Published in final edited form as:

Stat Med. 2018 February 10; 37(3): 390–404. doi:10.1002/sim.7513.

Covariate adjustment using propensity scores for dependent censoring problems in the accelerated failure time model

Youngjoo Cho¹, Chen Hu², and Debashis Ghosh^{*,3}

¹Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY 14642

²Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21205 and NRG Oncology, Statistics and Data Management Center, Philadelphia, PA 19103

³Department of Biostatistics and Informatics, University of Colorado, Aurora, CO 80045

Abstract

In many medical studies, estimation of the association between treatment and outcome of interest is often of primary scientific interest. Standard methods for its evaluation in survival analysis typically require the assumption of independent censoring. This assumption might be invalid in many medical studies, where the presence of dependent censoring leads to difficulties in analyzing covariate effects on disease outcomes. This data structure is called ‘semicompeting risks data,’ for which many authors have proposed an artificial censoring technique. However, confounders with large variability may lead to excessive artificial censoring, which subsequently results in numerically unstable estimation. In this paper, we propose a strategy for weighted estimation of the associations in the accelerated failure time model. Weights are based on propensity score modeling of the treatment conditional on confounder variables. This novel application of propensity scores avoids excess artificial censoring caused by the confounders and simplifies computation. Monte Carlo simulation studies and application to AIDS and cancer research are used to illustrate the methodology.

Keywords

Informative Censoring; Observational study; Perturbation; Resampling

1 Introduction

In many chronic disease clinical studies, the so-called “semi-competing risks” data [1] arise frequently. For example, in cancer studies, disease progression or cancer recurrence may be censored by death without progression; in HIV studies, virologic failure may be censored by dropouts that are often dependent on HIV viral load changes.

*Corresponding author, debashis.ghosh@ucdenver.edu.

Much research has been conducted on semicompeting risks data, a nice summary of which is given in [2]. As the name suggests, the time to the event of interest may be dependently censored but not vice versa. One of the approaches to estimation is based on an artificial censoring technique proposed by Lin et al. [3] and Peng and Fine [4] to handle dependent censoring. The artificial censoring is a tool to adjust for the discrepancy of censoring between treatment group and control group and often used in conjunction with accelerated failure time (AFT) regression model. The artificial censoring technique may censor originally uncensored observations to enable unbiased comparison of the distribution of the time to the event of interest [5]. This can be easily extended to accommodate an arbitrary number of covariates. One advantage of artificial censoring lies in the fact that regression parameters can be estimated based on modified estimating equations without parametric assumptions between event types. However, when a large number of covariates or continuous covariates are present, based on the expressions of the artificial censoring in [3] and [4], excessive artificial censoring may lead to unstable estimation and incorrect inference results. For example, estimators could have seemingly small but potentially underestimated variability. It seems that having small variability of the estimators may be desirable, but it implies that the variability of the estimators are not properly estimated because of the excessive artificial censoring. As we show in the simulation studies at Section 6, excessive artificial censoring can lead to incorrect coverage.

This artificial censoring issue can be especially problematic in non-randomized observational studies, where the evaluation of the association between the treatment and the time to primary event lies on properly adjusting for potential confounders. Similar issues arise in subgroup analysis from randomized studies. For example, the Radiation Therapy Oncology Group (RTOG) conducted a multicenter, randomized phase III trial to evaluate the optimal treatment schedule involving a combination of androgen suppression (hormone therapy) and radiotherapy (RT) for clinically localized prostate cancer patients [6]. After the trial was closed, there was emerging evidence suggesting that a subset of study participants, the so-called “intermediate risk” patients, may have benefited the most from the experimental therapy, a more intensified RT regimen by irradiating the whole pelvic (WPRT) instead of the prostate only (PORT). These patients are considered intermediate risk because their disease prognosis in terms of cancer recurrence is not viewed as favorably as low-risk patients but is better than that of high-risk patients. It was hypothesized that WPRT is more potent to eliminate the residual microscopic disease and the increased risk of lymph node involvement. There are, however, at least two main challenges to properly assess the association between treatment and disease progression. First, with a median age of 70, the disease progression may be dependently censored by death. Second, this patient group encompasses a broad range of men with rather heterogeneous clinicopathological characteristics and outcomes, which makes it important to properly adjust for these potential confounders, such as prostate specific antigen (PSA) level, Gleason score (GS), tumor size, tumor stage and age.

Motivated by the RTOG 9413 dataset, our interest is to estimate the association of treatment and time to nonterminal event by adjusting for confounders. Our approach to the problem involves propensity scores, proposed by Rosenbaum and Rubin [7]. While their use has been of interest in causal inference, they also satisfy a balancing property that corresponds to the

distribution of confounders being equal for both treatment and control groups given the propensity score [8]. Thus, the propensity score can provide sufficient information to balance the confounders between treatment group and control group. This can be an important tool for reducing the artificial censoring needed for estimating the association in an asymptotically unbiased manner, as illustrated in numerical examples (Section 5 and 6).

In this paper, we propose methodology for estimation of the association of treatment and time to nonterminal event in observational studies adjusting for confounders using the propensity score. The goal is to improve estimation of the association between treatment and time to nonterminal event. The paper is organized as follows. In Section 2, we introduce the data structure and modeling assumptions, including a brief review about propensity scores. Section 3 reviews the methodology for estimation of regression parameters with semi-competing risks data. Theoretical results regarding estimation and inference using the proposed method are demonstrated in Section 4. In Section 5, we apply the proposed methodology to the RTOG prostate cancer trial as well as a HIV study previously analyzed in the literature. Simulation studies are shown in Section 6. Section 7 concludes with remarks and discussion.

2 Preliminaries

2.1 Data and Model

Let X be time to the event of interest, D time to the dependent censoring, and C time to independent censoring. Define $I(A)$ to be the indicator function for the event A , and let $a \wedge b$ be the minimum of a and b . Define $\mathbf{W} \equiv (Z, \mathbf{V}^T)^T$ to be a vector of k variables, where $\mathbf{V}$ represents the confounder variables and Z is a binary treatment variable. Define

$$\tilde{X} = X \wedge D \wedge C, \quad \tilde{D} = D \wedge C \quad \delta = I(X \leq \tilde{D}), \quad \xi = I(D \leq C).$$

We assume all times are log-transformed. The data consist of n independent observations $(\tilde{X}_i, \tilde{D}_i, \mathbf{W}_i, \delta_i, \xi_i)$, $i = 1, \dots, n$. The model is given by

$$\begin{pmatrix} X_i = \boldsymbol{\theta}_0^T \mathbf{W}_i + \varepsilon_i^X \\ D_i = \boldsymbol{\eta}_0^T \mathbf{W}_i + \varepsilon_i^D \end{pmatrix} \quad i = 1, \dots, n, \quad (1)$$

where $\boldsymbol{\beta}_0 = (\boldsymbol{\theta}_0^T, \boldsymbol{\eta}_0^T)^T$ is a $2k \times 1$ vector and $\varepsilon_i \equiv (\varepsilon_i^X, \varepsilon_i^D)$ are error terms with unknown common bivariate distribution F . Let $\boldsymbol{\theta}_0 \equiv \left\{ \theta_0^{tr}, (\boldsymbol{\theta}_0^{cf})^T \right\}^T$ and $\boldsymbol{\eta}_0 \equiv \left\{ \eta_0^{tr}, (\boldsymbol{\eta}_0^{cf})^T \right\}^T$, where θ_0^{tr} and η_0^{tr} are the subcomponents $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ corresponding to Z . Similarly, $\boldsymbol{\theta}_0^{cf}$ and $\boldsymbol{\eta}_0^{cf}$ are the components of $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ corresponding to $\mathbf{V}$. We can rewrite (1) as

$$\begin{pmatrix} X_i = \theta_0^{tr} Z_i + (\boldsymbol{\theta}_0^{efd})^T \mathbf{V}_i + \varepsilon_i^X \\ D_i = \eta_0^{tr} Z_i + (\boldsymbol{\eta}_0^{efd})^T \mathbf{V}_i + \varepsilon_i^D \end{pmatrix}, \quad i=1, \dots, n,$$

We formulate the model only in the upper wedge where $X < D$ and C is independent with (X, D) given $\mathbf{W}$, but X and D can be dependent given $\mathbf{W}$ [1,4].

2.2 Propensity score

In the causal inference literature, the main interest is to establish, identify and estimate causal effects of interventions on outcomes. Suppose that we have a binary treatment variable Z and confounders $\mathbf{V}$. In randomized studies, randomization ensures the possibility of a causal interpretation for treatment effects. In observational studies, the existence of the confounders prevents straightforward estimation of causal effects. Rosenbaum and Rubin [7] proposed the propensity score to estimate the causal effect of the treatment for the observational studies. The propensity score is defined as the probability of being in the treatment group given the confounders, given by

$$e^*(\mathbf{V}) \equiv P(Z=1|\mathbf{V}).$$

The propensity score has many desirable properties. It balances the confounders between the treatment and the control groups. With this balancing property, we can obtain conditional independence between the treatment and the confounders.

While the propensity score has been motivated in the literature on causal inference, we note that in this paper, we do not attempt to infer causality. We address the causality issue in the Discussion section. It is the balancing property of propensity scores that we will use in this paper in order to adjust for confounders. Based on model (1), ε_j are independent of Z_j and identically distributed. In our case, based on the balancing property of propensity score, we can derive the following important result.

Theorem 1—Given the propensity score $e^*(\mathbf{V}_i)$, $(X_i - \theta_0^{tr} Z_i, D_i - \eta_0^{tr} Z_i)$ are independent of Z_i and identically distributed.

The proof of this theorem can be found in Section A of the Supplementary Materials. This theorem implies that given the propensity score, we can establish i.i.d terms with omitted confounders. This theorem enables us to estimate the parameters for treatment variable in an asymptotically unbiased manner.

3 Proposed Methodology

As mentioned in the Introduction, artificial censoring techniques have been important tools for obtaining asymptotically unbiased estimators of regression coefficients for the time to event of interest in the presence of dependent censoring. Using artificial censoring

techniques, Lin et al. [3] and Peng and Fine [4] proposed two different estimating functions. Lin et al. [3] propose the estimating function

$$U_n^{L,full}(\beta) = n^{-1/2} \sum_{i=1}^n \tilde{\delta}_i^{*,full}(\beta) \left[\mathbf{W}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_i^{*,full}(\beta) \geq \tilde{X}_i^{*,full}(\beta) \mathbf{W}_j\}}{\sum_{j=1}^n I\{\tilde{X}_i^{*,full}(\beta) \geq \tilde{X}_i^{*,full}(\beta)\}} \right],$$

where

$$d^{full}(\beta) = \max_{1 \leq i \leq n} \{0, (\boldsymbol{\theta} - \boldsymbol{\eta})^T \mathbf{W}_i\}$$

$$\tilde{X}_i^{*,full}(\beta) = (X_i - \boldsymbol{\theta}^T \mathbf{W}_i) \wedge \{\tilde{D}_i - \boldsymbol{\eta}^T \mathbf{W}_i - d^{full}(\beta)\}$$

$$\tilde{\delta}_i^{*,full}(\beta) = I\left[(X_i - \boldsymbol{\theta}^T \mathbf{W}_i) \leq \{\tilde{D}_i - \boldsymbol{\eta}^T \mathbf{W}_i - d^{full}(\beta)\}\right].$$

The estimating function by Peng and Fine [4] is

$$U_n^{P,full}(\beta) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} (\mathbf{W}_i - \mathbf{W}_j) \phi_{ij}^{full}(\beta),$$

where

$$d_{ij}^{full}(\beta) = \max\{0, (\boldsymbol{\theta} - \boldsymbol{\eta})^T \mathbf{W}_i, (\boldsymbol{\theta} - \boldsymbol{\eta})^T \mathbf{W}_j\}$$

$$\tilde{X}_{i(j)}^{*,full}(\beta) = (X_i - \boldsymbol{\theta}^T \mathbf{W}_i) \wedge \{\tilde{D}_i - \boldsymbol{\eta}^T \mathbf{W}_i - d_{ij}^{full}(\beta)\}$$

$$\tilde{\delta}_{i(j)}^{*,full}(\beta) = I\left[(X_i - \boldsymbol{\theta}^T \mathbf{W}_i) \leq \{\tilde{D}_i - \boldsymbol{\eta}^T \mathbf{W}_i - d_{ij}^{full}(\beta)\}\right]$$

$$\phi_{ij}^{full}(\beta) = \tilde{\delta}_{i(j)}^{*,full}(\beta) I\left\{\tilde{X}_{i(j)}^{*,full}(\beta) \leq \tilde{X}_{j(i)}^{*,full}(\beta)\right\} - \tilde{\delta}_{j(i)}^{*,full}(\beta) I\left\{\tilde{X}_{j(i)}^{*,full}(\beta) \leq \tilde{X}_{i(j)}^{*,full}(\beta)\right\}.$$

As can be seen the previous estimating functions, Lin et al. [3] used a single comparison of each residual time and Peng and Fine [4] compared different pairs of residual times for the artificial censoring. In [3], the same degree of the artificial censoring is applied to every residual time so that the estimator from this approach may be inefficient when many

confounders are included in the model. The method of [4] is clearly better than that in [3] in the sense that their artificial censoring is smaller than that in [3], but including continuous confounders with large variability may still cause excessive artificial censoring in their method.

We now propose an estimation procedure to avoid excessive artificial censoring by using the propensity scores. In this case, our goal is to estimate the parameter associated with binary exposure on the nonterminal event. The key idea is to only use the treatment variable in a weighted estimation procedure, where the weights adjust for confounding. Note that without the weights, the estimation procedure will yield a biased estimator. We assume that the propensity score is modeled by logistic regression with parameter vector $\alpha = (\alpha_1, \dots, \alpha_k)$.

Let α_0 be the true value of α and $H_i = (1, V_i^T)^T$, $i = 1, \dots, n$. The estimator of α_0 can be obtained by solving $G_n(\alpha)$, where

$$G_n(\alpha) = n^{-1/2} \sum_{i=1}^n H_i \left\{ Z_i - \frac{\exp(\alpha^T H_i)}{1 + \exp(\alpha^T H_i)} \right\} \quad (2)$$

Define the propensity score model to be

$$e_i(\alpha) = P(Z_i = 1 | H_i) = \left\{ \frac{\exp(\alpha^T H_i)}{1 + \exp(\alpha^T H_i)} \right\}.$$

The true value of propensity score is expressed by $e_i(\alpha_0)$ for each subject i , $i = 1, \dots, n$. The weight is defined by

$$w_i(\alpha) = \frac{Z_i}{e_i(\alpha)} = \frac{1 - Z_i}{1 - e_i(\alpha)}. \quad (3)$$

This weight takes value $1/e_i(\alpha)$ if $Z_i = 1$ and $1/(1 - e_i(\alpha))$ otherwise. In the causal inference literature, the typical technique is to apply the weight $w_i(\alpha)$ to estimate the average treatment effect [8–11]. We will use this weight in our proposed estimating procedure. Let η^{tr} and θ^{tr} be the association parameters for Z with respect to D and X , respectively. By using the weight, the proposed estimating function for η^{tr} is

$$S_n(\eta^{tr}, \alpha) = n^{-1/2} \sum_{i=1}^n \xi_i w_i(\alpha) \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\alpha) Z_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\alpha)} \right], \quad (4)$$

where $\tilde{D}_i^*(\eta^{tr}) = \tilde{D}_i - (\eta^{tr}) Z_i$. Let $\beta^{tr} = (\eta^{tr}, \theta^{tr})^T$. The proposed estimating function for β^{tr} , a modification of [3], is

$$U_n^L(\beta^{tr}, \alpha) = n^{-1/2} \sum_{i=1}^n \tilde{\delta}_i^*(\beta^{tr}) w_i(\alpha) \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta^{tr}) \geq \tilde{X}_i^*(\beta^{tr})\} w_j(\alpha) Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta^{tr}) \geq \tilde{X}_i^*(\beta^{tr})\} w_j(\alpha)} \right],$$

(5)

where

$$d(\beta^{tr}) = \max_{1 \leq i \leq n} \{0, (\theta^{tr} - \eta^{tr}) Z_i\}$$

$$\tilde{X}_i^*(\beta^{tr}) = (X_i - \theta^{tr} Z_i) \wedge \{\tilde{D}_i - \eta^{tr} Z_i - d(\beta^{tr})\}$$

$$\tilde{\delta}_i^*(\beta^{tr}) = I[(X_i - \theta^{tr} Z_i) \leq \{\tilde{D}_i - \eta^{tr} Z_i - d(\beta^{tr})\}].$$

Similarly, the proposed estimating function based on the work of [4] is

$$U_n^P(\beta^{tr}, \alpha) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} (Z_i - Z_j) w_i(\alpha) w_j(\alpha) \phi_{ij}(\beta^{tr}), \quad (6)$$

where

$$d_{ij}(\beta^{tr}) = \max \{0, (\theta^{tr} - \eta^{tr}) Z_i, (\theta^{tr} - \eta^{tr}) Z_j\}$$

$$\tilde{X}_{i(j)}^*(\beta^{tr}) = (X_i - \theta^{tr} Z_i) \wedge \{\tilde{D}_i - \eta^{tr} Z_i - d_{ij}(\beta^{tr})\}$$

$$\tilde{\delta}_{i(j)}^*(\beta^{tr}) = I[(X_i - \theta^{tr} Z_i) \leq \{\tilde{D}_i - \eta^{tr} Z_i - d_{ij}(\beta^{tr})\}]$$

$$\phi_{ij}(\beta^{tr}) = \tilde{\delta}_{i(j)}^*(\beta^{tr}) I\{\tilde{X}_{i(j)}^*(\beta^{tr}) \leq \tilde{X}_{j(i)}^*(\beta^{tr})\} - \tilde{\delta}_{j(i)}^*(\beta^{tr}) I\{\tilde{X}_{j(i)}^*(\beta^{tr}) \leq \tilde{X}_{i(j)}^*(\beta^{tr})\}.$$

Let $\gamma = (\alpha^T, \eta^{tr}, \theta^{tr})^T$, $\gamma_0 = (\alpha_0^T, \eta_0^{tr}, \theta_0^{tr})^T$ and $\beta_0^{tr} = (\eta_0^{tr}, \theta_0^{tr})^T$. In this case, γ_0 is a parameter that satisfies $E\{G_n(\alpha_0)\} = 0$, $E\{S_n(\eta_0^{tr}, \alpha_0)\} = 0$, $E\{U_n^L(\beta_0^{tr}, \alpha_0)\} = 0$ and

$E \left\{ U_n^P(\beta_0^{tr}, \alpha_0) \right\} = 0$. Now we propose two unbiased estimating functions for associate parameters of binary treatment from [3] and [4]. We solve

$$Q_n^L(\gamma) = \left\{ G_n^T(\alpha), S_n(\eta^{tr}, \alpha), U_n^L(\beta^{tr}, \alpha) \right\}^T = 0$$

$$Q_n^P(\gamma) = \left\{ G_n^T(\alpha), S_n(\eta^{tr}, \alpha), U_n^P(\beta^{tr}, \alpha) \right\}^T = 0 \quad (7)$$

to obtain estimators of γ_0 . Solutions for γ can be obtained by solving the estimating equations sequentially. First we estimate the propensity scores for all observations, denoted as $\{e_i(\hat{\alpha})\}_{i=1}^n$, where $\hat{\alpha}$ is the root of $G_n(\alpha) = 0$. Next, we solve $S_n(\eta^{tr}, \hat{\alpha}) = 0$ by including the estimated weights. Denote the estimator of η^{tr} be as $\hat{\eta}^{catr}$. Incorporating $\hat{\eta}^{catr}$ and the estimated weights, for $Q_n^L(\gamma)$ and $Q_n^P(\gamma)$, the two estimators of θ_0^{tr} are obtained through solving $U_n^L(\theta^{tr}, \hat{\eta}^{catr}, \hat{\alpha}) = 0$ and $U_n^P(\theta^{tr}, \hat{\eta}^{catr}, \hat{\alpha}) = 0$. These solutions are denoted by $\hat{\theta}^{Lcatr}$ and $\hat{\theta}^{Pcatr}$ for the approaches from [3] and [4], respectively.

This methodology works because the propensity score adjusts for the distribution of confounders between the treatment and control groups. Although only the treatment variable is utilized in the construction of the estimating functions, the proposed weights contain information from confounders which lead to their proper adjustment. Based on the model assumptions, $(X_i - \theta_0^T W_i, D_i - \eta_0^T W_i)$, $i = 1, \dots, n$ have a common joint distribution not depending on W_i . By Theorem 1, we show that $(X_i - \theta_0^{tr} Z_i, D_i - \eta_0^{tr} Z_i)$, $i = 1, \dots, n$ also have a common joint distribution, which do not depend on Z_i given the propensity score $e_i(\alpha_0)$ because the distribution of V is same for both treatment group and control group given the same propensity score value. Moreover, note that $d(\cdot)$ and $d_{ij}(\cdot)$, $i, j = 1, \dots, n$ are the artificial censoring quantities based on [3] and [4]. Therefore, the artificial censoring quantities are increasing functions of the number and variability of confounders. Since only the treatment variable is utilized in the estimation procedure, it is expected to have very small artificial censoring compared to estimation procedures in [3] and [4] which employ all variables. This point is seen in the numerical study in Section 6.

Another advantage of the proposed method is ease of computation. Numerically, this involves solving a one-dimensional equation, which is much faster and easier to do than the multidimensional case.

4 Theoretical Results and Inference

We now determine the asymptotic behavior of the estimated association parameter between the treatment and outcome of events. Note that our estimation procedure is similar to one assuming a misspecified model. Several authors [12–14] have shown that the estimated

treatment effect converges to a constant value (which may not be the same as true value in underlying true model) in probability. Here, we show that our estimator converges to γ_0 , true value of treatment effect underlying true model (1).

Let $\Lambda_{10}(\cdot|\mathbf{V})$ and $\Lambda_{20}(\cdot|\mathbf{V})$ be the true baseline hazard function for $D_i - \eta_0^{tr} Z_i$ and for

$X_i - \theta_0^{tr} Z_i$ censored by the $D_i - \eta_0^{tr} Z_i - d(\beta_0^{tr})$ given the true propensity score $e(\alpha_0)$, respectively. Martingales for the dependent censoring and the event of interest [3] are given by

$$M_{1i}^w(t; \eta_0^{tr}, \alpha_0) = w_i(\alpha_0) \left[\xi_i I \left\{ \tilde{D}_i^* \left(\eta_0^{tr} \right) \leq t \right\} - \int_{-\infty}^t I \left\{ \tilde{D}_i^* \left(\eta_0^{tr} \right) \geq u \right\} \lambda_{10}(u|\mathbf{V}) du \right].$$

$$M_{2i}^w(t; \beta_0^{tr}, \alpha_0) = w_i(\alpha_0) \left[\tilde{\delta}_i^* \left(\beta_0^{tr} \right) I \left\{ \tilde{X}_i^* \left(\beta_0^{tr} \right) \leq t \right\} - \int_{-\infty}^t I \left\{ \tilde{X}_i^* \left(\beta_0^{tr} \right) \geq u \right\} \lambda_{20}(u|\mathbf{V}) du \right].$$

For the [4] type estimating function, given the propensity scores, we have the same setup as in the proof of Appendix of [4].

Let $\hat{\gamma}^L = (\hat{\alpha}^T, \hat{\eta}^{catr}, \theta^{Lcatr})^T$ and $\hat{\gamma}^P = (\hat{\alpha}^T, \hat{\eta}^{catr}, \theta^{Pcatr})^T$ be a solution of $Q_n^L(\gamma) = 0$ and $Q_n^P(\gamma) = 0$, respectively. Now we propose asymptotic properties of $\hat{\gamma}^L$ and $\hat{\gamma}^P$. Their proofs are located in the Section A of the online Supplementary Materials.

Theorem 2

By the conditions (A.1) – (A.10) and (B.1) – (B.2) in the online Supplementary Materials, $\hat{\gamma}^L$ and $\hat{\gamma}^P$ are strongly consistent for γ_0 .

Theorem 3

By the conditions (A.1) – (A.10) and (B.1) – (B.2) in the online Supplementary Materials, and by Theorem 2, $n^{1/2}(\hat{\gamma}^L - \gamma_0)$ has an asymptotic normal distribution with mean 0 and covariance matrix $(\Lambda_0^L)^{-1} \Omega_0^L (\Lambda_0^L)^{-1}$, where Λ_0^L is a nonsingular matrix, and Ω_0^L is the limiting covariance matrix of $Q_n^L(\gamma_0)$. Similarly, $n^{1/2}(\hat{\gamma}^P - \gamma_0)$

has an asymptotic normal distribution with mean 0 and covariance matrix

$(\Lambda_0^P)^{-1} \Omega_0^P (\Lambda_0^P)^{-1}$, where Λ_0^P is a nonsingular matrix, and Ω_0^P is the limiting covariance matrix of $Q_n^P(\gamma_0)$.

We may estimate the covariance matrix by extending the approach of [16]. In this case, the first step is to estimate Ω_0^L and Ω_0^P . Let $\hat{\beta}^{Lcatr} = (\hat{\eta}^{catr}, \hat{\theta}^{Lcatr})$ and $\hat{\beta}^{Pcatr} = (\hat{\eta}^{catr}, \hat{\theta}^{Pcatr})$. Adapting the work of [3] and [4], we now consider the following weighted empirical influence functions for $Q_n^L(\gamma_0)$ and $Q_n^P(\gamma_0)$.

$$\hat{v}_{1i} = \mathbf{H}_i \left\{ Z_i - \frac{\exp(\hat{\alpha}^T \mathbf{H}_i)}{1 + \exp(\hat{\alpha}^T \mathbf{H}_i)} \right\}$$

$$\begin{aligned} \hat{v}_{2i}^{(1)} = w_i(\hat{\alpha}) \left(\xi_i \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \geq \tilde{D}_i^*(\hat{\eta}^{catr})\} w_j(\hat{\alpha}) Z_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \geq \tilde{D}_i^*(\hat{\eta}^{catr})\} w_j(\hat{\alpha})} \right] - \sum_{l=1}^n \frac{w_l(\hat{\alpha}) \xi_l I\{\tilde{D}_i^*(\hat{\eta}^{catr}) \geq \tilde{D}_l^*(\hat{\eta}^{catr})\}}{\sum_{j=1}^n w_j(\hat{\alpha}) I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \geq \tilde{D}_l^*(\hat{\eta}^{catr})\}} \right. \\ \left. \times_i \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \geq \tilde{D}_l^*(\hat{\eta}^{catr})\} w_j(\hat{\alpha}) Z_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \geq \tilde{D}_l^*(\hat{\eta}^{catr})\} w_j(\hat{\alpha})} \right] \right) \end{aligned}$$

$$\begin{aligned} \hat{v}_{2i}^{(2)} = w_i(\hat{\alpha}) \left(\tilde{\delta}_i^*(\hat{\beta}^{Lcatr}) \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \geq \tilde{X}_i^*(\hat{\beta}^{Lcatr})\} w_j(\hat{\alpha}) Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \geq \tilde{X}_i^*(\hat{\beta}^{Lcatr})\} w_j(\hat{\alpha})} \right] \right. \\ \left. - \sum_{l=1}^n \frac{w_l(\hat{\alpha}) \tilde{\delta}_l^*(\hat{\beta}^{Lcatr}) I\{\tilde{X}_i^*(\hat{\beta}^{Lcatr}) \geq \tilde{X}_l^*(\hat{\beta}^{Lcatr})\}}{\sum_{j=1}^n w_j(\hat{\alpha}) I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \geq \tilde{X}_l^*(\hat{\beta}^{Lcatr})\}} \right. \\ \left. \times_i \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \geq \tilde{X}_l^*(\hat{\beta}^{Lcatr})\} w_j(\hat{\alpha}) Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \geq \tilde{X}_l^*(\hat{\beta}^{Lcatr})\} w_j(\hat{\alpha})} \right] \right) \end{aligned}$$

$$\hat{v}_{3i}^{(3)} = \frac{2}{n-1} \sum_{j=1}^n w_i(\hat{\alpha}) w_j(\hat{\alpha}) (Z_i - Z_j) \phi_{ij}(\hat{\beta}^{Pcatr}).$$

Let $\hat{v}_{2i}^L = \{\hat{v}_{2i}^{(1)}, \hat{v}_{2i}^{(2)}\}^T$ and $\hat{v}_{2i}^P = \{\hat{v}_{2i}^{(1)}, \hat{v}_{2i}^{(2)}\}^T$. Moreover, let $\hat{v}_i^L = \{\hat{v}_{1i}^T, (\hat{v}_{2i}^L)^T\}^T$ and $\hat{v}_i^P = \{\hat{v}_{1i}^T, (\hat{v}_{2i}^P)^T\}^T$. The estimator of Ω_0^L and Ω_0^P are

$$\hat{\Omega}^L = \frac{1}{n} \sum_{i=1}^n \hat{v}_i^L (\hat{v}_i^L)^T \quad \hat{\Omega}^P = \frac{1}{n} \sum_{i=1}^n \hat{v}_i^P (\hat{v}_i^P)^T. \quad (8)$$

It is important to consider the variability arising from the propensity score modeling. For nonrandomized studies, the true propensity scores adjust for the imbalance between confounders. However, the estimated propensity scores is random so that the variability from propensity score modeling needs to be accounted for. Since the empirical influence functions contain weights based on the estimated propensity score, the weights impact the variance of

$\left(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr}\right)^T$. Ignoring these weights results in inflation of variance of the estimators of interest $\left(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr}\right)^T$ because if the propensity score is treated as known, it leads to a decrease in precision for $\left(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr}\right)^T$. This issue is well explained in [15].

Now let us consider estimation of the covariance matrix of $\hat{\gamma}^L$ and $\hat{\gamma}^P$. In this case, the main issue is estimation of Λ_0^L and Λ_0^P . As discussed in [3] and [4], direct estimation of Λ_0^L and Λ_0^P involves the estimation of unknown hazard function of error terms, which is numerically unstable. Although the estimating functions are continuous with respect to α , the derivatives of estimating functions with respect to α have a very complicated form. The resampling approach from [16] is an appealing way to estimate the covariance matrix. Their approach is to solve a stochastic equation a large number of times and to use the solutions to estimate the covariance matrix. This approach does not require estimating the asymptotic slope matrix Λ_0 , so it is a suitable approach for nonsmooth estimating equations such as those in this paper. We have extended the approach of [16] to estimate the covariance of $\hat{\gamma}$. As discussed before, it is important to incorporate the sampling variability of $\hat{\alpha}$ when quantifying the variability of $\left(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr}\right)^T$. Let $U_n^L(\gamma) = \left[S_n^T(\eta^{tr}, \alpha), \left\{ U_n^L(\beta^{tr}, \alpha) \right\} \right]^T$. Given the data, the following stochastic equations are solved:

$$\begin{pmatrix} G_n(\alpha) = -n^{1/2} \sum_{i=1}^n \hat{v}_{1i} A_i \\ U_n^L(\gamma) = -n^{1/2} \sum_{i=1}^n \hat{v}_{2i}^L A_i \end{pmatrix}, \quad (9)$$

where A_i , $i = 1, \dots, n$ are standard normal random variables. Then the covariance matrix is estimated by solving the first and second equations in (9) in many times: Step 1: we solve $G_n(\alpha)$ to obtain $\hat{\alpha}^*$; Step 2: we update $w_i(\hat{\alpha}^*)$, $i = 1, \dots, n$ with $\hat{\alpha}^*$ from Step 1 and solve

$U_n^L(\gamma)$ to obtain $\hat{\tau}^{L*} = \left\{ \hat{\eta}^{catr*}, \hat{\theta}^{Lcatr*} \right\}^T$ with updated weights $w_i(\hat{\alpha}^*)$ by plugging in $U_n^L(\gamma)$. By repeating these two steps a sufficiently large number of times, we obtain

$\hat{\gamma}^{L*} = \left\{ (\alpha^*)^T, (\hat{\tau}^{L*})^T \right\}^T$ and the covariance matrix can be estimated based on $\hat{\gamma}^{L*}$. We apply the same procedure to [4] type estimator. We solve

$$\begin{pmatrix} G_n(\alpha) = -n^{1/2} \sum_{i=1}^n \hat{v}_{1i} A_i \\ U_n^P(\gamma) = -n^{1/2} \sum_{i=1}^n \hat{v}_{2i}^P A_i \end{pmatrix}, \quad (10)$$

By solving (10), we repeat Step 1 and Step 2 with the [4] type estimator, and then we

estimate covariance matrix with $\hat{\gamma}^{P*} = \left\{ (\alpha^*)^T, (\hat{\tau}^{P*})^T \right\}^T$. Note that in this procedure, the only random part is A_i while the observed data are treated as fixed [3]. Moreover, the variability of $\hat{\alpha}$ is incorporated in the estimators of interest through the updating weights. If we ignore the variability of the propensity score, we do not solve the first equation of (9) and only solve the second equation by using $w_i(\hat{\alpha})$, $i = 1, \dots, n$.

The following theorem proves the proposed estimation procedure for covariance matrix by solving equation (9) converges to the covariance matrix in Theorem 2, given the observed data. The detailed proof may be found in Section A of Supplementary Materials.

Theorem 4

Let $\hat{\gamma}^*$ be either $\hat{\gamma}^{L*}$ or $\hat{\gamma}^{P*}$. By the conditions (A.1) – (A.10) and (B.1) – (B.2) in the Supplementary Materials, and Theorem 3, conditional on observed data, the asymptotic distribution of $n^{1/2}(\hat{\gamma}^* - \hat{\gamma})$ is same as the unconditional distribution of $n^{1/2}(\hat{\gamma} - \gamma_0)$.

In addition to estimating the asymptotic covariance matrix numerically, in some cases it may be more convenient to use a nonparametric bootstrap to ease the potentially over-complicated numerical issues from computing Λ_0^L and Λ_0^P . After bootstrapping at the individual subject level, we solve the proposed estimating equations based on the resultant datasets as delineated above, and we may estimate the covariance of model parameters from solutions of estimating functions (2), (4), and (5) or (2), (4), and (6) based on a large amount of bootstrapped samples.

The resampling and the bootstrap approaches are similar, but there are two main differences. First, the resampling method does not resample from original data itself while the bootstrap method does. Moreover, the bootstrap method does not require simulation of standard normal random variable and computation of the empirical influence functions.

5 Real Data Analysis

5.1 Subgroup Analysis of Prostate Cancer Trial

As described in the Introduction, the primary interest of this subgroup analysis is to compare the time to disease progression between two types of radiotherapy (RT), irradiating the whole pelvic (WPRT) and the prostate only (PORT), among the intermediate risk patients. Time to disease progression is defined as the time from randomization to the first occurrence of local progression, regional/nodal failure, distant metastasis or biochemical failure. Patient and pretreatment tumor characteristics that may potentially impact disease progression and death and are used include: age, Karnofsky performance status (KPS), prostate specific antigen (PSA), Gleason Score, tumor size and T stage. Age, PSA and tumor size were analyzed as continuous variables, while the other covariates were treated as categorical variables. A total of 677 patients who met the intermediate risk criteria, including 354 received WPRT and 323 received PORT, were included in this analysis. There were 219

subjects experienced both disease progression and death, 137 experienced progression and censored later, 177 died without progression, and 144 censored without any events.

We first consider two conventional methods for semi-competing risks: fitting the model employing both treatment and confounders (Full model) and fitting the model using treatment variable only using approaches in [3] and [4] (Unadjusted). In all cases, 200 times of resampling and bootstrap runs are tried. When fitting the Full model, in 200 resampling runs, 69 resampling runs give nonconvergence, which implies that the approaches in [3] and [4] are not stable.

Table 1 shows the results. The standard error of the original [3] and [4] estimators is computed by using the resampling approach from [16]. For our proposed method, all 200 resampling runs are employed. We computed the standard error of estimators in three ways: the resampling method which ignores variability of the propensity score (Naive), the bootstrap method (Bootstrap) and the resampling method incorporating variability of the propensity score (Resamp). The proposed method is more stable than the approaches in [3] and [4].

To examine how many observations are artificially censored, we computed the artificial censoring rate. Let CR_D be the censoring rate for D , which is defined as $1 - \sum_{i=1}^n \xi_i/n$ and CR_X be the censoring rate for X , which is defined as $1 - \sum_{i=1}^n \delta_i/n$. Let ACR_{FL} be the artificial censoring rate from [3] approach and ACR_{FP} be the one from [4] approach considering both treatment and all confounders. Let $\hat{\beta}^{LF}$ and $\hat{\beta}^{PF}$ are estimators from approaches in [3] and [4] from including both treatment and all confounders in the model. Using arguments in [18], mathematical definitions of these quantities are

$$ACR_{FL} = 1 - \frac{\sum_{i=1}^n \tilde{\delta}_i^{full*} \left(\hat{\beta}^{LF} \right)}{\sum_{i=1}^n \delta_i}$$

$$ACR_{FP} = 1 - \frac{\sum_{i=1}^n \sum_{j \neq i} \tilde{\delta}_{i(j)}^{full*} \left(\hat{\beta}^{PF} \right)}{(n-1) \sum_{i=1}^n \delta_i},$$

where $\tilde{\delta}_i^{full*}$ and $\tilde{\delta}_{i(j)}^{full*}$ are artificial censoring indicators using both treatment and all confounders using approaches from [3] and [4], respectively. Let ACR_{AL} be the artificial censoring rate from proposed method from [3] approach and ACR_{AP} be the one from the proposed method of [4]. Then

$$ACR_{AL} = 1 - \frac{\sum_{i=1}^n \tilde{\delta}_i^* \left(\hat{\beta}^{Lcatr} \right)}{\sum_{i=1}^n \delta_i}$$

$$ACR_{AP} = 1 - \frac{\sum_{i=1}^n \sum_{j \neq i} \tilde{\delta}_{i(j)}^* \left(\hat{\beta}^{P_{catr}} \right)}{(n-1) \sum_{i=1}^n \delta_i}.$$

Based on the above quantities, the artificial censoring rates of the full model by [3] and [4] are 0.632 and 0.196, respectively. The artificial censoring rates from the proposed approaches from Section 3 are 0.034 and 0.048, so more uncensored observations are retained than in the full model approach.

5.2 AIDS Trial

We now consider the AIDS Clinical Trial Group (ACTG) Study 364 [4, 17]. This dataset was previously analyzed by [4]. While the event of interest was when HIV RNA level became greater than 2000 copies per ml, subjects may also withdraw from the study follow-up. Moreover, it was speculated that the level of HIV RNA can serve as a useful marker for study withdrawal. Although this HIV dataset is from a randomized study, including key covariates may increase the precision of the estimated treatment estimates comparing with modeling treatment assignment alone. In the HIV study, there are two continuous covariates of interest, baseline RNA level (after log-transform) and CD4 count. Variation of CD4 count is very large. The range of CD4 count is 1589.5, suggesting that including CD4 count in the modeling could lead to excessive artificial censoring and efficiency loss in estimation. There were three treatment groups, NFV only, EFV only, and NFV+EFV, in the original dataset. For the purposes of illustration, we combine the EFV only and NFV+EFV groups and treat NFV as the reference level. This would result in excessive artificial censoring when conventional approaches were used. We also removed one case whose CD4 count equals zero.

Similar to the prostate cancer data analysis, we considered the following models: the proposed approach with propensity score adjustment, the conventional approaches (Lin and P&F) with the adjustment for continuous covariates (Full model) or without no covariate adjustment (Unadjusted). As shown in Table 2, when using the proposed approach, ignoring the variability of the propensity scores (Naive estimate) results in a slightly larger standard error than that with resampling or bootstrap method. Interestingly, we can see that the standard errors by ignoring variability of the propensity score in the proposed approach are almost same as ones when employing the treatment variable only without propensity score adjustment. Moreover, we can see the standard errors from the proposed approaches are smaller than those from the unadjusted model, suggesting the incorporation of additional covariates through propensity scores improves the estimation precision of the treatment effect. Of note, due to the variability of RNA level and CD4 count, the artificial censoring rates are 100% when using the conventional approaches ([3] and [4]). On the other hand, the artificial censoring rates by using the proposed method is 0.069 for the approach based on [3] type and 0.013 for [4] type, respectively.

6 Simulation Studies

6.1 Assuming the true propensity score model

Next, we performed some simulations to explore the finite-sample properties of the proposed methodology. We first considered a situation in which data are generated according to model (1). We simulate a confounder $V \sim N(0, 4)$ and a treatment variable Z as Bernoulli random

variable with probability $\exp(\alpha_0^T H) [1 + \exp(\alpha_0^T H)]^{-1}$, where $H = (1, V^T)^T$ and $\alpha_0 = (\alpha_1, \alpha_2) = (0, 0.5)^T$. Then the error distribution ϵ is bivariate normal with mean $(0, 1.2)^T$ and

covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\rho = 0, 0.25, 0.5$. Independent censoring times are simulated as $C \sim \log(U)$, where U has uniform distribution on $(0, 200)$. The true values of parameters are $\theta_0 = (0.5, 1)^T$ and $\eta_0 = (1, 0.5)^T$. 500 datasets are simulated and for each simulated dataset, 500 bootstrap runs and 500 resampling runs are generated.

We calculated bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage probability (CP) of the proposed estimators from considering both treatment and confounder. For estimators considering both treatment and confounder, we use original [3] and [4] approaches. In this case, as with the real data analysis case, the resampling approach in [16] is employed for the standard error calculation and the coverage probability is based on the empirical distribution based on resampling runs. The coverage probability is based on the empirical distributions of resampling runs or bootstrap runs. As in the real data analysis case, when implementing the proposed method, we calculated the standard error of the proposed estimators in three ways: the resampling method which ignores variability of the propensity score (se1), the bootstrap method (se2) and the resampling method incorporating variability of the propensity score (se3). As expected, the standard error (se1) is large and the coverage is conservative for the resampling-based estimator which ignores the variability of propensity score. We thus only show results using the other two approaches computing standard errors, Bootstrap (se2) and Resampling (se3).

Table 3 shows the numerical results for the treatment variable considering both treatment and confounder and the proposed methodology with $N = 250$. Numerical results for $N = 500$ are shown in the Supplementary Materials. Henceforth the procedure considering both treatment and confounder is denoted as “the full model procedure”. The estimators from utilizing all both treatment and confounder for the dependent censoring and the event of interest by [3] and [4] are denoted as $\hat{\eta}^F, \hat{\theta}^{LF}$ and $\hat{\theta}^{PF}$, respectively. Moreover, the estimators for Z from the full model procedure for the dependent censoring and the event of interest by [3] and [4] are denoted as $\hat{\eta}^{F,tr}, \hat{\theta}^{LF,tr}$ and $\hat{\theta}^{PF,tr}$, respectively. Since the simulation results are based on the joint distributions of estimators for dependent censoring from [3] and [4], simulation runs are removed if the standard errors of any one of these estimators in the simulation runs are zero. When $N = 250$, for $\rho = 0, 0.25$ and 0.5 , 34, 62 and 127 results from simulation runs are removed because the standard errors of the estimators by [3] for either confounder or treatment are zero. However, for the proposed method, all 500 runs for the estimation of the association parameter of the treatment and the outcome and its standard error are used.

If the simulation runs corresponding to only the estimators for dependent censoring and the estimator for the event of interest by [4] in the full model procedure are considered, there are no simulation runs which have standard error zero. It can be seen that the method of [3] is problematic in this case. This is due to the excessive artificial censoring as computed.

The numerical results show that the proposed method works well. The mean of the standard errors when considering estimated propensity score as the truth is high so that coverage probability is conservative. However, the bootstrap and resampling approaches provide the desired coverage. When comparing the coverage of estimators between the full model procedure and resampling and bootstrap procedures, the proposed approaches show better coverage relative to the procedure in [3].

Moreover, Table 3 shows artificial censoring proportion. To examine the effect of the artificial censoring, entire simulation runs are included in this calculation. When considering both the treatment and confounder, the artificial censoring rate by the approach in [3] is high. However, the artificial censoring rate of the proposed method based on [3] is small, which implies that loss of observations for the proposed methodology is much smaller than that in the full model procedure.

6.2 Not assuming the true propensity score model

We also performed numerical studies when we fit a misspecified propensity score model. We generate $\mathbf{J} = (J_1, J_2)^T$ from a bivariate normal distribution with mean $(0, 0)^T$ and covariance matrix $\begin{pmatrix} 4 & a \\ a & 1 \end{pmatrix}$, where $a = 0, 1$. Then we define $V = J_1$ and $Z = I(J_2 > 0)$. Other parameter settings are same as before except $\rho = 0, 0.25$. Three scenarios are considered:

(Case 1) Dependent censoring with confounder ($a = 1$ and $\rho = 0.25$)

(Case 2) Independent censoring with confounder ($a = 1$ and $\rho = 0$)

(Case 3) Dependent censoring with randomized study ($a = 0$ and $\rho = 0.25$)

The results for $N = 500$ are shown in Section B of the Supplementary Materials. When applying the full model approach, 54 runs, 43 runs, and 96 runs are removed for Case 1, Case 2 and Case 3, respectively in calculation of standard errors on the full model due to either standard error of treatment or confounder being zero. As with the simulation using the true propensity score, the standard errors are computed from the proposed resampling and bootstrap. The coverage probability is based on empirical distributions of resampling runs or bootstrap runs.

The numerical results in Table 4 indicate that the proposed method works well. The nonparametric bootstrap and new resampling approach reflected variation of estimated propensity score into $\hat{\beta}^{Lcatr}$ and $\hat{\beta}^{Pcatr}$. Our methodology even works well for the randomized study. One of the interesting features is that more simulation runs are lost in a randomized study than those in an observational study. Even though our primary goal is to estimate the association between treatment and outcome of interest in observational and non-randomized studies where potential confounders having large variations, it also works well

for randomized studies including continuous covariates with large variance. We also calculate the artificial censoring proportion. As for the simulation study assuming that logistic regression model is true, even in the case not assuming propensity score model, the artificial censoring proportion from the proposed method is significantly lower than that of the full model case.

Although the performance is similar to that in the previous section, the bias of the proposed estimator is higher than when logistic regression model is correctly specified. This phenomenon is expected because the subsequent estimating function will no longer be unbiased.

6.3 Impact of a confounder with high variance : simulation using HIV dataset

In these simulation studies, the method from [4] has demonstrated better performance than that in [3]. In this section, we show that if the variability of confounders is large, even the method in [4] does not provide correct coverage for estimators. We performed a simulation study based on the HIV dataset. First, by using estimates from real data analysis, the time to event of interest and time to dependent censoring are generated from

$$\begin{pmatrix} X=1.101 \times Z - 0.739 \times V_1 - 8 \times 10^{-6} \times V_2 + \varepsilon^X \\ D=0.782 \times Z - 0.538 \times V_1 - 8 \times 10^{-4} \times V_2 + \varepsilon^D \end{pmatrix},$$

where Z indicates the binary treatment defined in the real data analysis, V_1 is logarithm of RNA value, and V_2 is CD4 count. Let $L^* = (1, V_1^T, V_2^T)^T$. In this case, Z is generated as

Bernoulli(p), where $p = \frac{\exp(\alpha_0^T L^*)}{1 + \exp(\alpha_0^T L^*)}$ and α_0 are logistic regression coefficients $(1.145, -0.133, 8.3 \times 10^{-5})^T$ from the real data study. Based on the residual values, we generate error values from a bivariate normal distribution with mean $(7.88, 8.2)^T$ and covariance matrix $\begin{pmatrix} 1.02 & 0.42 \\ 0.42 & 0.42 \end{pmatrix}$. censoring Independent time C has a uniform(3,10) distribution. In each simulation run, 150 observations (without replacement) are selected in observations. For each simulated dataset, 500 resamplings and bootstraps are used.

As in the real data study, including confounders results in excessive artificial censoring. Only 22 runs using [3] give nonzero standard errors and no simulation using [4] provide nonzero standard errors for both treatment and confounders. Thus, we cannot obtain the correct coverage for both the treatment and confounders using their approaches. Table 5 shows the results of the simulation study using the proposed method. Standard errors are computed in the same three ways as with the other simulation studies. The proposed methodology indicates that the standard errors of estimators are properly computed and achieve correct coverage.

The artificial censoring rates of the full model for [3] and [4] are 0.988 and 1, respectively. Compared to the artificial censoring rates in the full model, the artificial censoring rates by

the proposed method based on [3] and [4] are 0.071 and 0.048, respectively. In this study, the method in [4] does not provide the correct coverage for the estimate. The result of the simulation study shows the effectiveness of our approach when the variability of confounders is extremely large.

7 Discussion

In this paper, we have proposed a methodology for estimating association between treatment and nonterminal event under a semicompeting risks data structure in the context of an observational or non-randomized study. In medical studies, it is also common that the event of interest occurs several times. This type of data structure is referred to as recurrent events. Recurrent events in the presence of dependent censoring have been extensively studied by [19], [20], and [18]. Our methodology could be extended to this situation and is currently under investigation.

Our approach is similar to that in [12] in the sense that the purpose is to estimate treatment effects by using weights. However, the approach of [12] assumes independent censoring and uses probability of censoring as weight, while our approach allows for dependent censoring and uses the propensity score as weight. Moreover, our proposed method estimates the association parameter in the true model, while the estimator in [12] targets the parameter under a misspecified model.

Our approach does not provide information on how confounders affect (X, D) . This is a limitation of our approach. For causal inference with uncensored data, it has been shown that even if the outcome model is misspecified, the treatment effect is unbiased when the propensity score is correctly specified in a linear model [21]. In their approach, their model includes the propensity score as a covariate. Following their approach, one remedy is to specify the model as

$$\begin{pmatrix} X_i = \theta_0^{tr} Z_i + \varphi_0 e_i(\alpha_0) + \zeta_i^X \\ D_i = \eta_0^{tr} Z_i + \varphi_0 e_i(\alpha_0) + \zeta_i^D \end{pmatrix} \quad i=1, \dots, n$$

where $(\zeta_i^X, \zeta_i^D)^T$, $i=1, \dots, n$ are error terms. This approach enables us to perform inference about the confounders via the propensity score. Research for modeling the influence of confounders and treatment variables with semi-competing risks data is a topic for future research.

In this paper, our approach requires modeling the propensity score and is done using a logistic regression model. As can be seen in the simulation study, misspecification of propensity score model may impact the performance of the proposed estimator. In this case, both the outcome and propensity score models are misspecified. The performance when both models are misspecified is worse than the case when the propensity score model is correctly specified. This phenomenon is shown in [21] for uncensored data. There are other ways to construct propensity score by nonparametric methods, such as classification and regression trees [22]. Recently, [10] and [11] proposed combining propensity scores from logistic

regression and nonparametric learning method for causal inference. It can be also interesting topic to compare performance between parametric modeling and nonparametric method, and methodology of [10] and [11].

In this paper, ideas from causal inference were used to motivate the estimation procedure. However, it is important to note that it is controversial to make any formal causal interpretation. If borrowing from the results in [23], we can show that there exists a joint distribution of (X^*, D^*) given treatment variable such that X^* and D^* have the same values as X and D but X^* and D^* are independent. If we employ the potential outcomes framework to establish causal relationship between treatment and event of interest, we need several strong and untestable assumptions. Robins [24] proposes assumptions to make causal inference in randomized studies. Similar to [24], we now describe the necessary assumptions for causal interpretation in semicompeting risks data.

1. Strong ignorability assumption: Potential outcomes are statistically independent with treatment assignment given confounders. For uncensored data, this can be expressed mathematically as

$$\{Y(1), Y(0)\} \perp Z | \mathbf{V}.$$

This assumption implies that if possible confounding mechanisms are known and can be measured using $\mathbf{V}$, then the treatment assignment does not affect potential outcomes. With the usual right censored data for a single event and independent censoring assumption, this assumption has to be modified. Let $\{D(1), D(0), C(1), C(0)\}$ be times to death and censoring for the treatment group and control groups, respectively. It is reasonable to assume that potential censoring times for treatment group and control group are equal as C . Then extending this strong ignorability assumption for uncensored data to censored data with independent censoring assumption yields [25]

$$\{D(1), D(0)\} \perp (Z, C) | \mathbf{V}. \quad (11)$$

In semicompeting risks data, we can extend (10) with time to event of interest. Let $X(1)$ and $X(0)$ be time to event of interest for treatment group and control group, respectively. Then

$$\{X(1), X(0), D(1), D(0)\} \perp (Z, C) | \mathbf{V}. \quad (12)$$

2. Noninteraction assumption: the error distribution is independent of treatment assignment.
3. Rank preservation assumption: if subject i experiences the terminal event earlier than j under treatment, hypothetically, subject i experiences the terminal event earlier than j under control [2].

The first assumption is needed to be able to interpret the parameters of interest as a causal effect. This assumption is not made in this paper. The second one is a general model assumption for bivariate AFT model. This is made in [3] and [4], including this paper. The third assumption, which is also not assumed in this paper, is sensible because the adjustment for the dependent censoring is required to establish the causal argument. Of note, the third assumption is much stronger than the first and second ones. If there exists a scientific example to reasonably satisfy this third assumption, then our approach can be a tool to provide causal interpretation.

Although our proposed estimator achieves the desired numerical characteristics, there is instability induced by weighting by propensity score. This instability can be a serious problem if the propensity score is close to zero or one. Including the propensity score as a covariate along with not weighting the estimating function might be a solution for this issue. Alternatively, we can adopt the convex optimization approach of [32] to obtain weight that avoid extreme values.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research is supported by National Institutes of Health grant R01 CA129102 and U10 CA180822. The authors wish thank to the referees and the associate editor, whose comments have significantly improved the paper.

References

1. Fine JP, Jiang H, Chappell R. On semicompeting risks data. *Biometrika*. 2001; 88(4):907–919. DOI: 10.1093/biomet/88.4.907
2. Varadhan R, Xue Q-L, Bandeen-Roche K. Semicompeting risks in aging research: methods, issues and needs. *Lifetime Data Analysis*. 2014; 20(4):538–562. DOI: 10.1007/s10985-014-9295-7 [PubMed: 24729136]
3. Lin DY, Robins JM, Wei LJ. Comparing two failure time distributions in the presence of dependent censoring. *Biometrika*. 1996; 83(2):381–393. DOI: 10.1093/biomet/83.2.381
4. Peng L, Fine JP. Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American statistical Association*. 2006; 101(475):1085–1093. DOI: 10.1198/016214506000000131
5. Ding AA, Shi G, Wang W, Hsieh J-J. Marginal regression analysis for semicompeting risks data under dependent censoring. *Scandinavian Journal of Statistics*. 2009; 36(3):481–500. DOI: 10.1111/j.1467-9469.2008.00635.x
6. Roach M 3rd, DeSilvio M, Lawton C, Uhl V, Machtay M, Seider MJ, Rotman M, Jones C, Asbell SO, Valicenti RK, Han S, Thomas CR Jr, Shipley WS. Phase III trial comparing whole-pelvic versus prostate-only radiotherapy and neoadjuvant versus adjuvant combined androgen suppression: Radiation Therapy Oncology Group 9413. *Journal of Clinical Oncology*. 2003; 21(10):1904–1911. DOI: 10.1200/JCO.2003.05.004 [PubMed: 12743142]
7. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70(1):41–55. DOI: 10.1093/biomet/70.1.41
8. Williamson EJ, Forbes A, White IR. Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Statistics in Medicine*. 2014; 33(5):721–737. DOI: 10.1002/sim.5991 [PubMed: 24114884]

9. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*. 2004; 23(19):2937–2960. DOI: 10.1002/sim.1903 [PubMed: 15351954]
10. Zhu, Y. Ph D Thesis. The Pennsylvania State University; 2013. Data-adaptive approaches to modeling propensity scores in causal inference problems.
11. Zhu Y, Ghosh D, Mitra N, Mukherjee B. A data-adaptive strategy for inverse weighted estimation of causal effects. *Health Services and Outcomes Research Methodology*. 2014; 14(3):69–91. DOI: 10.1007/s10742-014-0124-y
12. Boyd AP, Kittelson JM, Gillen DL. Estimation of treatment effect under non-proportional hazards and conditionally independent censoring. *Statistics in Medicine*. 2012; 31(28):3504–3515. DOI: 10.1002/sim.5440 [PubMed: 22763957]
13. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. *Biometrika*. 1986; 73(2): 363–369. DOI: 10.1093/biomet/73.2.363
14. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American statistical Association*. 1989; 84(408):1074–1078. DOI: 10.1080/01621459.1989.10478874
15. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*. 1992; 48(2):479–495. [PubMed: 1637973]
16. Parzen MI, Wei LJ, Ying Z. A resampling method based on pivotal estimating functions. *Biometrika*. 1994; 81(2):341–350. DOI: 10.1093/biomet/81.2.341
17. Albrecht MA, Bosch RJ, Hammer SM, Liou S-H, Kessler H, Para MF, Eron J, Valdez H, Dehlinger M, Katzenstein DA, for The AIDS Clinical Trials Group 364 Study Team. Nelfinavir, Efavirenz, or both after the failure of Nucleoside treatment of HIV infection. *New England Journal of Medicine*. 2001; 345(6):398–407. DOI: 10.1056/NEJM200108093450602 [PubMed: 11496850]
18. Hsieh J-J, Ding AA, Wang W. Regression analysis for recurrent events data under dependent censoring. *Biometrics*. 2011; 67(3):719–729. DOI: 10.1111/j.1541-0420.2010.01497.x [PubMed: 21039394]
19. Ghosh D, Lin DY. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*. 2003; 59(4):877–885. DOI: 10.1111/j.0006-341X.2003.00102.x [PubMed: 14969466]
20. Ghosh D. Semiparametric analysis of recurrent events: artificial censoring, truncation, pairwise estimation and inference. *Lifetime Data Analysis*. 2010; 16(4):509–524. DOI: 10.1007/s10985-009-9150-4 [PubMed: 20063182]
21. Vansteelandt S, Daniel RM. On regression adjustment for the propensity score. *Statistics in Medicine*. 2014; 33(23):4053–4072. DOI: 10.1002/sim.6207 [PubMed: 24825821]
22. Breiman, L., Friedman, J., Stone, CJ., Olshen, RA. *Classification and regression trees*. Chapman and Hall/CRC; 1984.
23. Ghosh D. A causal framework for surrogate endpoints with semicompeting risks data. *Statistics and probability letters*. 2012; 82(11):1898–1902. DOI: 10.1016/j.spl.2012.06.010 [PubMed: 22899873]
24. Robins JM. An analytic method for randomized trials with informative censoring: Part 1. *Lifetime Data Analysis*. 1995; 1(3):241–254. DOI: 10.1007/BF00985759 [PubMed: 9385104]
25. Bai X, Tsiatis AA, O'Brien SM. Doubly robust estimators of treatment specific survival distributions in observational studies with stratified sampling. *Biometrics*. 2013; 69(4):830–839. DOI: 10.1111/biom.12076 [PubMed: 24117096]
26. Ying Z. A large sample study of rank estimation for censored regression data. *The Annals of Statistics*. 1993; 21(1):76–99.
27. Ferguson, TS. *A Course in Large Sample Theory*. Chapman and Hall; London: 1996.
28. Ghosh, D. Ph D thesis. University of Washington; 2000. Nonparametric and semiparametric analysis of recurrent events in the presence of terminal events and dependent censoring.
29. Stefanski LA, Boos DD. The calculus of M-estimation. *The American Statistician*. 2002; 56(1):29–38. DOI: 10.1198/000313002753631330
30. Honoré BE, Powell JL. Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*. 1994; 64(1–2):241–278. DOI: 10.1016/0304-4076(94)90065-5

31. Fleming, TR., Harrington, DP. Counting Processes and Survival Analysis. second. Wiley; New York: 2005.
32. Zubizarreta JR. Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*. 2015; 110(511):910–922. DOI: 10.1080/01621459.2015.1023805

Table 1

Point estimates and standard error (SE) of θ in prostate cancer dataset

Death			First disease occurrence (Lin type)			First disease occurrence (P&F type)		
Est	SE	Art cen	Est	SE	Art cen	Est	SE	Art cen
Full model	0.076	0.019	-0.059	0.004	0.632	-0.062	0.011	0.196
	0.076	0.065	-0.05	0.119	0.031	-0.091	0.101	0.048
Proposed	Naive	0.062		0.119			0.101	
	Boot	0.064	0.055	-0.077	0.11	0.034	-0.112	0.096
	Resamp		0.059		0.115		0.097	

Lin type : [3] type approach P&F type : [4] type approach

Lin : method in [3] P&F : method in [4]

Table 2

Point estimates and standard error (SE) in HIV data analysis

Withdrawal			First viologic failure (Lin type)			First viologic failure (P&F type)		
	Est	SE	Est	SE	Art cen	Est	SE	Art cen
Full model	0.782	0.016	1.101	0	1	1.174	0	1
Unadjusted	0.749	0.151	1.021	0.24	0.089	0.818	0.237	0.039
Proposed	Naive	0.15		0.241			0.239	
	Boot	0.745	0.149	0.936	0.221	0.069	0.766	0.217
	Resamp		0.148	0.229			0.225	

Lin type : [3] type approach P&F type : [4] type approach

Lin : method in [3] P&F : method in [4]

Table 3

Bias, empirical standard deviation (EMPSD), mean of standard error (SEE), 95% coverage (CP) and censoring rates (ACR , CP_D , CP_X) for estimators of parameters of Z when $N = 250$ when assuming logistic regression model

		Bias	EMPSD	SEE			CP			ACR ^c	CR _D ^a	CR _X ^b
				Boot	Resamp	Boot	Resamp	Boot	Resamp			
$\rho = 0$	Full	$\hat{\eta}^{F,tr}$	0.001	0.152	-	0.154	-	0.946	-			
		$\hat{\theta}^{LF,tr}$	0.007	0.187	-	0.132	-	0.833	0.779			
		$\hat{\theta}^{PF,tr}$	-0.01	0.149	-	0.154	-	0.957	0.148			
Proposed		$\hat{\eta}^{catr}$	0.006	0.196	0.191	0.192	0.924	0.934	-		0.09	0.214
		$\hat{\theta}^{Lcatr}$	-0.011	0.242	0.241	0.25	0.946	0.94	0.066			
		$\hat{\theta}^{Pcatr}$	-0.015	0.192	0.197	0.219	0.954	0.94	0.065			
$\rho = 0.25$	Full	$\hat{\eta}^{F,tr}$	0.004	0.159	-	0.155	-	0.961	-			
		$\hat{\theta}^{LF,tr}$	-0.01	0.199	-	0.122	-	0.753	0.843			
		$\hat{\theta}^{PF,tr}$	0.001	0.153	-	0.151	-	0.938	0.144			
Proposed		$\hat{\eta}^{catr}$	0.014	0.187	0.191	0.192	0.944	0.946	-		0.09	0.195
		$\hat{\theta}^{Lcatr}$	0.009	0.241	0.238	0.247	0.936	0.938	0.064			
		$\hat{\theta}^{Pcatr}$	-0.0004	0.19	0.194	0.216	0.956	0.956	0.064			
$\rho = 0.5$	Full	$\hat{\eta}^{F,tr}$	-0.001	0.167	-	0.156	-	0.962	-		0.09	0.167

	Bias	EMPSD	SEE			CP			ACR^c	CR_D^a	CR_X^b
			Boot	Resamp	Boot	Resamp	Boot	Resamp			
$\hat{\eta}^{LF, tr}$	-0.021	0.192	-	0.109	-	0.702	-	0.906			
$\hat{\theta}^{PF, tr}$	0.004	0.149	-	0.15	-	0.952	-	0.133			
Proposed	0.015	0.187	0.191	0.193	0.944	0.948	-				
$\hat{\eta}^{Lcatr}$	0.014	0.237	0.235	0.245	0.94	0.944	0.065				
$\hat{\theta}^{Pcatr}$	0.003	0.182	0.188	0.211	0.95	0.952	0.066				

$\hat{\eta}^{F, tr}$: the estimator of dependent censoring for Z in the full model
 $\hat{\theta}^{LF, tr}, \hat{\theta}^{PF, tr}$: the estimator of the event of interest by original [3] and [4] for Z in the full model
 $\hat{\eta}^{catr}$: the estimator of dependent censoring by the proposed method
 $\hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr}$: the estimator of the event of interest by [3] and [4] type approaches in the proposed method
 ACR : the artificial censoring rate
 CRD : the censoring rate subject to the independent censoring
 CRX : the censoring rate subject to the dependent censoring

Table 4

Bias, empirical standard deviation (EMPSD), mean of standard error (SEE), 95% coverage (CP) and censoring rates (ACR , CP_D , CP_X) for estimators of parameters of Z when $N = 250$ when not assuming logistic regression model

		Bias	EMPSD	SEE			CP			ACR	CP_D	CP_X
				Boot	Resamp	Boot	Resamp	Boot	Resamp			
Case 1	Full	$\hat{\eta}_{F,tr}$	0.166	–	0.153	–	0.951	–	–	–	–	–
		$\hat{\theta}_{LF,tr}$	0.203	–	0.123	–	0.78	–	0.839	–	–	–
		$\hat{\theta}_{PF,tr}$	0.151	–	0.151	–	0.953	–	0.143	–	–	–
	Proposed	$\hat{\eta}_{catr}$	0.183	0.186	0.186	0.958	0.958	–	–	0.088	0.194	–
		$\hat{\theta}_{Lcatr}$	0.232	0.227	0.234	0.94	0.948	–	0.062	–	–	–
		$\hat{\theta}_{Pcatr}$	0.19	0.191	0.21	0.968	0.974	–	0.064	–	–	–
	Full	$\hat{\eta}_{F,tr}$	0.187	–	0.153	–	0.95	–	–	–	–	–
		$\hat{\theta}_{LF,tr}$	0.189	–	0.137	–	0.84	–	0.783	–	–	–
		$\hat{\theta}_{PF,tr}$	0.155	–	0.154	–	0.937	–	0.149	–	–	–
	Proposed	$\hat{\eta}_{catr}$	0.191	0.186	0.187	0.936	0.94	–	–	0.088	0.215	–
Case 2	Full	$\hat{\eta}_{F,tr}$	0.191	0.186	0.187	0.936	0.94	–	–	–	–	–
		$\hat{\theta}_{LF,tr}$	0.243	0.229	0.238	0.93	0.94	–	0.062	–	–	–
		$\hat{\theta}_{Pcatr}$	0.204	0.194	0.213	0.944	0.95	–	0.064	–	–	–
	Proposed	$\hat{\eta}_{catr}$	0.204	0.194	0.213	0.944	0.95	–	0.064	–	–	–
		$\hat{\theta}_{Lcatr}$	0.204	0.194	0.213	0.944	0.95	–	0.064	–	–	–
		$\hat{\theta}_{Pcatr}$	0.204	0.194	0.213	0.944	0.95	–	0.064	–	–	–
	Full	$\hat{\eta}_{F,tr}$	0.269	–	–0.143	–	0.941	–	–	–	0.078	0.2
		$\hat{\theta}_{LF,tr}$	0.269	–	–0.143	–	0.941	–	–	–	0.078	0.2
		$\hat{\theta}_{Pcatr}$	0.269	–	–0.143	–	0.941	–	–	–	0.078	0.2
	Proposed	$\hat{\eta}_{catr}$	0.269	–	–0.143	–	0.941	–	–	–	0.078	0.2

	Bias	SEE		CP		ACR	CR _D	CR _X
		EMPSD	Boot	Resamp	Boot	Resamp		
$\hat{\theta}^{LF, tr}$	-0.005	0.204	-	0.125	-	0.787	0.887	
$\hat{\theta}^{PF, tr}$	-0.002	0.144	-	0.14	-	0.931	0.158	
Proposed	0.009	0.152	0.154	0.154	0.952	0.958	-	
$\hat{\eta}^{catr}$	0.004	0.189	0.18	0.181	0.94	0.934	0.054	
$\hat{\theta}^{Lcatr}$	0.012	0.161	0.157	0.157	0.946	0.94	0.053	
$\hat{\theta}^{Pcatr}$								

$\hat{\eta}^{F, tr}$: the estimator of dependent censoring for Z in the full model
 $\hat{\theta}^{LF, tr}, \hat{\theta}^{PF, tr}$: the estimator of the event of interest by original [3] and [4] for Z in the full model
 $\hat{\eta}^{catr}$: the estimator of dependent censoring by the proposed method
 $\hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr}$: the estimator of the event of interest by [3] and [4] type approaches in the proposed method
ACR: the censoring rate subject to the independent censoring
CR_D: the censoring rate subject to the dependent censoring

Table 5

Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for simulation study by using HIV dataset

	Bias	EMPSD	SEE (Naive)	SEE (Bootstrap)	SEE (Resamp)	CP (Naive)	CP (Bootstrap)	CP (Resamp)
$\hat{\eta}^{catr}$	-0.001	0.166	0.182	0.169	0.174	0.964	0.954	0.946
$\hat{\theta}^{Lcatr}$	-0.006	0.23	0.261	0.236	0.239	0.96	0.948	0.946
$\hat{\theta}^{Pcatr}$	-0.007	0.216	0.251	0.224	0.224	0.976	0.944	0.934

Estimators $\hat{\eta}^{catr}$: the proposed estimator of the dependent censoring; $\hat{\theta}^{Lcatr}$: the proposed estimator using [3] approach; $\hat{\theta}^{Pcatr}$: the proposed estimator using [4] approach