NIH Public Access
Author Manuscript
Stat Med. Author manuscript; available in PMC 2014 October 01.

Published in final edited form as:
Stat Med. 2012 December 30; 31(30): 4255–4268. doi:10.1002/sim.5511.

Contrasting treatment-specific survival using double-robust estimators

Min Zhang[a,*] and **Douglas E. Schaubel**[a]

[a] Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, USA.

Abstract

In settings where a randomized trial is infeasible, observational data are frequently used to compare treatment-specific survival. The average causal effect (ACE) can be used to make inference regarding treatment policies on patient populations, and a valid ACE estimator must account for imbalances with respect to treatment-specific covariate distributions. One method through which the ACE on survival can be estimated involves appropriately averaging over Cox-regression-based fitted survival functions. A second available method balances the treatment-specific covariate distributions through Inverse Probability of Treatment Weighting (IPTW), then contrasts weighted nonparametric survival function estimators. Since both methods have their advantages and disadvantages, we propose methods which essentially combine both estimators. The proposed methods are double-robust, in the sense that they are consistent if at least one of the two working regression models (i.e., logistic model for treatment, and Cox model for death hazard) is correct. The proposed methods involve estimating the ACE with respect to restricted mean survival time, defined as the area under the survival curve up to some pre-specified time point. Asymptotic results are derived and evaluated through simulation. We apply the proposed methods to estimate the ACE of donation-after-cardiac-death kidney transplantation using data obtained from multiple centers in the Netherlands.

Keywords

Average causal effect; Cox regression; Double robust estimator; Inverse weighting; Right censoring; Restricted mean lifetime

1. Introduction

Observational data are frequently used to compare treatment-specific survival in settings where a randomized clinical trial is infeasible. Even in cases where a randomized trial to compare treatments is feasible, observational studies may be an attractive alternative since much greater sample sizes can be obtained at considerably less cost and effort. Methods applicable to observational data include those which accommodate imbalances with respect to the treatment-specific distributions of pre-treatment patient characteristics. Covariate

Copyright © 2011 John Wiley & Sons, Ltd.

[*]Correspondence to: Min Zhang, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, USA. mzhang@umich.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

adjustment may also be desired for randomized trials, to adjust for chance imbalances in adjustment factors and perhaps to increase precision.

Despite the value of observational studies, the randomized trial rightfully serves as the gold standard. This has important implications from at least two angles. First, in analyzing observational data, one is motivated to compute an estimator whose target (e.g., difference in treatment-specific means) would be obtained in the setting where treatment was randomized. Second, there is an incentive to estimate quantities that would be estimated in the context of a randomized trial. For example, in the case of censored survival data, the analysis of data from a randomized trial would likely consist of plots of treatment-specific Kaplan-Meier or Nelson-Aalen survival curves. Survival probability is easily intuited by non-statisticians. A related measure is the area under the survival curve. In particular, if T represents failure time with survival function, $P(T > t) = S(t)$, then mean survival time is equal to the area under the entire survival curve, given by $E[T] = \int_0^\infty S(t)\,dt$. The most popular methods of estimating $S(t)$ in settings where covariate adjustment is unnecessary are nonparametric; i.e., the afore-listed Kaplan-Meier [1] and Nelson-Aalen methods. In cases where covariate effects are modeled, the Cox [2] model (a semiparametric approach) has dominated the hazard regression applications in the biomedical literature almost since its inception. The nonparametric aspects of each of the three afore-listed methods result in inference which is restricted to the $(0, \tau]$ time interval, where τ is the maximum observation time. Although mean survival time may be of most inherent interest, since inference is on a restricted range anyway, restricted mean lifetime, $E[min\{T, L\}] = \int_0^L S(t)\,dt$ is a useful and practical alternative metric; e.g., see Karrison[3], Karrison [4], Zucker [5], Andersen, Hansen and Klein [6]; Schaubel et al [7]; Meier et al [8], Andersen and Perme [9]. Restricted mean lifetime has a straightforward interpretation (i.e., expected number of time units lived out of the next L) and is the measure of interest in this report.

With respect to restricted mean lifetime, the average causal effect (ACE) is the area between the average treatment-specific survival curves (out to $t = L$), with the averaging (for both treatments) being with respect to the marginal covariate distribution. As will be explicitly developed later, this quantity is the same as the area between the unadjusted treatment-specific survival curves in the setting of a randomized study. There are different ways to estimate the pertinent treatment-specific average survival functions. Various authors have advocated fitting Cox models, then explicitly averaging over fitted survival curves; e.g., Karrison[3], Zucker[5], and Chen and Tsiatis[10]. An alternative method involves Inverse Probability of Treatment Weighting (IPTW) [11, 12, 13, 14, 15]. In IPTW, each subject is weighted by the inverse of the probability of being assigned the treatment they actually received. The weighted treatment-specific samples have a covariate distribution that equals that of the margin (across both treatments). Hence, differences between integrated IPTW versions non-parametric estimators serve as estimators of the ACE of restricted mean lifetime. For example, Hubbard et al [16] developed IPTW methods to contrast survival curves, while Wei [15] proposed various measures based on $S(t)$ using the IPTW method, including restricted mean lifetime. Instead of inverse probability weighting, a popular alternative is to match by the probability of being treated ([17, 18], which we do not consider in this article.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In this report, we propose a semiparametric double-robust estimator of the ACE on restricted mean lifetime. The proposed method can be viewed as a hybrid of the approaches of Chen and Tsiatis [10] and Wei [15]. A logistic model is assumed for treatment assignment, and a Cox model is used for the death hazard conditional on treatment and the adjustment covariates; both are working models. The proposed methods are double-robust in the sense that consistent estimation of the ACE is obtained if at least one of the two working models is correct (Bang and Robins [19], Kang and Schafer [20], Robins, Rotnitzky and Zhao [21]).

The data which motivate our methods originate from a multi-center study of kidney transplant patients from the Netherlands [22]. The practice of transplanting kidneys from deceased donors following cardiac death (so-called DCD kidneys) is controversial. The practice happens to be a lot more frequent in the Netherlands than other parts of the world (e.g., the united States). We sought to estimate the average causal effect of DCD kidney transplantation versus the its alternative, which we refer to collectively as non-DCD transplantation.

The remainder of this report is organized as follows. In the next section, we formalize the ideas outlined above, and describe the proposed methods and corresponding asymptotic properties. The finite-sample applicability of the procedures is assessed through simulation in Section 3. The proposed methods are then used in Section 4 to analyze the kidney transplant data described above. Section 5 concludes the report with some discussion.

2. Method

Suppose we are interested in comparing two groups, with group denoted by A (with A=0 or 1) in terms of the mean of the restricted lifetime up to time L. If we denote the survival time by T , then the restricted lifetime is defined as $\min(T, L)$ and restricted mean lifetime can be represented as $E\{\min(T, L)\}$. In the setup we consider in this article, treatment groups are not randomized, and therefore some sort of adjustment for imbalance in baseline covariates, Z, is required. As in almost all studies involving time to an event, survival time is subject to right censoring, denoted by C. We assume that censoring is conditionally independent of death time given treatment; i.e., $T \perp\!\!\!\perp C|A$, an assumption that we discuss further in Section 5. We define the observed possibly censored lifetime as $U = \min(T, C)$ and the indicator for not being censored as $\Delta = I(T \leq C)$. The observed data for each subject i are $(A_i, Z_i, U_i, \Delta_i)$, which are assumed to be independent and identically distributed across $i = 1, \ldots, n$. We denote the observed counting process of event and the at-risk process by $N_i(t) = I(U_i \leq t, \Delta_i = 1)$ and $Y_i(t) = I(U_i \geq t)$, respectively. For simplicity of presentation, we define $A_{ij} = I(A_i = j)$, $N_{ij}(t) = A_{ij}N_i(t)$, and $Y_{ij} = A_{ij}Y_i(t)$.

The quantity that we would like to infer in the comparison of two treatments (the ACE) is the difference in restricted mean lifetimes had all subjects in the population received treatment $A = 1$ as opposed to that had all subjects received $A = 0$. Specifically, denoting the potential (counterfactual) lifetime of a subject, if possibly contrary to facts, s/he received treatment $A = j$ by T^j, $j = 0, 1$, the restricted mean lifetime for treatment $A = j$ is $\mu_j = E\{\min(T^j, L)\}$ and the treatment effect can be contrasted through the difference, $\delta = \mu_1 - \mu_0$. It can be shown that restricted mean lifetime can be represented as the area under the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

survival curve up to L, $\mu_j = \int_0^L S^j(t)\,dt$, where $S^j(t)$ is the marginal survival function of T^j. Therefore, treatment-specific restricted mean lifetimes and their difference represent a cumulative measure of treatment effects. Throughout this article, we will make the Stable Unit Treatment Value Assumption (SUTVA; [23, 24]) which assumes that there is no interference between subjects and that the observed survival time T for a subject receiving treatment $A = j$ is equal to her/his potential lifetime under the treatment j; i.e., $T = AT^1 + (1 - A)T^0$. This assumption usually holds in a study with no interference between subjects, as in the application we will consider. However, this assumption may not hold, for example, in studies involving infectious disease as subjects interfere with each other and an subject's potential response under one treatment may be influenced by treatments other subjects receive.

The average causal effect δ is defined in terms of potential outcomes, which are not observed for all subjects. Nevertheless, inference on the hypothetical quantity δ has to be based on observed data. In observational data, the distribution of baseline covariates Z among subjects in one group is possibly different from that in the other group, or equivalently, Z is not independent of A. If these covariates also predict potential survival times, i.e., Z is not independent of (T^0, T^1), then both treatment difference and differences in Z contribute to the observed difference in survival times between two groups, i.e., the effect of treatment on survival is confounded by imbalance in covariates. Put another way, treatment assignment is not independent of the potential lifetimes due to the mutual correlation with covariates Z, also referred to as confounders. This nonindependence introduces some difficulty in making causal inference based on observational data; in contrast with randomized study where by design treatment assignment is independent of potential lifetimes. A key assumption that allows causal inference on observational data possible is that, conditional on Z, treatment assignment can be viewed as random in the sense that it is independent of potential lifetimes; i.e., $A \perp\!\!\!\perp (T^0, T^1) | Z$. This assumption, required by both the Chen and Tsiatis [10] method and the IPTW method of Wei [15], is referred to as the "strong ignorability" assumption of treatment [17], or the "no unmeasured confounders" assumption [12]. Note that this condition is also well-studied in the economics and social science literature; for example, [25, 26]. This assumption states that the dependence of treatment assignment on the potential outcomes can be completely eliminated by the observed variables Z. This assumption cannot be tested statistically and can only be justified based on knowledge on the subject matter. In medical applications, as treatment decisions made by patients or their caregivers are usually based on information available at the time of the decision-making, such as demographics, comorbidities, severity and past treatments of the subjects. If such information is also captured in the data at hand, then this assumption is plausible.

In the next two paragraphs, we describe the Chen and Tsiatis method [10] and the IPTW method of Wei [15], in order to later establish the relationship between each of these two methods and our proposed method. Much of the notation introduced here will be needed later in the development of the proposed method. The method of Chen and Tsiatis [10] removes confounding by first estimating the treatment effect conditional on covariates Z. This part is straightforward since, conditional on Z, treatment A can be viewed as randomly

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

assigned. The conditional treatment effect measures the difference in survival had subjects with covariates $Z = z$ been assigned to $A = 1$ versus $A = 0$, since $f(T|A = j, Z = z) = f(T^j|A = j, Z = z) = f(T^j|Z = z)$ for $j = 0, 1$, where f denotes "distribution of" and the two equalities are due to the SUTVA and no unmeasured confounders assumptions, respectively. Note, this result implies $P(T^j > t|Z) = P(T > t|A = j, Z)$, which relates the conditional survival functions of potential lifetimes to those of the observed lifetimes. Next, the average causal effect of A can be estimated by the average of the conditional effects across the distribution of Z. Specifically, the Chen and Tsiatis method [10] posits treatment-stratified Cox models [2] for T given (A, Z),

$$\lambda_{ij}(t) \equiv \lambda(t|A_i = j, Z_i) = \lambda_{0j}(t)\, e^{\beta_j^T Z_i}, \quad j = 0, 1, \quad (1)$$

where $\lambda(t|A, Z)$ denotes the conditional hazard function given A and Z, and $\lambda_{0j}(t)$ are unspecified treatment-specific baseline hazard functions. Inference on this model can be carried out by standard survival analysis techniques. For example, β_j can be estimated by the maximum partial likelihood estimator, $\hat{\beta}_j$ [2, 27], while the baseline cumulative hazard function $\Lambda_{0j}(t) \equiv \int_0^t \lambda_{0j}(u)\, du$ can be consistently estimated by the Breslow estimator [28], denoted by $\hat{\Lambda}_{0j}(t) \equiv \int_0^t \frac{\sum_{i=1}^n dN_{ij}(t)}{\sum_{i=1}^n Y_{ij}(t) exp\left(\hat{\beta}_j^T Z_i\right)}$. The treatment effect conditional on (which equals the difference in restricted mean lifetime) can be estimated by $\int_0^L \left\{ e^{-\hat{\Lambda}_{i1}(t)} - e^{-\hat{\Lambda}_{i0}(t)} \right\} dt$, where $e^{-\hat{\Lambda}_{ij}(t)}$ estimates the conditional survival function $P(T_i > t|A_i = j, Z_i)$, with $\hat{\Lambda}_{ij}(t) = e^{\hat{\beta}_j^T Z_i} \hat{\Lambda}_{0j}(t)$. Finally, the average causal treatment effect δ is estimated by averaging the conditional treatment effects across all Z_i for $i = 1, \ldots, n$; i.e., $\tilde{\delta} = n^{-1} \sum_{i=1}^n \int_0^L \left\{ e^{-\hat{\Lambda}_{i1}(t)} - e^{-\hat{\Lambda}_{i0}(t)} \right\} dt$. Note that $n^{-1} \sum_{i=1}^n e^{-\hat{\Lambda}_{ij}(t)}$ estimates the marginal survival function $S^j(t)$ and therefore $\tilde{\delta}$ is also integrated difference in estimated marginal survival functions.

Instead of going through the treatment effect (and treatment-specific survival functions) conditional on covariates, a different strategy is to estimate the average treatment effect directly through weighted nonparametric estimators. In particular, the IPTW method [15] removes confounding by building up the whole population that could have received the treatment, say, $A = j$, by inverse weighting the individuals in group j with the probability of being in that group conditional on covariates. For example, if a subject in treatment j has covariates $Z = z$ and the probability of receiving the treatment $P(A = j|Z = z)$, then this subject actually represents $1/P(A = j|Z = z)$ individuals in the population of interest that could have received the treatment and inverse weighting by this probability builds up the whole population. Specifically, as marginal survival function can be estimated through the Nelson-Aalen estimator of cumulative hazard function, the IPTW estimator for the marginal cumulative hazard function of T^j, $\Lambda_j(t)$, can be viewed as a weighted Nelson-Aalen estimator; i.e.,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\hat{\Lambda}_j^{IPTW}(t) = \int_0^t = \frac{\sum_{i=1}^n w_{ij}\left(\hat{\theta}\right) dN_i(t)}{\sum_{i=1}^n w_{ij}\left(\hat{\theta}\right) Y_i(t)}, \quad (2)$$

where $w_{ij}\left(\hat{\theta}\right) = I(A_i = j)/p_{ij}\left(\hat{\theta}\right)$ and $p_{ij}\left(\hat{\theta}\right)$ is an estimator of $P(A_i = j|Z_i)$. Note that if the weights, w_{ij}, are set to 1, then (2) reduces the usual Nelson-Aalen estimator of the cumulative hazard function of T. If, instead, $p_{ij}\left(\hat{\theta}\right)$ is set to 1, then (2) reduces to the Nelson-Aalen estimator conditional on each treatment. The probability $P(A_i = j|Z)$ can estimated be by fitting a logistic regression model, which assumes

$$logit\{P(A_i = 1|Z_i)\} = \theta^T X_i, \quad (3)$$

where X_i includes an intercept and (possibly transformed) elements of Z_i and logit(u) = log $\{u/(1-u)\}$.

The Chen and Tsiatis [10] method builds models for survival time and the resulting estimators for $S^j(t)$, j = 0, 1, and δ are consistent for the true average causal effect if the assumed model (1) is correct. The IPTW method builds a logistic model for treatment assignment and the resulting estimators are consistent if the assumed logistic model (3) is correct. In our proposed method, we propose to build models for both the treatment assignment and survival time. The strategy is to attempt to either model the treatment assignment correctly, allowing one to balance the distribution of covariates between treatments , or to model the survival process correctly. If at least one of the models are correct, then the ACE can be estimated consistently. Therefore, in the following, we will refer to models (1) and (3) as working models as they are not necessarily believed to be true. As the IPTW method, the proposed method estimates treatment effect through estimating the marginal cumulative hazard functions, $\Lambda_j(t)$. Specifically, the proposed estimator for $\Lambda_j(t)$ is

$$\hat{\Lambda}_j(t) = \int_0^t \frac{n^{-1}\sum_{i=1}^n e^{-\hat{\Lambda}_{ij}(u)} d\hat{\Lambda}_{ij}(u) + n^{-1}\sum_{i=1}^n \left[w_{ij}\left(\hat{\theta}\right)\left\{ e^{\hat{\Lambda}_j^C(u)} dN_{ij}(u) - e^{-\hat{\Lambda}_{ij}(u)} d\hat{\Lambda}_{ij}(u)\right\}\right]}{n^{-1}\sum_{i=1}^n e^{-\hat{\Lambda}_{ij}(u)} + n^{-1}\sum_{i=1}^n \left[w_{ij}\left(\hat{\theta}\right)\left\{ e^{\hat{\Lambda}_j^C(u)} Y_{ij}(u) - e^{-\hat{\Lambda}_{ij}(u)}\right\}\right]}, \quad (4)$$

where $\hat{\Lambda}_j^C(u)$ is the Nelson-Aalen estimator of the treatment-specific cumulative hazard function (given $A = j$) of C,

$$\hat{\Lambda}_j^C(t) = \int_0^t \frac{\sum_{i=1}^n dN_{ij}^C(u)}{\sum_{i=1}^n Y_{ij}(u)},$$

with $N_{ij}^C(t) = A_{ij} I(U_i \le t, \Delta_i = 0)$. Consequently, one can estimate $S^j(t)$ by $\hat{S}^j(t) = e^{-\hat{\Lambda}_j(t)}$, and μ_j by $\hat{\mu}_j = \int_0^L \hat{S}^j(u)\, du$. Finally, the proposed estimator for δ is given by $\hat{\delta} = \hat{\mu}_1 = \hat{\mu}_0$.

Before introducing the main theorem regarding asymptotic properties of the proposed estimators, let us first heuristically explain how to understand this method and why it is

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

expected to posses, the double- robust property. Note that Kang and [20] provide an exposition of double robust estimators in the context of continuous outcomes. As $\Lambda_j(t)$ can be expressed as $-\int_0^L S^j(u)^{-1} dS^j(u)$, our proposed method estimates $\Lambda_j(t)$ by estimating $S^j(u)$ and $-dS^j(u)$ respectively. Specifically, the denominator inside the integral in (4) estimates $S^j(u)$ and the numerator estimates $-dS^j(u)$. Considering the denominator, the first term, $n^{-1}\sum_{i=1}^{n} e^{-\hat{\Lambda}_{ij}(u)}$, is the Chen and Tsiatis estimator of $S^j(u)$ and as explained earlier it is an average of estimators of the conditional survival functions $P(T_i > u|A_i = j, Z_i)$ across $i = 1, \ldots, n$. Suppose hypothetically that every subject in the population received treatment j and, in addition, that censoring does not exist. In this setting, $I(T_i \geq u)$ would serve as the response, such that the residual for each subject at time u, in the fit of the model (1), is given by $I(T_i \geq u) - e^{-\hat{\Lambda}_{ij}(u)}$. If the model for survival is correctly specified, on average residuals are close to zero; average of residuals estimates the bias of the Chen and Tsiatis estimator. In reality, because the subjects who actually received treatment j are not representative of the whole population due to lack of randomization, the average of residuals among those who actually received $A = j$ does not directly estimate the bias. Applying the idea of the IPTW method, it is easy to see that the bias can be consistently estimated from residuals on those who are actually in group j by inverse weighting their contributions by the corresponding probability of being in group j. In addition, in reality, even for subjects in group j, $I(T_i \geq u)$, and correspondingly the residuals, are not observed for all of them due to censoring, and instead one only observes $Y_i(u)$. A solution to this is that one further weights $Y_i(u)$ by the probability of not being censored. Therefore, the bias can be estimated by the second term in the denominator, i.e., $n^{-1}\sum_{i=1}^{n} w_{ij}(\hat{\theta})\left\{e^{\hat{\Lambda}_j^C(u)} Y_i(u) - e^{-\hat{\Lambda}_{ij}(u)}\right\}$, where $e^{\hat{\Lambda}_j^C(u)}$ estimates $P(C_i > u|A_i)^{-1}$, the inverse probability of remaining uncensored as of time u $A_i = j$. To summarize, if model (1) is correct, the first term of the denominator estimates $S^j(u)$ and the second term estimates zero; if model (1) is possibly wrong but model (3) is correct, then the second term estimates the bias of the first term and again the denominator consistently estimates $S^j(u)$. Similarly one can apply the same idea to the numerator, where the first term can be viewed as an estimator for $-dS^j(u)$ and the second term is either estimator of zero or the bias of the first term. As a result, the proposed estimator (4) is expected to be consistent for the true ACE if at least one of the working models are correct.

We point out that the denominator of (4) is itself a double-robust estimator of $S^j(u)$. In our proposed method, we do not use it directly due to the following considerations. The denominator can be written equivalently as $n^{-1}\sum_{i=1}^{n}\left[w_{ij}(\hat{\theta}) e^{\Lambda_j^C(u)} Y_{ij}(u)\right] + n^{-1}\sum_{i=1}^{n}\left[\left\{1 - w_{ij}(\hat{\theta})\right\} e^{-\Lambda_{ij}(u)}\right]$, which can be viewed as an augmented IPTW estimator of $S^j(u)$ with the first term being an IPTW estimator second term as the augmentation term [29]. From this perspective, the denominator builds upon the idea that $E\{Y_i(t)/P(C_i > t)\} = P(T_i > t)$ in a one sample setting. Few practitioners would use this method in the one-sample setting, in part because the resulting survival curve is not monotone. The two most popular estimators in the one-sample setting are the Kaplan-Meier and Nelson-Aalen estimators. Our method builds on the latter (and the Kaplan-Meier estimator is asymptotically equivalent to the Nelson- Aalen

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

estimator). One can interpret our estimator as a modified Nelson-Aalen estimator, with the modifications being to incorporate adjustment covariates and double-robustness. In addition, simulation studies show that using the denominator to estimate $S^j(u)$ directly and then to estimate μ_j may lead to considerably more bias than the proposed method in finite samples.

The asymptotic properties of the proposed estimators for μj and δ are summarized by the following theorem, the proof of which is outlined in the Appendix.

Theorem 1: Under conditions (a) – (f) listed in the Appendix, as n $\to \infty$, if at least one of the working models specified in (1) and (3) is correct then $\hat{\mu}_j$ *converges in probability to* μ_j *and* $n^{\frac{1}{2}}\left(\hat{\mu}_j - \mu_j\right)$ is asymptotically normal with mean zero and variance $E\left(\phi_{ij}^2\right)$, *where* $\phi_{ij} = -\int_0^L S^j(u)\,\varphi_{ij}(u)\,du$,

$$\begin{aligned}\varphi_{ij}(t) = & B_j^T\left(t;\beta_j^*,\theta^*\right) V^{-1}(\theta^*) X_i\left\{A_{ij} - p_{ij}(\theta^*)\right\} \\ & + F_j^T\left(t;\beta_j^*\right)\Omega_j^{-1}\left(\beta_j^*\right) U_{ij}\left(\beta_j^*\right) \\ & + \int_0^t G_j\left(u,t;,\beta_j^*,\theta\right)\frac{dM_{ij}^*(u)}{r_j^{(0)}(u;\beta_j)} \\ & + \int_0^t H_j\left(u,t;\beta_j^*,\theta^*\right)\frac{dM_{ij}^C(u)}{E\left\{Y_{ij}(u)\right\}} \\ & + \int_0^t \frac{w_{ij}(\theta^*)\,e^{\Lambda_j^C(u)} dM_{ij}^\dagger(u) - \left\{w_{ij}(\theta^*) - 1\right\} e^{-\Lambda_{ij}^*(u)}\left\{d\Lambda_{ij}^*(u) - d\Lambda_j(u)\right\}}{D_j\left(u;\beta_j^*,\theta^*\right)},\end{aligned}$$

θ^* *and* β_j^* *are the asymptotic limiting values of* $\hat{\theta}$ *and* $\hat{\beta}_j$, respectively, which may or may not equal to the respective true values, depending on whether the corresponding assumed model is correct, and $dM_{ij}^\dagger(u) = dN_{ij}(u) - Y_{in}(u)\,d\Lambda_j(u)$, *with* $B_j\left(t;\beta_j^*,\theta^*\right)$, $F_j\left(t;\beta_j^*,\theta^*\right)$, $G_j\left(u,t;\beta_j^*,\theta^*\right)$, $H_j\left(u,t;\beta_j^*,\theta^*\right)$ defined in the Appendix. In addition, under the same conditions, $\hat{\delta}$ *converges in probability to* δ *and* $n^{\frac{1}{2}}\left(\hat{\delta} - \delta\right)$ is asymptotically normal with mean zero and variance $E(\varphi_{i1} - \varphi_{i0})^2$.

In the above theorem, φ_{ij} and φ_{ij} seem complicated and this is because it is stated without explicitly assuming which working model is correctly specified. If one or both of the working models are correctly specified, some of the terms in $\varphi_{ij}(t)$ and φ_{ij} are identically zero. For example, if the model (1) is correct, then $B_j\left(t;\beta_j^*,\theta^*\right)$ is equal to zero, and if the model (3) is the true model, then $F_j\left(t;\beta_j^*,\theta^*\right)$ and $G_j\left(u,t;\beta_j^*,\theta^*\right)$ are identically zero. Variance for $\hat{\mu}_j$ and $\hat{\delta}$ can be consistently estimated through $n^{-1}\sum_{i=1}^n \hat{\phi}_{ij}^2$ and $n^{-1}\sum_{i=1}^n\left(\hat{\phi}_{i1} - \hat{\phi}_{i0}\right)^2$, respectively, where $\hat{\phi}_{ij}$ is obtained by replacing limiting values in $\hat{\phi}_{ij}$ with their empirical counterparts. Although $\hat{\phi}_{ij}$ seems complicated, variance estimators

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

can actually be computed very fast. SAS code for implementing the proposed methods and the variance estimators are available at http://www-personal.umich.edu/~mzhangst/.

3. Simulation Studies

In this section, we report results from simulation studies to evaluate the finite sample properties of the proposed method. Results are based on 1000 Monte Carlo data sets with a sample size of n=600.

In our simulated data, baseline covariates, Z_1, Z_2, Z_3, are generated as normal with zero mean and unit variance, and the correlation between Z_1 and Z_3 is 0.2. Each covariate is truncated at -4 and 4 to be consistent with the regularity conditions listed in the Appendix. We then generated treatment indicator, A, according to a logistic regression model with $\text{logit}\{P(A = 1|Z)\} = -0.5Z_1 - 0.5Z_2$. Lifetime, T, was generated as exponential with rate $\exp(-2.5 - 1.5Z_1 - Z_2 - 0.7Z_3)$ for $A = 0$ and $\exp(-3 - Z_1 - 0.9Z_2 - Z_3)$ for $A = 1$ respectively. Note that covariates Z_1 and Z_2 predict both treatment assignment and survival time, therefore they serve as confounders. Finally, censoring time C was generated as exponential with rate $\exp(-4.5)$, which lead to approximately 25% censoring.

We compare the proposed method with the other methods introduced previously: the method of Chen and Tsiatis [10], where one models the relationship of survival time to covariates by treatment-specific Cox models; and the IPTW method of Wei [15] wherein one instead models the treatment assignment with covariates using a logistic regression model. Each of the three estimators are evaluated under settings where the assumed models for survival time and treatment assignment are both correct or both incorrect or only one of them is correct. Specifically, for the $T|A, Z$ model used in both the proposed and Chen and Tsiatis (2001) methods, the correct model was fitted using covariates (Z_1, Z_2, Z_3) for each treatment, while the incorrect model was fitted using (Z_1, Z_3). For the $A|Z$ model used in the proposed and the IPTW methods, the correct model was fitted using (Z_1, Z_2), whereas the incorrect model was fitted using Z_1 only.

Restricted mean lifetimes and their difference were estimated with L set to 10 and 20. Table 1 and Table 2 summarize results for estimating μ_1 and δ, respectively, and results for μ_0 are very similar and therefore are not reported. Under all scenarios in which at least one of the working models is correctly specified, the proposed estimators perform well, which is consistent with the purported double-robust property of the proposed methods. Specifically, the proposed estimators are approximately unbiased for the true parameters and the 95% coverage probabilities achieve the nominal level. In contrast, the Chen and Tsiatis and IPTW estimators perform well when the corresponding assumed model is correct; however, large biases and small coverage probabilities are observed if the assumed model is incorrect.

As mentioned before, the denominator of (4) itself can be used as an estimator of the survival function $S^j(u)$, which can be integrated to estimate μ_j as well. Although asymptotically it is also double-robust in the sense that it is consistent for the truth when at least one of the working models are correct, our simulation studies show that it has considerable larger bias than the proposed method(see Table 1).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

4. Application

We applied the proposed methods to compare survival following deceased-donor kidney transplantation among patients receiving a transplant through donation-after-cardiac-death (DCD) versus the remainder (referred to here as non-DCD). Data were provided by Eurotransplant, the Dutch Organ Transplant Registry and the Bureau of Geneaology. Specifically, dates of registration on the kidney waiting list and, where applicable, kidney transplantation, were collected by Eurotransplant. Information regarding donor and recipient characteristics, were provided by The Dutch Organ Transplant Registry, as well as date of death. Data from the Bureau of Geneaology served as the basis for verifying mortality information.

A total of n=1,139 patients were included in the analysis; 459 of which were DCD kidney transplants, and 680 non-DCD transplants. The mean age at transplant was approximately 49 years, and there were 88 observed deaths. As indicated above, the groups being compared were DCD (j=1) versus non-DCD (j=0). Adjustment covariates included age, sex, vascular disease (as a primary renal diagnosis), panel reactive antibodies, expanded criteria donor, method of first dialysis, and years on dialysis prior to transplant.

Under the proposed method (see Table 3), mean 5-year post-transplant survival time is estimated to be $\hat{\mu}_1 = 4.64$ years for the DCD group and $\hat{\mu}_0 = 4.54$ years for the non-DCD group, for difference of $\hat{\delta} = 0.10$ years (p=0.23). Therefore there appears to be no difference in 5-year restricted mean lifetime between recipients of DCD versus non-DCD kidneys. Results were similar based on the method of Chen & Tsiatis (2001) and the IPTW approach. In terms of precision, the lowest estimated standard error was from the Chen & Tsiatis (2001) method; since this method also estimated the largest difference $\hat{\delta}$, it also yielded the lowest p value, albeit still non-significant (p=0.06). However, the validity of this method requires the Cox model to be correct, unlike the proposed method, which only requires that either the Cox model or the logistic model is correct.

5. Discussion

We propose a semiparametric double-robust estimator of the mean difference in treatment-specific restricted mean survival time. The proposed method uses working models for treatment assignment and the death hazard, and is consistent if at least one of the two working models is correct. Asymptotic properties of the proposed estimator are derived and shown through simulation to be applicable to practical-sized samples.

We compare our method to two existing methods through which differences in restricted mean lifetimes can be estimated. The method of Chen and Tsiatis fits group-specific Cox models, then averages over the fitted values to obtain the average causal effect (ACE). The IPTW method used inverse probability of treatment weighting to estimate the ACE. Our proposed method can be viewed as a combination of Chen and Tsiatis and the IPTW methods. We obtain consistency if the hazard regression model (from Chen and Tsiatis method) or the group assignment model (used in IPTW method) are correctly specified.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

A potential disadvantage of our method is that censoring times are assumed to be conditionally independent of the death times, given only on the treatment indicator. The analogous assumption by Chen and Tsiatis requires conditional independence given both treatment and adjustment covariates, a much less stringent assumption. However, in many observational studies it is quite reasonable to assume that the adjustment covariates do not predict the censoring hazard. For example, in retrospective cohort studies (e.g., particularly those based on registry or other pre-collected databases), censoring may be primarily administrative; i.e., the date the database was closed, a fixed calendar date which is external to the patients, let alone their adjustment covariate pattern. In the IPTW method of Wei [15], the assumption on the censoring distribution is the same as ours, although the IPTW method does not involve inverse probability of censoring weighting. It should be noted that the IPTW method of Wei has been extended to handle dependent censoring easily, under the "nounmeasured-confounders-for-censoring" assumption, by inverse probability of censoring weighting (IPCW), wherein one models the probability of censoring conditional on baseline and/or time-dependent covariates and modifies the weight function $w_{ij}\left(\hat{\theta}\right)$ by further weighting it by the inverse probability of remaining uncensored; see Schaubel and Wei[31]. However, it is not straightforward to extend the proposed method to accommodate dependent censoring since, in addition to modifying the weight, one needs to modify the second term in both the numerator and the denominator of (4) as well. Therefore, we do not consider this more general assumption on censoring in this article.

The application of the proposed methods implies that 5-year restricted mean post-transplant survival time is no different for patients receiving a kidney transplant through donation after cardiac death (DCD). The importance of this finding is tied to the potential to increase the deceased-donor kidney pool by increasing DCD transplantation. A natural question is whether a difference in restricted mean lifetime would be observed in the presence of a similar study with longer post-transplant follow-up. In fact, in a large percentage of practical settings, it would be preferable to use mean survival time (i.e., without inference being restricted to the $(0, L]$ time interval). The development of robust methods for estimating and contrasting mean lifetime would be valuable.

In this article, we considered estimation of the average causal treatment effect for the entire population of interest. A different but also relevant quantity is the so-called average treatment effect for the treated (ATT)[30], which targets the question of whether or not the treatment actually worked among treated subjects. In terms of restricted mean lifetimes, the ATT is $E\{\min(T^1, L) - \min(T^0, L)|A = 1\}$ and estimating ATT requires a weaker ignorability assumption than the strong ignorability assumption assumed before; i.e., $T^0 \perp\!\!\!\perp A|Z$. We expect that the ideas underlying our proposed method could be extended to estimate the ATT, which would be interesting for future research.

Acknowledgments

This work was supported in part by National Institutes of Health grant 5R01-DK070869. The authors thank Maastricht University Medical Center, for access to the kidney transplant data. They also thank three anonymous Referees whose comments strengthened the manuscript considerably.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Appendix

1. Conditions and Preparation Results

We introduce the following notation that will be used in the proof of Theorem 1:

$$R_j^{(d)}(t;\beta) = n^{-1}\sum_{i=1}^{n} Y_{ij}(t)\, Z_i^{\otimes d} exp\left\{\beta^T Z_i\right\}, \quad r_j^{(d)}(t;\beta) = E\left\{R_j^{(d)}(t;\beta)\right\},$$
$$\bar{Z}_j(t;\beta) = \frac{R_j^{(1)}(t;\beta)}{R_j^{(0)}(t;\beta)}, \quad \bar{z}_j(t;\beta) = \frac{r_j^{(1)}(t;\beta)}{r_j^{(0)}(t;\beta)},$$
$$d\Lambda_{0j}^*(t) = \frac{E\{dN_{ij}(t)\}}{r_j^{(0)}(t;\beta^*)}, \quad d\Lambda_{ij}^*(t) = exp\left(\beta_j^{*T} Z_i\right) d\Lambda_{0j}^*(t)$$
$$dM_{ij}^*(t) = dN_{ij}(t) - Y_{ij}(t)\, d\Lambda_{ij}^*(t)$$
$$\Omega_j(\beta) = \int_0^\tau \left\{\frac{r_j^{(2)}(t;\beta)}{r_j^{(0)}(;\beta)} - \bar{z}_j(t;\beta)^{\otimes 2}\right\} E\{Y_{ij}(t)\,\lambda_{ij}(t)\}\, dt,$$

and

$$V(\theta) = E\left[\frac{exp\left(\theta^T X\right) X^{\otimes 2}}{\{1+exp(\theta^T X)\}^2}\right],$$

for d = 0, 1, 2, where for a column vector a, $a^{\otimes 2} = aa^T$, $a^{\otimes 1} = a$, and $a^{\otimes 0} = 1$.

We assume the following regularity conditions for $i = 1, \ldots, n$, and $j = 0, 1$:

- **a.** $P(U_i \geq \tau) > 0$.
- **b.** Z_i is bounded almost surely.
- **c.** $\Lambda_{0j}(\tau) < \infty$.
- **d.** β_j^* is the unique solution to

  $$\int_0^\tau E\{Y_{ij}(t)\, Z_i \lambda_{ij}(t)\}\, dt - \int_0^\tau \bar{z}_j(t;\beta)\, E\{Y_{ij}(t)\,\lambda_{ij}(t)\}\, dt = 0,$$

  and $\Omega_j(\beta^*)$ is positive definite.
- **e.** θ^* is the unique maximizer to $E\{A\theta^T X - \log(1 + e^{\theta^T X})\}$ and $V(\theta^*)$ is positive definite.
- **f.** $P(A_i = j|Z_i)$ is bounded away from 0.

With regard to model (1), it was shown that (Lin and Wei, 1989), under the assumed regularity conditions, $\hat{\beta}_j \xrightarrow{p} \beta_j^*$, with $\beta_j^* = \beta_j$ if model (1) is correct, and $n^{\frac{1}{2}}\left(\hat{\beta}_j - \beta_j^*\right)$ is asymptotically normal with

$$n^{\frac{1}{2}}\left(\hat{\beta}_j - \beta_j^*\right) = \Omega_j^{-1}\left(\beta_j^*\right) n^{-\frac{1}{2}} \sum_{i=1}^{n} U_{ij}\left(\beta_j^*\right) + o_p(1),$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

where $U_{ij}\left(\beta_j^*\right)=\int_0^\tau\left\{Z_i-\bar{z}_j\left(t;\beta_j^*\right)\right\}dM_{ij}^*(t)$. It is then straightforward to show that

$$n^{\frac{1}{2}}\left\{\hat{\Lambda}_{ij}(t)-\Lambda_{ij}^*(t)\right\}=K_{ij}^T\left(t;\beta_j^*\right)\Omega_j^{-1}\left(\beta_j^*\right)n^{-\frac{1}{2}}\sum_{i=1}^n U_{ij}\left(\beta_j^*\right)+e^{\beta_j^{*T}Z_i}n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t\frac{dM_{ij}^*(u)}{r_j^{(0)}\left(u;\beta_j^*\right)}+o_p(1),$$

where $K_{ij}\left(t;\beta_j^*\right)=\int_0^t\left\{Z_i-\bar{z}_j\left(u;\beta_j^*\right)\right\}d\Lambda_{ij}^*(u)$.

As for model (3), under the assumed regularity conditions, $\hat{\theta}\xrightarrow{p}\theta^*$ with $\theta^*=\theta$ if model (3) is true, and

$$n^{\frac{1}{2}}\left(\hat{\theta}-\theta^*\right)=V^{-1}(\theta^*)\,n^{-\frac{1}{2}}\sum_{i=1}^n X_i\left\{A_i-expit\left(\theta^{*T}X_i\right)\right\}+o_p(1)$$

for $j = 0, 1$ (Zeng and Chen,2009).

Finally, regarding censoring, it is standard result (e.g., Fleming and Harrington, 1991) that

$$n^{\frac{1}{2}}\left\{\hat{\Lambda}_j^C(t)-\Lambda_j^C(t)\right\}=n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t\frac{dM_{ij}^C(u)}{E\{Y_{ij}(u)\}}+o_p(1),$$

where $dM_{ij}^C(u)=dN_{ij}^C(u)-Y_{ij}(u)\,d\Lambda_j^C(u)$

2. Consistency

If model (3) is correct, then $\hat{\theta}\xrightarrow{p}\theta$ and $p_{ij}\left(\hat{\theta}\right)\xrightarrow{p}p_{ij}(\theta)=P(A_{ij}=1|Z_i)$. Considering the denominator of (4), it can be rewritten as $n^{-1}\sum_{i=1}^n\left[w_{ij}\left(\hat{\theta}\right)e^{\hat{\Lambda}_j^C(u)}Y_{ij}(u)-\left\{w_{ij}\left(\hat{\theta}\right)-1\right\}e^{-\hat{\Lambda}_{ij}(u)}\right]$ and it converges in probability to

$$\begin{aligned}&E\left\{\frac{A_{ij}}{p_{ij}(\theta)}e^{\Lambda_j^C(u)}Y_{ij}(u)\right\}-E\left[\left\{\frac{A_{ij}}{p_{ij}(\theta)}-1\right\}e^{-\Lambda_{ij}^*(u)}\right]\\ =\;&E\left[E\left\{\frac{A_{ij}}{p_{ij}(\theta)}e^{\Lambda_j^C(u)}Y_{ij}(u)\,|Z_i\right\}\right]-E\left\{E\left[\left\{\frac{A_{ij}}{p_{ij}(\theta)}-1\right\}e^{-\Lambda_{ij}^*(u)}|Z_i\right]\right\}\\ =\;&E\left\{P(T_i>u|A_{ij}=1,Z_i)\right\}\equiv S^j(t).\end{aligned}$$

Similarly, one can obtain that the numerator converges in probability to $-dS^j(u)$ uniformly in $u\in[0,\ \tau]$. Combining results, we obtain that $\hat{\Lambda}_j(t)\xrightarrow{p}\Lambda_j(t)$ uniformly in $t\in[0,\ \tau]$. Therefore, by the continuous mapping theorem, $\hat{S}^j(t)\xrightarrow{p}S^j(t)$ uniformly in $t\in[0,\ \tau]$ and in addition, $\hat{\mu}_j$ and $\hat{\delta}$ are consistent for μ_j and δ respectively.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

If model (1) is correct, then $\hat{\beta}_j \xrightarrow{p} \beta_j^*$ and $\hat{\Lambda}_{ij}(t) \xrightarrow{p} \Lambda_{ij}(t)$ uniformly in $t \in [0, \tau]$. The denominator of (4) converges in probability to

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n} e^{-\hat{\Lambda}_{ij}(t)} + n^{-1}\sum_{i=1}^{n} w_{ij}\left(\hat{\theta}\right)\left\{e^{\hat{\Lambda}_j^C(u)}Y_{ij}(u) - e^{-\hat{\Lambda}_{ij}(u)}\right\} \\ \xrightarrow{p} \; & S^j(u) + E\left[w_{ij}(\theta^*)\left\{e^{\Lambda_j^C(u)}Y_{ij}(u) - e^{-\Lambda_{ij}(u)}\right\}\right] \\ = \; & S^j(u) + E\left\{E\left[w_{ij}(\theta^*)\left\{e^{\Lambda_j^C(u)}Y_{ij}(u) - e^{-\Lambda_{ij}(u)}\right\}|Z_i\right]\right\} \\ = \; & S^j(u) + E\left\{\frac{p_{ij}(\theta)}{p_{ij}(\theta^*)} \times 0\right\} = S^j(u). \end{aligned}$$

Similarly, the numerator converges in probability to $-dS^j(u)$ uniformly in $u \in [0, \tau]$. Therefore, the proposed estimators for $\Lambda_{ij}(t)$, μ_j and δ are consistent for the true values when model (1) is correct.

Therefore, the proposed estimators are consistent for the true values when at least one of the working models is correct.

3. Asymptotic Normality

In the proofs of asymptotic normality, we do not specify explicitly which working model is correct and we denote that $\hat{\beta}_j \xrightarrow{p} \beta_j^*$, $\hat{\Lambda}_{ij}(t) \xrightarrow{p} \Lambda_{ij}^*(t)$, and $\hat{\theta} \xrightarrow{p} \theta^*$. Let us first consider $n^{\frac{1}{2}}\left\{\hat{\Lambda}_j(t) - \Lambda_j(t)\right\}$ which, as we will show, can be approximated by a scaled summation of independent and identically distributed variates. We make the following decomposition:

$$n^{\frac{1}{2}}\left\{\hat{\Lambda}_j(t) - \Lambda_j(t)\right\} = n^{\frac{1}{2}}\left\{\hat{\Lambda}_j\left(t;\hat{\theta},\hat{\Lambda}_{ij},\hat{\Lambda}_j^C\right) - \hat{\Lambda}_j\left(t;\theta^*,\hat{\Lambda}_{ij},\hat{\Lambda}_j^C\right)\right\} \quad (5)$$

$$+ n^{\frac{1}{2}}\left\{\hat{\Lambda}_j\left(t;\theta^*,\hat{\Lambda}_{ij},\hat{\Lambda}_j^C\right) - \hat{\Lambda}_j\left(t;\theta^*,\hat{\Lambda}_{ij}^*,\hat{\Lambda}_j^C\right)\right\} \quad (6)$$

$$+ n^{\frac{1}{2}}\left\{\hat{\Lambda}_j\left(t;\theta^*,\hat{\Lambda}_{ij}^*,\hat{\Lambda}_j^C\right) - \hat{\Lambda}_j\left(t;\theta^*,\Lambda_{ij}^*,\Lambda_j^C\right)\right\} \quad (7)$$

$$+ n^{\frac{1}{2}}\left\{\hat{\Lambda}_j\left(t;\theta^*,\Lambda_{ij}^*,\Lambda_j^C\right) - \Lambda_j(t)\right\}. \quad (8)$$

By Taylor series expansion and substituting preparation results presented previously, after a lot of algebra, we obtain that

$$(5) = B_j^T\left(t;\beta_j^*,\theta^*\right) V^{-1}(\theta^*)\, n^{-\frac{1}{2}}\sum_{i=1}^{n} X_i\left\{A_{ij} - p_{ij}(\theta^*)\right\} + o_p(1),$$

where

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$
\begin{aligned}
&B_j\left(t;\beta_j^*,\theta^*\right)\\
&=\int_0^t \frac{E\left[\left\{e^{\Lambda_j^C(u)}dN_{ij}(u)-e^{-\Lambda_{ij}^*(u)}d\Lambda_{ij}^*(u)\right\}w_{ij}(\theta^*)(-1)\{1-p_{ij}(\theta^*)\}X_i\right]}{D_j\left(u;\beta_j^*,\theta^*\right)}\\
&+\int_0^t\left\{E\left[\left\{e^{\Lambda_j^C(u)}Y_{ij}(u)-e^{-\Lambda_{ij}^*(u)}\right\}w_{ij}(\theta^*)\{1-p_{ij}(\theta^*)\}X_i\right]\right\}\frac{dQ_j\left(u;\beta_j^*,\theta^*\right)}{D_j^2\left(u;\beta_j^*,\theta^*\right)},D_j\left(u;\beta_j^*,\theta^*\right)\\
&=E\left[w_{ij}(\theta^*)e^{\Lambda_j^C(u)}Y_{ij}(u)-\{w_{ij}(\theta^*)-1\}e^{-\Lambda_{ij}^*(u)}\right]dQ_j\left(u;\beta_j^*,\theta^*\right)\\
&=E\left[w_{ij}(\theta^*)e^{\Lambda_j^C(u)}dN_{ij}(u)-\{w_{ij}(\theta^*)-1\}e^{-\Lambda_{ij}^*(u)}d\Lambda_{ij}^*(u)\right],
\end{aligned}
$$

$$
(6)=F_j^T\left(t;\beta_j^*,\theta^*\right)\Omega_j^{-1}\left(\beta_j^*\right)n^{-\frac{1}{2}}\sum_{i=1}^n U_{ij}\left(\beta_j^*\right)+n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t G_j\left(u,t;\beta_j^*,\theta^*\right)\frac{dM_{ij}^*(u)}{r_j^{(0)}\left(u;\beta_j^*\right)}+o_p(1),
$$

where

$$
\begin{aligned}
F_j\left(t;\beta_j^*,\theta^*\right)=&\int_0^t\frac{E\left[\{w_{ij}(\theta^*)-1\}e^{-\Lambda_{ij}^*(u)}d\Lambda_{ij}^*(u)\left\{K_{ij}\left(u;\beta_j^*\right)-Z_i+\bar{z}_j\left(u;\beta_j^*\right)\right\}\right]}{D_j\left(u;\beta_j^*,\theta^*\right)}-\int_0^t E\left[\{w_{ij}(\theta^*)-1\}e^{-\Lambda_{ij}^*(u)}K_{ij}\left(u;\beta_j^*\right)\right]\frac{dQ_j\left(u;\beta_j^*,\theta^*\right)}{D_j^2\left(u;\beta_j^*,\theta^*\right)},\\
G_j\left(u;\beta_j^*,\theta^*\right)=&\int_u^t\frac{E\left[\{w_{ij}(\theta^*)-1\}e^{-\Lambda_{ij}^*(s)}e^{\beta_j^{*T}Z_i}d\Lambda_{ij}^*(s)\right]}{D_j\left(s;\beta_j^*,\theta^*\right)}-\frac{E\left[\{w_{ij}(\theta^*)-1\}e^{-\Lambda_{ij}^*(u)}e^{\beta_j^{*T}Z_i}\right]}{D_j\left(u;\beta_j^*,\theta^*\right)}-\int_u^t E\left[\{w_{ij}(\theta^*)-1\}e^{-\Lambda_{ij}^*(s)}e^{\beta_j^{*T}Z_i}\right]\Big\}\frac{dQ}{D}
\end{aligned}
$$

and

$$
(7)=n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t H_j\left(u,t;\beta_j^*,\theta^*\right)\frac{dM_{ij}^C(u)}{E\{Y_{ij}(u)\}}+o_p(1),
$$

where

$$
H_j\left(u,t;\beta_j^*,\theta^*\right)=\int_u^t\frac{E\left\{w_{ij}(\theta^*)e^{\Lambda_j^C(s)}dN_{ij}(s)\right\}}{D_j\left(s;\beta_j^*,\theta^*\right)}-\int_u^t\left\{\frac{Q_j\left(s;\beta_j^*,\theta^*\right)}{D_j^2\left(s;\beta_j^*,\theta^*\right)}\times E\left[w_{ij}(\theta^*)Y_{ij}(s)e^{\Lambda_j^C(s)}\right]\right\}.
$$

Finally, as for the last term, it is straightforward to show that

$$
(8)=n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t\frac{w_{ij}(\theta^*)e^{\Lambda_j^C(u)}dM_{ij}^\dagger(u)-\{w_{ij}(\theta^*)-1\}e^{-\Lambda_{ij}^*(u)}\left\{d\Lambda_{ij}^*(u)-d\Lambda_j(u)\right\}}{D_j\left(u;\beta_j^*,\theta^*\right)}+o_p(1),
$$

where $dM_{ij}^\dagger(u)=dN_{ij}(u)-Y_{ij}(u)\,d\Lambda_j(u)$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Combining the above results, we have shown that we can represent $n^{\frac{1}{2}}\left\{\hat{\Lambda}_j(t)-\Lambda_j(t)\right\}$ as $n^{-\frac{1}{2}}\sum_{i=1}^{n}\varphi_{ij}(t)$ plus a term that converges in probability to zero, where

$$\begin{aligned}\varphi_{ij}(t)=&B_j^T\left(t;\beta_j^*,\theta^*\right)V^{-1}(\theta^*)X_i\left\{A_{ij}-p_{ij}(\theta^*)\right\}\\&+F_j^T\left(t;\beta_j^*\right)\Omega_j^{-1}\left(\beta_j^*\right)U_{ij}\left(\beta_j^*\right)\\&+\int_0^t G_j\left(u,t;\beta_j^*,\theta^*{}_j\right)\frac{dM_{ij}^*(u)}{r_j^{(0)}(u;\beta_j)}\\&+\int_0^t H_j\left(u,t;\beta_j^*,\theta^*\right)\frac{dM_{ij}^C(u)}{E\left\{Y_{ij}(u)\right\}}\\&+\int_0^t\frac{w_{ij}(\theta^*)e^{\Lambda_j^C(u)}dM_{ij}^{\dagger}(u)-\left\{w_{ij}(\theta^*)-1\right\}e^{-\Lambda_{ij}^*(u)}\left\{d\Lambda_{ij}^*(u)-d\Lambda_j(u)\right\}}{D_j\left(u;\beta_j^*,\theta^*\right)}.\end{aligned}$$

When one of the working models is correct, using similar techniques used in proving consistency of $\hat{\Lambda}_{ij}(t)$ for $\Lambda_{ij}(t)$, it can be shown that $\varphi_{ij}(t)$ has mean zero and are identically and independently distributed across $i = 1, \ldots, n$.

Considering the estimation of μ_j, $n^{\frac{1}{2}}\left(\hat{\mu}_j-\mu_j\right)$ can be written as

$$\begin{aligned}n^{\frac{1}{2}}\left(\hat{\mu}_j-\mu_j\right)=&\ n^{\frac{1}{2}}\int_0^L\hat{S}^j(u)-S^j(u)\,du\\=&\ n^{\frac{1}{2}}\int_0^L e^{-\hat{\Lambda}_j(u)}-e^{-\Lambda_j(u)}du\\=&\ n^{-\frac{1}{2}}\sum_{i=1}^{n}\phi_{ij}+o_p(1),\end{aligned}$$

where $\phi_{ij}=-\int_0^L S^j(u)\varphi_{ij}(u)\,du$. When at least one of the two working models is correct, the φ_{ij} variates are independent and identically distributed with mean 0. Therefore, $n^{\frac{1}{2}}\left(\hat{\mu}_j-\mu_j\right)$ converges to a normal distribution with mean 0 and variance $E\left(\phi_{ij}^2\right)$. It then follows that $n^{\frac{1}{2}}\left(\hat{\delta}-\delta\right)$ is also asymptotically normal with mean 0 and variance $E(\varphi_{i1}-\varphi_{i0})^2$ and $n^{\frac{1}{2}}\left(\hat{\delta}-\delta\right)=n^{-\frac{1}{2}}\sum_{i=1}^{n}(\phi_{i1}-\phi_{i0})+o_p(1)$.

References

1. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of American Statistical Association. 1958; 53:457–481.
2. Cox DR. Regression models and life tables (with Discussion). Journal of the Royal Statistical Society, Series B. 1972; 34:187–200.
3. Karrison T. Restricted mean life with adjustment for covariates. Journal of the American Statistical Association. 1987; 82:1169–1176.
4. Karrison TG. Use of Irwin's restricted mean as an index for comparing survival in different treatment groupsInterpretation and power considerations. Controlled Clinical Trials. 1997; 18:151–167. [PubMed: 9129859]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

5. Zucker DM. Restricted mean life with covariates: modification and extension of a useful survival analysis method. Journal of the American Statistical Association. 1998; 93:702–709.
6. Andersen PK, Hansen MG, Klein JP. Regression Analysis of Restricted Mean Survival Time Based on Pseudo-Observations. Lifetime Data Analysis. 2004; 10:335–350. [PubMed: 15690989]
7. Schaubel DE, Guidinger MK, Biggins SW, Kalbfleisch JD, Pomfret EA, Sharma P, Merion RM. Survival benefit-based deceased-donor liver allocation. American Journal of Transplantation. 2009; 9(4 Pt 2):970–981. [PubMed: 19341419]
8. Meier P, Karrison T, Chappell R, Xie H. The Price of KaplanMeier. Journal of the American Statistical Association. 2004; 99:890–896.
9. Andersen PK, Perme MK. Pseudo-observations in survival analysis. Statistical Methods in Medical Research. 2010; 19:71–99. [PubMed: 19654170]
10. Chen P, Tsiatis AA. Causal inference on the difference of the restricted mean life between two groups. Biometrics. 2001; 57:1030–1038. [PubMed: 11764241]
11. Rotnitzky A, Robins JM. Inverse Probability Weighting in Survival Analysis. The Encyclopedia of Biostatistics. 2005; 4:2619–2625.
12. Robins JM, Hernán M, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000; 11:550–560. [PubMed: 10955408]
13. Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. Statistics in Medicine. 2005; 24:3089–3110. [PubMed: 16189810]
14. Cole SR, Hernàn MA. Adjusted survival curves with inverse probability weights. Computer Methods and Programs in Biomedicine. 2004; 75:45–49. [PubMed: 15158046]
15. Wei, G. Doctoral Dissertation. Department of Biostatistics, university of Michigan; 2008. Semiparametric methods for estimating cumulative treatment effects in the presence of non-proportional hazards and dependent censoring..
16. Hubbard, A.; van der Laan, MJ.; Robins, JM. Statistical Models in Epidemiology, the Environment and Clinical trials. Vol. 116. Springer Verlag; 1999. Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies.; p. 135-178.
17. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.
18. Ho DE, Imai K, Gary King, Stuart EA. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political science. 2007; 15:199–236.
19. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics. 2005; 61:962–972. [PubMed: 16401269]
20. Kang JDY, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating population means from incomplete data. Statistical Science. 2007; 26:523–539.
21. Robins JM, Rotnitzky A, Zhao LP. Estiamtion of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association. 1994; 89:846–866.
22. Snoeijs MG, Schaubal DE, Hene R, Hoitsma AJ, Idu MM, IJzermans JN, Ploeg RJ, Ringers J, Christiaans MH, Buurman WA, van Heurn LWE. Kidneys from donors after cardiac death provide survival benefit. Journal of the American Society of Nephrology. 2010; 21:10151021.
23. Rubin DB. Basu D. Discussion of ”Randomization Analysis of Experimental Data in the Fisher Randomization Test. Journal of the American Statistical Association. 1980; 75:591–593.
24. Rubin DB. Which ifs have causal answers? Discussion of Holland's “Statistics and causal inference.”. Journal of the American Statistical Association. 1986; 81:961–962.
25. Heckman JJ, Ichimura H, Todd PE. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. The Review of Economic Studies. 1997; 64:605–654.
26. Heckman JJ. The Scientific Model of Causality. Sociological Methodology. 2005; 35:1–97.
27. Cox DR. Partial likelihood. Biometrika. 1975; 62:269–275.
28. Breslow NE. Contribution to the discussion on the paper by D. R. Cox, regression and life tables. Journal of the Royal Statistical Society, Series B. 1972; 34:216–217.
29. Tsiatis, AA. Semiparametric Theory and Missing Data. Springer; New York: 2006.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

30. Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: an application to data on right hear catherization. Health Services & Outcomes Research Methodology. 2001; 2:259–278.

31. Schaubel DE, Wei G. Double inverse weighted estimation of cumulative treatment effects under non-proportional hazards and dependent censoring. Biometrics. 2011; 67:29–38. [PubMed: 20560935]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Estimation of restricted mean lifetime for $A = 1$.

Method	*T* model	*A* model	True	BIAS	ESD	ASE	CP
		L=10					
Proposed	T	T	6.849	0.002	0.201	0.193	0.942
	T	F		−0.002	0.201	0.191	0.936
	F	T		0.002	0.205	0.196	0.937
	F	F		−0.302	0.218	0.207	0.696
Denominator	T	T		0.028	0.205		
	T	F		0.028	0.203		
	F	T		0.029	0.209		
	F	F		−0.275	0.220		
IPTW		T		0.001	0.212	0.204	0.938
		F		−0.304	0.227	0.217	0.715
Chen & Tsiatis	T			0.006	0.194	0.182	0.929
	F			−0.378	0.212	0.202	0.535
		L=20					
Proposed	T	T	11.488	0.017	0.426	0.418	0.941
	T	F		0.008	0.422	0.410	0.938
	F	T		0.019	0.436	0.424	0.938
	F	F		−0.686	0.453	0.438	0.652
Denominator	T	T		0.071	0.438		
	T	F		0.071	0.427		
	F	T		0.073	0.447		
	F	F		−0.630	0.457		
IPTW		T		0.016	0.454	0.449	0.946
		F		−0.690	0.474	0.463	0.684
Chen & Tsiatis	T			0.021	0.410	0.395	0.938
	F			−0.810	0.438	0.426	0.533

T model: indicates whether the model for *T* is true or false; *A* model: indicates whether the model for *A* is true or false. Bias is the Monte Carlo Bias; ESD is the Monte Carlo standard deviation of estimates; ASE is the Monte Carlo average of estimated standard errors; CP is the coverage probability of nominal 95% Wald confidence intervals.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Estimation of difference in restricted mean lifetimes. Entries as in Table 1.

Method	*T* model	*A* model	True	BIAS	ESD	ASE	CP
		L=10					
Proposed	T	T	0.871	−0.010	0.220	0.222	0.945
	T	F		−0.022	0.217	0.228	0.953
	F	T		−0.009	0.229	0.230	0.953
	F	F		−0.686	0.260	0.259	0.245
IPTW		T		−0.013	0.248	0.256	0.960
		F		−0.690	0.279	0.285	0.329
Chen & Tsiatis	T			−0.004	0.208	0.204	0.943
	F			−0.726	0.256	0.256	0.188
		L=20					
Proposed	T	T	1.682	−0.009	0.454	0.453	0.946
	T	F		−0.031	0.449	0.465	0.956
	F	T		−0.005	0.470	0.467	0.953
	F	F		−1.45	0.537	0.533	0.208
IPTW		T		−0.013	0.510	0.518	0.956
		F		−1.460	0.580	0.586	0.295
Chen & Tsiatis	T			0.002	0.420	0.417	0.950
	F			−1.500	0.518	0.520	0.167

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Estimation of five year restricted mean lifetimes for DCD and non-DCD kidney recipients and their difference. Standard error for each estimator is reported in parenthesis and P value is for comparison of the mean restricted lifetimes between the two groups.

Method	$\hat{\mu}_0$	$\hat{\mu}_1$	δ	P value
IPTW of Wei	4.53 (0.057)	4.66 (0.068)	0.14 (0.088)	0.11
Chen & Tsiatis	4.50 (0.058)	4.66 (0.063)	0.16 (0.084)	0.06
Proposed	4.54 (0.057)	4.64 (0.068)	0.105 (0.088)	0.23