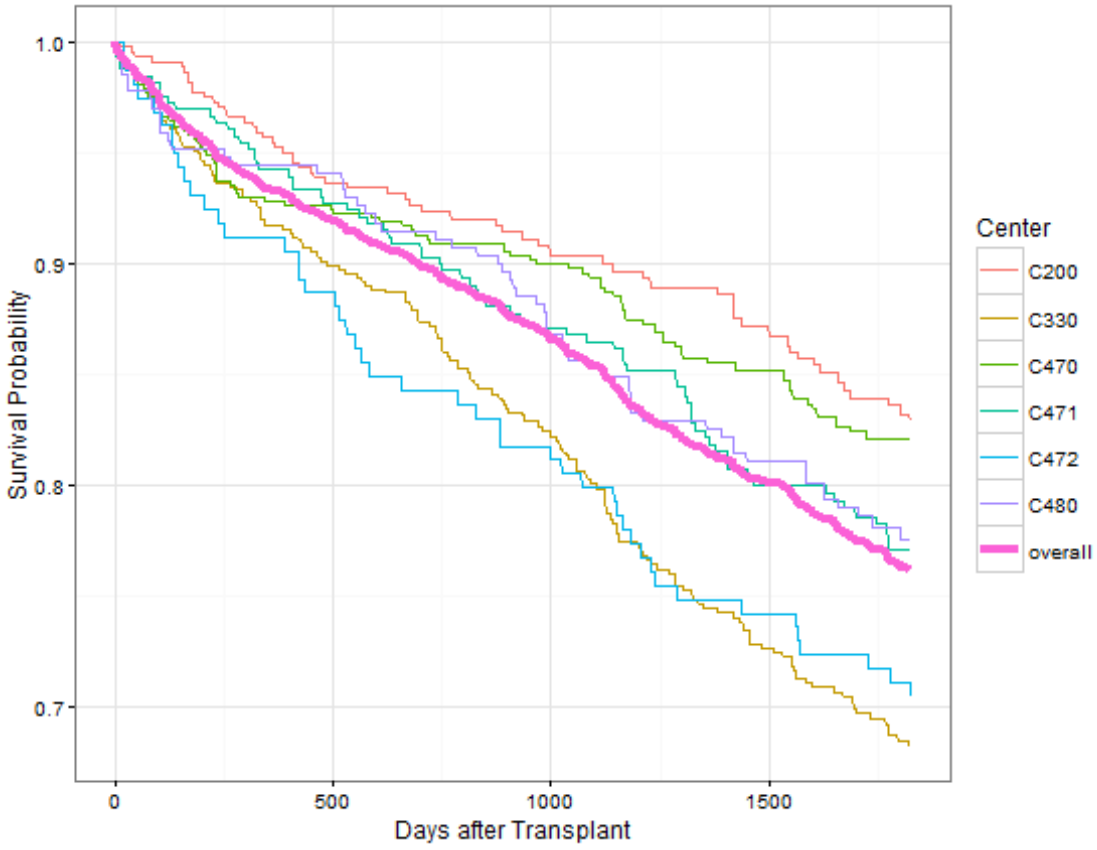

Figure 3.1: 5-year post-transplant survival by center

Table 3.3: Analysis of Region 10 data (n=6,273): Evaluation of the effect on survival (and RMST) of kidney transplantation

t(years)	Quantity	C220	C330	C470	C471	C472	C480
1	$S^1(t)$	0.955	0.917	0.928	0.942	0.912	0.944
3		0.923	0.872	0.909	0.903	0.843	0.915
5		0.903	0.800	0.894	0.865	0.799	0.852
1	$S^0(t)$	0.910	0.897	0.916	0.908	0.874	0.917
3		0.822	0.785	0.829	0.800	0.758	0.838
5		0.744	0.677	0.739	0.704	0.646	0.764
1	$\delta(t)$	0.044	0.020	0.013	0.035	0.038	0.027
3		0.101	0.087	0.080	0.103	0.085	0.077
5		0.159	0.124	0.155	0.160	0.152	0.088
5	$\mu_1(t)$(days)	1668	1515	1636	1606	1505	1607
5	$\mu_0(t)$(days)	1441	1342	1434	1398	1299	1473
5	$\Delta(t)$(days)	227	173	202	209	206	133

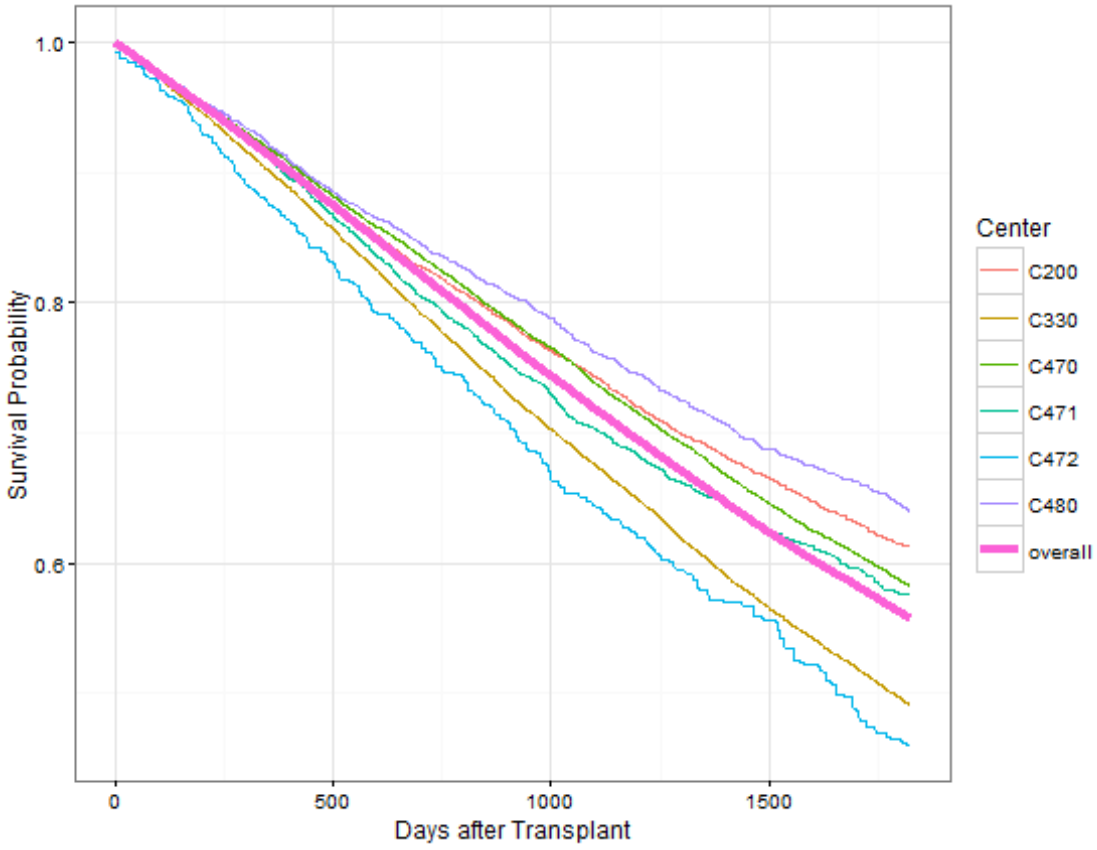

Figure 3.2: 5-year survival for waitlisted patients by center

Table 3.4: Analysis of Region 10 data (n=6,273) by center on 5-year survival

Quantity	C220	C330	C470	C471	C472	C480
$\mu_1 - \bar{\mu}_1$	79.84*	-78.33*	46.34*	19.30*	-81.12*	9.33*
	(1.96)	(1.93)	(1.99)	(2.64)	(4.60)	(2.89)
$\mu_0 - \bar{\mu}_0$	36.01*	-65.46*	25.48*	0.68	-75.41*	72.24*
	(2.59)	(1.90)	(1.70)	(4.53)	(6.62)	(3.16)
$\Delta - \bar{\Delta}$	43.83*	-12.87*	20.86*	18.61*	-5.70	-62.91*
	(3.18)	(2.58)	(2.75)	(5.23)	(6.34)	(4.22)

had they, contrary to fact, not received treatment. The estimation of treated survival probability is carried out in the same way as described in Chapter II. To estimate the survival of treated patients in the absence of treatment, we proposed a semiparametric method based on the conditional survival probability. Specifically, we group patients based on their pre-treatment prognostic score and then obtain the individual conditional survival function based on marginal survival probability estimated from the corresponding prognostic group. The limiting distribution and asymptotic variances were derived. For computational conveniences, a bootstrap method is employed to estimate the standard error. The proposed methods were proved through simulations to work well even in small datasets.

Similar to methods proposed in Chapter II, this method is also non-parametric since the assumed model only contribute to the prognostic score used for grouping, which relaxes the assumption of proportionality between pre- and post-treatment hazards. To quantify the treatment effect, we use the difference in survival functions and restricted mean survival time. Compared to what was proposed in Chapter II, the method of conditional survival probability does not rely on big data sets anymore. To illustrate this strength, we applied the proposed method to center-level SRTR data. We obtained the center-specific survival curves in Region 10 and compare them to that of the overall population. According to our results, the benefit of transplantation (difference in RMST) for 5-year survival varies from 133 days to 227 days. Among all 6 centers, center 200 has the greatest treatment effect as it has the longest expected 5-year post-tranplant survival and a relative long treatment-free survival as well. Center 472 has both relative low treatment-free survival and post-treatment survival but has second best treatment effect. Therefore when evaluating the benefit of transplant of each center, treatment-free survival can be an important

factor. In this case, it is more appropriate to use δ as a major evaluation criteria. For a specific center, Δ represents the contrasts of restricted mean survival time between treated and untreated group and eliminates some confounding factors such as patients quality.

CHAPTER IV

Matching Methods for Evaluating the Effect of A Time-dependent Treatment on the Survival Function in the Presence of Time-dependent Covariates

4.1 Introduction

In many clinical and epidemiology settings, data are available on various longitudinal covariates, collected for each patient as the study unfolds over time. In the case in which all covariates are collected at baseline (time 0), methods proposed in Chapter II and Chapter III are valid for estimating the effect of a time-dependent treatment. However, in some studies (e.g., liver transplantation), several important time-dependent predictors are collected, such as Model for End-stage Liver Disease (MELD) score. The proposed method is in fact motivated by the end-stage liver disease (ESLD) setting. In the United States, chronic end-stage liver disease patients are sequenced on the waitlist in decreasing order of MELD score, which is a very strong predictor of pretransplant mortality. Transplantation generally results in the dependent censoring of pretransplant death, since MELD scores predict both waitlist mortality and transplant rates. Patients may also be removed from the waitlist (or made temporally inactive) and hence ineligible to receive a transplant. Ignoring these important time-dependent variables when estimating the treatment effect will generally lead to a biased estimate of the effect on survival of liver transplantation.

Several methods for estimating the effect of a time-dependent treatment in the presence of time-dependent covariates have been proposed in the literature. Various authors have proposed methods based on partly conditional modeling (Zheng and Heagerty, 2005; Gong and Schaubel, 2013) method and the closely related concept of landmark analysis (Feurer et al., 1992; van Houwelingen, 2007; van Houwelingen and Putter, 2012; Parast, Tian and Cai, 2014), although only Feurer et al (1992) explicitly considered treatment effects. Gong and Schaubel (2016) proposed to use partly conditional hazard regression to model each of pretreatment and posttreatment survival, then estimate the treatment effect nonparametrically. A disadvantage of Gong & Schaubel (2016) is its reliance on the correct specification of several semiparametric models.

In this chapter, methods from Chapter II are extended to the setting in which data are available on time-dependent covariates. The objective is to estimate the treatment effect on survival in a way that appropriately incorporates the time-dependent factors. In contrast to the methods listed in the proceeding paragraph, the proposed methods in this chapter use prognostic score matching in place of the projection of fitted survival curves from the model.

The remainder of the chapter proceeds as follows. In Section 2 the notation and proposed models are introduced. Section 3 lays out our estimation methods. Section 4 presents the numerical evaluation of the methods. Section 5 shows the real data application of the methods.

4.2 Methods

4.2.1 Notation

We first define the parameter of interest in the causal inference framework. Typically, this framework considers the setting wherein each individual has two potential

outcomes, corresponding to the two possible treatment regimes (e.g., treated and untreated). In the counterfactual world, let $D_i^1(T_i)$ denote the potential death time (measured from time 0) if patient i is treated at T_i. The counterfactual quantity $D_i^0(T_i)$ denotes the potential death time if, contrary to fact, patient i never received treatment. By definition, both $D_i^1(T_i)$ and $D_i^0(T_i)$ are greater than T_i and the counterfactuals are meaningfully defined only for individuals that receive treatment.

Next, we define notation for the observed data. Let D_i denote death time for subject i. The obsevation time is denoted by $U_i = D_i \wedge C_i$, with $a \wedge b = \min\{a, b\}$. The death indicator is given by $\Delta_i = I(D_i < C_i)$. The at-risk indicator is defined as $Y_i(t) = I(U_i \geq t)$ and the treatment indicator is defined by $\Delta_i^T = I(T_i < U_i)$. We also define $Y_i^1(t) = I(U_i \geq t, T < t)$, which equals 1 when subject i is at risk at time t and has already initiated treatment. Correspondingly, we define the post-treatment counting process increment, $dN_i^1(t) = Y_i^1(t)\, dN_i(t)$.

Due to complexity in our new covariates setting, we need to define notation additional to those aforementioned. The covariate vector, which contains some time-varying variables, is denoted by $\boldsymbol{Z}_i(s)$. The patient's covariate history up to time s is given by $\mathcal{H}_i(s) = \{Z_i(u); 0 \leq u < s\}$. We assume that $D_i^0(T_i)$ and $D_i^1(T_i)$ are conditionally independent given T_i and the observed covariates $\mathcal{H}_i(T_i)$, known as the strong ignorability assumption [31]. For a patient with treatment time $T_i = s$, we are interested in the average difference between $(D_i^1 - s)_+$ and $(D_i^0 - s)_+$ given $[\mathcal{H}_i(s), T_i = s]$.

Analogous to the setup in Chapter II, for a patient initiating treatment at time $T = s$, there are two death times of interest; the post-treatment residual death time $\widetilde{D}^1(s) = (D^1(s) - s)_+$ and residual death time that would have observed in the absence of treatment $\widetilde{D}^0(s) = (D^0(s) - s)_+$. At time of treatment $T = s$, we observe

$\mathcal{H}(s)$. Conditional on $[\mathcal{H}(s), T = s]$, we contrast the following survival functions:

$$S^1(t; s|\mathcal{H}(s), T = s) = P\left\{\widetilde{D}^1(s) > t|\mathcal{H}(s), T = s\right\}$$

$$S^0(t; s|\mathcal{H}(s), T = s) = P\left\{\widetilde{D}^0(s) > t|\mathcal{H}(s), T = s\right\}.$$

For fixed $L > 0$, restricted mean survival times (RMST) are given by

$$\mu_1(L; s|\mathcal{H}(s), T = s) = \int_0^L S^1(t; s|\mathcal{H}(s), T = s)dt$$

$$\mu_0(L; s|\mathcal{H}(s), T = s) = \int_0^L S^0(t; s|\mathcal{H}(s), T = s)dt.$$

The contrast in survival function δ is defined as

$$\delta(t; s|\mathcal{H}(s), T = s) = S^1(t; s|\mathcal{H}(s), T = s) - S^0(t; s|\mathcal{H}(s), T = s),$$

while a contrast in RMST is defined as

$$\Delta(L; s|\mathcal{H}(s), T = s) = \mu_1(L; s|\mathcal{H}(s), T = s) - \mu_0(L; s|\mathcal{H}(s), T = s).$$

Average survival functions are then defined as

$$S^1(t) = E[S^1(t; T|\mathcal{H}(T), T)]$$

$$S^0(t) = E[S^0(t; T|\mathcal{H}(T), T)]$$

where, in both cases, the expectation is taken with respect to the joint distribution of $[\mathcal{H}(T), T]$. Correspondingly, average RMST are given by:

$$\mu_1(L) = E[\mu_1(L; s|\mathcal{H}(s), T = s)] = \int_0^L S^1(t)dt$$

$$\mu_0(L) = E[\mu_0(L; s|\mathcal{H}(s), T = s)] = \int_0^L S^0(t)dt.$$

The ATT can be then defined in terms of mean difference in survival probability as

$$\delta(t) = E[\delta(t; T|\mathcal{H}(T), T)] = S_1(t) - S_0(t)$$

and, in terms of mean difference in RMST, by

$$\Delta(L) = E[\Delta(L|\mathcal{H}(T), T)] = \mu_1(L) - \mu_0(L) = \int_0^L \delta(t)dt.$$

Since in our proposed methods, we are using the risk class of each treated patients instead of using $(\mathcal{H}(T), T)$ explicitly, $S^1(t)$ and $S^0(t)$ can also be represented as

$$S^1(t) = E[S^1(t; T|G_i^1(T_i))]$$

$$S^0(t) = E[S^0(t; T|G_i^0(T_i))],$$

where $G_i^1(T_i)$ and $G_i^0(T_i)$ are the risk class index for posttreatment and treatment-absent survival for patient i, respectively. Hence, instead of estimating $\delta(t) = E\{\delta_i(t|\mathcal{H}(T_i), T_i)\}$, we are instead estimating the very closely related quantity $\delta(t) = E\{\delta_i(t|\boldsymbol{G}_i(T_i))\}$, where $\boldsymbol{G}_i(T_i) = [G_i^0(T_i), G_i^1(T_i)]'$.

Next, we will describe the proposed methods for estimating $\delta(t)$ and $\Delta(L)$.

4.2.2 Estimation of $S^1(t)$

Our proposal for estimating $S^1(t)$ is an extension of that proposed in Chapter II. The presence of time-dependent predictors motivates us to build a new post-treatment prognostic score model. Let $\lambda_1(t; s|\mathcal{H}(s), T = s)$ denote the conditional post-treatment hazard function corresponding to $S_1(t; s|\mathcal{H}(s), T = s)$. We assume the following model,

$$\lambda_1(t; s|\mathcal{H}(s), T_i = s) = h\left\{\lambda_0^1(t), \boldsymbol{\beta}_1^{'}\boldsymbol{Z}_{i1}(s) + \boldsymbol{\beta}_2'\boldsymbol{g}(s)\right\}, \tag{4.1}$$

where the covariate $\boldsymbol{Z}_{i1}(s)$ is chosen to summarize the pre-treatment history and is fixed at treatment time $T_i = s$ and $\lambda_0^1(t)$ is the baseline hazard function. We take the linear predictor $\left\{\boldsymbol{\beta}_1^{'}\boldsymbol{Z}_{i1}(s) + \boldsymbol{\beta}_T\boldsymbol{g}(s)\right\}$ from (4.1) as a prognostic score for each treated patient with respect to post-treatment survival. The next step is to group

treated patients based on this post-treatment prognostic score by simply building grids or using quantiles, as described in Chapter II.

In the absence of censoring, we could average with respect to the empirical distribution of $\{T_i, \mathcal{H}_i(T_i)\}$. In our case, treatment times are subject to right censoring, therefore this averaging generally depends on the C_i distribution. This implies inverse weighting the observed treatment times using w_i, such that the inverse weighted distribution reflects data which would have been obtained in the absence of censoring. To estimate w_i, we assume the following proportional hazards model for C_i,

$$\lambda_i^C(t) = \lambda_0^C(t) \exp\left\{\boldsymbol{\beta}_C' \boldsymbol{Z}_i(0)\right\}, \tag{4.2}$$

where we assume C_i is administrative censoring and only dependent on baseline covariates. Observed data used to fit model 4.2 include $\{U_i, I(C_i < D_i), \boldsymbol{Z}_i(0)\}$, with $\boldsymbol{\beta}_C$ and cumulated hazard $\Lambda_0^C(t)$ estimated through unweighted Cox regression. The weight is given by

$$w_i = \frac{\Delta_i^T}{P(C > T_i | T_i, \boldsymbol{Z}_i(0))}. \tag{4.3}$$

Suppose there are K groups after building grids. We set $\widehat{S}_k^1(t)$ be the estimator for post-treatment survival function for the kth group, which can be estimated by the Nelson-Aalen method. We define the group indicator $G_{ik}^1 = I(\text{patient } i \text{ is in group } k)$, with the proposed estimator of $S^1(t)$ given by

$$\widehat{S}^1(t) = \left[\sum_{i=1}^n w_i\right]^{-1} \times \sum_{i=1}^n w_i \sum_{k=1}^K G_{ik}^1 \widehat{S}_k^1(t). \tag{4.4}$$

4.2.3 Estimation of $S^0(t)$

In this section, we will introduce a semiparametric estimator for $S_0(t)$. Since in practice we do not observe $(D^0(T) - T)_+$ for patients with $T < D$, the basic idea is

to first obtain a pertinent estimator $\widehat{S}_i^0(t)$ for each treated patient based on 'similar' matched patients, then compute $\widehat{S}^0(t)$ as an appropriately weighted average of the $\widehat{S}_i^0(t), i = 1, \ldots n$. As an analog to $S^1(t)$, we propose to use the following estimator,

$$\widehat{S}^0(t) = \left(\sum_{i=1}^{n} w_i \right)^{-1} \sum_{i=1}^{n} w_i \widehat{S}_i^0(t),$$

where w_i is inherited from each corresponding treated patient and hence has the same definition as in Section 4.2.2.

We begin by describing the assumed hazard model for survival in the absence of treatment. We let $\lambda^0(t; s|\mathcal{H}(s), T = s)$ denote the hazard function corresponding to $S^0(t; s|\mathcal{H}(s), T = s)$, for which we assume the following model

$$\lambda_i^0(t; s|\mathcal{H}_i(s)) = \lambda_{00}(t) \exp\left\{ \boldsymbol{\beta}_0' \boldsymbol{Z}_{i0}(s) \right\}, \tag{4.5}$$

where $\boldsymbol{Z}_{i0}(s)$ is chosen such that $\lambda_0(t; s|\mathcal{H}_i(s)) = \lambda_0(t; s|\boldsymbol{Z}_{i0}(s))$. Model (4.5) is a partly conditional model in the sense that it conditions on the information which is "frozen" at time s while hazard at time $s + t$ is of interest. Here, we propose to estimate $\boldsymbol{\beta}_0$ by stratifying the model based on calender time cross-sections (Gong and Schaubel, 2013). To begin, we choose a set of K calender dates $\{CS_1, \ldots, CS_K\}$. For calendar date CS_k, we select cross section of treatment-eligible patients who were not treated. For patient i, follow-up time as of calender date CS_k is denoted by s_{ik}. Therefore, a patient selected into cross-section CS_k must, as follow-up time s_{ik} be: alive, uncensored and untreated at s_{ik}. Following Gong and Schaubel (2013), we estimate $\boldsymbol{\beta}_0$ through the following stratified model

$$\lambda_{0k}(t; s|\mathcal{H}_i(s_{ik})) = \lambda_{00k}(t) \exp\left\{ \boldsymbol{\beta}_0' \boldsymbol{Z}_{i0}(s_{ik}) \right\}, \tag{4.6}$$

where $\boldsymbol{\beta}_0$ is the same parameter as in model (4.5).

We compare treated patient,i, and a potential control, k, with respect to treatment-free prognostic score through the difference in prognostic score:

$$\psi_{i:k}(T_i) = \left\{ \boldsymbol{\beta}_0' \boldsymbol{Z}_{i0}(T_i) - \boldsymbol{\beta}_0' \boldsymbol{Z}_{k0}(T_i) \right\}.$$

Patient k can be considered a suitable match for treated patient i if $\psi_{i:k}(T_i)$ falls into a small caliper, $\psi_{i:k}(T_i) \in [-\epsilon, \epsilon]$,where ϵ is a predefined small number. Note that if we use different models (e.g., additive hazards model[19]) to obtain the prognostic score, then the criteria to select matched patients is identical as the one for proportional hazard models. To be matched to treated patient i, a control patient needs to be both not-yet-treated and at-risk, in addition to being prognostically similar to patient i with respect to residual death time hazard.

By matching qualified patients, we obtain matched sets corresponding to each treated patient. Within each matched set, patients have approximate homogeneity with respect to pre-treatment death risk and can be viewed as the counterfactual cases corresponding to that specific treated patient. Using the selected matching patients, we can estimate the survival probability $S_i^0(t)$ for each treated patient i.

To estimate $\widehat{S}_i^0(t)$, unlike the analog on treated side, survival in the absence of treatment for treated patient i is subject to dependent censoring. For a matched patient k for treated patient i, we anticipate that $\mathcal{H}_k(T_i + t)$ would be predictive of both the treatment hazard and pre-treatment death hazard at time (T_i+t). However, in the matching process, we only conditioned on $\mathcal{H}_k(T_i)$. For matched patient k, $(D_k^0 - T_i)_+$ can be either censored by $(C_k - T_i)_+$ or $(T_k - T_i)_+$, and both represent violations of independent censoring. As we assume censoring only dependent on baseline covariates as in model (4.2), we will only consider the dependent censoring caused by treatment assignment. The potential bias due to such dependent censoring can be corrected through a variant of Inverse Probability of Censoring Weighting

(IPCW; Robins and Rotnitzky, 1992). For this propose, we fit the following treatment model:

$$\lambda_i^T(t|\mathcal{H}_i(t)) = \lambda_0^T(t)\exp\left\{\boldsymbol{\beta}_T'\boldsymbol{Z}_i(t)\right\}. \tag{4.7}$$

We define $\widetilde{Y}_{i:k}^0(t) = Y_k^0(T_i+t)$ and $\widetilde{N}_{i:k}^0 = dN_k^0(T_i+t)$ and $I_{i:k}$ being the indicator for not-yet-treated patient k matched to patient i. The weight of patient k is given by

$$\begin{aligned} w_{i:k}^0(t) &= \frac{I_{i:k}\tilde{Y}_{i:k}^0(t)}{P(T_k > T_i + t | T_k > T_i, T_i, \boldsymbol{Z}_k)} \\ &= I_{i:k}\tilde{Y}_{i:k}^0(t)\exp\left\{\Lambda_k^T(T_i+t) - \Lambda_k^T(T_i)\right\}, \end{aligned}$$

where $\Lambda_k^T(t) = \int_0^t \lambda_0^T(u)\exp\left\{\boldsymbol{\beta}_T'\boldsymbol{Z}_k(u)\right\}du$. The quantity $w_{i:k}^0(t)$ can be thought of as the inverse of the conditional probability of remaining untreated and uncensored at time (T_i+t), given that the subject was untreated and treatment-eligible at time T_i. Our proposed estimator of $S_i^0(t)$ is given by $\widehat{S}_i^0(t) = \exp\left\{-\widehat{\Lambda}_i^0(t)\right\}$, where

$$\widehat{\Lambda}_i^0(t) = \sum_{k=1}^n \int_0^t \left\{\sum_{k=1}^n w_{i:k}^0(u)\right\}^{-1} w_{i:k}^0(u)d\tilde{N}_{i:k}^0(u) \tag{4.8}$$

Note that, in the treated group, every treated patient is unique but in the untreated group, the same patient can appear in multiple matched sets.

4.2.4 Variance Estimator for $\widehat{S}^1(t)$

We now provide heuristic arguments leading to a variance estimator for $S^1(t)$.

As n goes to ∞, we let the number of groups K goes to ∞ as well but at a slower rate than n such that the the number of individuals in each group will also go to infinity. As we assume post-treatment survival only depends on the variables at time of treatment, individuals in the same group can be viewed as identical with respect to conditional post-treatment death hazard, corresponding to $S_1(t;s|\mathcal{H}(s), T=s)$. If there is no censoring, by the definition of $S^1(t)$, $\widehat{S}^1(t) = n^{-1}\sum_{k=1}^K n_k\widehat{S}_k^1(t)$. When

$K \to \infty$, $\widehat{S}^1(t) = n^{-1}\sum_{i=1}^{n} S_i^1(t)$, by the Uniform Weak Law of Large Number (UWLLN), $\widehat{S}^1(t) \to S^1(t)$ in probability for all $S^1(t)$.

When there is no censoring , $w_{k\bullet}\left\{\sum_{k=1}^{K} w_{k\bullet}\right\}^{-1}$ can be viewed as a density function of a certain function of $(T, \boldsymbol{Z})$ given prognostic score $\boldsymbol{\beta}_1'\boldsymbol{Z}_i(T_i)+\beta_2' g(T_i)$ falling in kth interval. We denote this density as $f_\phi(k) = w_{k\bullet}\left\{\sum_{k=1}^{K} w_{k\bullet}\right\}^{-1}$. As $n \to \infty$, $K \to \infty$, such that each interval of the prognostic score will be close to a value on the domain. Thus we can re-write $f_\phi(k)$ as $f_\phi(x)$ with $x \in (-\infty, \infty)$. Therefore we have

$$\widehat{S}^1(t) = \int_{-\infty}^{\infty} f_\phi(x)\widehat{S}_x^1(t)dx = \int_{-\infty}^{\infty} \widehat{S}_x^1(t)dF_\phi(x), \tag{4.9}$$

where F_ϕ is the CDF of ϕ.

In order to study the asymptotic distribution of $\widehat{S}^1(t)$, we need to define an additional set of notation. Let $Y_{ik}(t) = G_{ik}^1 Y_i(t)$, $dM_{ik}^1(t) = G_{ik}^1\{dN_i(t) - Y_i(t)d\Lambda_k^1(t)\}$. We also define $\rho = E(\Delta_i^T)$ which can be estimated by $n^{-1}\sum_{i=1}^{n}\Delta_i^T$. We then have

$$n^{1/2}\rho^{1/2}\left\{\widehat{\Lambda}_k^1(t) - \Lambda_k^1(t)\right\} = n^{1/2}\rho^{1/2}\sum_{i=1}^{n}\varphi_{ik}^1(t),$$

where $\widehat{\Lambda}_j^1(t)$ is the estimator of post-treatment cumulated hazard function for group k, and $\varphi_{ik}^1(t) = \int_0^t \pi^{-1}(u)dM_{ik}^1(u)$, where $\pi(u) = P(U \geq u)$. Under mild regularity conditions, $\{\varphi_{1k}^1(t), \ldots, \varphi_{nk}^1(t)\}$ are independent and identically distributed mean 0 variates. As a result, $n^{1/2}\left\{\widehat{\Lambda}_k^1(t) - \Lambda_k^1(t)\right\}$ converges to asymptotically to a mean-zero normal distribution with variance $E\left[\varphi_{1k}^1(t)^2\right]$ by the Multivariate Central Limit Theorem. By applying the Functional Delta Method, we obtain that $n^{1/2}\left\{\widehat{S}_k^1(t) - S_k^1(t)\right\}$ is also asymptotically mean-zero Normal with variance estimator,

$$\widehat{\sigma}_k^2(t) = n^{-1}\sum_{i=1}^{n}\left\{\widehat{S}^1(t)\widehat{\varphi}_{ik}^1(t)\right\}^2,$$

where $\widehat{\varphi}_{ik}^1(t) = \int_0^t \widehat{\pi}^{-1}(u)d\widehat{M}_{ik}^1(u)$ and $d\widehat{M}_{ik}^1(u) = G_{ik}\{dN_i^1(u) - Y_{ik}^1(u)d\Lambda^1(u)\}$

As discussed above, $\widehat{S}^1(t)$ should converge in probability to its limiting value $S^1(t) = \int_{-\infty}^{\infty} S_x^1(t) dF_\phi(x)$, in the discretized case we define $S(t) = \sum_{k=1}^{K} f_\phi(k) S_k^1(t)$.

Next, we consider the asymptotic distribution of $\widehat{S}^1(t)$. We can write,

$$\begin{aligned}
\widehat{S}^1(t) - S^1(t) &= \sum_{k=1}^{K} f_\phi(k) \left\{ \widehat{S}_k^1(t) - S_k^1(t) \right\} \\
&= \sum_{k=1}^{K} f_\phi(k) \left\{ - \sum_{i=1}^{n_T} S_k(t) \varphi_{ik}^1(t) \right\} \\
&= - \sum_{k=1}^{K} f_\phi(k) S_k^1(t) \sum_{i=1}^{n} \varphi_{ik}^1(t) \\
&= - \sum_{i=1}^{n} \sum_{k=1}^{K} f_\phi(k) S_k^1(t) \varphi_{ik}^1(t) \\
&= - \sum_{i=1}^{n} \varphi_{i\bullet}^1(t),
\end{aligned}$$

where $\varphi_{i\bullet}^1(t) = \sum_{k=1}^{K} f_\phi(k) S_k^1(t) \varphi_{ik}^1(t)$. Since $\varphi_{i\bullet}^1(t)$ are also independent and mean-zero, $n^{1/2} \left\{ \widehat{S}^1(t) - S^1(t) \right\}$ converges in distribution to a zero-mean Normal with a variance that can be consistently estimated by

$$\widehat{\sigma}_1^2 = n^{-1} \sum_{i=1}^{n} (\widehat{\varphi}_{i\bullet}^1)^2,$$

where we define

$$\widehat{\varphi}_{i\bullet}(t) = \sum_{k=1}^{K} f_\phi(k) \widehat{S}_k^1(t) \widehat{\varphi}_{ik}^1(t).$$

4.2.5 Variance Estimator for $\widehat{S}^0(t)$

As long as $\widehat{S}_i^0(t) = \exp\left\{ -\widehat{\Lambda}_i^0(t) \right\}$ is an consistent estimator for $S_i^0(t)$, $\widehat{S}^0(t)$ will converge in probability to

$$S^0(t) = \int_{-\infty}^{\infty} f_\phi(x) S_x^0(t) dx = \int_{-\infty}^{\infty} S_x^0(t) dF_\phi(x), \tag{4.10}$$

with $f_\phi(x)$ defined as in the previous section. A discretized version of (4.10) can be expressed as $S^0(t) = \sum_{k=1}^{K} f_\phi(k) S_k^0(t)$, where $f_\phi(k) = w_{k\bullet}$.

By using the matching on the treatment-free prognostic score, qualified patients selected should have identical treatment-free death hazard to the specific treated patient. We use a variant of IPCW to remedy the issue of dependent censoring for not-yet-treated patients, such that $\widehat{\Lambda}_i^0(t)$ is IPCW-adjusted Nelson-Aalen estimator for $\Lambda_i^0(t)$.

Next we derive the limiting distribution for $\widehat{S}^0(t)$. We start with the limiting distribution for $\widehat{S}_i^0(t)$. Let $G_{ik}^0 = I(G_i^0 = k)$. Correspondingly we define $Y_{ik}^0 = G_{ik}^0 Y_i^0(t)$ and $dM_{ik}^0(t) = G_{ik}^0 \{dN_i^0(t) - Y_i^0(t)d\Lambda^0(t)\}$. Similar to the process for $\widehat{S}_k^1(t)$, analogous arguments lead to

$$n^{1/2}\left\{\widehat{S}_k^0(t) - S_k^0(t)\right\} = -n^{1/2}\sum_{i=1}^{n} S_k^0(t)\varphi_{ik}^0(t)$$

asymptotically. Different from the treatment side, where each subject k can appear only once, a given subject i in the treatment-free side can be matched to sevreal treated patients. As such, the asymptotically independent terms with respect to the treatment-free side are given by

$$\varphi_{ik}^0(t) = \int_0^t \pi^{-1}(u)dM_{ik}^0(u),$$

such that $n^{1/2}\left\{\widehat{S}_k^0(t) - S_k^0(t)\right\}$ converges in distribution to a zero-mean Normal with a variance that can be consistently estimated by

$$(\widehat{\sigma}_k^0)^2(t) = n^{-1}\sum_{i=1}^{n}\left\{\widehat{S}_k^0(t)\widehat{\varphi}_{ik}^0\right\}^2,$$

where we define $\widehat{\varphi}_{ik}^0 = \int_0^t \widehat{\pi}^{-1}(u)d\widehat{M}_{ik}^0(u)$ with $d\widehat{M}_{ik}^0(u) = G_{ik}^0\left\{dN_i^0(u) - Y_i^0(u)d\widehat{\Lambda}^0(u)\right\}$.

Similar to $\widehat{S}^1(t)$, $\widehat{S}^0(t) - S^0(t) = -\sum_{i=1}^n \varphi_{i\bullet}^0$ with $\varphi_{i\bullet}^0(t) = \sum_{k=1}^{n_T} f_\phi(k)S_k^0(t)\varphi_{ik}^0(t)$. Therefore $n^{1/2}\left\{\widehat{S}^0(t) - S^0(t)\right\}$ is asymptotically mean-zero Normal with variance estimator,

$$\widehat{\sigma}_0^2(t) = n^{-1}\sum_{i=1}^{n}(\widehat{\varphi}_{i\bullet}^0)^2,$$

where we define $\widehat{\varphi}^0_{i\bullet}(t) = \sum_{k=1}^{n_T} f_\phi(k)\widehat{S}^0_k(t)\widehat{\varphi}^0_{ik}(t)$.

Combining the results above, we can represent

$$n^{1/2}\left\{\widehat{\delta}(t) - \delta(t)\right\} = n^{1/2}\sum_{i=1}^{n}\left\{\varphi^1_{i\bullet}(t) - \varphi^0_{i\bullet}(t)\right\},$$

where $\varphi^1_{i\bullet}(t) - \varphi^0_{i\bullet}(t)$ components are independent and identically distributed with mean 0. Note that the grouping of combinations by subject implicitly accounts for the possibility that patients may contribute follow-up on both the $S^0(t)$ and $S^1(t)$ sides. The quantity $n^{1/2}\left\{\widehat{\delta}(t) - \delta(t)\right\}$ converges asymptotically to a Normal variate with mean 0 and a variance that can be consistently estimated by

$$\widehat{\sigma}^2_\delta(t) = n^{-1}\sum_{i=1}^{n}\left\{\varphi^1_{i\bullet}(t) - \varphi^0_{i\bullet}(t)\right\}^2.$$

For computational convenience, we propose to use the bootstrap method to compute the asymptotic variance, as the point estimators are fast to compute. We evaluate the performance of bootstrap in the next section.

4.3 Simulation

Simulations were conducted to evaluate the proposed methods in finite samples. The treatment time T was generated from an exponential distribution with hazard $\lambda_{T0}\exp\{\beta_{T1}Z_1 + \beta_{T2}Z_2 + \beta_{T3}Z_3(t) + \beta_{T4}Z_4(t)\}$ while treatment-free death times D^0 were generated as exponential with hazard $\lambda_{D0}\exp\{\beta_{D1}Z_1 + \beta_{D2}Z_2 + \beta_{D3}Z_3(t) + \beta_{D4}Z_4(t)\}$. Here, Z_1, Z_2, Z_3 and Z_4 are confounders that affect both T and D^0. Censoring times C were generated from an exponential distribution with hazard $\lambda_{C0}\exp\{\beta_{C1}Z_1 + \beta_{C2}Z_2\}$. Times between treatment and death $(D^1 - T)_+$ were generated from exponential distribution with rate $\lambda_{10}\exp\{\beta_{10}Z_1 + \beta_{11}Z_2 + \beta_{12}Z_3(T) + \beta_{13}Z_4(T) + \beta_{15}T\}$, where we set $\beta_{10} = \beta_{D1}$, $\beta_{11} = \beta_{D2}$,$\beta_{13} = \beta_{D3}$ and $\beta_{14} = \beta_{D4}$. Baseline covariates Z_1 and Z_2 were generated from a Uniform(-1,1). Baseline Z_3 and Z_4 were generated

by Uniform (0,1) and Uniform (-0.5, 1.5), respectively. Z_3 and Z_4 have increments that follow Uniform (0,1) and Uniform (-0.5, 1.5) respectively at each time unit. We denote the actual death times as D, for treated patients $D = D^1 = T + (D^1 - T)_+$ and for untreated patients $D = D^0$. There were $n = 5000$ subjects in all simulation configurations, with each data configuration replicated 500 times. To obtain the estimated standard error we bootstrap 25 datasets for each run. The bootstrap sample size is 2500 to reduce computation time ([2],[3]), with the subsequent standard error estimator approximately re-scaled.

In practice, we observe the minimum of T, D^0 and C. In simulations, however, we always observe T, D^0, $(D^1 - T)_+$ and C for all patients. True values of $S_1(t)$, $S_0(t)$, $\delta(t)$ and $\Delta(L)$ were obtained using monte-carlo on these counterfactuals. Naturally, for the purpose of computing $\widehat{S}_0(t)$, only $[(D \wedge C \wedge T), N^D(D \wedge C \wedge T)]$ were used; similarly, only $[(D \wedge C - T)_+, N^D(D \wedge C)]$ were used for subjects with $(D \wedge C > T)$ for the purpose of computing $\widehat{S}_1(t)$. Hence, $\widehat{\delta}(t)$ and $\widehat{\Delta}(L)$ were, for each replicate, only based on data that would in reality be observed.

After generating the data, the partly conditional model,

$$\lambda_i^0(t; s|\mathcal{H}_i(s)) = \lambda_{00}(t) \exp\left\{\beta_{00}Z_1 + \beta_{01}Z_2 + \beta_{02}Z_3(s) + \beta_{03}Z_4(s)\right\},$$

was fitted to obtain the treatment-free prognostic score. To fit the conditional model, we generated a random variable that follows Uniform (0,25) for every individual as their listing calender time and set cross section $CS = \{5, 10, 15, 20, 15\}$. For post-treatment survival, the model

$$\lambda_i^1(t; s|\mathcal{H}(s), T_i = s) = \lambda_{10} \exp\left\{\beta_{10}Z_1 + \beta_{11}Z_2 + \beta_{12}Z_3(s) + \beta_{13}Z_4(s) + \beta_{15}s\right\}$$

was fitted to obtain the post-treatment prognostic score. The estimator of $\Lambda^T(t)$ was

calculated from a fully time-dependent Cox model,

$$\lambda_i^T(t|\mathcal{H}_i(t)) = \lambda_0^T \exp\left\{\beta_{T1}Z_1 + \beta_{T2}Z_2 + \beta_{T3}Z_3(t) + \beta_{T4}Z_4(t)\right\}.$$

The caliper width ε for matching was set to 0.05. In estimating $S^1(t)$, patients are grouped by half deciles of prognostic score. For all simulations configurations we set $\tau_T = 10$.

In the first set of simulations, we examine the bias, empirical standard deviation and bootstrap standard deviation of the proposed estimators with various treatment effect under light censoring with $\lambda_{C0} = 0.015$, where around 15% of individuals get censored. We set $\lambda_{0D} = 0.15, 0.25$ for moderate and negative treatment effect. The remaining parameters are set equal across the three scenarios: $\lambda_{T0} = 0.01$, $\beta_{D1} = \beta_{T1} = \beta_{C1} = \log 2$, $\beta_{14} = \log(3)/100$, $\beta_{D2} = \beta_{T2} = \beta_{C2} = \log 3$,$\beta_{D3} = \beta_{T3} = \log 3/10$, $\beta_{D4} = \beta_{T4} = \log 2/10$ and $\lambda_{10} = 0.02$.

In the second set of simulations, we examine the properties of the proposed estimators under moderate censoring. The parameter setting is same as the first set of simulations, except for the change in the censoring parameter to $\lambda_{C0} = 0.02$, which results in approximately 30% censoring.

Results for the first and second set of simulations are shown in Table 4.1 and Table 4.2, respectively. In both tables, the absolute bias of $S_1(t)$ $S_0(t)$ and $\delta(t)$ range from 0.001 to 0.017. The bias of μ_0, μ_1 and Δ is slightly larger, but considering the scale of these quantities, the bias is still negligible. The bootstrap standard error (BSE) was generally close to empirical standard deviation (ESD). The coverage probability is around 0.95, except for $S_0(15)$. Some degree of under-coverage is observed, but not in unacceptable amounts. The estimation of $S_0(t)$ is more sensitive to censoring percentage, as the censoring percentage does not affect the results of $S_1(t)$ very much, while the bias of $\widehat{S}_0(t)$ becomes more pronounced with larger censoring percentage.

Table 4.1: Simulation results: Light censoring

Setting	t	Quantity	True	BIAS	ESD	BSE	CP
Moderate	3	$S^1(t)$	0.819	0.002	0.014	0.014	0.94
treatment	6		0.682	0.001	0.019	0.017	0.94
effect	10		0.547	-0.001	0.021	0.019	0.93
	3	$S^0(t)$	0.771	0.005	0.006	0.006	0.93
	6		0.572	0.008	0.010	0.009	0.91
	10		0.360	0.014	0.011	0.011	0.86
	3	$\delta(t)$	0.048	0.004	0.014	0.014	0.97
	6		0.111	0.008	0.020	0.018	0.92
	10		0.187	0.015	0.022	0.020	0.90
	10	$\mu_0(t)$	7.411	0.089	0.075	0.069	0.90
	10	$\mu_1(t)$	6.508	0.009	0.134	0.126	0.92
	10	$\Delta(t)$	0.904	0.080	0.142	0.134	0.91
Negative	3	$S^1(t)$	0.800	0.002	0.013	0.014	0.92
treatment	6		0.656	0.002	0.018	0.017	0.93
effect	10		0.517	-0.001	0.019	0.019	0.92
	3	$S^0(t)$	0.838	0.004	0.005	0.006	0.93
	6		0.679	0.008	0.009	0.009	0.93
	10		0.484	0.016	0.011	0.011	0.86
	3	$\delta(t)$	-0.037	0.002	0.014	0.014	0.97
	6		-0.022	0.007	0.019	0.018	0.92
	10		0.033	0.014	0.021	0.020	0.90
	10	$\mu_0(t)$	7.356	0.078	0.069	0.069	0.90
	10	$\mu_1(t)$	7.197	0.015	0.127	0.123	0.93
	10	$\Delta(t)$	-0.159	0.063	0.137	0.132	0.91

Table 4.2: Simulation results: Moderate censoring

Setting	t	Quantity	True	BIAS	ESD	BSE	CP
Moderate	3	$S^1(t)$	0.819	0.001	0.014	0.014	0.94
treatment	6		0.682	0.001	0.018	0.017	0.94
effect	10		0.547	0.002	0.019	0.019	0.93
	3	$S^0(t)$	0.771	0.006	0.007	0.006	0.93
	6		0.572	0.009	0.010	0.009	0.91
	10		0.360	0.017	0.011	0.011	0.87
	3	$\delta(t)$	0.048	0.006	0.014	0.014	0.96
	6		0.111	0.008	0.018	0.018	0.92
	10		0.187	0.015	0.019	0.020	0.90
	10	$\mu_0(t)$	7.411	0.092	0.078	0.069	0.91
	10	$\mu_1(t)$	6.508	0.011	0.123	0.126	0.92
	10	$\Delta(t)$	0.904	0.081	0.136	0.134	0.91
Negative	3	$S^1(t)$	0.800	0.000	0.013	0.014	0.92
treatment	6		0.656	-0.002	0.017	0.017	0.93
effect	10		0.517	-0.001	0.019	0.019	0.92
	3	$S^0(t)$	0.838	0.004	0.006	0.006	0.93
	6		0.679	0.008	0.009	0.009	0.91
	10		0.484	0.016	0.012	0.011	0.86
	3	$\delta(t)$	-0.037	0.004	0.013	0.014	0.97
	6		-0.022	0.010	0.017	0.018	0.92
	10		0.033	0.017	0.019	0.020	0.90
	10	$\mu_0(t)$	7.356	0.063	0.071	0.067	0.90
	10	$\mu_1(t)$	7.197	-0.010	0.117	0.123	0.93
	10	$\Delta(t)$	-0.159	0.073	0.122	0.132	0.91

4.4 Analysis of Liver Transplant Data

We applied our proposed methods in real data to estimate the effect of deceased-donor liver transplantation (LT) ($j = 1$) on survival compared to the absence of LT ($j = 0$) among waitlisted patients, by Model for End-stage Liver Disease (MELD) score. Data were obtained from Scientific Registry of Transplant Recipients. The SRTR data system includes data on all donors, wait-listed candidates, and transplant recipients in the United States, as submitted by the members of the Organ Procurement and Transplantation Network (OPTN), and has been described elsewhere. The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors.

The study population included $n = 108,236$ patients aged ≥ 18 and listed between 01/01/2005 and 12/31/2016, of which 58,941 received deceased donor liver transplant. We excluded patients who were Status 1 (acute liver failure) or previously transplanted. Follow-up time begins at the date when the patient got listed and ends at earliest of death, loss to follow-up, or the end of the observation period (12/31/2016). Cross-section dates are chosen by every 2 years, which leads to 1/1/2007, 1/1/2009, 1/1/2011/, 1/1/2013 and 1/1/2015. The transplant hazard model, $\lambda_i^T(t|\mathcal{H}_i(t)) = \lambda_0^T(t)\exp\left\{\boldsymbol{\beta}_T'\boldsymbol{Z}_i(t)\right\}$, was adjusted by age, gender, race/ethnicity, diagnosis, body mass index, blood type, albumin, dialysis, diabetes, ascites, hepatic encephalopathy, allocation MELD score, serum sodium, international normalized ratio.The pre-transplant model for $\lambda_i^0(t)$ is adjusted by age, gender, race/ethnicity, diagnosis, body mass index, blood type, albumin, dialysis, diabetes, ascites, hepatic encephalopathy, lab MELD score, serum sodium, serum creatinine,

serum bilirubin, international normalized ratio and time on wait-list. In the post-transplant model, $\lambda_1(t; s|\mathcal{H}(s), T_i = s) = \lambda_0^1(t) \exp\left\{\boldsymbol{\beta}_1' \boldsymbol{Z}_{i1}(s) + \boldsymbol{\beta}_2' \boldsymbol{g}(s)\right\}$, $\boldsymbol{Z}_{i1}(s)$ include treatment time T_i, age, gender, race/ethnicity, diagnosis, body mass index, blood type, albumin, dialysis, diabetes, ascites, hepatic encephalopathy, lab MELD score, serum sodium, serum creatinine, serum bilirubin and international normalized ratio.

In this application, we set τ_T and τ_1 to 3 years and 5 years, respectively. A total of 58,941 patients received LT (54.45%) and 41,055 deaths were observed. We set the caliper width of prognostic score matching $\epsilon = 0.01$.

We show the results for lower MELD score groups. The estimated survival curves on [0,5] year interval of the two groups are presented in Figure 4.1, by MELD score category. Note that the MELD score categories refer to MELD at transplant. Within a MELD category, $\widehat{S}^1(t)$ can be interpreted as the average survival probability, with t representing residual time post-transplant. Analogously, $\widehat{S}^0(t)$ is the average survival that would have resulted in the absence of liver transplantation, among patients who received a liver transplant. The survival curve of transplant group remain similar across different MELD score groups. The survival for the waitlist group decreases strongly with the increase of MELD score.

In Table 4.3, we list estimates of difference in survival probability, $\widehat{\delta}(t)$ for $t = 1, 3, 5$, as well as $\widehat{\Delta}(5)$, the difference in 5-year restricted mean residual survival time. MELD 12-14 group gains the most benefit from the transplantation ($\Delta(5) = 0.864$ year) which was caused by the low survival probability in this MELD group. All the groups have significant transplant effect on 5-year survival.

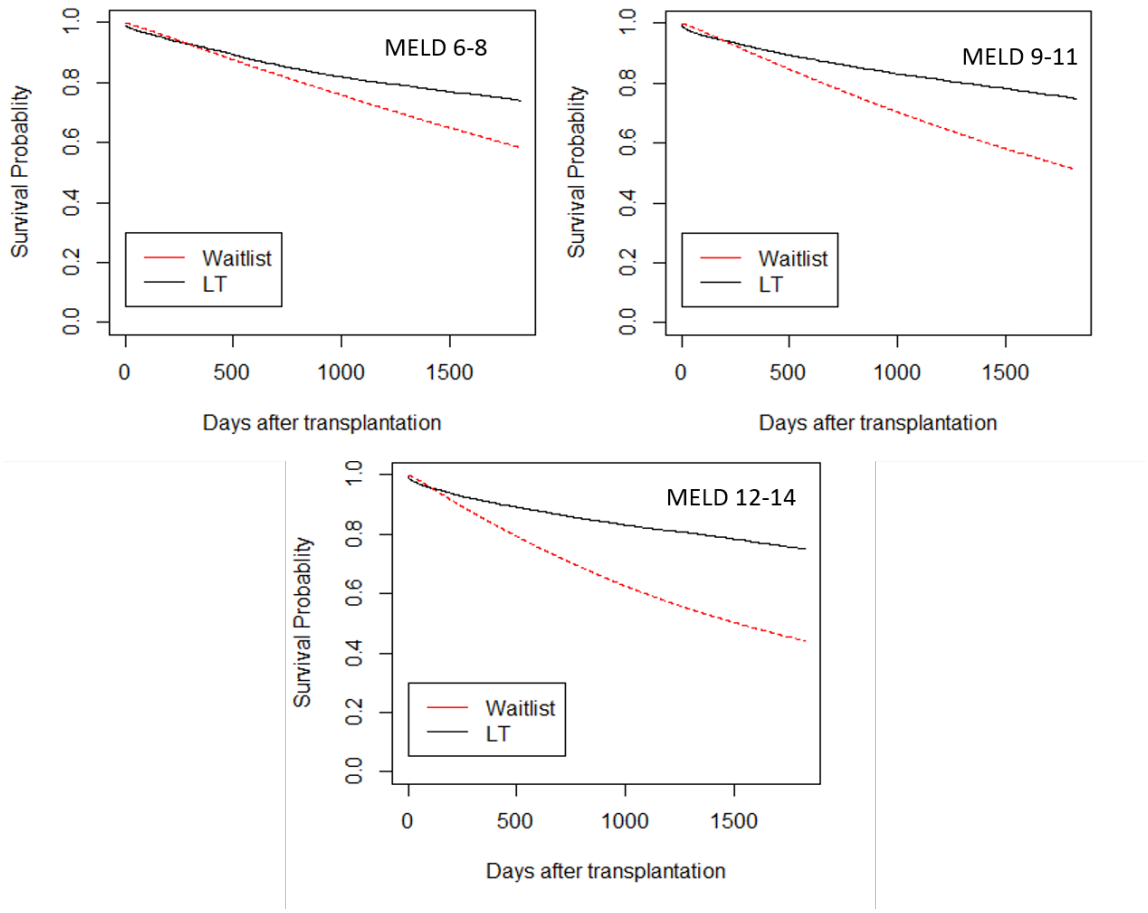

Figure 4.1: Analysis of SRTR data: estimated survival curves

Table 4.3: Analysis of SRTR data: estimating the effect of liver transplantation on the transplanted (with 95% confidence interval in parentheses), by MELD score at transplant

MELD Score	$\widehat{\delta}(1)$	$\widehat{\delta}(3)$	$\widehat{\delta}(5)$	$\widehat{\Delta}(5)$ (year)
6-8	0.007	0.072	0.156	0.294
	(0.000,0.014)	(0.061, 0.083)	(0.140, 0.172)	(0.254, 0.336)
9-11	0.026	0.143	0.237	0.557
	(0.018, 0.033)	(0.130, 0.156)	(0.222, 0.252)	(0.509, 0.606)
12-14	0.065	0.223	0.309	0.864
	(0.057, 0.073)	(0.213, 0.233)	(0.296, 0.322)	(0.824, 0.904)

4.5 Discussion

In this chapter, we proposed methods to estimate the average causal effect among the treated of a time-dependent treatment with the presence of time-dependent variables. In particular, the proposed treatment effect contrasts the post-treatment survival with the survival function that applied to treated patients had they, contrary to fact, not received treatment. To estimate the survival of treated patients in the absence of treatment, we proposed a matching method to create a group of patients that is considered as counterfactual version for each treated patient and then properly average over the survival functions. The limiting distribution and asymptotic variances were derived. For computational conveniences, a bootstrap method is employed to estimate the standard error. The proposed methods were shown through simulations to work well in finite sample sizes. The proposed methods were applied to quantify the survival benefit of deceased donor liver transplantation among the transplanted, by Model for End-stage Liver Disease (MELD) score.

Chapter IV is a generalization from Chapter II in the presence of time-dependent variables and, as such inherits some basic characteristics of the methods in Chapter II. The proposed method is non-parametric in the sense that the assumed models only contribute to the prognostic score used for matching. For the treatment effect, we target directly at survival function and restricted mean lifetime, which is more flexible compared to a hazard ratio. Due to the nature of non-parametric estimation, the proposed methods do not require that pre- and post-treatment hazards are proportional or have any particular relationship to each other. With the presence of time-varying variables, we incorporate the partly conditional model as a working model for the pretreatment survival.

The results from SRTR show that higher MELD score groups benefit more from the transplantation, which is intuitive because higher MELD score usually indicate higher death mortality. Transplantation in higher mortality groups will bring patients larger difference in mean survival time, but it does not necessarily suggest patients not to get transplanted until their MELD score gets higher.

There are now many methods available for evaluating a time-dependent treatments. Marginal Structural Models (MSM; [12],[13]) are not well-suited to our setup due to the potential for treatment to interact with timevarying covariates. Structural Nested Failure Time Models (SNFTMs;[27],[14]) are an alternative. These methods either do not target at survival functions or do not estimate average treatment effect among the treated. Gong and Schaubel (2017) proposed methods that use partly conditional modeling to estimate the treatment effect. Compared to Gong and Schaubel (2017), the proposed methods only use the partly conditional model to obtain the prognostic score instead of using them as treatment effect. However, Gong and Schaubel (2017) rely more heavily on correct specification of the model.

Due to the nature of matching, the proposed methods have the advantages of handling covariates of higher dimensions, and greater robustness towards model misspecification. On the other hand, matching also relies on several assumptions such as no unmeasured confounding and overlapping support between treated and untreated groups. In the process of matching, caliper width is subject to change depending on the application at hand. One needs to ensure that every treated subject gets a sufficient number of matched subjects while ensuring that the matched sets consist only of patients similar enough to the index treated subject.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Aalen O.O. (1989). A linear regression model for the analysis of life times. *Statistics in medicine* **8(8)**, 907–925.

[2] Bickel, Peter J., Gotze, F. and van Zwet W. R. (1997) Resampling fewer than n observations: gains, losses and remedies for losses. *Statistica Sinica* **7**, 1-31.

[3] Bickel, Peter J. and Sakov, Anat.(2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica* **18**, 967-985.

[4] Chen, P. and Tsiatis, A. A. (2001). Causal inference on the difference of the restricted mean life between two groups. *Biometrics* **57**, 10301038.

[5] Cox D.R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society,Series B* **34**,187-200.

[6] Cox D.R. and Oakes D. The analysis of survival data. Chapman and Hall, New York (1984).

[7] Efron B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics* **7**, 126.

[8] Fleming T.R. and Harrington D.P. Counting processes and survival analysis. New York: John Wiley and Sons (1991).

[9] Gong Q., Schaubel D.E. (2013). Partly conditional estimation of the effect of a timedependent factor in the presence of dependent censoring *Biometrics* **69(2)**, 338-347.

[10] Gong Q., Shcaubel D.E. (2016). Estimating the average treatment effect on survival based on observational data and using partly conditional modeling *Biometrics* **73(1)**, 134–144.

[11] Hansen B.B. (2008). The prognostic analogue of the propensity score. *Biometrics* **95(2)**, 481–488.

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

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

[14] Hernan M.A., Cole S.R., Margolick J, Cohen M. and Robins J.M. (2005). Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and drug safety* **14**, 477–491.

[15] Hernan, M. A., Cole, S. R., Margolick, J., Cohen, M., and Robins, J.M. (2005). Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and Drug Safety* **14**, 477491.

[16] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data: 2nd Edition*, New York: Wiley.

[17] Kaplan E.L. and Meier P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **282**, 457481.

[18] Li Y., Schaubel D.E. and He K. (2014). Matching methods for obtaining survival functions to estimate the effect of a time-dependent treatment. *Statistics in Biosciences* **6(1)**, 105–126.

[19] Lin D.Y. and Ying Z. (1994). Semiparametric analysis of additive risk model. *Biometrika*, **81**, 61-71.

[20] Lin D.Y. and Ying Z. (1995). Semiparametric analysis of the additive-multiplicative intensity models for counting processes. *Ann. Statist.*, **23**, 1712-1734.

[21] Lok J., Gill R., van der Vaart A and Robins J. (2004). Estiamting the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statistica Neerlandica* **58**, 271–295.

[22] Lu B. (2005). Propensity score matching with time-dependent covariates. *Biometrics* **61**, 721-728.

[23] Petersen M.L., Deeks S.G., Martin J.N. and van der Laan M.J. (2007). History-adjusted marginal structural models for estimating time-varying effect modification. *American Journal of Epidemiology* **166**, 185- 193.

[24] Pollard, David. Section 8: Uniform laws of large numbers. Empirical Processes, 39–42, Institute of Mathematical Statistics and American Statistical Association, Haywood CA and Alexandria VA (1990).

[25] Prentice R.L. and Breslow N.E. (1978). Retrospective studies and failure time models. *Biometrika* **65**, 153-158.

[26] Rao P.S., Schaubel D.E., Guidinger M.K., Andreoni K.A., Wolfe R.A., Merion R.M., Port F.K. and Sung R.S. (2009). A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. *Transplantation* **88(2)**, 231–236.

[27] Robins J.M., Hernn M.A. and Brumback B. (2000). Marginal structural models and causal inference in epidemiology.*Epidemiology* **11**, 550-560.

[28] Robins J.M., Rotnitzky A. (1992). Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers. In: Jewell N.P., Dietz K., Farewell V.T. (eds) AIDS Epidemiology. Birkhuser, Boston, MA.

[29] Robins J.M., Blevins D., Ritter G. and Wulfsohn M. (1992). G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology* **3**, 319–336.

[30] Rosenbaum P. and Rubin D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.

[31] Rubin D.B. (1974). Estimating casual effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688-701.

[32] Schaubel D.E., Wolfe R.A. and Port F.K. (2006). A sequential stratification method for estimating the effect of an time-dependent experimental treatment in observational studies. *Biometrics* **62**, 910917.

[33] Schaubel D.E., Wolfe R.A., Sima C.S. and Merion R.M. (2009). Estimating the effect of a time-dependent treatment in by levels of an internal time-dependent covariate. *Journal of the American Statistical Association* **104**, 49-59

[34] Sharma, P., Schaubel, D. E., Goodrich, N. P., and Merion, R. M. (2015). Serum sodium and the survival benefit of liver transplantation. *Liver Transplantation* **21**, 308313.

[35] Stuart E.A. (2010). Matching Methods for Causal Inference: A review and a look forward. *Statistical Science* **25(1)**, 121.

[36] Van Houwelingen, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics* **34**, 7085.

[37] van der Vaart, A.W. Asymptotic statistics. Cambridge University Press (2000).

[38] Zhang, M. and Schaubel, D. E. (2011). Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics* **67**, 740749.

[39] Zheng, Y. Y. and Heagerty, P. J. (2005). Partly conditional survival models for longitudinal data. *Biometrics* **61**, 379391.