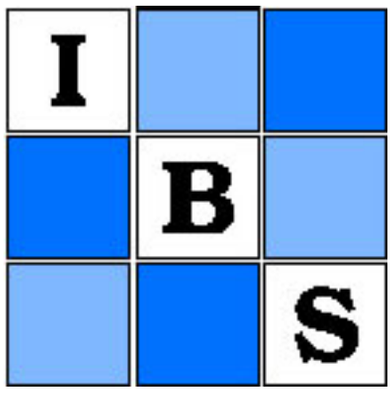

BIOMETRICS 57, 1030–1038
December 2001

Causal Inference on the Difference of the Restricted Mean Lifetime Between Two Groups

Pei-Yun Chen* **and Anastasios A. Tsiatis****

Department of Statistics, North Carolina State University,
Raleigh, North Carolina 27695, U.S.A.
**email:* pchen2@stat.ncsu.edu
***email:* tsiatis@stat.ncsu.edu

SUMMARY. When comparing survival times between two treatment groups, it may be more appropriate to compare the restricted mean lifetime, i.e., the expectation of lifetime restricted to a time L, rather than mean lifetime in order to accommodate censoring. When the treatments are not assigned to patients randomly, as in observational studies, we also need to account for treatment imbalances in confounding factors. In this article, we propose estimators for the difference of the restricted mean lifetime between two groups that account for treatment imbalances in prognostic factors assuming a proportional hazards relationship. Large-sample properties of our estimators based on martingale theory for counting processes are also derived. Simulation studies were conducted to compare these estimators and to assess the adequacy of the large-sample approximations. Our methods are also applied to an observational database of acute coronary syndrome patients from Duke University Medical Center to estimate the treatment effect on the restricted mean lifetime over 5 years.

KEY WORDS: Causal inference; Cox's proportional hazard model; Martingale process; Observational study; Restricted lifetime; Stochastic integral; Survival analysis.

1. Introduction

Comparison of the distribution of survival times for patients assigned to different treatment groups is often of interest in medical research. If the treatments were assigned to patients randomly and if there was no censoring, simple sample averages could be used to estimate the mean lifetime of the two groups. It is often the case that survival times are censored because individuals may still be alive at the time of analysis or there was loss to follow-up. When survival data are censored, it may be infeasible to estimate the mean survival time, in which case it may be more appropriate to compare the restricted mean lifetime, or the expectation of lifetime restricted to some time L. In some cases, when the effect of treatment is expected to benefit patients for only a fixed time, the comparison of restricted mean lifetime may be preferred even if censoring is not an issue.

In a randomized study with censored survival data, the restricted mean lifetime could be estimated for each treatment group by finding the area under the Kaplan–Meier curve up to time L and the difference between the treatment-specific estimates could be used as the basis of a test statistic. In an observational study, however, a direct comparison may be misleading because the patients that receive one treatment generally differ systematically from those that receive the other treatment. The more appropriate analysis would account for treatment imbalances by important prognostic factors. For example, in an observational study of patients with acute coronary syndrome who received their initial heart catheterization between 1986 and 1997 from Duke University Medical Center, one question of interest is the comparison of the restricted mean lifetime of patients treated medically versus those treated with percutaneous coronary intervention (PCI). The data set comprises 6033 patients, of which 3868 were treated with PCI and 2165 were medically treated (MED), and includes follow-up times, status at the end of follow-up (dead or censored), and the patients' baseline demographics, medical histories, and catheterization results for acute coronary syndrome.

In examining the data, we found substantial treatment imbalance in some of the important prognostic factors. For example, the average percentage of ejection fraction is 49.92 (SE = 0.29) for the MED group and 55.02 (SE = 0.21) for the PCI group. Also, 18.29% of the MED patients had congestive heart failure prior to entry into this study as compared with 7.65% of the PCI patients. On average, patients who receive medical therapy are prognostically poorer than those who receive PCI. Consequently, it is important that we account for treatment imbalances when comparing the restricted mean lifetime between those two groups. This example will be discussed later in greater detail.

Karrison (1987) developed a method for a covariate-adjusted comparison of two groups with respect to the restricted mean lifetime. He assumed that the failure time distribution was related to covariates through a proportional hazards

model. In his model, he allowed different baseline hazards but assumed the same multiplicative effect of covariate on the hazard function for both groups. The baseline hazard was also assumed to be piecewise exponential.

Zucker (1998) implemented Karrison's basic approach for estimating the restricted mean lifetime by assuming the stratified Cox model. Different baseline hazard functions for treatment groups were assumed and were left unspecified and a constant regression coefficient was used across the covariates. The method was also extended to achieve robustness against misspecification in the underlying model.

In both Karrison (1987) and Zucker's (1998) articles, the covariates are treated as fixed rather than random. Hence, the variation of the covariates was not taken into consideration when deriving the asymptotic variance of their estimators for the difference of the restricted mean lifetime. We will show that this generally underestimates the variance of the estimator for treatment difference.

For the one-sample problem, Shen and Fleming (1997) proposed an estimator for the mean survival function based on Cox's regression model that incorporates covariates in a heterogeneous population. The estimator is shown to be more efficient than the Kaplan–Meier estimate. However, they did not consider treatment comparisons with covariate imbalance.

Because, in an observational study, treatment comparisons may be difficult to make because of confounding, we take the point of view described by Rubin (1974, 1978) of estimating the average causal treatment difference in restricted lifetime. This causal treatment effect is defined through counterfactual random variables. We demonstrate that the covariate-adjusted comparison of restricted mean lifetime proposed by Zucker (1998) and Karrison (1987) and the methods in this article are estimators of the average causal treatment difference in restricted lifetime. Another approach that has been advocated by Hubbard, van der Laan, and Robins (1999) for estimating the average causal treatment difference in survival in observational studies is through the use of inverse probability weighted estimators. In this approach, both the censoring distribution and the probability of receiving one of the treatments (propensity score) are modeled as a function of the covariates. The inverse of the product of these is used to derive weighted estimating equations that will yield estimators of the average causal treatment effect. They also show how to use the observed covariates to obtain locally efficient estimators. The advantage of this method is that the relationship of survival as a function of the covariates can be left arbitrary, whereas the methods of this article model this relationship directly. However, to implement the methods of Hubbard et al., one needs to model the relationship of censoring and treatment assignment as a function of the covariates, whereas these relationships can be left arbitrary using the methods advocated here.

The article is organized as follows. The definition of average causal treatment effect, counterfactual random variables, and the assumptions necessary to conduct such an analysis are reviewed in Section 2. The basic strategy for computing the difference in the restricted mean lifetime is derived in Section 3. In Section 4, we give results of several simulation studies intended to illustrate the accuracy of the large-sample approximations. In Section 5, we applied our methods to estimate the treatment effects on the restricted mean lifetime from an observational study of patients with coronary artery disease. The discussion and some additional remarks are given in the last section. Technical derivations are summarized in the Appendix.

2. Notation and Assumptions

Suppose we are interested in comparing the mean of the restricted lifetime up to time L for two treatment groups. Specifically, if the random variable T denotes the survival time for a randomly selected individual in our population, then the restricted lifetime for that individual is given by $T_L = \min(T, L)$. Since we will only be considering restricted lifetimes, then by convention, we will refer to the restricted lifetime by T (suppressing the subscript L). Thus, from now on, T is a positive random variable bounded by L.

We first review the definition of average causal treatment effect. Let us consider two treatments referred to as treatments 0 and 1. Using the notion of counterfactual random variables (potential outcomes) as described by Rubin (1974, 1978), we define T^0 to correspond to the restricted lifetime of a randomly chosen individual in our population if, possibly contrary to fact, he/she received treatment 0. Similarly, we define T^1 as the restricted lifetime if he/she received treatment 1. These are hypothetical quantities since an individual can receive only one treatment. Nonetheless, the average causal treatment effect is defined as

$$\delta = \mathrm{E}\left(T^1\right) - \mathrm{E}\left(T^0\right) = \int_0^L \left\{\mathrm{P}\left(T^1 \geq u\right) - \mathrm{P}\left(T^0 \geq u\right)\right\} du.$$

In actuality, the experimental sample will receive only one treatment, 0 or 1, and this will be indicated by the treatment indicator $A = (0, 1)$, and the observed response (restricted lifetime) is equal to $T = T^0 I(A = 0) + T^1 I(A = 1)$, where $I(\cdot)$ denotes the indicator function. It is important to understand under what conditions the average causal treatment effect can be identified from the observable data. In an observational study, patients receiving treatment 1 may not be prognostically similar to those receiving treatment 0. However, if prognostic factors Z can be identified that are believed to explain the prognostic variation and if, in addition, we believe that treatment choice depends on Z, then conditional on Z, it may be reasonable to assume that treatment assignment is random. Specifically, we may assume that the counterfactual random variables (T^0, T^1) are conditionally independent of treatment assignment A given Z. Rubin (1978) refers to this as the strong ignorability assumption. With this assumption, the average causal treatment effect δ can be identified by the distribution of the observable data (T, A, Z). This follows because the distribution of the counterfactual random variable T^0 is equal to

$$\begin{aligned} \mathrm{P}\left(T^0 \geq u\right) &= \mathrm{E}_Z\left\{\mathrm{P}\left(T^0 \geq u \mid Z\right)\right\} && (1)\\ &= \mathrm{E}_Z\left\{\mathrm{P}\left(T^0 \geq u \mid A = 0, Z\right)\right\} && (2)\\ &= \mathrm{E}_Z\left\{\mathrm{P}\left(T \geq u \mid A = 0, Z\right)\right\}. && (3) \end{aligned}$$

Equation (2) follows because of the strong ignorability assumption. In equation (1), the expectation is taken with respect to the marginal distribution of the covariates Z. To emphasize this, we subscript the expectation as E_Z. It is important to note that, in equation (3), the expectation is again

taken with respect to the marginal distribution of Z and not the conditional distribution of $Z \mid A = 0$. Similarly,

$$\mathrm{P}\left(T^1 \geq u\right) = \mathrm{E}_Z\{\mathrm{P}(T \geq u \mid A = 1, Z)\}.$$

Therefore, the average causal treatment effect is given by

$$\delta = \int_0^L \left[\mathrm{E}_Z\{\mathrm{P}(T \geq u \mid A = 1, Z)\} - \mathrm{E}_Z\{\mathrm{P}(T \geq u \mid A = 0, Z)\}\right] du.$$

In most studies, we are not able to observe the survival times of all the patients due to various types of censoring. To accommodate censoring, we define a potential time to censoring, which we denote by C. The data we observe in such a study is $X = \min(T, C)$ and $\Delta = I(T \leq C)$. We will make the usual assumption that censoring is noninformative conditional on the covariates and treatment assignment, namely, that C is conditionally independent of (T^0, T^1) given (A, Z). This last assumption also implies that C is conditionally independent of T given (A, Z). We also assume that $\mathrm{P}(C \geq L) > 0$. Therefore, the typical observable data from an observational study can be summarized as n independent random vectors $\{X_i = \min(T_i, C_i), \Delta_i = I(T_i \leq C_i), A_i, Z_i; i = 1, \ldots, n\}$.

The question we address is how to estimate the average causal treatment difference in restricted lifetime from a sample of censored survival data given above. This will be accomplished by deriving estimates for the $\mathrm{P}(T \geq u \mid A, Z)$ using different modeling assumptions. By (3), it is natural to estimate $S_0(u) = \mathrm{P}(T^0 \geq u)$ by $\hat{S}_0(u) = n^{-1} \Sigma_{i=1}^n \hat{S}_0(u \mid Z_i)$, where $\hat{S}_0(u \mid Z_i)$ is an estimator for $S_0(u \mid Z_i) = \mathrm{P}(T \geq u \mid A = 0, Z_i)$. Similarly, $\hat{S}_1(u) = n^{-1} \Sigma_{i=1}^n \hat{S}_1(u \mid Z_i)$ is an estimator for $S_1(u) = \mathrm{P}(T^1 \geq u)$. Then an estimator for δ is given by

$$\hat{\delta} = \int_0^L \{\hat{S}_1(u) - \hat{S}_0(u)\} du. \tag{4}$$

Note that the averaging is performed over all covariate vectors Z_i, $i = 1, \ldots, n$, across both treatment groups. Intuitively, to adjust for confounding, we must ensure that comparable empirical distributions for the covariates are used in each treatment group. The causal model above makes it clear that the comparable distribution of covariates to be used is that from both treatment groups combined. The models we will consider are proportional hazards models. These are described in detail in the next section. Proportional hazards models, first introduced by Cox (1972), are the most widely used models for censored survival data. We focus on these models because they are flexible, robust (semiparametric), and software is widely available to obtain estimates.

3. Estimators for the Treatment Difference in Restricted Mean Lifetime

In order to obtain the estimator given by (4), we need to model the relationship of the survival distribution to treatment and covariates, i.e., $\mathrm{P}(T \geq u \mid A, Z)$, and find estimates for the parameters in this model. In this article, we will consider two different models, both of which are proportional hazards models. The first model, which we refer to as model 1, assumes

$$\lambda(t \mid A = 0, Z) = \lambda_0(t) e^{\beta_0^T Z}$$

and

$$\lambda(t \mid A = 1, Z) = \lambda_1(t) e^{\beta_1^T Z},$$

where $\lambda(t \mid A, Z)$ denotes the conditional hazard function at time t given treatment A and covariates Z. The treatment-specific baseline hazard functions $\lambda_1(t)$ and $\lambda_0(t)$ are left unspecified. We also consider a second proportional hazards model, referred to as model 2, which assumes

$$\lambda(t \mid A, Z) = \lambda(t) e^{\gamma_0 A + \gamma_1^T Z} = \lambda(t) e^{\gamma^T W},$$

where $\gamma = (\gamma_0, {\gamma_1}^T)^{\mathrm{T}}$ and $W = (A, Z^T)^{\mathrm{T}}$.

We note that, in model 1, both the baseline hazard and the regression coefficients associated with Z are allowed to vary by treatment. This is more general than the model by Karrison (1987) and Zucker (1998), which assumed that the regression coefficients associated with the covariate Z were the same for both treatments. In contrast, model 2 is more restrictive than any of the others in that both the baseline hazard and regression coefficients associated with Z are the same for both treatments. We consider these two models in order to assess the trade-off between increased variance associated with model complexity and the increased bias associated with model misspecification.

Throughout the remainder of the article, it will be convenient to use counting process notation and the associated martingale properties of these counting processes as given by Fleming and Harrington (1991) to define the estimators and derive their large-sample properties. Toward that end, we define the following notation. The counting process that counts the number of observed deaths that occur before or at time t for individual i is given by $N_i(t) = I(X_i \leq t, \Delta_i = 1)$, and the at-risk process for the ith individual is given by $Y_i(t) = I(X_i \geq t)$. Consider the filtration $\mathcal{F}(t)$ to be the increasing sequence of σ-algebras generated by $\sigma\{I(X_i \leq u, \Delta_i = 1), I(X_i \leq u, \Delta_i = 0), A_i, Z_i; u \leq t; i = 1, \ldots, n\}$, which corresponds to the information regarding the observed deaths and censoring prior to and including time t and all the A_i and Z_i. With respect to this filtration, we define the martingale process

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(u) Y_i(u) du,$$

where $\lambda_i(u) = \{(1 - A_i)\lambda_0(u) e^{\beta_0^T Z_i} + A_i \lambda_1(u) e^{\beta_1^T Z_i}\}$ for model 1 and $\lambda_i(u) = \lambda(u) e^{\gamma^T W_i}$ for model 2. We also define $M(t) = \Sigma_{i=1}^n M_i(t), N(t) = \Sigma_{i=1}^n N_i(t)$, and $Y(t) = \Sigma_{i=1}^n Y_i(t)$.

For model 1, the estimators $\hat{\beta}_0$ and $\hat{\Lambda}_0(u)$ are obtained by using the maximum partial likelihood estimator and the Breslow (1972) estimator for the cumulative hazard function using data only for individuals in treatment group 0. The Breslow estimator for $\Lambda_0(u) = \int_0^u \lambda_0(t) dt$ is

$$\hat{\Lambda}_0(u) = \int_0^u \frac{\sum_{i=1}^n (1 - A_i) dN_i(t)}{\sum_{i=1}^n (1 - A_i) e^{\hat{\beta}_0^T Z_i} Y_i(t)}.$$

A similar approach can be used for treatment group 1. Consequently, the estimator for $S_0(u \mid Z_i) = \mathrm{P}(T \geq u \mid A = 0, Z_i)$ is given by $\hat{S}_0(u \mid Z_i) = \exp\{-\hat{\Lambda}_0(u)\exp(\hat{\beta}_0^T Z_i)\}$. Analogous arguments using only treatment group 1 are used to define the estimator $\hat{S}_1(u \mid Z_i) = \exp\{-\hat{\Lambda}_1(u)\exp(\hat{\beta}_1^T Z_i)\}$ for $S_1(u \mid Z_i)$. Therefore, $\hat{\delta}_1$, the estimator for the average causal treatment effect in restricted lifetime for model 1, is given by equation (4), where $\hat{S}_0(u) = n^{-1}\sum_{i=1}^n \hat{S}_0(u \mid Z_i)$ and $\hat{S}_1(u) = n^{-1}\sum_{i=1}^n \hat{S}_1(u \mid Z_i)$.

For model 2, we again use the maximum partial likelihood estimator for γ and the Breslow estimator for $\Lambda(u)$, where

$$\hat{\Lambda}(u) = \int_0^u \frac{\sum_{i=1}^n dN_i(t)}{\sum_{i=1}^n e^{\hat{\gamma}^T W_i} Y_i(t)}.$$

All the data from both treatment groups are used to estimate the common parameters γ and $\Lambda(u)$. Here again, the estimator for the average causal treatment effect, $\hat{\delta}_2$, is given by equation (4), but $\hat{S}_0(u \mid Z_i)$ is computed as $\exp[-\{\hat{\Lambda}(u)\exp(\hat{\gamma}_1^T Z_i)\}]$ and $\hat{S}_1(u \mid Z_i)$ is computed as $\exp[-\{\hat{\Lambda}(u)\exp(\hat{\gamma}_0 + \hat{\gamma}_1^T Z_i)\}]$.

In order to derive tests for the null hypothesis of no treatment effect, $\delta = 0$, or construct confidence intervals for δ, we need to establish the asymptotic normality of $\hat{\delta}$ and find estimators for the asymptotic variance. It is often the case that estimators are asymptotically linear, i.e.,

$$n^{1/2}(\hat{\delta} - \delta) = n^{-1/2}\sum_{i=1}^n \varphi_i + o_p(1),$$

where φ_i are independent and identically distributed random variables such that $\mathrm{E}(\varphi_i) = 0$ and $\mathrm{E}(\varphi_i^2) < \infty$ and $o_p(1)$ corresponds to a term that converges in probability to zero as n tends to infinity. φ_i is referred to as the ith influence function for $\hat{\delta}$. The asymptotic properties for the estimator can be obtained through the asymptotic properties of the influence functions, namely, an estimator that can be represented as above will be asymptotically normal and the asymptotic variance will be equal to $\mathrm{E}(\varphi_i^2)$. We shall argue that our estimators $\hat{\delta}_1$ and $\hat{\delta}_2$ from models 1 and 2 are indeed asymptotically linear and derive their corresponding influence functions.

Since, for either $\hat{\delta}_1$ or $\hat{\delta}_2$ (generically referred to as $\hat{\delta}$) we can express $\hat{\delta} = \int_0^L \{\hat{S}_1(u) - \hat{S}_0(u)\}du$, we will derive the influence function for $\hat{\delta}$ as follows: We first derive the influence functions for $\hat{S}_0(u)$ and $\hat{S}_1(u)$, which are then used to derive the influence functions for $\int_0^L \hat{S}_0(u)du$ and $\int_0^L \hat{S}_1(u)du$. The influence function for $\hat{\delta}$ is then the difference between the influence function for $\int_0^L \hat{S}_1(u)du$ and that for $\int_0^L \hat{S}_0(u)du$. Recall that the estimator $\hat{S}_0(u) = n^{-1}\sum_{i=1}^n \hat{S}_0(u \mid Z_i)$. Hence,

$$\begin{aligned} n^{1/2}&\int_0^L \{\hat{S}_0(u) - S_0(u)\}du \\ &= n^{-1/2}\sum_{i=1}^n \int_0^L \{\hat{S}_0(u \mid Z_i) - S_0(u \mid Z_i)\}du \quad (5) \\ &\quad + n^{-1/2}\sum_{i=1}^n \int_0^L \{S_0(u \mid Z_i) - S_0(u)\}du. \quad (6) \end{aligned}$$

The arguments given from here on will apply to model 1. The corresponding derivations for model 2 are very similar and hence are omitted. The details for model 2 can be obtained from the authors upon request.

Because $\hat{S}_0(u \mid Z_i)$ is a function of the Breslow estimator for the cumulative hazard function and the maximum partial likelihood estimator for the regression parameters in a proportional hazards model, we can take advantage of the martingale representation of these quantities to approximate (5) by

$$\begin{aligned} n^{-1/2}\sum_{i=1}^n \int_0^L &\left[g_0^T\{Z_i - \mu_0(t,\beta_0)\} - \frac{h_0(t)}{s_0^{(0)}(t,\beta_0)}\right] \\ &\times (1 - A_i)dM_i(t) + o_p(1), \end{aligned} \quad (7)$$

where g_0, $\mu_0(t,\beta_0)$, $h_0(t)$, and $s_0^{(0)}(t,\beta_0)$ are all deterministic quantities defined in the Appendix. Combining the results from equations (5), (6), and (7), we derive the ith influence function of $\int_0^L \hat{S}_0(u)du$ as $\varphi_{0,i}$, where

$$\varphi_{0,i} = \int_0^L \left[g_0^T\{Z_i - \mu_0(t,\beta_0)\} - \frac{h_0(t)}{s_0^{(0)}(t,\beta_0)}\right](1 - A_i)dM_i(t) \quad (8)$$

$$+ \int_0^L \{S_0(u \mid Z_i) - S_0(u)\}du. \quad (9)$$

Similarly, we can identify the ith influence function for $\int_0^L \hat{S}_1(u)du$, $\varphi_{1,i}$ for treatment group 1. Hence, the ith influence function for $\hat{\delta}_1$ is then equal to $\varphi_{1,i} - \varphi_{0,i}$, given by

$$\int_0^L \left[g_1^T\{Z_i - \mu_1(t,\beta_1)\} - \frac{h_1(t)}{s_1^{(0)}(t,\beta_1)}\right] A_i dM_i(t) \quad (10)$$

$$- \int_0^L \left[g_0^T\{Z_i - \mu_0(t,\beta_0)\} - \frac{h_0(t)}{s_0^{(0)}(t,\beta_0)}\right] \times (1 - A_i)dM_i(t) \quad (11)$$

$$+ \int_0^L [\{S_1(u \mid Z_i) - S_0(u \mid Z_i)\} - \{S_1(u) - S_0(u)\}]\,du. \quad (12)$$

The statistics in (10) and (11) are stochastic integrals of a counting process martingale for which standard formulas can be used to find the variance. Also, because (12) is a function of Z_i only, it is $\mathcal{F}(0)$ measurable and hence uncorrelated with (10) and (11). Furthermore, (10) is also uncorrelated with (11) by $A_i(1 - A_i) = 0$. Hence, the asymptotic variance of $\hat{\delta}_1$ is $\mathrm{var}\{(10)\} + \mathrm{var}\{(11)\} + \mathrm{var}\{(12)\}$, which equals

$$\begin{aligned} &(1 - \pi)g_1^T\Sigma_1 g_1 + \int_0^L \frac{h_1^2(t)\lambda_1(t)}{s_1^{(0)}(t,\beta_1)}\,dt \\ &\quad + \pi g_0^T\Sigma_0 g_0 + \int_0^L \frac{h_0^2(t)\lambda_0(t)}{s_0^{(0)}(t,\beta_0)}\,dt \\ &\quad + \mathrm{var}\left[\int_0^L \{S_1(u \mid Z_i) - S_0(u \mid Z_i)\}du\right]. \end{aligned}$$

The detailed derivation for the asymptotic variance of $\hat{\delta}$ and its consistent estimator are given in the Appendix.

In examining the variance of $\hat{\delta}$, we note that var{(12)} is the variation induced by the covariates. This term was not included by Karrison (1987) or Zucker (1998). Consequently, the variance estimates provided previously underestimate the true sampling variance. However, in the case where

$$H_0^*: S_1(u \mid Z) = S_0(u \mid Z) \quad \text{for all } Z, \tag{13}$$

which is referred to as the strong null hypothesis, this last term in the variance vanishes. We refer to (13) as the strong null hypothesis because this assumption implies the weaker null hypothesis,

$$H_0: \delta = 0. \tag{14}$$

However, we wish to emphasize that the converse is not necessarily true and that, when the null hypothesis H_0 is true without (13) holding, then the variance of $\hat{\delta}$ would be underestimated if the variation induced by the covariates is not accounted for. We examine the impact of this source of variation in the simulation studies of the next section.

4. Simulation Studies

Monte Carlo simulations were carried out to assess the performance of our proposed estimators in several cases. We compare the performance of our estimators $\hat{\delta}_1$ and $\hat{\delta}_2$ with the one using the difference of the area under the Kaplan–Meier curves from time 0 to L between the treatment groups, which we denote by $\hat{\delta}_{KM}$. This last estimator clearly does not account for possible treatment imbalance in prognostic factors.

4.1 *Under the Strong Null Hypothesis*

We consider a single confounding variable Z that follows the standard normal distribution. We assume treatment assignment is related to Z through a logistic regression model, namely, $P(A = 1 \mid Z) = e^Z/(1 + e^Z)$. To simulate the strong null hypothesis, we assume, given Z, that T^0 and T^1 are generated from the same distribution with hazard function equal to e^{1+4Z}. The censoring time C is generated from an exponential distribution with hazard .1 and is generated independently of the other variables. Five thousand simulations were used with a sample size of 400. We choose L such that $P(C \geq L) = .3$; hence, $L = 12.04$ in this simulation. The results, given in Table 1, are summarized in terms of the sampling bias, the sampling standard error (SSE), the sampling average of the standard error estimates (SEE), and the coverage probability (CP). Also, in order to assess the impact of not accounting for the variation due to the covariates, we derived SEE* and CP*, which denote, respectively, the sampling average of the standard error estimates and the coverage probability without taking the variation of the covariate into consideration.

Because the data were generated under the strong null hypothesis (13), the true causal treatment effect is $\delta = 0$. Also, for this scenario, the survival data follow the more restrictive assumptions of model 2. From the third column of Table 1, we note that both $\hat{\delta}_1$ and $\hat{\delta}_2$ are unbiased and their SEE is in good agreement with the SSE. However, the SEE of $\hat{\delta}_2$ is about half that of $\hat{\delta}_1$, illustrating the superiority of the second estimator when the assumptions of the more restrictive model 2 are met. As pointed out in the previous section, when the conditional distribution of the survival time given covariates is identical by treatment, then there is no increase in the variance due to the variability in the covariates. This is illustrated by the close agreement of SEE and SEE* as well as CP and CP* being close to the nominal level .95 in this case. The estimator $\hat{\delta}_{KM}$ is severely biased due to confounding. Since the data were generated so that the hazard increased as Z increased and individuals with greater covariate value were more likely to be assigned to treatment group 1, then without adjusting for the confounding variable Z, we would expect treatment group 1 to have smaller average survival times compared with treatment group 0. This, of course, is the bias that is reflected in $\hat{\delta}_{KM}$.

Table 1
Five thousand Monte Carlo simulations of sample size of 400 *under different scenarios;* * *indicates value found without taking the variation of the covariate into consideration*

		Strong null hypothesis $\delta = 0$	Alternative hypothesis 1 $\delta = 3.0662$	Alternative hypothesis 2 $\delta = 1.7128$
$\hat{\delta}_1$	Bias	.0289	.0085	.0050
	SSE	.2297	.2744	.2151
	SEE	.2302	.2780	.2174
	SEE*	.2294	.2416	.1779
	CP	.9470	.9524	.9486
	CP*	.9448	.9216	.8936
$\hat{\delta}_2$	Bias	.0019	.7960	.6560
	SSE	.1124	.2047	.1745
	SEE	.1125	.2276	.1953
	SEE*	.1123	.1598	.1179
	CP	.9520	.0404	.0420
	CP*	.9500	.0076	.0060
$\hat{\delta}_{KM}$	Bias	−3.0417	−3.3024	−6.1366
	SSE	.5114	.5479	.5339
	SEE	.5136	.5475	.5329
	CP	.0000	.0000	.0000

4.2 *Under the Alternative Hypothesis*

For the second simulation, the distribution of T^0 given Z is generated from a proportional hazards model with hazard function equal to e^{1+4Z} and the distribution of T^1 given Z is generated from a different proportional hazards model with hazard function equal to e^{-2+3Z}. In this scenario, the assumptions of model 1 are satisfied but not those of model 2. The censoring time C has hazard .1 and is independent of all other variables. L is also chosen to be 12.04. Again, 5000 simulations were used with a sample size of 400. In this case, the true δ is equal to 3.0662. From the fourth column of Table 1, we can see that $\hat{\delta}_1$ performs well. It is unbiased and the SSE is well approximated by the SEE. In contrast, $\hat{\delta}_2$ has substantial bias. Again, $\hat{\delta}_{KM}$ is severely biased due to confounding. Also, in this scenario, the theory suggests a contribution to the variance of the estimator due to the variation of the covariate Z. This is indeed the case, as is illustrated by the underestimate of SEE* and the coverage probability CP*, which is smaller than the nominal level .95.

To further investigate this issue, we generated the covariate Z from a normal distribution with mean zero and standard

Table 2

Five thousand Monte Carlo simulations of sample size of 400 when there was no confounding

		Bias	SSE	SEE	CP
	Z_1	−.0055	.2807	.2833	.9500
	Z_2	−.0666	.7692	.7762	.9426
	Z_3	−.0294	.5224	.5311	.9536
$\hat{\delta}_1$	Z_1,Z_2	−.0081	.3369	.3378	.9518
	Z_1,Z_3	−.0055	.2822	.2834	.9504
	Z_2,Z_3	−.0660	.7711	.7759	.9414
	Z_1,Z_2,Z_3	−.0076	.3390	.3382	.9518
$\hat{\delta}_{KM}$		−.0207	.5240	.5311	.9538

deviation two, keeping all other variables as before. The results of this scenario are summarized in the last column in Table 1. The true δ equals 1.7128 under this situation. Again, $\hat{\delta}_1$ is an unbiased estimator and performs better than $\hat{\delta}_2$. We note, however, that the underestimation of SSE* and the coverage probability CP* are worse than in the previous case because of the greater variation in the covariates.

4.3 *Simulations When There Is No Confounding*

In order that a variable be a confounder, it must be related to both survival and treatment assignment. In order to assess the performance of our estimator when there is no confounding, we conducted another set of simulation experiments. We considered three covariates, i.e., Z_1, Z_2, and Z_3, generated independently from standard normal distributions. The survival distribution was only dependent on Z_1, namely, we generated T^0 and T^1 from exponential distributions with hazard functions equal to e^{1+4Z_1} and e^{-2+3Z_1}, respectively. Treatment assignment was only related to Z_2 through $\mathrm{P}(A = 1 \mid Z_2) = e^{5Z_2}/(1 + e^{5Z_2})$. Hence, Z_1 was related to survival but was balanced across treatment groups, Z_2 did not affect survival but was related to treatment assignment, and Z_3 was not related to either survival or treatment. Consequently, this is an example where none of the covariates are confounders and the unadjusted estimator of treatment difference $\hat{\delta}_{KM}$ should be unbiased. Since the data were generated under model 1, we only considered the estimator $\hat{\delta}_1$ using regression models that included different combinations of Z_1, Z_2, and Z_3. Five thousand simulations were used with a sample size of 400 and $L = 12.04$. The results are summarized in Table 2.

All the estimators of treatment difference from the various models and $\hat{\delta}_{KM}$ were unbiased and had coverage probabilities near .95. However, we notice that including Z_1, the variable that affects survival, in the model gave us substantial increase in efficiency, whereas including Z_2, the variable that affects treatment assignment but not survival, gave us a decrease in efficiency. Including Z_3, the variable that was not related to either survival or treatment assignment, had no effect on efficiency. It is clear from these results that one should include variables that affect survival in the model. If these variables also affect treatment assignment, then adjusting for such variables is necessary to avoid bias. However, adjusting for prognostic variables will increase efficiency whether they are related to treatment assignment or not. In contrast, variables that do not affect prognosis should not be included in models because they can only increase the variance of the estimator.

Table 3

Estimates of 5-year difference of the restricted mean lifetime between an initial treatment of PCI and MED therapy

	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_{KM}$
Estimate	.1760	.1725	.3621
Standard error	.0377	.0355	.0419

5. Example

We applied our methods to the data from an observational study of acute coronary syndrome patients from Duke University Medical Center to estimate the difference of the restricted mean lifetime over 5 years between an initial treatment of percutaneous coronary intervention (PCI) versus medical therapy (MED). Among the 6033 patients in this study, 3786 patients have been followed for 5+ years or died prior to the end of 1998. The rest have censored survival times.

We first identified some important prognostic factors, which include age, race, history of smoking, hypertension, diabetes, congestive heart failure, number of previous myocardial infarctions, maximum percent stenosis of main artery, and ejection fraction, by using Cox's proportional hazard model. Using both models 1 and 2, described in Section 3, we derived the estimator for the treatment difference in mean lifetime restricted to 5 years, $\hat{\delta}_1$, $\hat{\delta}_2$, respectively, and the standard error of these estimators. These are contrasted with the unadjusted estimator based on the difference in the area under the Kaplan–Meier curves up to 5 years for the two treatments.

The results of our analysis are given in Table 3. The estimates $\hat{\delta}_1$ and $\hat{\delta}_2$ are close to each other, with $\hat{\delta}_2$ having slightly smaller standard error. Both estimators indicate that the restricted mean lifetime for the PCI group is significantly higher than the MED group by about 0.17 years. The estimator based on the difference of the area under the Kaplan–Meier curves, $\hat{\delta}_{KM}$, is 0.3621 years, which is substantially different than $\hat{\delta}_1$ and $\hat{\delta}_2$.

Figure 1 shows the estimators for the survival function using the Kaplan–Meier estimator and our estimators $\hat{S}_1(u)$ and $\hat{S}_0(u)$ under the two model assumptions for both treatment groups. We note that the survival curves for our estimators under the two model assumptions are very close to each other for each treatment group. Model diagnostics, not presented here, confirmed the adequacy of model 2, which explains the agreement of the results from the two models. The Kaplan–Meier estimator, which is unadjusted for covariates, has a substantially larger difference than our estimators. A careful examination of the distribution of covariates by treatment suggests that patients assigned medication are prognostically worse on average. Thus, we would expect that adjusting for prognostic factors would result in a smaller treatment difference.

Based on this study population, we draw the conclusion that the restricted mean lifetime up to 5 years for PCI patients is estimated to be .1725 years higher than patients in the MED group, with a standard error of .0355 years.

6. Discussion

In this article, we presented two estimators for the average causal treatment difference in restricted lifetime that can be used in observational studies with confounding variables. We

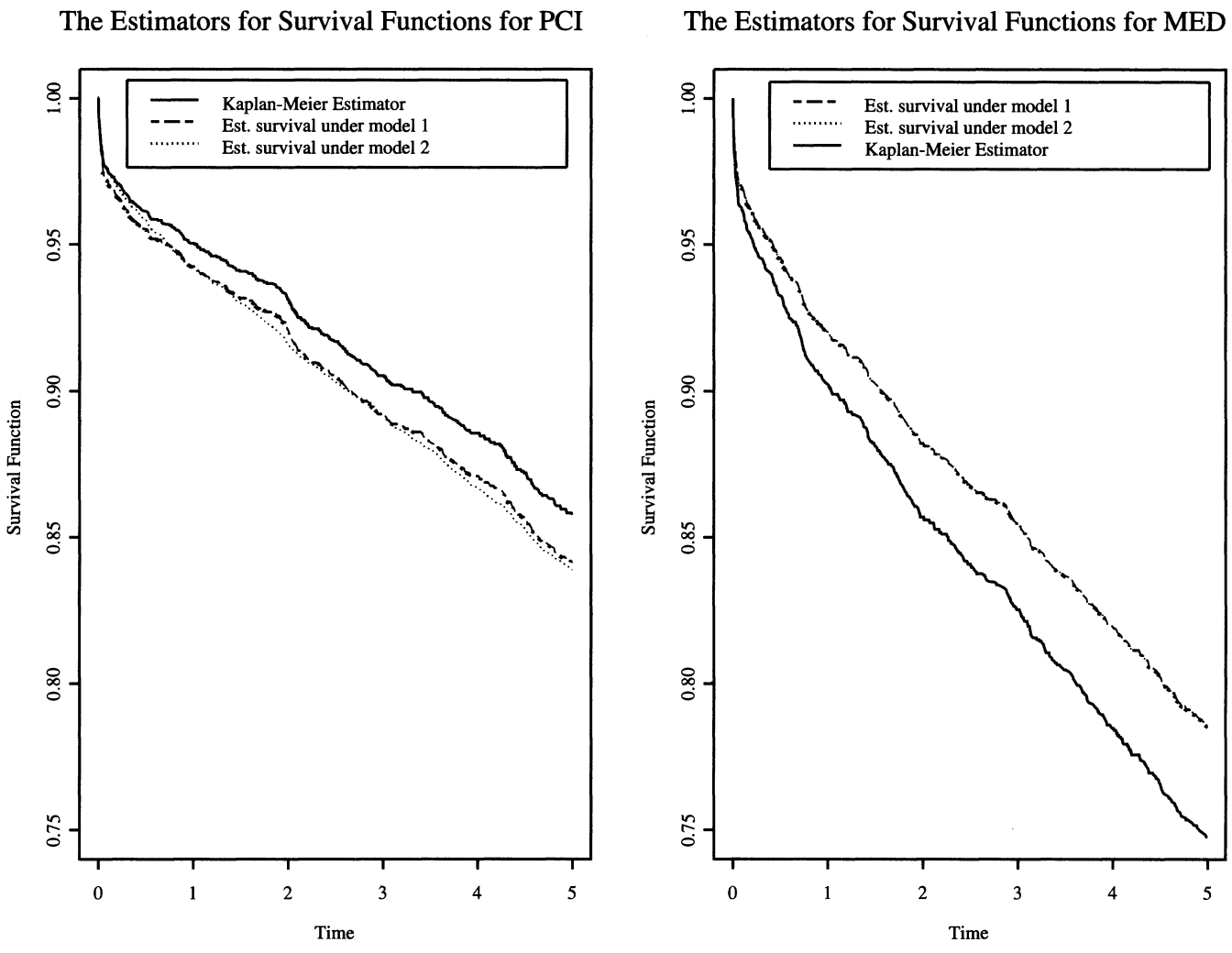

Figure 1. The estimators for survival functions for both treatment groups. The estimators for MED group under model 1 and model 2 assumptions are very close to each other.

also derived the large-sample properties of these estimators and a formula for standard error that properly accounts for all sources of variation, including that induced through the covariates.

An interesting observation made by one of the referees is that, when the strong null hypothesis (13) is true, the variance induced through the covariates disappears (i.e., var$\{(12)\}$ = 0). Therefore, a Wald test of the strong null hypothesis that used an estimated variance without including this term would still have the correct level and would always be larger and hence more powerful than the corresponding Wald test that included an estimate for var$\{(12)\}$ as part of the variance estimate. However, this would not be true for the more general null hypothesis (14), where not accounting for var$\{(12)\}$ may result in a test with the incorrect level. Moreover, if interest were in estimating the average causal treatment effect δ, then not accounting for var$\{(12)\}$ would generally lead to confidence intervals that are too narrow.

Our simulation studies suggest that the large-sample properties are well approximated if the modeling assumptions are correct but can lead to bias when the assumptions are violated. Our studies also show that the variance of the estimator can be reduced substantially by not including extraneous parameters and variables that do not affect the survival distribution. We speculate that building models hierarchically and choosing the one with the fewest parameters that meets some reasonable information criterion of fit may be a reasonable strategy. This will be an issue of future study.

Although we presented two different models, the strategy outlined for finding estimators and their standard errors could be used with a variety of proportional hazards models. This can include models with interaction terms of treatment and covariates as well as interaction of functions of time and covariates and treatment. Such models may be part of the model-building process.

Acknowledgements

This research was supported by grant AI-31789 from the National Institute of Allergy and Infectious Disease and grant

CA-51962 from the National Cancer Institute. We are also very grateful for the helpful comments made by the associate editor and the referees.

Résumé

Pour comparer les durées de survie de deux groupes de traitement, en prenant en compte les censures, il peut être plus approprié de comparer les durées de vie moyennes restreintes, c'est à dire l'espérance des durées de vie restreintes à un délai L, plutôt que les durées de vie moyenne. Quand les traitements ne sont pas attribués par tirage au sort, comme dans les petites enquêtes d'observation par exemple, il faut aussi prendre en compte les déséquilibres de distribution des facteurs de confusion entre traitements. Dans cet article nous proposons des estimateurs de la différence des moyennes restreintes entre les deux groupes, ajustées sur des facteurs pronostiques déséquilibrés, sous les hypothèses d'un modèle de taux proportionnels. Les propriétés de notre estimateur, pour des échantillons de grande taille, sont aussi établies à partir de la théorie des martingales appliquées aux processus de comptage. Des études de simulation ont été réalisées pour comparer ces estimateurs et pour établir l'adéquation des approximations pour les grands échantillons. Notre méthode a aussi été appliquée à une série de données provenant du Duke University Medical Center, concernant des malades atteints de syndrome coronaire aigu, pour estimer un effet traitement sur des moyennes de survie restreintes à 5 ans.

References

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

Breslow, N. E. (1972). Contribution to the discussion on the paper by D. R. Cox, regression and life tables. *Journal of the Royal Statistical Society, Series B* **34,** 216–217.

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

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

Hubbard, A. E., van der Laan, M. J., and Robins, J. M. (1999). Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies. In *Statistical Models in Epidemiology, the Environment and Clinical Trials*, IMA Volumes in Mathematics and Its Applications, Vol. 116. M. E. Halloran and D. Berry (eds), 135–178. New York: Springer Verlag.

Karrison, T. (1987). Restricted mean life with adjustment for covariates. *Journal of the American Statistical Association* **82,** 1169–1176.

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

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

Shen, Y. and Fleming, T. R. (1997). Large sample properties of some survival estimators in heterogeneous samples. *Journal of Statistical Planning and Inference* **60,** 123–138.

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.

Received January 2001. *Revised June* 2001. *Accepted June* 2001.

Appendix

Finding the Influence Function for $\hat{\delta}_1$ and Its Asymptotic Variance

The following are some basic definitions:

$$S_0^{(k)}(t,\beta_0) = n^{-1}\sum_{i=1}^{n}(1-A_i)Z_i^{\otimes k}e^{\beta_0^T Z_i}Y_i(t), \qquad k=0,1,2,$$

where $a^{\otimes 0}=1$, $a^{\otimes 1}=a$, and $a^{\otimes 2}=aa^T$,

$$s_0^{(k)}(t,\beta_0) = \mathrm{E}\left\{(1-A)Z^{\otimes k}e^{\beta_0^T Z}Y(t)\right\},$$

$$\bar{Z}_0(t,\beta_0) = \frac{S_0^{(1)}(t,\beta_0)}{S_0^{(0)}(t,\beta_0)},$$

$$\mu_0(t,\beta_0) = \frac{s_0^{(1)}(t,\beta_0)}{s_0^{(0)}(t,\beta_0)}.$$

First, we identify the influence function for $\hat{S}_0(u)$ by noting that

$$\begin{aligned} n^{1/2}\{\hat{S}_0(u)-S_0(u)\} &= n^{-1/2}\sum_{i=1}^{n}\{\hat{S}_0(u\mid Z_i)-S_0(u\mid Z_i)\} \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\{S_0(u\mid Z_i)-S_0(u)\}. \end{aligned}$$

Applying the asymptotic results of Andersen and Gill (1982) for Cox's model, we get

$$\begin{aligned} n^{1/2}\{\hat{S}_0(u)-S_0(u)\} &= n^{-1/2}\sum_{i=1}^{n}b_0^T(u)\int_0^L\{Z_i-\mu_0(t,\beta_0)\}(1-A_i)dM_i(t) \\ &\quad - n^{-1/2}\sum_{i=1}^{n}c_0(u)\int_0^u {s_0^{(0)}(t,\beta_0)}^{-1}(1-A_i)dM_i(t) \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\{S_0(u\mid Z_i)-S_0(u)\}+o_p(1), \end{aligned}$$

where

$$\pi = \mathrm{P}(A=0),$$

$$c_0(u) = \mathrm{E}\left\{S_0(u\mid Z)e^{\beta_0^T Z}\right\},$$

$$\begin{aligned} b_0(u) = (\pi\Sigma_0)^{-1}\int_0^u \Big[\mu_0(t,\beta_0)\mathrm{E}\left\{S_0(u\mid Z)e^{\beta_0^T Z}\right\} \\ - \mathrm{E}\left\{ZS_0(u\mid Z)e^{\beta_0^T Z}\right\}\Big]\lambda_0(t)dt, \end{aligned}$$

and Σ_0 is the information matrix for a single observation in treatment group 0.

Integrating the above quantity from zero to L yields

$$\begin{aligned} n^{1/2}&\int_0^L \{\hat{S}_0(u) - S_0(u)\}du \\ &= n^{-1/2}\sum_{i=1}^n g_0^T \int_0^L \{Z_i - \mu_0(t,\beta_0)\}(1-A_i)dM_i(t) \\ &\quad - n^{-1/2}\sum_{i=1}^n \int_0^L \frac{h_0(t)}{s_0^{(0)}(t,\beta_0)}(1-A_i)dM_i(t) \\ &\quad + n^{-1/2}\sum_{i=1}^n \int_0^L \{S_0(u \mid Z_i) - S_0(u)\}du + o_p(1), \end{aligned}$$

where $g_0 = \int_0^L b_0(u)du$ and $h_0(t) = \int_t^L c_0(u)du$.

Hence, the ith influence function for $\int_0^L \hat{S}_0(u)du$ is

$$\varphi_{0,i} = g_0^T \int_0^L \{Z_i - \mu_0(t,\beta_0)\}(1-A_i)dM_i(t) \quad \text{(A.1)}$$

$$- \int_0^L \frac{h_0(t)}{s_0^{(0)}(t,\beta_0)}(1-A_i)dM_i(t) \quad \text{(A.2)}$$

$$+ \int_0^L \{S_0(u \mid Z_i) - S_0(u)\}du. \quad \text{(A.3)}$$

Since Z_i is $\mathcal{F}(0)$ measurable, the martingale stochastic integrals (A.1) and (A.2) are uncorrelated with (A.3). Furthermore,

$$\begin{aligned} &\operatorname{cov}\{(\text{A.1}),(\text{A.2})\} \\ &\quad = g_0^T \mathrm{E}\left[\int_0^L \frac{h_0(t)\{Z_i - \mu_0(t,\beta_0)\}}{s_0^{(0)}(t,\beta_0)} \right. \\ &\qquad \left. \times (1-A_i)\lambda_0(t)e^{\beta_0^T Z_i}Y_i(t)dt\right] = 0 \end{aligned}$$

since $\mathrm{E}[\{Z_i - \mu_0(t,\beta_0)\}(1-A_i)e^{\beta_0^T Z_i}Y_i(t)] = 0$.

By applying standard formulas for the variance of stochastic integrals of martingale processes, the variance for $\varphi_{0,i}$ can be shown to equal

$$\begin{aligned} \operatorname{var}(\varphi_{0,i}) &= \operatorname{var}\{(\text{A.1})\} + \operatorname{var}\{(\text{A.2})\} + \operatorname{var}\{(\text{A.3})\} \\ &= \pi g_0^T \Sigma_0 g_0 + \int_0^L \frac{h_0^2(t)\lambda_0(t)}{s_0^{(0)}(t,\beta_0)}dt \\ &\quad + \operatorname{var}\left\{\int_0^L S_0(u \mid Z_i)du\right\}. \end{aligned}$$

A consistent estimator for the asymptotic variance is obtained naturally by

$$\begin{aligned} \widehat{\operatorname{var}}(\varphi_{0,i}) &= \frac{n_0}{n}\hat{g}_0^T\hat{\Sigma}_0\hat{g}_0 + \int_0^L \frac{\hat{h}_0^2(t)}{S_0^{(0)}(t,\hat{\beta}_0)}d\hat{\Lambda}_0(t) \\ &\quad + \widehat{\operatorname{var}}\{\int_0^L S_0(u \mid Z_i)du\}, \end{aligned}$$

where

$$n_0 = \sum_{i=1}^n I(A_i = 0),$$

$$\begin{aligned} \hat{\Sigma}_0 &= {n_0}^{-1}\sum_{i=1}^n \int_0^L (1-A_i) \\ &\quad \times \left[\frac{S_0^{(2)}(t,\hat{\beta}_0)}{S_0^{(0)}(t,\hat{\beta}_0)} - \left\{\frac{S_0^{(1)}(t,\hat{\beta}_0)}{S_0^{(0)}(t,\hat{\beta}_0)}\right\}^{\otimes 2}\right] \\ &\quad \times dN_i(t), \end{aligned}$$

$$\begin{aligned} \hat{g}_0 &= {n_0}^{-1}\hat{\Sigma}_0^{-1}\sum_{i=1}^n \int_0^L \int_0^u \hat{S}_0(u \mid Z_i)e^{\hat{\beta}_0^T Z_i} \\ &\quad \times \left\{\bar{Z}_0(t,\hat{\beta}_0) - Z_i\right\} d\hat{\Lambda}_0(t)\,du, \end{aligned}$$

$$\hat{h}_0(t) = n^{-1}\sum_{i=1}^n \int_t^L \hat{S}_0(u \mid Z_i)e^{\hat{\beta}_0^T Z_i}du,$$

and $\widehat{\operatorname{var}}\{\int_0^L S_0(u \mid Z_i)du\}$ is obtained using the moment estimator

$$n^{-1}\sum_{i=1}^n \left\{\int_0^L \hat{S}_0(u \mid Z_i)du - n^{-1}\sum_{j=1}^n \int_0^L \hat{S}_0(u \mid Z_j)du\right\}^2.$$

By analogy, we will have a similar result for the ith influence function $\varphi_{1,i}$ for $\int_0^L \hat{S}_1(u)du$ and its asymptotic variance for the treatment 1 group. The ith influence function for $\hat{\delta}_1$ is then equal to

$$\begin{aligned} &\varphi_{1,i} - \varphi_{0,i} \\ &\quad = \int_0^L \left[g_1^T\{Z_i - \mu_1(t,\beta_1)\} - \frac{h_1(t)}{s_1^{(0)}(t,\beta_1)}\right] A_i dM_i(t) \quad \text{(A.4)} \\ &\qquad - \int_0^L \left[g_0^T\{Z_i - \mu_0(t,\beta_0)\} - \frac{h_0(t)}{s_0^{(0)}(t,\beta_0)}\right] \\ &\qquad\quad \times (1-A_i)dM_i(t) \quad \text{(A.5)} \\ &\qquad + \int_0^L \left[\{S_1(u \mid Z_i) - S_0(u \mid Z_i)\}\right. \\ &\qquad\quad \left. - \{S_1(u) - S_0(u)\}\right]du. \quad \text{(A.6)} \end{aligned}$$

Since Z_i is $\mathcal{F}(0)$ measurable, (A.6) is uncorrelated with (A.4) and (A.5). Also, (A.4) is uncorrelated with (A.5) by $A_i(1-A_i) = 0$. Then the asymptotic variance of $\hat{\delta}_1$ is given by

$$\begin{aligned} &(1-\pi)g_1^T\Sigma_1 g_1 + \int_0^L \frac{h_1^2(t)\lambda_1(t)}{s_1^{(0)}(t,\beta_1)}dt \\ &\quad + \pi g_0^T\Sigma_0 g_0 + \int_0^L \frac{h_0^2(t)\lambda_0(t)}{s_0^{(0)}(t,\beta_0)}dt \\ &\quad + \operatorname{var}\left[\int_0^L \{S_1(u \mid Z_i) - S_0(u \mid Z_i)\}du\right]. \end{aligned}$$

All the terms above can be estimated consistently using the obvious sample analogs.