# Estimating Differences in Restricted Mean Lifetime Using Observational Data Subject to Dependent Censoring

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

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.
[*]*email:* mzhangst@umich.edu

Summary. In epidemiologic studies of time to an event, mean lifetime is often of direct interest. We propose methods to estimate group- (e.g., treatment-) specific differences in restricted mean lifetime for studies where treatment is not randomized and lifetimes are subject to both dependent and independent censoring. The proposed methods may be viewed as a hybrid of two general approaches to accounting for confounders. Specifically, treatment-specific proportional hazards models are employed to account for baseline covariates, while inverse probability of censoring weighting is used to accommodate time-dependent predictors of censoring. The average causal effect is then obtained by averaging over differences in fitted values based on the proportional hazards models. Large-sample properties of the proposed estimators are derived and simulation studies are conducted to assess their finite-sample applicability. We apply the proposed methods to liver wait list mortality data from the Scientific Registry of Transplant Recipients.

Key words: Counterfactual; Cumulative treatment effect; Inverse weighting; Proportional hazards model.

## 1. Introduction

Often in clinical and epidemiologic studies, groups of subjects are compared with respect to their survival times. Since any study is of finite duration, the time until the event of interest may be censored. Typically in observational studies, the factor of interest is not randomized (e.g., method of treatment) and may not even be assigned (e.g., race, gender, diagnosis), necessitating some form of covariate adjustment, such as that obtained through regression modeling. Since its development, the proportional hazards model (Cox, 1972) has dominated the biomedical literature as the method of choice for the regression modeling of censored data.

The popularity of the Cox model among practitioners and, by now, clinical investigators makes it an attractive means of comparing groups in observational studies. In Cox regression, the impact of each covariate is usually summarized by its effect on the hazard function. However, when comparing groups of subjects, investigators are often more interested in differences in mean lifetime than ratios of hazard functions. The survival time distribution may be heavily right skewed. Moreover, for semi- or nonparametric modeling, the mean is not well estimated; for example, the estimated survival function need not drop to 0. Therefore, a frequently employed alternative is the restricted mean lifetime; that is, for fixed $L > 0$, if $T$ denotes survival time, then the restricted mean lifetime is defined as $E\{\min(T, L)\}$. Restricted mean lifetime can also be expressed as $\int_0^L P(T > t)dt$, the area under the survival curve over $(0, L]$, a quantity easily understood by clinical investigators. For example, if $L = 5$ years, one could interpret $E\{\min(T, L)\}$ as the average number of years lived out of the next 5. Restricted mean lifetime is typically of greater interest to clinicians than the usual Cox metric, the hazard ratio. In fact, in certain settings $E\{\min(T, L)\}$ may be of more interest

than $E(T)$ itself. For example, in the context of pediatric liver transplantation, it is almost always that case that a child receiving a liver transplant will need a second liver transplant in the next 10 years. Hence, using $T$ to represent posttransplant survival time, survival after the first 10 years of posttransplant follow-up could not be realistically assumed to be due to the initial liver transplant; making $E\{\min(T, 10)\}$ of greater relevance than $E(T)$.

This article is motivated by the desire to compare wait list survival among end-stage liver disease (ESLD) patients listed for liver transplantation. A frequent cause of chronic liver disease is Hepatitis C virus (HCV), the primary diagnosis for approximately 40% of ESLD cases. Liver transplantation is the preferred treatment for ESLD, but there are far more patients awaiting transplantation than there are available donor organs. The principle underlying the current system for allocating deceased-donor livers in the United States is that priority for transplantation should be based on a patients' death rate in the absence of liver transplantation. Specifically, the patients most likely to die on the wait list should get top priority for transplantation. Currently, patients on the liver transplant wait list are sequenced in a decreasing order of Model of End Stage Liver Disease (MELD) score (Weisner et al., 2001). The MELD score is a function of three laboratory measures indicative of liver function, but does not consider underlying liver disease. It is suspected that HCV+ patients have lower wait list survival than HCV− patients. However, few studies have directly compared mean wait list survival time by diagnosis group. To our knowledge, no published analysis has compared mean wait list survival times (i.e., survival, in the absence of liver transplantation) between HCV+ and HCV− patients. Therefore, our objective is to estimate the difference between wait list lifetime between HCV+ and

HCV− patients, adjusting for baseline (i.e., time 0) characteristics (e.g., age, gender, race, MELD score).

Comparison of liver wait list survival times is complicated by the potential for dependent censoring. Specifically, death on the liver wait list is censored by the receipt of a liver transplant, and such censoring is not independent of the survival time that would have been observed on the wait list, even conditional on the baseline adjustment covariates. A given patient's MELD score typically changes over time. The updating of MELD scores is mandatory, meaning that a longitudinal sequence of MELD scores is observed for each patient. As discussed by several previous authors in the context of causal inference (e.g., Robins, 2000; Hernan, Brumback, and Robins, 2000, 2001), a comparison of survival time by HCV status should not adjust for internal time-dependent covariates, as defined by Kalbfleisch and Prentice (2002). The fact that time-dependent MELD strongly affects both wait list mortality and censoring (liver transplantation) means that lack of its adjustment (i.e., by adjusting for baseline values only) will result in the dependent censoring of wait list death time via liver transplantation.

Various authors have proposed methods for comparing restricted mean survival time in the context of Cox regression (e.g., Karrison, 1987; Zucker, 1998; Chen and Tsiatis, 2001). For example, the method of Chen and Tsiatis (2001) proposes fitting separate group-specific Cox models, then averaging over the fitted restricted mean lifetimes, with the averaging being with respect to the covariate distribution of the entire study sample. These approaches have several nice properties. First, it is not required that treatment-specific hazards be proportional. Second, an ''overall'' treatment effect estimator is obtained, without assuming that treatment-specific adjustment covariate effects are equal. Third, the target treatment effect is interpretable as an average over a well-defined covariate distribution. However, each of the afore-listed methods assumes that censoring is independent of the survival time, conditional on baseline adjustment covariates.

We propose methods for estimating group-specific differences in restricted mean lifetime for the setting in which survival time is dependently censored. The structure of the hazard model we assume is very flexible, allowing for group-specific baseline hazards and regression coefficients. In its most general form, this amounts to fitting separate models for each treatment group. The dependent censoring is overcome through the well-established inverse probability of censoring weighting (IPCW); see Robins and Rotnitzky (1992); Robins (1993); Robins and Finkelstein (2000).

The remainder of this article is organized as follows. In Section 2, we set up the notation and formalize the problem of interest. The proposed methods are described in Section 3. Asymptotic properties are listed in Section 4, with corresponding proofs given in the Web Appendix. In Section 5, a simulation study is presented. The proposed methods are applied to national liver wait list data in Section 6. The article concludes with a discussion in Section 7.

## 2. Notation and Data Structure

Suppose we are interested in comparing two groups ($A = 0, 1$), which are not randomized, in terms of restricted mean lifetime up to time $L$. We denote the survival time by $T$. As in almost all studies involving time to an event, the event time $T$ may be censored due to various reasons. Two types of censoring are considered in the following development. We let $C_1$ denote censoring which is independent conditional on baseline covariates, $Z$, and group indicator, $A$; for example, censoring due to the end of study. We let $C_2$ denote dependent censoring; that is, censoring that is not independent of $T$ given $(Z, A)$. For example, in the context of the data that motivated our work, a patient's waitlist mortality is censored if and when the patient receives a liver transplant and the transplant hazard and wait list mortality hazards may be correlated, even conditional on $(Z, A)$, through mutual dependence on time-dependent covariates (e.g., MELD score). In notation, we assume that $C_1 \perp\!\!\!\perp T | Z, A$, where $\perp\!\!\!\perp$ denotes "independent of"; while $C_2$ is not assumed to satisfy this condition. In practice, one observes the minimum of the survival time and time to censoring. We therefore let $U = T \wedge C_1 \wedge C_2$ represent the observation time and define indicators for observing the failure time and dependent censoring times: $\Delta_1 = I(T \leq C_1 \wedge C_2)$ and $\Delta_2 = I(C_2 \leq T \wedge C_1)$, respectively. We let $X^\dagger(t)$ represent the time-dependent covariate at time $t$. Note that $X^\dagger(0)$ would include the elements of $Z$ and potentially other factors predictive of $C_2$. We let $\widetilde{X}(t) = \{X^\dagger(u); u \in [0, t)\}$ denote the history of all baseline and time-dependent covariates up to just before time $t$. The observed data may be summarized as $O_i = \{A_i, U_i, \Delta_{1i}, \Delta_{2i}, Z_i, \widetilde{X}_i(U_i)\}$, with the $O_i$ assumed to be independent and identically distributed (i.i.d.) across $i = 1, \ldots, n$. Note that the set of observed variates is redundant, in the sense that $\widetilde{X}(t)$ includes all baseline covariates $Z$; however, this representation is convenient for presentation purposes.

To define the parameter of interest, we follow the potential outcome framework studied by Rubin (1974, 1978) and adopted by Chen and Tsiatis (2001). Let $T^0$ denote the potential (or counterfactual) lifetime for a randomly selected subject from the population if, possibly contrary to fact, s/he were in group 0, and similarly $T^1$ the potential random variable corresponding to group 1. In reality, $T^0$ and $T^1$ are never observed simultaneously for a subject and the (possibly) observed survival time $T$ relates the two-dimensional potential outcomes $(T^0, T^1)$ through $T = I(A = 0)T^0 + I(A = 1)T^1$. The group-specific difference in restricted mean lifetime is defined as

$$\delta = E\{\min(T^1, L)\} - E\{\min(T^0, L)\}$$
$$= \int_0^L \{P(T^1 > t) - P(T^0 > t)\} \, dt. \tag{1}$$

Under the assumption that $(T^0, T^1) \perp\!\!\!\perp A | Z$, it follows that $P(T > t | A = j, Z) = P(T^j > t | A = j, Z)$ for $j = 0, 1$. Although defined through potential outcomes, the parameter of interest, $\delta$, can then be expressed in terms of observable variates as

$$\delta = \int_0^L E_Z \{P(T > t | A = 1, Z) - P(T > t | A = 0, Z)\} \, dt, \tag{2}$$

where the expectation $E_Z$ is taken in respect to the marginal distribution of $Z$. When $A$ is an indicator of treatment, $\delta$ has the interpretation of average causal treatment effect. We set $S_j(t|Z) = P(T > t | A = j, Z)$ and $S_j(t) = E_Z\{S_j(t|Z)\}$ for $j = 0, 1$.

Estimators for $\delta$ are proposed by Chen and Tsiatis (2001) under the assumption of independent censoring. That is, in our notation, it was assumed that censoring $C_2$ does not exist. Using sample averages to estimate expectations, if one can obtain estimators for $S_0(t|Z)$ and $S_1(t|Z)$, say, $\widehat{S}_0(t|Z)$ and $\widehat{S}_1(t|Z)$, respectively, then a natural estimator for $\delta$ is given by

$$\widehat{\delta} = \int_0^L n^{-1} \sum_{i=1}^{n} \{\widehat{S}_1(t|Z_i) - \widehat{S}_0(t|Z_i)\}\, dt. \qquad (3)$$

## 3. Methods

As argued in the previous section, we wish to model the conditional survival function of $T$ given baseline covariates and group indicator, $S_j(t|Z), j = 0, 1$. Because of its flexibility and popularity in practice, we adopt Cox's proportional hazards model (Cox, 1972, 1975) and, in the following development, we work with the more general model where both baseline hazards functions and regression coefficients of $Z$ are allowed to vary by group. Specifically, it is assumed that

$$\lambda_{ij}(t) \equiv \lambda(t|Z_i, A_i = j) = \lambda_{0j}(t) \exp\left(\beta_j^T Z_i\right), \quad j = 0, 1, \quad (4)$$

where $\lambda(t|Z_i, A_i = j)$ denotes the conditional hazard function given baseline covariates $Z_i$ and membership in group $A_i = j$, while $\lambda_{0j}(t)$ is the unspecified baseline hazard function for group $A = j$.

We now describe how to estimate parameters for model (4). To begin, suppose that $C_{2i}$ was not dependent censoring, but was instead another form of censoring that was conditionally independent of $T$ given $(Z, A)$. In this case, $\beta_j$ could be consistently estimated by the maximum partial likelihood estimator, $\widehat{\beta}_j^*$, which could be computed as the root of the estimating equation,

$$\sum_{i=1}^{n} \int_0^\tau \left\{ Z_i - \frac{\displaystyle\sum_{\ell=1}^{n} Z_\ell \exp\left(\beta_j^T Z_\ell\right) Y_{\ell j}(t)}{\displaystyle\sum_{\ell=1}^{n} \exp\left(\beta_j^T Z_\ell\right) Y_{\ell j}(t)} \right\} dN_{ij}(t) = 0, \quad (5)$$

where $\tau$ satisfies $P(U \geq \tau) > 0$ and, in practice, can be set to the maximum observation time in the study; $N_{ij}(t) = I(A_i = j)I(U_i \leq t, \Delta_{1i} = 1)$; and $Y_{ij}(t) = I(A_i = j)I(U_i \geq t)$. Additionally, $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(u) du$ could be consistently estimated by the Breslow estimator,

$$\widehat{\Lambda}_{0j}^*(t) = \int_0^t \frac{\displaystyle\sum_{i=1}^{n} dN_{ij}(t)}{\displaystyle\sum_{i=1}^{n} \exp\left(\widehat{\beta}_j^{*T} Z_i\right) Y_{ij}(t)}. \qquad (6)$$

From a different perspective, $\widehat{\Lambda}_{0j}^*(t)$ and $\widehat{\beta}_j^*$ are the solutions to the following estimating equations;

$$\sum_{i=1}^{n} \int_0^t dM_{ij}(u; \beta, \Lambda) = 0 \qquad (7)$$

$$\sum_{i=1}^{n} \int_0^\tau Z_i\, dM_{ij}(t; \beta, \Lambda) = 0, \qquad (8)$$

respectively, where $dM_{ij}(t; \beta, \Lambda) = dN_{ij}(t) - Y_{ij}(t) e^{\beta^T Z_i} d\Lambda(t)$ and $dM_{ij}(t) \equiv dM_{ij}(t; \beta_j, \Lambda_{0j})$. Under the assumption that $(C_{1i}, C_{2i}) \perp\!\!\!\perp T_i | (A_i, Z_i)$, we have $E\{dM_{ij}(u)|Z_i, A_i\} = 0$ and $E\{Z_i dM_{ij}(u)|Z_i, A_i\} = 0$ such that (7) and (8) have mean zero at the true parameter values.

However, as stated previously, although $C_{1i} \perp\!\!\!\perp T_i | (A_i, Z_i)$, it is not the case that $C_{2i} \perp\!\!\!\perp T_i | (A_i, Z_i)$. As a result, neither (7) nor (8) have mean zero at the truth, meaning that consistent estimators of $\beta_j$ and $\Lambda_{0j}(t)$ cannot be obtained from (5) and (6) respectively. We assume that the dependence of $C_{2i}$ and $T_i$ occurs through (and only through) the time-dependent process, $\widetilde{X}_i(U_i)$; that is, observed data. That is, we assume that $C_{2i}$ is conditionally independent of $T_i$ given $\{Z_i, A_i, \widetilde{X}_i(U_i)\}$; a condition we can express formally as follows:

$$\begin{aligned} \lim_{h \to 0} h^{-1} P\{&t \leq U_i < t + h, \Delta_{2i} = 1 | U_i \geq t, A_i, \widetilde{X}_i(t), T_i\} \\ &= \lim_{h \to 0} h^{-1} P\{t \leq U_i < t + h, \Delta_{2i} = 1 | U_i \geq t, A_i, \widetilde{X}_i(t)\}. \end{aligned} \qquad (9)$$

Assumption (9) is the critical "no unmeasured confounders" (Rubin, 1977; Robins, 1993) for censoring assumption. In our setting, the assumption essentially states that the hazard of being censored by $C_2$ at time $t$ depends only on observed data up to time $t$ and not additionally on future possibly unobserved data. We define the hazard function for the dependent censoring time, $C_{2i}$ for a subject in group $j$ as

$$\lambda_{ij}^C(t) = \lim_{h \to 0} h^{-1} P\{t \leq U_i < t + h, \Delta_{2i} = 1 | U_i \geq t, A_i = j, \widetilde{X}_i(t)\},$$

then set $\Lambda_{ij}^C(t) = \int_0^t \lambda_{ij}^C(u) du$.

We return now to the issue of estimating the parameters in (4). Reconsidering (7) and (8), although $E\{dM_{ij}(t)|Z_i, A_i\} \neq 0$, under (9), it can be shown that $E[\exp\{\Lambda_{ij}^C(t)\} dM_{ij}(t)| \quad Z_i, A_i, \widetilde{X}_i(t)] = 0$ (Robins and Finkelstein, 2000) and, after iterating the expectation, that $E[\exp\{\Lambda_{ij}^C(t)\} dM_{ij}(t)|Z_i, A_i] = 0$. More generally, it can be shown that $E\{W_{ij}(t) dM_{ij}(t)|Z_i, A_i\} = 0$, where $W_{ij}(t) = \exp\{\Lambda_{ij}^C(t)\} \kappa(t; Z_i, A_i)$, where the function $\kappa(t; Z_i, A_i)$ acts as a stabilization factor. Similarly, it can be shown that $E\{W_{ij}(t) Z_i\, dM_{ij}(t)|Z_i, A_i\} = 0$. Combining these zero-mean properties suggests the following set of IPCW estimating equations,

$$\sum_{i=1}^{n} \int_0^t W_{ij}(u) dM_{ij}(u; \beta, \Lambda) = 0 \qquad (10)$$

$$\sum_{i=1}^{n} \int_0^\tau W_{ij}(t) Z_i\, dM_{ij}(t; \beta, \Lambda) = 0. \qquad (11)$$

Substituting the solution to (10) into (11) then reorganizing algebraically suggests that $\beta_j$ be estimated by the solution to

$$\begin{aligned} \sum_{i=1}^{n} \int_0^\tau &\left\{ Z_i - \frac{\displaystyle\sum_{\ell=1}^{n} W_{\ell j}(t) Y_{\ell j}(t) Z_\ell \exp\left(\beta_j^T Z_\ell\right)}{\displaystyle\sum_{\ell=1}^{n} W_{\ell j}(t) Y_{\ell j}(t) \exp\left(\beta_j^T Z_\ell\right)} \right\} \\ &\times W_{ij}(t) dN_{ij}(t) = 0, \end{aligned} \qquad (12)$$

and that the weighted Breslow estimator,

$$\widehat{\Lambda}_{0j}(t) = \int_0^t \frac{\sum_{i=1}^{n} W_{ij}(u)dN_{ij}(u)}{\sum_{i=1}^{n} W_{ij}(u)Y_{ij}(u)\exp\left(\widehat{\beta}_j^T Z_i\right)}, \qquad (13)$$

be used to estimate $\Lambda_{0j}(t)$ for $j = 0, 1$.

The estimators in (12) and (13) are IPCW estimators (Robins and Rotnitzky, 1992; Robins, 1993; Robins and Finkelstein, 2000). The quantity $\exp\{\Lambda_{ij}^C(t)\}$ can be thought of heuristically as the inverse of the probability of not having been dependently censored as of time $t$. Note that $\widetilde{X}_i(t)$ is an internal time-dependent covariate (Kalbfleisch and Prentice, 2002); that is, a process generated by subject $i$ (as opposed to an external time-dependent covariate such as temperature or air quality). Therefore, $\exp\{\Lambda_{ij}^C(t)\}$ is not actually a probability, per se, but a product of conditional probabilities. Nonetheless, using Robins' Fundamental Identities (Robins and Rotnitzky, 1992; Robins and Finkelstein, 2000), it can be shown that the estimating function in (12) can be expressed as a (dependent censoring process) Martingale integral and hence has mean 0; a proof for which is outlined in Section 2 of the Web Appendix. The function $\kappa(t; A_i, Z_i)$ can be any function of $Z_i$ and $A_i$ (since these are conditioned upon by model (4) anyway) and is intended to stabilize the weighted estimators. In particular, $\exp\{\Lambda_{ij}^C(t)\}$ could be quite large toward the tail of the observation time distribution, which would result in weights that are quite large. One choice of $\kappa$ that has been suggested (e.g., Robins and Finkelstein, 2000; Hernán et al., 2000) is $\exp\{\Lambda_{ij}^C(t|Z_i, A_i)\}$. While $\Lambda_{ij}^C(t)$ would be based on a time-to-censoring model that used $\widetilde{X}_i(t-)$ as covariates, $\Lambda_{ij}^C(t|Z_i, A_i)$ would only use the baseline values. If censoring was in fact independent, then $W_{ij}(t)$ would tend toward 1. Another choice is $\kappa(t; A_i, Z_i) = 1$, which may be appropriate if censoring is light or moderate, in which case $W_{ij}(t)$ does not get unduly large. Hereafter, we refer to $\kappa(t; A_i, Z_i) = 1$ as the "unstabilized" estimator. Stabilized estimators are intended to be more efficient than the unstabilized version, at the expense of additional modeling effort.

In practice, $\Lambda_{ij}^C(t)$ in the weight function is unknown to us and therefore has to be modeled and estimated. To fit models for the dependent censoring time $C_{2i}$, one uses $U_i$ as the censored time variable but uses $\Delta_{2i}$ as the indicator for observing $C_{2i}$. Again, due to its flexibility, the proportional hazards model is a natural choice for the dependent censoring time, $C_{2i}$, with the model being conditional on the group indicator and both baseline and time-dependent covariates. To allow for more flexibility in modeling and hence robustness in estimating the weight, one could fit group-specific Cox models,

$$\lambda_{ij}^C(t) = \lambda_{0j}^C(t)\exp\left\{\theta_j^T X_i(t)\right\},$$

where $\lambda_{0j}^C(t), j = 0, 1$, are unspecified group-specific baseline hazard functions for $C_{2i}$ and $X_i(t)$ is a function of $\widetilde{X}_i(t)$ determined empirically (e.g., using standard model selection techniques, such as stepwise regression) to satisfy $\lambda_{ij}^C(t|\widetilde{X}_i(t)) = \lambda_{ij}^C(t|X_i(t))$. Based on the fitted model, one can estimate

$\Lambda_{ij}^C(t)$ by $\widehat{\Lambda}_{ij}^C(t)$, which can then be used to compute the estimated weight function, $\widehat{W}_{ij}(t)$.

Having estimated $\beta_j$ and $\Lambda_{0j}(t)$, one can correspondingly estimate $S_{ij}(t) \equiv S_j(t|Z_i)$ by

$$\widehat{S}_j(t|Z_i) \equiv \widehat{S}_{ij}(t) = \exp\left\{-\widehat{\Lambda}_{0j}(t)\exp\left(\widehat{\beta}_j^T Z_i\right)\right\}, \qquad j = 0, 1.$$

$$(14)$$

Finally, the proposed estimator for difference in restricted mean lifetime $\delta$ is then given by

$$\widehat{\delta} = \int_0^L \{\widehat{S}_1(t) - \widehat{S}_0(t)\}dt, \qquad (15)$$

where $\widehat{S}_j(t) = n^{-1}\sum_{i=1}^{n}\widehat{S}_{ij}(t)$ for $j = 0, 1$.

The key step in implementing the proposed method is to solve the weighted estimating equation (12), for which existing software may be exploited. For example, one can use `proc phreg` (SAS Institute; Cary, NC) with the counting process input format and the `weight` option. Correspondingly, $\widehat{\Lambda}_{0j}(t)$ and $\widehat{S}_{ij}(t)$ can be easily obtained. Estimating the variance is more involved, requiring additional programming (e.g., SAS's `proc iml`). For illustrative purpose, an SAS macro implementing the proposed methods is available at `http://www.sph.umich.edu/mzhangst/`.

## 4. Asymptotic Properties

In this section, we derive the asymptotic properties of the proposed estimators given by (15). To begin, we specify the regularity conditions, assumed to hold for $i = 1, \ldots, n$ and $j = 0, 1$.

(a) $\{A_i, Z_i, U_i, \Delta_{1i}, \Delta_{2i}, \widetilde{X}_i(U_i)\}$ are independent and identically distributed.

(b) $P(U_i \geq \tau) > 0$.

(c) $|Z_{ik}| < b_Z$ and $\int_0^\tau d|X_{ik}(t)| < b_X$, where $b_Z < \infty$ and $b_X < \infty$, and $k$ denotes the $k$th element.

(d) $\Lambda_{0j}(\tau) < \infty$; $\Lambda_{0j}^C(\tau) < \infty$.

(e) For $d = 0, 1, 2$,

$$\sup_{t \in [0, \tau]} \left\| R_j^{(d)}(t; \beta) - r_j^{(d)}(t; \beta) \right\| \xrightarrow{p} 0$$

$$\sup_{t \in [0, \tau]} \left\| R_{Cj}^{(d)}(t; \theta) - r_{Cj}^{(d)}(t; \theta) \right\| \xrightarrow{p} 0,$$

where we define

$$R_j^{(d)}(t; \beta) = n^{-1}\sum_{i=1}^{n} Y_{ij}(t)W_{ij}(t)Z_i^{\otimes d}\exp\{\beta^T Z_i\}$$

$$R_{Cj}^{(d)}(t; \beta) = n^{-1}\sum_{i=1}^{n} Y_{ij}(t)X_i^{\otimes d}\exp\{\theta^T X_i(t)\}$$

with $z^{\otimes 0} = 1, z^{\otimes 1} = z$ and $z^{\otimes 2} = zz^T$.

(f) The matrices $\Omega_j(\beta_j)$ and $\Omega_j^C(\theta_j)$ are assumed to be positive-definite, where

$$\Omega_j(\beta) = E\left[\int_0^\tau \left\{\frac{r_j^{(2)}(t;\beta)}{r_j^{(0)}(t;\beta)} - \overline{z}_j(t;\beta)^{\otimes 2}\right\}\right.$$
$$\left. \times r_j^{(0)}(t;\beta)\lambda_{0j}(t)\,dt\right]$$

$$\Omega_j^C(\theta) = E\left[\int_0^\tau \left\{\frac{r_{Cj}^{(2)}(t;\theta)}{r_{Cj}^{(0)}(t;\theta)} - \overline{x}_j(t;\theta)^{\otimes 2}\right\}\right.$$
$$\left. \times r_{Cj}^{(0)}(t;\theta)\lambda_{0j}^C(t)\,dt\right],$$

with $\overline{z}(t;\beta) = r_j^{(1)}(t;\beta)/r_j^{(0)}(t;\beta)$ and $\overline{x}(t;\theta) = r_{Cj}^{(1)}(t;\theta)/r_{Cj}^{(0)}(t;\theta)$.

(g) $P(A_i = j | Z_i) \in (0,1)$.

Variations on Condition (a) are possible, although at the expense of additional technical (e.g., Lindeberg-type) conditions. Condition (b) is a standard identifiability criterion. The boundedness implied by Condition (c) helps ensure the convergence of the several stochastic integrals used in the proofs; the same can be said for Condition (d). The second-derivative matrices in Condition (f) are at least nonnegative definite and will be positive-definite under any sensible specification of the covariate vectors. Condition (g) is the well-known positivity requirement from the causal inference literature. If it fails, $\delta$ fails to have a causal interpretation.

We describe the primary asymptotic result for our proposed unstabilized group effect estimator in the following theorem.

THEOREM 1: *Under conditions* $(a) - (g)$, *as* $n \to \infty, \widehat{\delta}$ *converges in probability to* $\delta$, *and for the unstabilized estimator,* $n^{1/2}(\widehat{\delta} - \delta)$ *converges to a zero-mean Normal with variance* $E\{(\phi_{i1} - \phi_{i0})^2\}$, *where*

$$\phi_{ij} = -E\left[Z_i^T \int_0^L \{\mu_{ij}(L) - \mu_{ij}(t)\}d\Lambda_{ij}(t)\right]\Omega_j^{-1}(\beta_j)U_{ij}(\beta_j)$$
$$- \int_0^L E\left[e^{\beta_j^T Z_i}\{\mu_{ij}(L) - \mu_{ij}(t)\}\right]d\Phi_{ij}(t) + (\mu_{ij} - \mu_j),$$

*where* $\mu_{ij}(L) = \mu_{ij}, \mu_{ij}(t) = \int_0^t S_{ij}(u)du, n^{\frac{1}{2}}\{\widehat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)\} = n^{-\frac{1}{2}}\sum_{i=1}^n \Phi_{ij}(t) + O_p(1)$ *and* $n^{\frac{1}{2}}(\widehat{\beta}_j - \beta_j) = \Omega_j(\beta_j)^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n U_{ij}(\beta_j) + O_p(1)$, *with* $\Phi_{ij}(t)$ *and* $U_{ij}(\beta_j)$ *defined in the Web Appendix.*

The variance can be consistently estimated by $n^{-1}\sum_{i=1}^n(\widehat{\phi}_{i1} - \widehat{\phi}_{i0})^2$, where $\widehat{\phi}_{ij}$ is obtained by replacing limiting values in $\phi_{ij}$ with their empirical counterparts. However, as shown in the Web Appendix, the computation of $\widehat{\phi}_{ij}$ is quite complicated owing to the complexity of $\widehat{U}_{ij}(\widehat{\beta}_j)$ and $\widehat{\Phi}_{ij}(t)$. As a result, estimating the variance through $\widehat{\phi}_{ij}$ is very inconvenient computationally. A computationally attractive alternative is to estimate $var(\widehat{\delta})$ by $n^{-1}\sum_{i=1}^n(\widehat{\phi}_{i1}^\dagger - \widehat{\phi}_{i0}^\dagger)^2$, where $\widehat{\phi}_{ij}^\dagger$ is obtained by replacing $\widehat{U}_{ij}(\widehat{\beta}_j)$ and $\widehat{\Phi}_{ij}(t)$ with

$\widehat{U}_{ij}^{\ddagger}(\widehat{\beta}_j)$ and $\widehat{\Phi}_{ij}^\dagger(t)$, respectively, where

$$\widehat{U}_{ij}^\dagger(\beta) = \int_0^\tau \{Z_i - \overline{Z}_j(t;\beta,W)\}\widehat{W}_{ij}(t)d\widehat{M}_{ij}(t)$$
$$\widehat{\Phi}_{ij}^\dagger(t) = \widehat{h}_j^T(t)\widehat{\Omega}_j(\widehat{\beta}_j)^{-1}\widehat{U}_{ij}^\dagger(\widehat{\beta}_j)$$
$$+ \int_0^t \widehat{W}_{ij}(s)R_j^{(0)}(s;\widehat{\beta}_j,\widehat{W})^{-1}d\widehat{M}_{ij}(s).$$

The key difference between $\widehat{\phi}_{ij}$ and

$$\widehat{\phi}_{ij}^\dagger = -n^{-1}\sum_{i=1}^n\left[Z_i^T\int_0^L\{\widehat{\mu}_{ij}(L) - \widehat{\mu}_{ij}(t)\}d\widehat{\Lambda}_{ij}(t)\right]\widehat{\Omega}_j^{-1}(\widehat{\beta}_j)$$
$$\times \widehat{U}_{ij}^\dagger(\widehat{\beta}_j) - n^{-1}\sum_{i=1}^n\int_0^L\left[e^{\widehat{\beta}_j^T Z_i}\{\widehat{\mu}_{ij}(L) - \widehat{\mu}_{ij}(t)\}\right]$$
$$\times d\widehat{\Phi}_{ij}^\dagger(t) + (\widehat{\mu}_{ij} - \widehat{\mu}_j),$$

is that the former accounts for the fact that $\widehat{W}_{ij}(t)$ is estimated, while $\widehat{\phi}_{ij}^\dagger$ is derived with $\widehat{W}_{ij}(t)$ treated as fixed.

We refer to $(\phi_{i1} - \phi_{i0})$ in Theorem 1 as the influence function of $\widehat{\delta}$, which satisfies $n^{\frac{1}{2}}(\widehat{\delta} - \delta) = n^{-\frac{1}{2}}\sum_{i=1}^n(\phi_{i1} - \phi_{i0}) + O_p(1)$. The asymptotic results stated in Theorem 1 are for the unstabilized estimator with the stabilization factor $\kappa(t; A_i, Z_i) = 1$. As equations (12) and (13) are unbiased estimating equations for general $\kappa(t; A_i, Z_i)$, by similar argument, it can be shown that consistency and asymptotic normality hold for estimators with $\kappa(t; A_i, Z_i) \neq 1$ but the form of influence function will be different and even more complicated. Therefore, treating the weights as fixed would be a practical way to estimate the variance of the proposed estimators. As demonstrated in the Web Appendix, for general $\kappa(t; A_i, Z_i), var(\widehat{\delta})$ can still be estimated by $n^{-1}\sum_{i=1}^n(\widehat{\phi}_{i1}^\dagger - \widehat{\phi}_{i0}^\dagger)^2$ with $\widehat{\phi}_{ij}^\dagger$ specified previously. According to Tsiatis (2006, Chapter 9.1), the influence function of $\widehat{\delta}$, if the weight is estimated, is the projection of the influence function when weight is known and fixed onto the orthogonal complement of the nuisance tangent space; for example, the spaces associated with the nuisance baseline hazard function and nuisance parameter $\theta_j$ in the model for $C_2$. As the true influence function is a projection, its variance is smaller than the influence function if weight is fixed and therefore the proposed variance estimator will be conservative in estimating the variance of $\widehat{\delta}$ where weight is actually estimated. This point is discussed by several previous authors (Hernán et al., 2000, 2001; Pan and Schaubel, 2008). The effect of estimating the weight is slight, as will be demonstrated empirically in the next section.

## 5. Simulation Study

We report on simulations to evaluate performance of the proposed methods. Results of two other methods are also reported: a naive method, which estimates $\delta$ by taking difference in areas under group-specific Kaplan–Meier curves from time 0 to $L$ and consequently ignores all possible confounding, and the method proposed by Chen and Tsiatis (2001), which adjusts for baseline covariates but not time-dependent confounders for censoring.

Data were generated under three scenarios, corresponding to different confounding mechanisms and, in each scenario, two different percentages of censoring were considered,

referred to as light censoring or heavy censoring cases. Specifically, in each scenario, for the light censoring case, about 20% subjects are censored by $C_2$ and about 5% are censored by $C_1$, and for the heavy censoring case, about 30% subjects are censored by $C_2$ and about 10% are censored by $C_1$. All reported results are based on 2000 Monte Carlo datasets, and $L$ is chosen to be 15.

In the first scenario, data were generated such that both baseline and time-dependent confounders exist. For each Monte Carlo dataset, a single baseline confounder $Z$ was generated as a truncated standard normal, truncated at $-4$ and $4$ on each side, and group indicator $A$ as Bernoulli with parameter $\exp(-0.6Z)/\{1 + \exp(-0.6Z)\}$. Survival time $T$ was generated by transforming $\epsilon_1 \sim$ Uniform $(0,1)$ using the inverse of the cumulative distribution function (cdf) of a Weibull distribution with shape parameter 1.25 and scale parameter $\exp(-0.3Z - 3.3)$ for group $A = 1$ or $\exp(-0.4Z - 3)$ for $A = 0$. We then generated dependent censoring $C_2$ such that it depends both on baseline and time-dependent covariates as follows. In order for a time-dependent covariate to be a confounder, it should be correlated both with $T$ and $C_2$ conditioning on $(A, Z)$. To achieve this, we first generated $X_t$ such that $X_t = -5\log\{A\epsilon_1 + (1-A)(1-\epsilon_1)\} + \epsilon_2$, where $\epsilon_2 \sim$ Uniform$(0,1)$, independent of all other variables, and then let $X(t) = I(X_t \geq t)$. Consequently, the time-dependent covariate $X(t)$ is correlated with survival time $T$ through their mutual relationships with $\epsilon_1$ and such correlation exits even conditioning on $(A, Z)$. Next, we generated dependent censoring time $C_2$ using a proportional hazards model with hazard rate $\exp\{\gamma_1 + 0.2A + 0.2Z + \gamma_2 X(t)\}$. This procedure ensures that $X(t)$ is a time-dependent confounder and $C_2$ follows a proportional hazards model with $Z$ and $X(t)$ as our model assumes. Finally, censoring time $C_1$ was generated as Weibull with shape and scale parameters 3 and $\exp(\gamma_3)$ respectively. The coefficients $(\gamma_1, \gamma_2, \gamma_3)$ are set to $(-5.1, 1.5, -11)$ for light

censoring case and to $(-4.45, 1.5, -9.7)$ for heavy censoring case.

In the second scenario, data were generated such that, conditioning on $(A, Z)$, time-dependent covariate $X(t)$ correlates only with survival time $T$ but not with censoring time $C_2$ and therefore it is not a confounder if $(A, Z)$ is properly adjusted for. We compare the proposed methods with the Chen and Tsiatis (2001) method, which should be consistent and asymptotically normal under this scenario. Data are generated similarly as before except for that $\gamma_2$ was set to 0, ensuring that $X(t)$ does not affect $C_2$. Specifically, we chose $(\gamma_1, \gamma_2, \gamma_3)$ equal to $(-4, 0, -11)$ for light censoring case and equal to $(-3.4, 0, -9.7)$ for heavy censoring case.

In the third scenario, we let the time-dependent covariate $X(t)$ be conditionally independent of survival time $T$ but still correlated with censoring time $C_2$ and, consequently, is not a confounder either. Data were generated the same as in scenario one except for that $X_t = -5\log\{A\epsilon_3 + (1-A)(1-\epsilon_3)\} + \epsilon_2$, where $\epsilon_3 \sim$ Uniform$(0,1)$, independent of all other variables. We evaluate how the proposed methods compare with those of Chen and Tsiatis (2001), which should be unbiased under this scenario. All coefficients are set equal to those used in scenario 1.

Tables 1 and 2 list the results of our simulation study. The proposed estimators (unstabilized or stabilized) are consistent for the true parameter under all scenarios, whereas the Chen and Tsiatis (2001) method leads to biased estimators under scenario 1, where a time-dependent confounder exists. The naive estimator is biased under all three scenarios, where baseline or/and time-dependent confounders exist. Although the variance is estimated by treating the estimated weight as fixed, coverage probabilities of the proposed estimators achieve the nominal level. On the other hand, due to the nonignorable bias, coverage probabilities of the other two estimators are very low, illustrating that the impact of confounders

**Table 1**
*Estimators for difference in restricted mean lifetime under scenarios 1–3, light censoring case (2000 Monte-Carlo datasets; sample size: 1000; True is the true value of parameter; MC Bias is the Monte Carlo bias; MC SD is the Monte Carlo standard deviation of estimates; Ave. SE is the Monte Carlo average of estimated standard errors; CP is the coverage probability of nominal 95% Wald confidence interval.)*

| Method | True | MC Bias | MC SD | Ave. SE | CP |
|---|---|---|---|---|---|
| Light Censoring, Scenario 1 | | | | | |
| Unstabilized inverse weighting | 1.172 | 0.021 | 0.319 | 0.327 | 0.950 |
| Stabilized inverse weighting | 1.172 | 0.015 | 0.317 | 0.327 | 0.952 |
| Chen & Tsiatis' method | 1.172 | 0.316 | 0.321 | 0.324 | 0.840 |
| naive | 1.172 | −0.429 | 0.325 | 0.332 | 0.755 |
| Light Censoring, Scenario 2 | | | | | |
| Unstabilized inverse weighting | 1.172 | 0.005 | 0.321 | 0.330 | 0.954 |
| Stabilized inverse weighting | 1.172 | 0.002 | 0.321 | 0.330 | 0.953 |
| Chen & Tsiatis' method | 1.172 | 0.005 | 0.323 | 0.329 | 0.953 |
| Naive | 1.172 | −0.729 | 0.326 | 0.336 | 0.422 |
| Light Censoring, Scenario 3 | | | | | |
| Unstabilized inverse weighting | 1.172 | −0.013 | 0.337 | 0.328 | 0.939 |
| Stabilized inverse weighting | 1.172 | −0.016 | 0.337 | 0.327 | 0.936 |
| Chen & Tsiatis' method | 1.172 | −0.017 | 0.336 | 0.326 | 0.934 |
| Naive | 1.172 | −0.746 | 0.341 | 0.332 | 0.392 |

**Table 2**
*Estimators for difference in restricted mean lifetime under scenarios* 1–3, *heavy censoring case* (*entries are as in Table* 1)

| Method | True | MC Bias | MC SD | Ave. SE | CP |
|---|---|---|---|---|---|
| | | Heavy Censoring, Scenario 1 | | | |
| Unstabilized inverse weighting | 1.172 | 0.040 | 0.331 | 0.342 | 0.948 |
| Stabilized inverse weighting | 1.172 | 0.050 | 0.334 | 0.344 | 0.947 |
| Chen & Tsiatis' method | 1.172 | 0.593 | 0.334 | 0.337 | 0.584 |
| Naive | 1.172 | −0.157 | 0.339 | 0.344 | 0.926 |
| | | Heavy Censoring, Scenario 2 | | | |
| Unstabilized inverse weighting | 1.172 | 0.006 | 0.329 | 0.329 | 0.959 |
| Stabilized inverse weighting | 1.172 | 0.003 | 0.328 | 0.346 | 0.960 |
| Chen & Tsiatis' method | 1.172 | 0.004 | 0.334 | 0.345 | 0.956 |
| Naive | 1.172 | −0.725 | 0.344 | 0.350 | 0.455 |
| | | Heavy Censoring, Scenario 3 | | | |
| Unstabilized inverse weighting | 1.172 | −0.009 | 0.350 | 0.345 | 0.936 |
| Stabilized inverse weighting | 1.172 | −0.014 | 0.348 | 0.343 | 0.941 |
| Chen & Tsiatis' method | 1.172 | −0.015 | 0.345 | 0.340 | 0.940 |
| Naive | 1.172 | −0.743 | 0.348 | 0.344 | 0.424 |

can be severe if they are not properly accounted for in the analysis.

Under scenarios 2 and 3, where there are no time-dependent confounders, the usual partial likelihood estimator is semi-parametric efficient in estimating coefficients of a Cox proportional hazards model. Therefore, the Chen and Tsiatis (2001) method can be used as a benchmark to evaluate the loss of efficiency of the proposed methods due to inverse probability weighting. Our results demonstrate that the stabilized version of the proposed estimator behaves very similarly to the estimator of Chen and Tsiatis (2001). This would be expected since the weights would tend toward 1 in this scenario, such that the loss of efficiency (corresponding to the unstabilized weighting) is only mild. In addition, simulation results suggest that the effect of the stabilization factor, $\kappa(t; Z, A)$, on the efficiency is more pronounced for estimators of the Cox regression parameter, compared to the estimators of $\delta$. Results in Table 3 show that, at least under the simulated scenarios, stabilization results in considerable efficiency gains for $\widehat{\beta}_j$, but only mild increases in precision for $\widehat{\delta}$.

## 6. Application
Data were obtained from the Scientific Registry of Transplant Recipients (SRTR). The study population ($n = 6371$) included all chronic liver disease patients initially wait listed for deceased-donor liver transplantation in the United States at age $\geq 18$ between March 1, 2002 and February 28, 2003. For each patient, the time origin ($t = 0$) was the date of wait listing. Patients were followed from that date until the earliest of death, receipt of a liver transplant, loss to follow-up, and the end of the observation period: December 31, 2008.

The event of interest was wait list mortality. Independent censoring consisted of random loss to follow-up and administrative censoring at the end of the observation period. Dependent censoring occurred through liver transplantation which,

although not preventing the observation of death, does preclude wait list death.

The objective of the analysis was to compare 5-year mean wait list survival time between Hepatitis C positive (HCV+) versus HCV− patients. HCV is a leading cause of chronic liver disease. Baseline adjustment covariates included the following factors, as measured at the time of wait initial listing: age, gender, race, region, MELD score, serum albumin, sodium, body mass index, diabetes, hospitalization status, ascites, dialysis, and encephalopathy. The MELD score is a log linear combination of serum creatinine, bilirubin, and international normalized ratio for prothrombin time. MELD has been shown to be a very strong predictor of wait list mortality. Currently, patients are ordered on the wait list in a decreasing order of MELD score, such that the higher the MELD score, the greater the liver transplant hazard.

MELD is time-dependent, since a patient's score will be updated regularly. Other time-dependent covariates include dialysis, serum albumin, sodium, and active and removal status. In the time-until-transplant model, each of these factors was represented in the time-dependent covariate vector. The Cox model for transplant was given by

$$\lambda_i^C(t) = \{1 - I_i(t)\}\{1 - R_i(t)\}\lambda_0^C(t)\exp\{\theta_0^T X_i(t)\}, \quad (16)$$

where $I_i(t)$ is an indicator for being inactivated from the wait and $R_i(t)$ is an indicator for being removed from the wait list at time $t$. It is not possible for a patient to receive a transplant while they are inactivated (usually temporary) or removed (typically permanent). To fit model (16), we deleted patient subintervals where either $I_i(t) = 1$ or $R_i(t) = 1$. After model (16) was fitted, the IPCW weight was then computed using $\int_1^t \{1 - I_i(s)\}\{1 - R_i(s)\}d\widehat{\Lambda}_1^C(s)$, such that the transplant hazard increment was set to 0 for each subinterval where the patient was either inactivated or removed.

The study population consisted of 2754 HCV+ ($j = 1$) and 3617 HCV− ($j = 0$) patients. There were a total of

**Table 3**

*Comparing efficiency of the unstabilized and stabilized inverse weighting methods in estimating δ in (1) and coefficients of Cox's proportional hazards models in (4) ($\widehat{\delta}$ is weighted estimator for δ; $\widehat{\beta}_j$ is weighted estimator for $\beta_j$ in (4), $j = 0, 1$; RE is the relative efficiency compared to the unstabilized inverse weighting method, calculated as the square of the ratio of the Monte Carlo standard deviation for the unstabilized inverse weighting estimator over that for the indicated estimator)*

| Method | $\widehat{\beta}_1$ MC SD | RE | $\widehat{\beta}_0$ MC SD | RE | $\widehat{\delta}$ MC SD | RE |
|---|---|---|---|---|---|---|
| | Light Censoring, Scenario 1 | | | | | |
| Unstabilized inverse weighting | 0.059 | 1 | 0.058 | 1 | 0.319 | 1 |
| Stabilized inverse weighting | 0.057 | 1.085 | 0.055 | 1.114 | 0.317 | 1.011 |
| | Light Censoring, Scenario 2 | | | | | |
| Unstabilized inverse weighting | 0.058 | 1 | 0.058 | 1 | 0.321 | 1 |
| Stabilized inverse weighting | 0.057 | 1.035 | 0.057 | 1.060 | 0.320 | 1.004 |
| | Light Censoring, Scenario 3 | | | | | |
| Unstabilized inverse weighting | 0.060 | 1 | 0.060 | 1 | 0.337 | 1 |
| Stabilized inverse weighting | 0.057 | 1.094 | 0.056 | 1.143 | 0.337 | 0.999 |
| | Heavy Censoring, Scenario 1 | | | | | |
| Unstabilized inverse weighting | 0.069 | 1 | 0.066 | 1 | 0.334 | 1 |
| Stabilized inverse weighting | 0.063 | 1.195 | 0.059 | 1.250 | 0.331 | 1.016 |
| | Heavy Censoring, Scenario 2 | | | | | |
| Unstabilized inverse weighting | 0.064 | 1 | 0.066 | 1 | 0.329 | 1 |
| Stabilized inverse weighting | 0.062 | 1.070 | 0.062 | 1.110 | 0.328 | 1.006 |
| | Heavy Censoring, Scenario 3 | | | | | |
| Unstabilized inverse weighting | 0.069 | 1 | 0.072 | 1 | 0.350 | 1 |
| Stabilized inverse weighting | 0.062 | 1.217 | 0.063 | 1.320 | 0.348 | 1.024 |
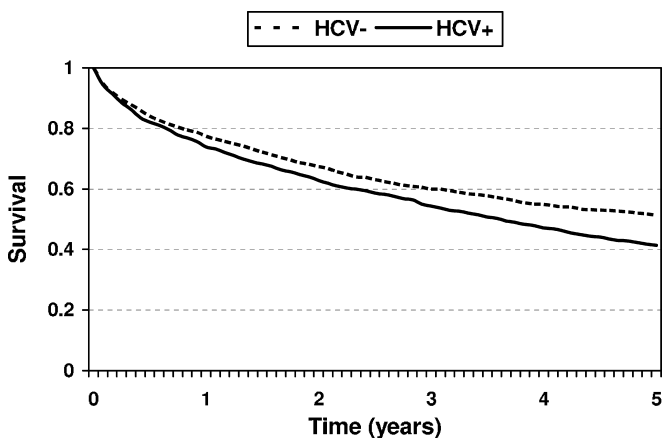


**Figure 1.** Average wait list survival probability for HCV− and HCV + patients at 5 years.

1849 wait list deaths, 3194 liver transplants, and 1328 independently censored subjects. Average survival curves are presented in Figure 1. Average wait list survival probability was 51% for HCV− patients at 5 years, compared to 41% for HCV+ patients. In Table 4, we compute 1-, 3-, and 5-year average restricted mean wait list lifetime for HCV+ and HCV− patients. For each of $L = 1, L = 3$, and $L = 5$ years, average restricted mean lifetime is significantly greater for HCV− compared to HCV+ patients. In Figure 2,

**Table 4**

*Average restricted mean wait list lifetime for HCV+ and HCV− patients restricted to $L = 1$, $L = 3$, and $L = 5$ years*

| $L$ | $\widehat{\mu}_1$ | $\widehat{\mu}_0$ | $\widehat{\delta}$ | SE $(\widehat{\delta})$ | p |
|---|---|---|---|---|---|
| 1 | 0.83 | 0.85 | −0.02 | 0.01 | 0.047 |
| 3 | 2.10 | 2.21 | −0.11 | 0.03 | 0.002 |
| 5 | 3.04 | 3.31 | −0.26 | 0.06 | <0.0001 |

we plot the point estimates and 95% confidence intervals for δ.

## 7. Discussion

In this article, we have proposed methods for estimating differences in restricted mean survival time between groups where group assignment is not randomized and, conditional on baseline covariates and group assignment, censoring may still be correlated with survival time. Differences in restricted mean lifetime may be of direct interest, and could also serve as a cumulative effect measure in settings where group-specific hazards are nonproportional. To be general, in our formulation, we considered that both conditionally independent and dependent censoring exist, which is often the case in practice; this formulation includes as a special case when only one type of censoring exits. The proposed methods employ two general approaches to account for two types of confounders. The proposed methods combine
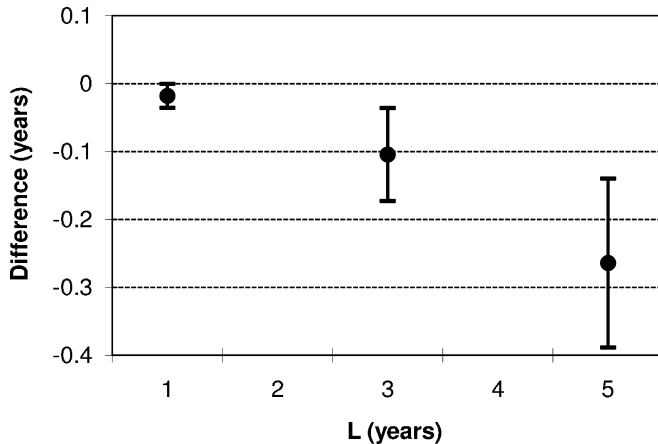
**Figure 2.** Point estimates and 95% confidence intervals for differences in average restricted mean wait list lifetime between HCV+ and HCV− patients.

inverse probability of censoring weighting (e.g., Robins and Rotnitzky, 1992) and the procedure of explicitly averaging over the marginal covariate distribution (e.g., Chen and Tsiatis, 2001).

In our proposed procedure, computation is simplified by treating the IPCW weights as fixed. Since the inverse weights are actually estimated using the data, treating them as fixed should result in conservative confidence intervals and hypothesis tests, as reported by several previous authors (e.g., Hernán et al., 2000, 2001; Pan and Schaubel, 2008). Our simulation results reveal the proposed standard error estimators and corresponding confidence intervals are quite accurate. An alternative to treating the weights as fixed would be to use the bootstrap, but this is much less-convenient computationally.

The methods we propose require that the IPCW weight be correctly specified. The degree of bias introduced by misspecifying the Cox model for censoring would be expected to increase with increasing proportion of dependent (relative to independent) censoring; increasing strength of association (i) between the death hazard and time-dependent confounders and (ii) between the dependent censoring hazard and time-dependent confounders; and of course degree to which the IPCW model is misspecified. Fortunately, one can readily evaluate the fit of the proportional hazards model through well-established techniques and using standard software.

There are several alternatives to our proposed approach, one being joint modeling. For example, one could combine a mixed model for the longitudinal measurements with a hazard model that uses time-dependent covariates (e.g., Proust-Lima and Taylor, 2009). Or, one could also model mean residual survival time directly using pseudo-observations (Andersen, Hansen, and Klein, 2004). It could well be the case that the preferred method is a function of the data configuration. A detailed comparison of the three types of approaches would be very useful to practitioners.

## 8. Supplementary Materials

The Web Appendix, referenced in Sections 3 and 4 is available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

REFERENCES

Andersen, P. K., Hansen, M. G., and Klein, J. P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis* **10,** 335–350.

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

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–200.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–275.

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

Hernán 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—Applications and Case Studies* **96,** 440–448.

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

Karrison, T. (1987). Restricted mean life with adjustment for covariates. *Journal of the American Statistical Association* **18,** 151–167.

Pan, Q. and Schaubel, D. E. (2008). Proportional hazards regression based on biased samples and estimated selection probabilities. *Canadian Journal of Statistics* **36,** 111–127.

Proust-Lima, C. and Taylor, J. M. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: A joint modeling approach. *Biostatistics* **10,** 535–549.

Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 24–23. Alexandria, Virginia: American Statistical Association.

Robins, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. *Proceedings of the American Statistical Association Section on Bayesian Statistical Science* **1999,** 6–10.

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

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology—Methodological Issues*. Boston: Birkhäuser, 297–331.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66,** 688–701.

Rubin, D. B. (1977). Inference and missing data. *Biometrika* **63,** 581–592.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* **6,** 34–58.

SAS Institute Inc. (2006) *SAS Online Doc 9.1.3.* Cary, North Carolina: SAS Institute, Inc.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data.* New York: Springer.

Wiesner, R. H., McDiarmid, S. V., Kamath, P. S., Edwards, E. B., Malinchoc, M., Kremers, W. K., Krom, R. A., and Kim, W. R. (2001). MELD and PELD: Application of survival models to liver allocation. *Liver Transplantation* **7,** 567–580.

Zucker, D. M. (1998). Restricted mean life with covariates: Modification and extension of a useful survival analysis method. *Journal of the American Statistical Association* **93,** 702–709.