2016

Covariate adjustment using propensity scores for dependent censoring

Youngjoo Cho, *University of Rochester*
Chen Hu, *Johns Hopkins University*
Debashis Ghosh

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

Youngjoo Cho[1], Chen Hu[2], and Debashis Ghosh*[3]

[1]Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY 14642
[2]Division 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
[3]Department of Biostatistics and Informatics, University of Colorado, Aurora, CO 80045

October 25, 2016

Abstract

In many medical studies, estimation of treatment effects is often of primary scientific interest. Standard methods for evaluating the treatment effect in survival analysis typically require the assumption of independent censoring. Such an 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 'semi-competing risks data'. In marginal modeling under semicompeting risks data, an artificial censoring technique is a promising approach to handle dependent censoring. However, continuous covariates with large variabilities may lead to excessive artificial censoring, which subsequently results in numerically unstable estimation. In this paper, we propose a strategy for weighted estimation of treatment effects 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 continuous covariates and

*corresponding author, debashis.ghosh@ucdenver.edu

simplifies computation. Monte Carlo simulation studies and application to AIDS and cancer research are used to illustrate the methodology.
Key words: Causal Inference, Observational study, Perturbation, Resampling

1 Introduction

In medical studies, researchers might collect information on nonterminal events which may be closely related to a terminal event. For example, in AIDS Clinical Trial Group (ACTG) Study 364 ([1], [2]), the event of interest is time when HIV RNA level is greater than 2000 copies per ml, but subjects were also followed for withdrawal from the study. The level of HIV RNA can serve as a useful marker as withdrawal of study [2]. Another example is the Radiation Therapy Oncology Group (RTOG) 9413 study ([3], [4]). In this study, the event of interest is the first occurrence of local failure, distant failure or biochemical failure. However, death also exists and may censor the event of interest. Such a data structure has been termed 'semicompeting risks data' [5].

Much research has been conducted on semicompeting risks data, a nice summary of which is given in [6]. 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. [7] 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 [8]. This can be easily extended to accommodate an arbitrary number of covariates. The advantage of artificial censoring lies in the fact that regression parameters can be estimated based on modified estimating equations without a parametric distributional assumption between event types. However, when a large number of covariates or continuous covariates are present, based on the expressions of the artificial censoring in [2] and [7], 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 small variability of the estimators from model using all covariates may be desirable, but it implies that the variability of the estimators are not properly estimated because of the excessive artificial censoring. As can be seen in the simulation studies at Section 6, excessive artificial censoring can lead to incorrect coverage.

In non-randomized observational studies, analyzing semicompeting risks data using [2] and [7] could be problematic due to excessive artificial censoring, as the

treatment effect evaluation lies on properly adjusting for potential confounders. Similar issues arise in subgroup analysis from randomized studies. The subgroup analysis of RTOG 9413 (Radiation Therapy Oncology Group) study can be a good example. RTOG 9413 was a multicenter randomized phase III trial for clinically localized intermediate-risk and/or high-risk prostate cancer patients. One of the primary objectives was to compare the efficacy combined androgen suppression (CAS) and whole pelvic radiotherapy (WPRT) followed by a boost to the prostate with CAS and prostate only radiotherapy (PORT). The protocol primary endpoint was progression-free survival, defined as time from randomization to the first occurrence of local progression, regional/nodal failure, distant failure, biochemical failure or death from any cause. While the initial reporting did not find that WPRT improves progression-free survival (PFS), a unplanned subgroup analysis reported in [3] suggested that WPRT may prolong PFS among intermediate risk patients (determined by the prostate specific antigen (PSA) and Gleason score (GS) at randomization). Given the nature of subgroup analysis and the fact that progression was dependently censored by death, it is of great interest to obtain treatment effect estimate for the time to the first occurrence of any disease failure (local, regional, distant, biochemical) within the semicompeting risks framework. In addition to initial PSA (ng/ml) and Gleason score, other potential prognostic variables include tumor size, T stage and age. The data being analyzed are based on the updated reporting [4].

Motivated by the RTOG 9413 dataset, our interest is to estimate treatment effect by adjusting confounders without excessive artificial censoring. Our approach to the problem involves propensity scores, proposed by Rosenbaum and Rubin [9]. While their use has been of interest in causal inference, they also satisfy a balancing property that corresponds to the distribution of covariates being equal for both treatment group and control group given the propensity score [10]. Thus, the propensity score can provide sufficient information to balance the covariates between treatment group and control group. This can be an important tool for reducing the artificial censoring needed for estimating a treatment effect in an asymptotically unbiased manner, as illustrated in numerical examples (Section 5 and 6).

In this paper, we propose methodology for estimation of treatment effects in observational studies adjusting for covariates using the propensity score. The paper is organized as follows. In Section 2, we introduce the data structure and modeling assumption, including a brief review about the propensity scores. Section 3 reviews the methodology for estimation of treatment effect. Theoretical results and details about inference using the proposed method are demonstrated in Section 4. In Section 5, we apply the proposed methodology to the ACTG 364 study and the RTOG 9413 study. Simulation studies are shown in Section 6.

Concluding remarks and discussion are in Section 7.

2 Preliminaries

2.1 Data and Model

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

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

All these times are log-transformed. The data consist of n independent observations $(\tilde{X}_i, \tilde{D}_i, \boldsymbol{W}_i, \delta_i, \xi_i)$, $i = 1, \ldots, n$. The model is

$$\begin{pmatrix} X_i = \boldsymbol{\theta}_0^T \boldsymbol{W}_i + \epsilon_i^X \\ D_i = \boldsymbol{\eta}_0^T \boldsymbol{W}_i + \epsilon_i^D \end{pmatrix} \quad i = 1, \ldots, n, \tag{1}$$

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

$$\left(\begin{array}{l} X_i = \theta_0^{tr} Z_i + (\boldsymbol{\theta}_0^{cfd})^T \boldsymbol{V}_i + \epsilon_i^X \\ D_i = \eta_0^{tr} Z_i + (\boldsymbol{\eta}_0^{cfd})^T \boldsymbol{V}_i + \epsilon_i^D \end{array} \right), \quad i = 1, \ldots, n.$$

We assume that the model is identifiable only in the upper wedge where $X < D$ and C is independent with (X, D) given $\boldsymbol{W}$, but X and D can be dependent given $\boldsymbol{W}$ ([2],[5]).

2.2 Propensity score

In the causal inference literature, the main interest is to establish, identify and estimate causal effects of interventions on outcomes. To describe the causal effect for binary treatment, the potential outcomes framework is a useful tool. Let Y be the continuous outcome without censoring and suppose that there are binary treatment variable Z and covariates $\boldsymbol{V}$. Under the potential outcome framework, we define $\{Y(1), Y(0)\}$ to be potential outcomes for a subject which belongs treatment group or control group. The average causal effect by the treatment is defined

by $E\{Y(1) - Y(0)\}$. Since each subject belongs to either the treatment group or control group, the outcome Y can be written as

$$Y = Y(1) \times Z + Y(0) \times (1 - Z).$$

In randomized studies, randomization ensures the possibility of a causal interpretation for treatment effects. However, in the observational studies, the existence of the confounders prevents straightforward estimation of causal effects. [9] proposed the propensity score to estimate the causal effect of the treatment for the observational studies. The propensity score is the probability of being in the treatment group given the confounders. The propensity score is defined by

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

The propensity score has many properties. It balances the confounders between the treatment and the control group. Given the same propensity score, the distribution of the confounders is equal between the treatment group and the control group. 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 will discuss the causality issue in the Discussion section. Rather, 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_i are independent of Z_i and identically distributed. In our case, based on the balancing property of propensity score, we can derive the following important result.

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

The proof of this theorem is located in the Section A of Supplementary Materials. This theorem implies that given the propensity score, we can establish i.i.d terms with omitting confounders. This theorem, as shown in the Supplementary material, enables us to estimate the treatment effect in an unbiased manner.

3 Proposed Methodology

As mentioned in the Introduction, the artificial censoring techniques are important tools for obtaining asymptotically unbiased estimators of regression coefficients for the time to event of interest in the presence of dependent censoring. Using the artificial censoring techniques, Lin et al. [7] and Peng and Fine [2] proposed two different estimating functions. Lin et al. [7] used a single comparison of each

residual time and Peng and Fine [2] compared different pairs of residual times for the artificial censoring. In [7], 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 covariates are included in the model. The method of [2] is clearly better than that in [7] in the sense that their artificial censoring is smaller than that in [7], but including continuous covariates with large variability may still cause the 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 treatment effect on the nonterminal event. The key idea is to only use the treatment variable in a weighted estimation procedure. Note that without the weights, the estimation procedure is problematic because we ignore the confounder. This weight is function of propensity score. We assume that the propensity score is modeled using logistic regression with parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$. Let $\boldsymbol{\alpha}_0$ be the true value of $\boldsymbol{\alpha}$ and $\boldsymbol{H}_i = (1, \boldsymbol{V}_i^T)^T, i = 1, \ldots, n$. Define the propensity score to be

$$e_i(\boldsymbol{\alpha}) = P(Z_i = 1|\boldsymbol{H}_i) = \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}.$$

Hence the true value of propensity score is expressed by $e_i(\boldsymbol{\alpha}_0)$ for each subject i. The weight is defined by

$$w_i(\boldsymbol{\alpha}) = \frac{Z_i}{e_i(\boldsymbol{\alpha})} + \frac{1 - Z_i}{1 - e_i(\boldsymbol{\alpha})}. \tag{2}$$

This weight takes value $1/e_i(\boldsymbol{\alpha})$ if $Z_i = 1$ and $1/(1 - e_i(\boldsymbol{\alpha}))$ otherwise. In the causal inference literature, the typical technique is to apply the weight $w_i(\boldsymbol{\alpha})$ to estimate the average treatment effect ([10],[11],[12],[13]). We would like to include this weight in the estimating functions. Let η^{tr} and θ^{tr} be the treatment effect parameters with respect to D and X, respectively. By using the weight, the proposed estimating function for η^{tr} is

$$S_n(\eta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \xi_i w_i(\boldsymbol{\alpha}) \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha})} \right], \tag{3}$$

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 [7], is

$$U_n^L(\beta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \tilde{\delta}_i^*(\beta^{tr}) w_i(\boldsymbol{\alpha}) \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta^{tr}) \geq \tilde{X}_i^*(\beta^{tr})\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta^{tr}) \geq \tilde{X}_i^*(\beta^{tr})\} w_j(\boldsymbol{\alpha})} \right], \tag{4}$$

where

$$d(\beta^{tr}) = \max_{1\le i\le n}\{0, (\theta^{tr} - \eta^{tr})Z_i\}$$
$$\tilde{X}_i^*(\beta^{tr}) = (X_i - \theta^{tr}Z_i) \wedge \{(D_i \wedge C_i) - \eta^{tr}Z_i - d(\beta^{tr})\}$$
$$\tilde{\delta}_i^*(\beta^{tr}) = I[(X_i - \theta^{tr}Z_i) \le \{(D_i \wedge C_i) - \eta^{tr}Z_i - d(\beta^{tr})].$$

Similarly, the proposed estimating function based on the work of [2] is

$$U_n^P(\beta^{tr}, \boldsymbol{\alpha}) = \frac{2n^{1/2}}{n(n-1)} \sum_{1\le i<j\le n} (Z_i - Z_j) w_i(\boldsymbol{\alpha}) w_j(\boldsymbol{\alpha}) \phi_{ij}(\beta^{tr}), \tag{5}$$

where

$$d_{ij}(\beta^{tr}) = \max\{0, (\theta^{tr} - \eta^{tr})Z_i, (\theta^{tr} - \eta^{tr})Z_j\}$$
$$\tilde{X}_{i(j)}^*(\beta^{tr}) = (X_i - \theta^{tr}Z_i) \wedge \{(D_i \wedge C_i) - \eta^{tr}Z_i - d_{ij}(\beta^{tr})\}$$
$$\tilde{\delta}_{i(j)}^*(\beta^{tr}) = I[(X_i - \theta^{tr}Z_i) \le \{(D_i \wedge C_i) - \eta^{tr}Z_i - d_{ij}(\beta^{tr})]$$
$$\phi_{ij}(\beta^{tr}) = \tilde{\delta}_{i(j)}^*(\beta^{tr}) I\{\tilde{X}_{i(j)}^*(\beta^{tr}) \le \tilde{X}_{j(i)}^*(\beta^{tr})\} - \tilde{\delta}_{j(i)}^*(\beta^{tr}) I\{\tilde{X}_{j(i)}^*(\beta^{tr}) \le \tilde{X}_{i(j)}^*(\beta^{tr})\}.$$

Let $\boldsymbol{G}_n(\boldsymbol{\alpha})$ be the score function for $\boldsymbol{\alpha}$, where

$$\boldsymbol{G}_n(\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{H}_i \left\{ Z_i - \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)} \right\} \tag{6}$$

Let $\boldsymbol{\gamma} = (\boldsymbol{\alpha}^T, \eta^{tr}, \theta^{tr}, \theta^{tr})^T$, $\boldsymbol{\gamma}_0 = (\boldsymbol{\alpha}_0^T, \eta_0^{tr}, \theta_0^{tr}, \theta_0^{tr})^T$ and $\beta_0^{tr} = (\eta_0^{tr}, \theta_0^{tr})^T$. In this case, $\boldsymbol{\gamma}_0$ is a parameter that satisfies $E\{\boldsymbol{G}_n(\boldsymbol{\alpha}_0)\} = 0$, $E\{S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$, $E\{U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$ and $E\{U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$, We solve

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}) = [\boldsymbol{G}_n^T(\boldsymbol{\alpha}), S_n^T(\eta^{tr}, \boldsymbol{\alpha}), \{U_n^L(\beta^{tr}, \boldsymbol{\alpha})\}^T, \{U_n^P(\beta^{tr}, \boldsymbol{\alpha})\}^T]^T = 0, \tag{7}$$

to obtain estimators of $\boldsymbol{\gamma}_0$. Solutions for $\boldsymbol{\gamma}$ can be obtained by solving the estimating equations sequentially. First we estimate the propensity scores for all observations, denoted as $\{e_i(\hat{\boldsymbol{\alpha}})\}_{i=1}^n$, where $\hat{\boldsymbol{\alpha}}$ is the root of $\boldsymbol{G}_n(\boldsymbol{\alpha}) = 0$. Next, we solve $S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}}) = 0$ by including the estimated weights. Denote the estimator of η_0^{tr} be as $\hat{\eta}^{catr}$. Incorporating $\hat{\eta}^{catr}$ and the estimated weights, the two estimators of θ_0^{tr} are obtained through solving $U_n^L(\theta^{tr}, \hat{\eta}^{catr}, \hat{\boldsymbol{\alpha}}) = 0$ and $U_n^P(\theta^{tr}, \hat{\eta}^{catr}, \hat{\boldsymbol{\alpha}}) = 0$. These solutions are denoted by $\hat{\theta}^{Lcatr}$ and $\hat{\theta}^{Pcatr}$ for the approaches from [2] and [7], 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, $(X_i - \boldsymbol{\theta}_0^T \boldsymbol{W}_i, D_i - \boldsymbol{\eta}_0^{tr} \boldsymbol{W}_i), i = 1, \ldots, n$ have a common joint distribution not depending on $\boldsymbol{W}_i$. By Theorem

1, we show that $(X_i - \theta_0^{tr} Z_i, D_i - \eta_0^{tr} Z_i), i = 1, \ldots, n$ also have a common joint distribution, which do not depend on Z_i given the propensity score $e_i(\boldsymbol{\alpha}_0)$ because the distribution of $\boldsymbol{V}$ is same for both treatment group and control group given the same propensity score value. Moreover, note that $d(\cdot)$ and $d_{ij}(\cdot), i, j = 1, \ldots, n$ are the artificial censoring quantities based on [2] and [7]. There, the artificial censoring quantities are increasing functions of the number and variability of covariates. 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 [2] and [7] 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

It is of interest to determine the asymptotic behavior of the estimated treatment effect. Note that our estimation procedure is similar to one assuming a misspecified model. Several authors ([14], [15], [16]) show 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. However, we show that our estimator converges to $\boldsymbol{\gamma}_0$, true value of treatment effect underlying true model (1).

Let $\lambda_{10}(\cdot|\boldsymbol{V})$ and $\lambda_{20}(\cdot|\boldsymbol{V})$ 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 $e_i(\boldsymbol{\alpha}_0)$, respectively. Martingales for the dependent censoring and the event of interest [7] are given by

$$M_{1i}^w(t; \eta_0^{tr}, \boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)[\xi_i I\{\tilde{D}_i^*(\eta_0^{tr}) \leq t\} - \int_{-\infty}^{t} I\{\tilde{D}_i^*(\eta_0^{tr}) \geq u\}\lambda_{10}(u|\boldsymbol{V})du].$$

$$M_{2i}^w(t; \beta_0^{tr}, \boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)[\tilde{\delta}_i^*(\beta_0^{tr}) I\{\tilde{X}_i^*(\beta_0^{tr}) \leq t\} - \int_{-\infty}^{t} I\{\tilde{X}_i^*(\beta_0^{tr}) \geq u\}\lambda_{20}(u|\boldsymbol{V})du].$$

For the [2] type estimating function, given the propensity scores, we have the same setup as in the proof of Appendix of [2].

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

Theorem 2. By the conditions (A.1) - (A.9) and (B.1) - (B.2) in the Supplementary Materials, $\hat{\boldsymbol{\gamma}}$ is strongly consistent for $\boldsymbol{\gamma}_0$.

Theorem 3. By the conditions (A.1) - (A.9) and (B.1) - (B.2) in the Supplementary Materials, and by Theorem 2, $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has an asymptotic normal distribution with mean 0 and covariance matrix $\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Lambda}_0^{-1}$. where $\boldsymbol{\Lambda}_0$ is a non-singular matrix, and $\boldsymbol{\Omega}_0$ is the limiting covariance matrix of $\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)$.

Note that convergence of the joint distribution of $(\hat{\boldsymbol{\alpha}}^T, \hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$ is proven. It is also possible to consider $\hat{\boldsymbol{\gamma}}^L = (\hat{\boldsymbol{\alpha}}^T, \hat{\eta}^{catr}, \hat{\theta}^{Lcatr})^T$ and $\hat{\boldsymbol{\gamma}}^P = (\hat{\boldsymbol{\alpha}}^T, \hat{\eta}^{catr}, \hat{\theta}^{Pcatr})^T$ separately, but this separation causes problems in statistical inference for $\hat{\boldsymbol{\alpha}}$ and $\hat{\eta}^{catr}$. This unified convergence result also implies that $\hat{\boldsymbol{\gamma}}^L$ and $\hat{\boldsymbol{\gamma}}^P$ are strongly consistent and asymptotically normal. The price for proving the joint convergence is that we need assumptions covering for both [2] and [7] estimating functions.

For inference, in addition to estimating the asymptotic covariance matrix numerically, it may be more convenient to use bootstrap to ease the potentially over-complicated numerical issues. After bootstrapping at the individual subject level, we solve the proposed estimating equations based on the resultant dataset as delineated above, and we may estimate the covariance of model parameters from solutions of estimating functions (3),(4),(5) and (6) based on a large amount of bootstrapped samples.

We may alternatively estimate the covariance matrix by extending approach of [17]. In this case, the first step is to estimate $\boldsymbol{\Omega}_0$. Let $\hat{\beta}^{Lcatr} = (\hat{\eta}^{catr}, \hat{\theta}^{Lcatr})$ and $\hat{\beta}^{Pcatr} = (\hat{\eta}^{catr}, \hat{\theta}^{Pcatr})$. Adapting the work of [2] and [7], we now consider the

following weighted empirical influence functions for $\boldsymbol{Q}_n(\boldsymbol{\gamma}_0^*)$.

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

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

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

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

Let $\hat{\boldsymbol{v}}_{2i} = \{\hat{v}_{2i}^{(1)}, \hat{v}_{2i}^{(2)}, \hat{v}_{2i}^{(3)}\}^T$. and $\hat{\boldsymbol{v}}_i = (\hat{\boldsymbol{v}}_{1i}^T, \hat{\boldsymbol{v}}_{2i}^T)^T$. The estimator of $\boldsymbol{\Omega}_0^*$ is

$$\hat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{v}}_i\hat{\boldsymbol{v}}_i^T. \tag{8}$$

It is important to consider the variability arising from the propensity score modeling. In nonrandomized studies, the true propensity scores adjust for the imbalance between confounders. However, the estimated propensity score is indeed a random variable, so the variability from propensity score modeling always exists. Since the empirical influence functions contain weights based on the estimated propensity score, the weights impact the variance of $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$. Ignoring these weights results in inflation of variance of the estimators of interest $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$ because if the propensity score is treated as known, it leads to a decrease in precision for $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$.

Now let us consider estimation of the covariance matrix of $\hat{\boldsymbol{\gamma}}$. In this case, the main issue is estimation of $\boldsymbol{\Lambda}_0$. As discussed in [2] and [7], direct estimation of $\boldsymbol{\Lambda}_0^*$ involves the estimation of unknown hazard function of error terms,

which is numerically unstable. Although the estimating functions are continuous with respect to $\boldsymbol{\alpha}$, the derivatives of estimating functions with respect to $\boldsymbol{\alpha}$ have a very complicated form. The resampling approach from [17] 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 $\boldsymbol{\Lambda}_0$, so it is a suitable approach for nonsmooth estimating equations such as those in this paper. We extended approach from [17] to estimate the covariance of $\hat{\boldsymbol{\gamma}}$. As discussed before, it is important to incorporate the sampling variability of $\hat{\boldsymbol{\alpha}}$ when quantifying the variability of $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$. Let $\boldsymbol{U}_n(\boldsymbol{\gamma}) = [S_n^T(\eta^{tr}, \boldsymbol{\alpha}), \{U_n^L(\beta^{tr}, \boldsymbol{\alpha})\}^T, \{U_n^P(\beta^{tr}, \boldsymbol{\alpha})\}^T]^T$. Given the data, the following stochastic equations are solved.

$$\left(\begin{array}{l} \boldsymbol{G}_n(\boldsymbol{\alpha}) = -n^{1/2}\sum_{i=1}^{n} \hat{\boldsymbol{v}}_{1i} A_i \\ \\ \boldsymbol{U}_n(\boldsymbol{\gamma}) = -n^{1/2}\sum_{i=1}^{n} \hat{\boldsymbol{v}}_{2i} A_i \end{array} \right), \tag{9}$$

where $A_i, i = 1, \ldots, n$ 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 $\boldsymbol{G}_n(\boldsymbol{\alpha})$ to obtain $\hat{\boldsymbol{\alpha}}^*$; Step 2: we update $w_i(\hat{\boldsymbol{\alpha}}^*), i = 1, \ldots, n$ with $\hat{\boldsymbol{\alpha}}^*$ from Step 1 and solve $\boldsymbol{U}_n(\boldsymbol{\gamma})$ to obtain $\hat{\boldsymbol{\tau}}^* = \{\hat{\eta}^{catr*}, \hat{\theta}^{Lcatr*}, \hat{\theta}^{Pcatr*}\}^T$ with updated weights $w_i(\hat{\boldsymbol{\alpha}}^*)$ by plugging in $\boldsymbol{U}_n(\boldsymbol{\gamma})$. By repeating these two steps a sufficiently large number of times, we obtain $\hat{\boldsymbol{\gamma}}^* = \{(\hat{\boldsymbol{\alpha}}^*)^T, (\hat{\boldsymbol{\tau}}^*)^T\}^T$ and the covariance matrix can be estimated based on $\hat{\boldsymbol{\gamma}}^*$. Note that in this procedure, the only random part is A_i while the observed data are treated as fixed [7]. Moreover, the variability of $\hat{\boldsymbol{\alpha}}$ is incorporated in the estimators of interest through the updating weights. If we ignore variability of the propensity score, we do not solve the first equation of (9) and only solve the second equation by using $w_i(\hat{\boldsymbol{\alpha}}), i = 1, \ldots, n$.

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. By the conditions (A.1) - (A.9) and (B.1) - (B.2) in the Supplementary Materials, and Theorem 3, conditional on observed data, the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}}^* - \hat{\boldsymbol{\gamma}})$ is same as the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$.

5 Real Data Analysis

5.1 RTOG (Radiation Therapy Oncology Group) 9413 dataset

We analyze two datasets alluded to in the Introduction: the RTOG (Radiation Therapy Oncology Group) 9413 dataset and the HIV dataset. In the RTOG study, the primary interest here is to compare the time to disease progression between Whole Pelvic Radiotherapy (WPRT) and Prostate Only Radiotherapy (PORT) among the intermediate risk patients, a subset of all randomized 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 (nadir+2ng/mL). 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, and the others as categorical variables. A total of 677 patients who met the intermediate risk criteria were included in this analysis. 200 times of resampling and bootstrap runs are tried. We first consider two conventional methods: fitting the model employing all covariates (Full model) and fitting the model using treatment variable only using approaches in [2] and [7] (Unadjusted). When employing all covariates to fit the model, in 200 resampling runs, 69 resampling runs give nonconvergence, which implies that the approaches in [2] and [7] are not stable.

(Table 1 should be here)

Table 1 shows the results. 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 [2] and [7].

To examine how many observations are artificially censored, we computed the artificial censoring rate. Let CR_D be the censoring rate subject to independent censoring, which is defined as $1 - \sum_{i=1}^{n} \xi_i/n$ and CR_X be the censoring rate subject to the dependent censoring, which is defined as $1 - \sum_{i=1}^{n} \delta_i/n$. Let ACR_{FL} be the artificial censoring rate from [7] approach and ACR_{FP} be the one from [2] approach considering all covariates. Let $\hat{\boldsymbol{\beta}}^{LF}$ and $\hat{\boldsymbol{\beta}}^{PF}$ are estimators from approaches in [2] and [7] from including all covariates in the model. Using arguments

in [18], mathematical definitions of these quantities are

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

where $\tilde{\delta}_i^{full*}$ and $\tilde{\delta}_{i(j)}^{full*}$ are artificial censoring indicators using all covariates using approaches from [2] and [7], respectively. Let ACR_{AL} be the artificial censoring rate from proposed method from [7] approach and ACR_{AP} be the one from the proposed method of [2]. Then

$$ACR_{AL} = 1 - \frac{\sum_{i=1}^{n} \tilde{\delta}_i^{*}(\hat{\beta}^{Lcatr})}{\sum_{i=1}^{n} \delta_i}$$
$$ACR_{AP} = 1 - \frac{\sum_{i=1}^{n} \sum_{j \neq i} \tilde{\delta}_{i(j)}^{*}(\hat{\beta}^{Pcatr})}{(n-1)\sum_{i=1}^{n} \delta_i}.$$

Based on the above quantities, the artificial censoring rates of the full model by [2] and [7] are 0.806 and 0.196, respectively. The artificial censoring rates from the proposed approaches from Section 3 are 0.034 and 0.048, which uses more uncensored observations than the full model approach.

5.2 HIV dataset

Although the 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. Range of CD4 count is 1589.5. Including CD4 count in 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 purpose of illustration, we combine EFV only with NFV+EFV and treat NFV as the reference level. Thus if one analyzes the dataset including the CD4 count, excessive artificial censoring occurs. For estimating treatment effect including this covariate, our proposed method is useful. In data analysis, one observed value is removed because it has CD4 count 0.

(Table 2 should be here)

Similar to the RTOG data analysis, we considered the following models: proposed approach with propensity score adjustment, conventional approaches (Lin and P&F) with covariate adjustment (Full model) or without 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 from the conventional approaches using all covariates are 1 for both the methods in [2] and [7]. On the other hand, the artificial censoring rates by using the proposed method is 0.069 for the approach based on [7] type and 0.013 for [2] type, respectively.

6 Simulation Studies

6.1 Assuming the true propensity score model

Next, we perform some simulations to explore the finite-sample properties of the proposed methodology. Note that true model is (1), so data is generated according to model (1). We generate a confounder $V \sim N(0, 4)$ and simulated treatment variable Z as Bernoulli random variable with probability $\exp(\boldsymbol{\alpha}_0^T \boldsymbol{H})[1 + \exp(\boldsymbol{\alpha}_0^T \boldsymbol{H})]^{-1}$, where $\boldsymbol{H} = (1, \boldsymbol{V}^T)^T$ and $\boldsymbol{\alpha}_0 = (\alpha_1, \alpha_2) = (0, 0.5)^T$. Then error variable $\boldsymbol{\epsilon}$ 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 $\boldsymbol{\theta}_0 = (1, 0.5)^T$ and $\boldsymbol{\eta}_0 = (0.5, 1)^T$. 500 datasets are simulated and with each simulated dataset, 500 bootstrap runs and 500 resampling runs mentioned in the previous section are tried.

We calculated bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage probability (CP) of estimators from considering all covariates and proposed estimator. In this case, the coverage probability is based on the empirical distribution based on resampling runs or bootstrap runs. As with the real data analysis case, in the proposed method, we calculated the standard error of the estimators in three ways : the resampling method which ignores vari-

ability of the propensity score (Naive), the bootstrap method (Bootstrap) and the resampling method incorporating variability of the propensity score (Resamp).

(Table 3 should be here)

Table 3 and 4 show the numerical results for treatment variable considering entire covariates and the proposed methodology. Henceforth the procedure considering entire covariates is denoted as "the full covariates procedure". The estimators from utilizing all covariates for the dependent censoring and the event of interest by [2] and [7] are denoted as $\hat{\boldsymbol{\eta}}^F$, $\hat{\boldsymbol{\theta}}^{LF}$ and $\hat{\boldsymbol{\theta}}^{PF}$, respectively. Moreover, the estimators for Z from utilizing all covariates for the dependent censoring and the event of interest by [2] and [7] 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, and the event of interest by approaches from [2] and [7], simulation runs are removed if the standard errors of any one of these estimators in the simulation runs are 0. 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 [7] for either confounder or treatment are 0. Similarly, when $N = 500$, 40, 77 and 168 results are omitted. This noncovergence problem is serious especially $\rho = 0.5$ and $N = 500$, 33.6% of entire simulation runs are abandoned, which is relatively large. However, for the proposed method, all 500 runs for the estimation of the treatment effect converge.

(Table 4 should be here)

If the simulation runs corresponding to only the estimators for dependent censoring and the estimator for the event of interest by [2] in full covariate procedure are considered, there are no simulation runs which have standard error zero. It can be seen that the method of [7] 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 true is high so that coverage probability is conservative. However, the bootstrap and resampling approaches provide desired coverage. When comparing coverage of estimators between the full covariate procedure and resampling and bootstrap procedures, the proposed approaches has better coverage relative to the procedure in [7]. Table 5 shows the artificial censoring proportion. To examine the effect of the artificial censoring, entire simulation runs are included in this calculation. When considering all covariates, the artificial censoring rate by the approach in [7] is high. However, the artificial censoring rate of the proposed method based on [7] is small, which implies that loss of observations for the proposed methodology is much smaller

than that in the full covariates procedure.

(Table 5 should be here)

We also performed numerical studies when we do not assume the correct logistic regression relationship between treatment variable and confounders. We generate $\boldsymbol{J} = (J_1, J_2)^T$ 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 are shown in Section B of the Supplementary Materials. The performance is similar to the case when the true relationship is parametric model, but the bias of the proposed estimator is higher than when logistic regression model is true for Case 1 and Case 2. This phenomenon is expected because the model for propensity score is misspecified, so it may impact performance of the proposed estimator.

6.2 Simulation study using HIV dataset

In these simulation studies, the method in [2] has better performance than that in [7]. Moreover, the method in [2] has good performance relative to the proposed approach. However, if the variability of covariates is large, even the method in [2] does not provide correct coverage for estimator. We perform a simulation study using the HIV dataset to show that Peng and Fine [2] estimator fails to provide proper standard error. First, by using estimates from real data analysis, the time to event of interest and time to dependent censoring are generated

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

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 $\boldsymbol{L}^* = (1, V_1^T, V_2^T)^T$. In this case, Z is generated as Bernoulli(p), where $p = \frac{\exp(\boldsymbol{\alpha}_0^T \boldsymbol{L}^*)}{1+\exp(\boldsymbol{\alpha}_0^T \boldsymbol{L}^*)}$ and $\boldsymbol{\alpha}_0$ 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 $\begin{pmatrix} 1.02 & 0.42 \\ 0.42 & 0.42 \end{pmatrix}$. Independent censoring time C has uniform distribution with minimum value 3 and maximum value 10. In each simulation run, 150 observations (without replacement) are selected in observations. 500 times of resampling and bootstrap runs are tried.

(Table 6 should be here)

As in the real data study, including all covariates results in excessive artificial censoring. Only 22 runs of estimators from [7] give nonzero standard errors and no simulation run of estimators using [2] provides nonzero standard errors for all covariates. Thus, we cannot obtain correct coverage for all covariates by using the approaches in [2] and [7]. Table 6 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 [2] and [7] 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 [2] and [7] are 0.071 and 0.048, respectively. In this study, the method in [2] 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 methodology for estimating treatment effects under a semicompeting risks data structure in the context of an observational study. In semicompeting risks data, only one nonterminal event is the event of interest. In medical studies, it is 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 [14] in the sense that the purpose is to estimate treatment effects by using weights. However, the approach of [14] assumes independent censoring and use probability of censoring as weight, while our approach allows dependent censoring and use propensity score as weight. Moreover, our proposed estimator is indeed the estimated true average treatment effect, but estimator in [14] targets treatment effect under misspecified model.

As can be seen, 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

$$\begin{pmatrix} X_i = \theta_0^{tr} Z_i + e_i(\boldsymbol{\alpha}_0) + \zeta_i^X \\ D_i = \eta_0^{tr} Z_i + e_i(\boldsymbol{\alpha}_0) + \zeta_i^D \end{pmatrix} i = 1, \ldots, n$$

where $(\zeta_i^X, \zeta_i^D)^T, i = 1, \ldots, n$ are error terms. This approach enables us to perform inference about confounder through the propensity score. Research for modeling the influence of confounders and treatment effect 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, [12] and [13] 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 [12] and [13].

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 we employ the approach 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 outcome approach to establish causal relationship between treatment and event of interest, we need several strong (and also untestable) assumptions. Robins [24] proposes assumptions to make causal inference in randomized studies. Similar to [24], we illustrate several assumptions for causal interpretation in semicompeting risks data.

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

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

This assumption implies that if possible confounding mechanisms are known, treatment assignment does not affect potential outcomes. In usual right censored data with single event and independent censoring assumption, this assumption should be modified due to the existence of censoring. Let $D(1)$ and $D(0)$ be time to death for treatment group and control group, $C(1)$ and $C(0)$ be time to censoring for treatment group and control group, 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]

$$\{D(1), D(0)\} \perp (Z, C)|\boldsymbol{V} \tag{10}$$

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,

$$\{X(1), X(0), D(1), D(0)\} \perp (Z, C)|\boldsymbol{V} \tag{11}$$

2. Noninteraction assumption: error distribution is independent of treatment assignment.

3. Rank preservation assumption: if subject i is dependently censored earlier than j under treatment, hypothetically, subject i is also dependently censored earlier than j under control [6].

The first assumption is made in [2] and [7], including this paper. However, the second assumption has much stronger than the first one. If there exists a scientific example to reasonably satisfy this second assumption, then our approach can be a tool to provide causal interpretation.

Acknowledgement

This research is supported by National Institutes of Health grant R01 CA129102 and U10 CA180822.

References

1. 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**: 398-407.

2. Peng L, Fine JP. Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American statistical Association* 2006; **101**: 1085-1093.

3. 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**: 1904-1911.

4. Roach M 3rd, Hunt D, Lawton CA, Hsu I, Lustig RA, Seider M, Christopher JU, Thomas CR, Shipley WU, Sandler H. Radiation Therapy Oncology Group (RTOG) 9413: A randomized trial comparing whole pelvic radiation therapy (WPRT) to prostate only (PORT) and neoadjuvant hormonal therapy (NHT) to adjuvant hormonal therapy (AHT). *International Journal of Radiation Oncology · Biology · Physics* 2013; **87**: S106-S107.

5. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika* 2001; **88**: 907-919.

6. Varadhan R, Xue Q-L, Bandeen-Roche K. Semicompeting risks in aging research: methods, issues and needs. *Lifetime Data Analysis*, 2014; **20**: 538-562.

7. Lin DY, Robins JM, Wei LJ. Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* 1996; **83**: 381-393.

8. 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**: 481–500.

9. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983; **70**: 41-55.

10. Williamson EJ, Forbes A, White IR. Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Statistics in Medicine*, 2014; **33**: 721-737.

11. 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**: 2937-2960.

12. Zhu Y. *Data-adaptive approaches to modeling propensity scores in causal inference problems.* Ph.D. Thesis, The Pennsylvania State University, 2013.

13. 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**: 69-91.

14. Boyd AP, Kittelson JM, Gillen DL. Estimation of treatment effect under non-proportional hazards and conditionally independent censoring. *Statistics in Medicine* 2012; **31**: 3504-3515.

15. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. *Biometrika*, 1986; **73**: 363-369.

16. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American statistical Association* 1989; **84**: 1074-1078.

17. Parzen MI, Wei LJ, Ying Z. A resampling method based on pivotal estimating functions. *Biometrika* 1994; **81**: 341-350.

18. Hsieh J-J, Ding AA, Wang W. Regression analysis for recurrent events data under dependent censoring. *Biometrics* 2011; **67**: 719-729.

19. Ghosh D, Lin DY. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* 2003; **59**: 877-885.

20. Ghosh D. Semiparametric analysis of recurrent events: artificial censoring, truncation, pairwise estimation and inference. *Lifetime Data Analysis* 2010; **16**: 509-524.

21. Vansteelandt S, Daniel RM. On regression adjustment for the propensity score. *Statistics in Medicine*, 2014; **33**: 4053-4072.

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**: 1898-1902.

24. Robins JM. An analytic method for randomized trials with informative censoring: Part 1. *Lifetime Data Analysis*, 1995; **1**: 241-254.

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**, 830-839.

26. Ying Z. A large sample study of rank estimation for censored regression data. *The Annals of Statistics*, 1993; **21**: 76-99.

27. Ferguson TS. *A Course in Large Sample Theory. Chapman and Hall*: London, 1996.

28. Ghosh D. *Nonparametric and semiparametric analysis of recurrent events in the presence of terminal events and dependent censoring.* Ph.D. thesis, University of Washington, 2000.

29. Stefanski LA, Boos DD. The calculus of M-estimation. *The American Statistician*, 2002; **56**: 29-38.

30. Honoré BE, Powell JL. Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics* 1994; **64**: 241-278.

31. Fleming TR, Harrington DP. *Counting Processes and Survival Analysis*, second edition, Wiley: New York, 2005.

Table 1: Point estimates of treatment effect and standard error (SE) in RTOG data analysis

	Proposed approach			
	Estimates	SE (Naive)	SE (Bootstrap)	SE (Resamp)
Death	0.064	0.062	0.055	0.059
First disease occurrence (Lin type)	-0.077	0.119	0.11	0.115
First disease occurrence (P&F type)	-0.112	0.101	0.096	0.097
	Conventional approach			
	Full model		Unadjusted	
	Estimates	SE	Estimates	SE
Death	0.076	0.019	0.076	0.065
First disease occurrence (Lin)	-0.059	0.004	-0.05	0.119
First disease occurrence (P&F)	-0.062	0.011	-0.091	0.101

Lin type : [7] type approach P&F type : [2] type approach
Lin : method in [7] P&F : method in [2]

Table 2: Point estimates and standard error (SE) in HIV data analysis

	Proposed approach			
	Estimates	SE (Naive)	SE (Bootstrap)	SE (Resamp)
Withdrawal	0.745	0.15	0.149	0.148
First viologic failure (Lin type)	0.936	0.241	0.221	0.228
First viologic failure (P&F type)	0.766	0.239	0.217	0.225
	Conventional approach			
	Full model		Unadjusted	
	Estimates	SE	Estimates	SE
Withdrawal	0.782	0.016	0.749	0.151
First viologic failure (Lin)	1.101	0	1.021	0.24
First viologic failure (P&F)	1.174	0	0.818	0.237

Lin type : [7] type approach P&F type : [2] type approach
Lin : method in [7] P&F : method in [2]

Table 3: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for estimators for treatment effect for Z including all covariates when $N = 250$ and $N = 500$

		$N = 250$				$N = 500$			
		Bias	EMPSD	SEE	CP	Bias	EMPSD	SEE	CP
$\rho = 0$	$\hat{\eta}^{F,tr}$	0.001	0.152	0.154	0.946	-0.008	0.131	0.109	0.952
	$\hat{\theta}^{LF,tr}$	0.007	0.187	0.132	0.833	0.012	0.135	0.125	0.891
	$\hat{\theta}^{PF,tr}$	-0.01	0.149	0.154	0.957	0.002	0.104	0.109	0.961
$\rho = 0.25$	$\hat{\eta}^{F,tr}$	0.004	0.159	0.155	0.961	-0.006	0.126	0.11	0.967
	$\hat{\theta}^{LF,tr}$	-0.01	0.199	0.122	0.753	-0.008	0.145	0.118	0.875
	$\hat{\theta}^{PF,tr}$	0.001	0.153	0.151	0.938	0.002	0.101	0.107	0.965
$\rho = 0.5$	$\hat{\eta}^{F,tr}$	-0.001	0.167	0.156	0.962	-0.011	0.123	0.11	0.958
	$\hat{\theta}^{LF,tr}$	-0.021	0.192	0.109	0.702	-0.035	0.153	0.108	0.834
	$\hat{\theta}^{PF,tr}$	0.004	0.149	0.15	0.952	0.0008	0.105	0.105	0.952

Estimators - $\hat{\eta}^{F,tr}$: the estimator of dependent censoring for Z in the full model ; $\hat{\theta}^{LF,tr}$: the estimator by [7] for Z in the full model; $\hat{\theta}^{PF,tr}$: the estimator by [2] for Z in the full model

Table 4: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for the proposed estimator when $N = 250$ and $N = 500$

$N = 250$									
		Bias	EMPSD	SEE			CP		
				Naive	Bootstrap	Resamp	Naive	Bootstrap	Resamp
$\rho = 0$	$\hat{\eta}^{catr}$	0.006	0.196	0.249	0.191	0.192	0.98	0.924	0.934
	$\hat{\theta}^{Lcatr}$	-0.011	0.242	0.406	0.241	0.25	1	0.946	0.94
	$\hat{\theta}^{Pcatr}$	-0.015	0.192	0.382	0.197	0.219	1	0.954	0.94
$\rho = 0.25$	$\hat{\eta}^{catr}$	0.014	0.187	0.25	0.191	0.192	0.98	0.944	0.946
	$\hat{\theta}^{Lcatr}$	0.009	0.241	0.405	0.238	0.247	1	0.936	0.938
	$\hat{\theta}^{Pcatr}$	-0.0004	0.19	0.381	0.194	0.216	0.998	0.956	0.956
$\rho = 0.5$	$\hat{\eta}^{catr}$	0.015	0.187	0.249	0.191	0.193	0.98	0.944	0.948
	$\hat{\theta}^{Lcatr}$	0.014	0.237	0.402	0.235	0.245	1	0.94	0.944
	$\hat{\theta}^{Pcatr}$	0.003	0.182	0.377	0.188	0.211	0.998	0.95	0.952
$N = 500$									
		Bias	EMPSD	SSE			CP		
				Naive	Bootstrap	Resamp	Naive	Bootstrap	Resamp
$\rho = 0$	$\hat{\eta}^{catr}$	0.013	0.137	0.176	0.134	0.134	0.986	0.95	0.952
	$\hat{\theta}^{Lcatr}$	0.008	0.17	0.284	0.166	0.168	0.998	0.934	0.942
	$\hat{\theta}^{Pcatr}$	-0.002	0.126	0.27	0.134	0.14	1	0.966	0.956
$\rho = 0.25$	$\hat{\eta}^{catr}$	0.02	0.135	0.175	0.133	0.133	0.988	0.934	0.934
	$\hat{\theta}^{Lcatr}$	0.017	0.166	0.282	0.163	0.164	0.996	0.94	0.928
	$\hat{\theta}^{Pcatr}$	0.001	0.131	0.268	0.131	0.136	1	0.956	0.948
$\rho = 0.5$	$\hat{\eta}^{catr}$	0.018	0.134	0.175	0.133	0.133	0.988	0.944	0.946
	$\hat{\theta}^{Lcatr}$	0.016	0.165	0.281	0.161	0.163	0.994	0.936	0.95
	$\hat{\theta}^{Pcatr}$	0.002	0.127	0.266	0.127	0.133	1	0.942	0.94

Estimators : $\hat{\eta}^{catr}$: the proposed estimator of the dependent censoring ; $\hat{\theta}^{Lcatr}$: the proposed estimator using [7] approach; $\hat{\theta}^{Pcatr}$: the proposed estimator using [2] approach

Table 5: Artificial censoring proportions when assuming logistic regression model is true

$N = 250$						
	CR_D[a]	CR_X[b]	ACR_{FL}[c]	ACR_{FP}[d]	ACR_{AL}[e]	ACR_{AP}[f]
$\rho = 0$	0.09	0.214	0.779	0.148	0.066	0.065
$\rho = 0.25$	0.09	0.195	0.843	0.144	0.064	0.064
$\rho = 0.5$	0.09	0.167	0.906	0.133	0.065	0.066
$N = 500$						
	CR_D[a]	CR_X[b]	ACR_{FL}[c]	ACR_{FP}[d]	ACR_{AL}[e]	ACR_{AP}[f]
$\rho = 0$	0.088	0.215	0.823	0.148	0.064	0.065
$\rho = 0.25$	0.089	0.193	0.883	0.142	0.066	0.068
$\rho = 0.5$	0.088	0.166	0.941	0.133	0.066	0.068

[a] the censoring rate subject to the independent censoring

[b] the censoring rate subject to the dependent censoring

[c] the artificial censoring rate from [7] approach considering all covariates

[d] the artificial censoring rate from [2] approach considering all covariates

[e] the artificial censoring rate from proposed method of [7] approach

[f] the artificial censoring rate from the proposed method of [2] approach

Table 6: 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 [7] approach; $\hat{\theta}^{Pcatr}$: the proposed estimator using [2] approach

Appendix

A. Technical Proofs

For proofs, we assume the following:

(A.1) The parameter space $\mathbf{\Gamma}$ is compact.

(A.2) Let

$$Q_1^{zw}(t;\eta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha})I\{\tilde{D}_1^*(\eta^{tr}) \geq t\}Z_1] \quad Q_1(t;\eta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha})I\{\tilde{D}_1^*(\eta^{tr}) \geq t\}]$$
$$Q_2^{zw}(t;\beta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha})I\{\tilde{X}_1^*(\beta^{tr}) \geq t\}Z_1] \quad Q_2(t;\beta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha})I\{\tilde{X}_1^*(\beta^{tr}) \geq t\}].$$

Denote the filtration as $\mathcal{F}_t = \{N_{1i}^w(u;\eta^{tr},\boldsymbol{\alpha}_0), N_{2i}^w(u;\eta^{tr},\boldsymbol{\alpha}_0), Y_{1i}^w(u;\eta^{tr},\boldsymbol{\alpha}_0), Y_{2i}^w(u;\beta^{tr},\boldsymbol{\alpha}_0), Z_i;\ i = 1,\ldots,n; 0 \leq u < t\}$, where

$$N_{1i}^w(t;\eta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{D}_i^*(\eta^{tr}) \leq t, \xi_i = 1\} = w_i(\boldsymbol{\alpha})\xi_i I\{\tilde{D}_i^*(\eta^{tr}) \leq t\}$$
$$N_{2i}^w(t;\beta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{X}_i^*(\beta^{tr}) \leq t, \tilde{\delta}_i^*(\beta^{tr}) = 1\} = w_i(\boldsymbol{\alpha})\tilde{\delta}_i^*(\beta^{tr})I\{\tilde{X}_i^*(\beta^{tr}) \leq t\}$$
$$Y_{1i}^w(t;\eta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{D}_i^*(\eta^{tr}) \geq t\}$$
$$Y_{2i}^w(u;\beta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{X}_i^*(\beta^{tr}) \geq t\}.$$

Define

$$m_1^w(t;\eta^{tr},\boldsymbol{\alpha}) = E\Big\{\int_{-\infty}^{t}\Big[Z_1 - \frac{E\{Q_1^{zw}(u;\eta^{tr},\boldsymbol{\alpha})\}}{E\{Q_1^w(u;\eta^{tr},\boldsymbol{\alpha})\}}\Big]dN_{11}^w(u;\eta^{tr},\boldsymbol{\alpha})\Big\}.$$
$$m_2^w(t;\beta^{tr},\boldsymbol{\alpha}) = E\Big\{\int_{-\infty}^{t}\Big[Z_1 - \frac{E\{Q_2^{zw}(u;\beta^{tr},\boldsymbol{\alpha})\}}{E\{Q_2^w(u;\beta^{tr},\boldsymbol{\alpha})\}}\Big]dN_{21}^w(u;\beta^{tr},\boldsymbol{\alpha})\Big\}$$

Moreover, $\boldsymbol{\gamma}_0 = (\boldsymbol{\alpha}_0, \eta_0^{tr}, \theta_0^{tr})^T$ is an interior point of $\mathbf{\Gamma}$.

(A.3) $\boldsymbol{W}$ has finite second moment.

(A.4) The solutions of $\boldsymbol{G}_n(\boldsymbol{\alpha}) = 0, S_n(\eta^{tr},\boldsymbol{\alpha}_0) = 0, U_n^L(\theta^{tr},\eta_0^{tr},\boldsymbol{\alpha}_0)$ and $U_n^P(\theta^{tr},\eta_0^{tr},\boldsymbol{\alpha}_0) = 0$ are unique.

(A.5) For $i = 1,\ldots,n$, there exists $s > 0$ and $r > 0$ such that $0 < s \leq w_i(\boldsymbol{\alpha}) \leq r < \infty$ for all $\boldsymbol{\alpha}$.

(A.6) For $i = 1,\ldots,n$, define

$$L_i^{(w,1)}(u) = Z_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^n I(\tilde{D}_j^*(\eta_0^{tr}) \geq u)w_j(\boldsymbol{\alpha}_0)}$$
$$L_i^{(w,2)}(u) = Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)}.$$

Then $L_i^{(w,1)}(\cdot)$ and $L_i^{(w,2)}(\cdot)$ are $\mathcal{F}^-$ predictable.

(A.7) Existence of limiting quantities : For every $u > 0$, there exist $\bar{z}^{(w,1)}(\cdot) > 0$ and $\bar{z}^{(w,2)}(\cdot) > 0$ such that

$$\bar{z}^{(w,1)}(u) = \lim_{n\to\infty} \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\} w_j(\boldsymbol{\alpha}_0)}$$

$$\bar{z}^{(w,2)}(u) = \lim_{n\to\infty} \frac{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta_0^{tr}) \geq u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta_0^{tr}) \geq u\} w_j(\boldsymbol{\alpha}_0)}.$$

(A.8) Let $\tilde{\lambda}(\beta^{tr}, \boldsymbol{\alpha}) = E\{n^{-1/2} U_n^P(\beta^{tr}, \boldsymbol{\alpha})\}$. Assume that $\tilde{\lambda}(\beta^{tr}, \boldsymbol{\alpha})$ is differentiable at β_0^{tr}, and both $\left.\frac{\partial \tilde{\lambda}(\beta^{tr}, \boldsymbol{\alpha})}{\partial \eta^{tr}}\right|_{\beta^{tr}=\beta_0^{tr}, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$ and $\left.\frac{\partial \tilde{\lambda}(\beta^{tr}, \boldsymbol{\alpha})}{\partial \theta^{tr}}\right|_{\beta^{tr}=\beta_0^{tr}, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$ are nonsingular.

(A.9) Condition 3 in [26] : C_is have uniformly bounded densities.

Note that (A.1), (A.2) and (A.3) are fundamental to proving asymptotic results. Condition 1 in [26] is satisfied in our case because we consider a binary treatment variable. (A.2), (A.3), (A.4), (A.6), (A.7) and (A.9) are needed for proving theoretical result based on [7] approach. (A.2), (A.4) and (A.8) are conditions needed for proving the theoretical results of Peng and Fine [2] type estimator. (A.5) guarantees that the weight neither diverges nor goes to 0. Moreover, we need assumptions dealing with the logistic regression model for propensity scores. These assumptions are from [27] (Chapter 17, p. 114), [12] and [13]. Let

$$\boldsymbol{H} = (1, \boldsymbol{V}^T)^T, \quad e(\boldsymbol{\alpha}) = \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H})}{1 + \exp(\boldsymbol{\alpha}^T \boldsymbol{H})}$$

(B.1) The parameter $\boldsymbol{\alpha}$ belongs to a compact subset of $\boldsymbol{\Gamma} \in \mathbb{R}^k$, where $\mathbb{R}^k$ is a k-dimensional Euclidean space. The likelihood

$$f(z, \boldsymbol{\alpha}) = e(\boldsymbol{\alpha})^z (1 - e(\boldsymbol{\alpha}))^{1-z},$$

is measurable in z for every $\boldsymbol{\alpha}$ in $\boldsymbol{\Gamma}$. Moreover, f is continuous in $\boldsymbol{\alpha}$ for every z.

(B.2) For all z and $\boldsymbol{\alpha}$,

$$\log\left(\frac{f(z, \boldsymbol{\alpha}|\boldsymbol{H})}{f(z, \boldsymbol{\alpha}_0|\boldsymbol{H})}\right) \leq h(z),$$

where $h(z)$ is a function satisfying $E_{\boldsymbol{\alpha}_0}|h(z)| < \infty$.

Now we proof the first theorem.

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

Proof. By balancing property of propensity score, $Z_i \perp \boldsymbol{V}_i | e(\boldsymbol{V}_i)$. Based on modeling assumption (1) in main paper, ϵ_i is independent of $\boldsymbol{W}_i$ and i.i.d. Hence ϵ_i is independent of Z_i. Note that $(X_i - \theta_0^{tr} Z_i, D_i - \eta_0^{tr} Z_i) = (\epsilon_i^X + \boldsymbol{\theta}_0^{cfd} \boldsymbol{V}_i, \epsilon_i^D + \boldsymbol{\eta}_0^{cfd} \boldsymbol{V}_i)$ Then given propensity score, the result follows. □

Now we can add one more assumption for modified error distribution that depends on confounders ([15],[16]).

(A.10) Conditions 2 and 4 in [26]. Let $\tilde{\epsilon}_i^X = X_i - \theta_0^{tr} Z_i$ and $\tilde{\epsilon}_i^D = D_i - \eta_0^{tr} Z_i$, for $i = 1, \ldots, n$. Denote $\tilde{\epsilon}_i = (\tilde{\epsilon}_i^X, \tilde{\epsilon}_i^D)^T, i = 1, \ldots, n$. Note that for Condition 2 in [26], in our context, given propensity score, the error density is the common joint density of $\{\tilde{\epsilon}_i\}_{i=1}^n$. Let $g(\cdot)$ be common density of $\{\tilde{\epsilon}_i | e^*(\boldsymbol{V}_i)\}_{i=1}^n$. We assume that $g(\cdot)$ and its derivative are bounded (Ghosh, 2000, Chapter 6), and

$$\int_{-\infty}^{\infty} \left(\frac{\dot{g}(t)}{g(t)} \right)^2 g(t) dt < \infty,$$

where $\dot{g}(t) = \dfrac{dg}{dt}$. Condition 4 in [26] and assumption (A.9) imply that $\sup_i E|\min\{\tilde{\epsilon}_i, C_i\}|^{\nu_0} < \infty$ for some $\nu_0 > 0$.

Now we will show that $\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)$ is unbiased estimating function, i.e., $E\{\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)\} = 0$.

Lemma 1. $E\{\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)\} = 0$

Proof. Recall that our proposed estimating function for dependent censoring is

$$S_n(\eta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} w_i(\boldsymbol{\alpha}) \left(Z_i - \frac{\sum_{j=1}^n w_j(\boldsymbol{\alpha}) I\{\tilde{D}_j^*(\eta) \geq t\} Z_j}{\sum_{j=1}^n w_j(\boldsymbol{\alpha}) I\{\tilde{D}_j^*(\eta) \geq t\}} \right) dN_{1i}(t; \eta)$$

where $N_{1i}(t; \eta) = I\{\tilde{D}_i^*(\eta) \leq t\}$ and

$$w_i(\boldsymbol{\alpha}) = \frac{Z_i}{e_i(\boldsymbol{\alpha}_0)} + \frac{1 - Z_i}{1 - e_i(\boldsymbol{\alpha}_0)}$$

Then

$$E\{S_n(\eta_0,\boldsymbol{\alpha}_0)\}=E\bigg[\int_{-\infty}^{\infty}f_D(t|\boldsymbol{V})S_C(t|\boldsymbol{V})w_1(\boldsymbol{\alpha}_0)\bigg(Z_1-\frac{E[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0^{tr})\geq t\}Z_1]}{E[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0^{tr})\geq t\}]}\bigg)\bigg]$$

Let $D_i^*(\eta_0)=D-\eta_0 Z$, $f_D(t|\boldsymbol{V})$ be common density of $D_i^*(\eta_0)$ and $S_C(t|\boldsymbol{V})$ be common survival function of $C_i-\eta_0 Z$. Given propensity score,

$$E\bigg[\int_{-\infty}^{\infty}E\bigg\{f_D(t|\boldsymbol{V})S_C(t|\boldsymbol{V})w_1(\boldsymbol{\alpha}_0)\bigg(Z_1-\frac{E[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0)\geq t\}Z_1]}{E[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0)\geq t\}]}\bigg)\bigg|e_1(\boldsymbol{\alpha}_0)\bigg\}\bigg]$$
$$=E\bigg[\int_{-\infty}^{\infty}f_D(t|\boldsymbol{V})S_C(t|\boldsymbol{V})E\bigg\{w_1(\boldsymbol{\alpha}_0)\bigg(Z_1-\frac{E[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0)\geq t\}Z_1]}{E[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0)\geq t\}]}\bigg)\bigg|e_1(\boldsymbol{\alpha}_0)\bigg\}\bigg]$$

Then we have

$$\frac{E\{E[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0)\geq t\}Z_1]|e_1(\boldsymbol{\alpha}_0)\}}{E\{E[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0)\geq t\}]|e_1(\boldsymbol{\alpha}_0)\}}=\frac{E[E\{Z_1w_1(\boldsymbol{\alpha}_0)\}E[I\{\tilde{D}_1^*(\eta_0)\geq t\}]|e_1(\boldsymbol{\alpha}_0)]}{E[E\{w_1(\boldsymbol{\alpha}_0)\}E[I\{\tilde{D}_1^*(\eta_0)\geq t\}]|e_1(\boldsymbol{\alpha}_0)]}$$
$$=\frac{P\{\tilde{D}_1^*(\eta_0)\geq t\}}{2P\{\tilde{D}_1^*(\eta_0)\geq t\}}=0.5$$

The second equality follows because

$$E\{w_1(\boldsymbol{\alpha}_0)\}=E\bigg(\frac{Z_1}{e_1(\boldsymbol{\alpha}_0)}+\frac{1-Z_1}{1-e_1(\boldsymbol{\alpha}_0)}\bigg)=2$$
$$E\{Z_1w_1(\boldsymbol{\alpha}_0)\}=E\bigg(\frac{Z_1^2}{e_1(\boldsymbol{\alpha}_0)}+\frac{Z_1(1-Z_1)}{1-e_1(\boldsymbol{\alpha}_0)}\bigg)=1$$

Hence

$$E\bigg[\int_{-\infty}^{\infty}f_D(t|\boldsymbol{V})S_C(t|\boldsymbol{V})E\bigg\{w_1(\boldsymbol{\alpha}_0)(Z_1-0.5)\bigg|e_1(\boldsymbol{\alpha}_0)\bigg\}\bigg]=$$
$$E\bigg[\int_{-\infty}^{\infty}f_D(t|\boldsymbol{V})S_C(t|\boldsymbol{V})(1-2\times 0.5)\bigg]=0$$

Thus $E\{S_n(\eta_0^{tr},\boldsymbol{\alpha}_0)=0$. Similarly, we can show that $E\{U_n^L(\beta_0^{tr},\boldsymbol{\alpha}_0)\}=0$. Recall that $\boldsymbol{\theta}_0=(\theta_0^{tr},(\boldsymbol{\theta}_0^{cfd})^T)^T$ and $\boldsymbol{\eta}_0=(\eta_0^{tr},(\boldsymbol{\eta}_0^{cfd})^T)^T$. Note that $\boldsymbol{\theta}_0^{cfd}$ is the true value of $\boldsymbol{\theta}_0$ corresponding to $\boldsymbol{V}$ and $\boldsymbol{\eta}_0^{cfd}$ is the true value of $\boldsymbol{\eta}_0$ corresponding $\boldsymbol{V}$. Under the true value of β_0^{tr} and $\boldsymbol{\alpha}_0$,

$$E\{U_n^P(\beta_0^{tr},\boldsymbol{\alpha}_0)\}=E[(Z_1-Z_2)w_1(\boldsymbol{\alpha}_0)w_2(\boldsymbol{\alpha}_0)\times\{P(\epsilon_1^X+(\boldsymbol{\theta}_0^{cfd})^T\boldsymbol{V}_1$$
$$\leq\{\epsilon_1^D+(\boldsymbol{\eta}_0^{cfd})^T\boldsymbol{V}_1-d_{12}(\beta_0^{tr})\}\wedge\{\epsilon_2^X+(\boldsymbol{\theta}_0^{cfd})^T\boldsymbol{V}_2\}\wedge\{\epsilon_2^D+(\boldsymbol{\eta}_0^{cfd})^T\boldsymbol{V}_2-d_{12}(\beta_0^{tr})\}$$
$$\wedge\{C_1-\eta_0^{tr}Z_1-d_{12}(\beta_0^{tr})\}\wedge\{C_2-\eta_0^{tr}Z_2-d_{12}(\beta_0^{tr})\}|e_1(\boldsymbol{\alpha}_0),e_2(\boldsymbol{\alpha}_0),Z_1,Z_2)$$
$$-P(\epsilon_2^X+(\boldsymbol{\theta}_0^{cfd})^T\boldsymbol{V}_2\leq\{\epsilon_2^D+(\boldsymbol{\eta}_0^{cfd})^T\boldsymbol{V}_2-d_{12}(\beta_0^{tr})\}\wedge\{\epsilon_1^X+(\boldsymbol{\theta}_0^{cfd})^T\boldsymbol{V}_1\}$$
$$\wedge\{\epsilon_1^D+(\boldsymbol{\eta}_0^{cfd})^T\boldsymbol{V}_1-d_{12}(\beta_0^{tr})\}\wedge\{C_1-\eta_0^{tr}Z_1-d_{12}(\beta_0^{tr})\}$$
$$\wedge\{C_2-\eta_0^{tr}Z_2-d_{12}(\beta_0^{tr})\}|e_1(\boldsymbol{\alpha}_0),e_2(\boldsymbol{\alpha}_0),Z_1,Z_2)\}].$$

By Theorem 1,

$$
\begin{aligned}
&P(\epsilon_1^X + (\boldsymbol{\theta}_0^{cfd})^T \boldsymbol{V}_1 \leq \{\epsilon_1^D + (\boldsymbol{\eta}_0^{cfd})^T \boldsymbol{V}_1 - d_{12}(\beta_0^{tr})\} \wedge \{\epsilon_2^X + (\boldsymbol{\theta}_0^{cfd})^T \boldsymbol{V}_2\} \\
&\wedge \{\epsilon_2^D + (\boldsymbol{\eta}_0^{cfd})^T \boldsymbol{V}_2 - d_{12}(\beta_0^{tr})\} \wedge \{C_1 - \eta_0^{tr} Z_1 - d_{12}(\beta_0^{tr})\} \\
&\wedge \{C_2 - \eta_0^{tr} Z_2 - d_{12}(\beta_0^{tr})\} | e_1(\boldsymbol{\alpha}_0), e_2(\boldsymbol{\alpha}_0), Z_1, Z_2) \\
&= P(\epsilon_2^X + (\boldsymbol{\theta}_0^{cfd})^T \boldsymbol{V}_2 \leq \{\epsilon_2^D + (\boldsymbol{\eta}_0^{cfd})^T \boldsymbol{V}_2 - d_{12}(\beta_0^{tr})\} \wedge \{\epsilon_1^X + (\boldsymbol{\theta}_0^{cfd})^T \boldsymbol{V}_1\} \\
&\wedge \{\epsilon_1^D + (\boldsymbol{\eta}_0^{cfd})^T \boldsymbol{V}_1 - d_{12}(\beta_0^{tr})\} \wedge \{C_1 - \eta_0^{tr} Z_1 - d_{12}(\beta_0^{tr})\} \\
&\wedge \{C_2 - \eta_0^{tr} Z_2 - d_{12}(\beta_0^{tr})\} | e_1(\boldsymbol{\alpha}_0), e_2(\boldsymbol{\alpha}_0), Z_1, Z_2).
\end{aligned}
$$

Thus $E\{U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$.

□

Theorem 2. Assuming conditions (A.1) - (A.9) and (B.1) - (B.2), $\hat{\boldsymbol{\gamma}}$ is strongly consistent for $\boldsymbol{\gamma}_0$.

Proof. Since the propensity model is true, by the conditions (B.1) and (B.2), $\hat{\boldsymbol{\alpha}}$ converges to $\boldsymbol{\alpha}_0$ almost surely. Define

$$
N_1^w(t; \eta^{tr}, \boldsymbol{\alpha}) = \sum_{i=1}^n N_{1i}^w(t; \eta^{tr}, \boldsymbol{\alpha}) \quad N_1^{zw}(t; \eta^{tr}, \boldsymbol{\alpha}) = \sum_{i=1}^n Z_i N_{1i}^w(t; \eta^{tr}, \boldsymbol{\alpha})
$$

$$
Q_{1n}^{zw}(t; \eta^{tr}, \boldsymbol{\alpha}) = \sum_{j=1}^n I\{\tilde{D}_j^*(\eta^{tr}) \geq t\} w_j(\boldsymbol{\alpha}) Z_j \quad Q_{1n}^w(t; \eta^{tr}, \boldsymbol{\alpha}) = \sum_{j=1}^n I\{\tilde{D}_j^*(\eta^{tr}) \geq t\} w_j(\boldsymbol{\alpha})
$$

$$
\tilde{N}_{1i}^w(t; \eta^{tr}, \boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha}) I\{\tilde{D}_i^*(\eta^{tr}) \geq t, \xi_i = 1\} \quad \tilde{N}_1^w(t; \eta^{tr}, \boldsymbol{\alpha}) = \sum_{i=1}^n \tilde{N}_{1i}^w(t; \eta^{tr}, \boldsymbol{\alpha})
$$

$$
\tilde{N}_1^{zw}(t; \eta^{tr}, \boldsymbol{\alpha}) = \sum_{i=1}^n \tilde{N}_{1i}^w(t; \eta^{tr}, \boldsymbol{\alpha}) Z_i.
$$

Then let

$$
S_n(t; \eta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^t \left[Z_i - \frac{Q_{1n}^{zw}(u; \eta^{tr}, \boldsymbol{\alpha})}{Q_{1n}^w(u; \eta^{tr}, \boldsymbol{\alpha})} \right] dN_{1i}^w(u; \eta^{tr}, \boldsymbol{\alpha})
$$

Note that the proposed estimating function for η^{tr} is $S_n(\infty; \eta^{tr}, \boldsymbol{\alpha})$. [26] argues that the expansion holds if $\boldsymbol{W}$ is treated as random. In his case, the full covariates are used for residual terms. In our case, due to Theorem 1, the argument of [26] still holds. Let $L(t; \eta^{tr}, \boldsymbol{\alpha})$ be any one of the empirical processes $\tilde{N}^w, \tilde{N}^{zw}, Q_{1n}^{zw}, Q_{1n}^w, N_1^w$ and N_1^{zw}. Take $0 \leq \zeta < 1, C_1^* > 0, K >$

$0, \omega > 0$. The approximation of Lemma 1 in [26] still holds for $L(t; \eta^{tr}, \boldsymbol{\alpha})$. For any neighborhood of η_0^{tr},

$$\sup_{|\eta^{tr}| \le C_1^*, EL(t;\eta^{tr},\boldsymbol{\alpha}_0) \le Kn^{1-\zeta}} |L(t; \eta^{tr}, \hat{\boldsymbol{\alpha}}) - E\{L(t; \eta^{tr}, \boldsymbol{\alpha}_0)\}| = o(n^{(1-\zeta)/2+\omega}). \quad (1)$$

Let C_1^*, C_2^*, ζ^* and ζ^{**} be positive constants. Denote $m_1(\eta^{tr}, \boldsymbol{\alpha}) = m_1(\infty; \eta^{tr}, \boldsymbol{\alpha})$ and $m_2(\beta^{tr}, \boldsymbol{\alpha}) = m_2(\infty; \beta^{tr}, \boldsymbol{\alpha})$. By [26], (1) leads to the conclusion that the estimating function $n^{-1/2} S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}})$ can be uniformly approximated by a nonrandom function $m_1^w(\eta^{tr}, \boldsymbol{\alpha}_0)$ by some order within the bounded area, i.e.,

$$\sup_{|\eta^{tr}| \le C_1^*} |n^{-1/2} S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}}) - m_1^w(\eta^{tr}, \boldsymbol{\alpha}_0)| = o(n^{-1/2+\zeta^*}),$$

Since $n^{-1/2} U_n^L(\beta^{tr}, \hat{\boldsymbol{\alpha}})$ is also a log-rank type estimating function, the approximation (1) is applicable and we can derive a uniform approximation by the nonrandom function $m_2^w(\beta^{tr}, \boldsymbol{\alpha}_0)$. Hence,

$$\sup_{||\beta^{tr}|| \le C_2^*} |n^{-1/2} U_n^L(\beta^{tr}, \hat{\boldsymbol{\alpha}}) - m_2^w(\beta^{tr}, \boldsymbol{\alpha}_0)| = o(n^{-1/2+\zeta^{**}}).$$

Denote any neighborhood of η_0^{tr} and β_0^{tr}, say $\mathcal{N}_0$ and $\mathcal{N}_1$, respectively. By using the strong consistency of $\hat{\boldsymbol{\alpha}}$ and uniqueness of $\boldsymbol{\alpha}_0$ based on the arguments in [26], we can show that

$$\sup_{\eta^{tr} \in \mathcal{N}_0} |n^{-1/2} S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}}) - m_1^w(\eta^{tr}, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0$$
$$\sup_{\beta^{tr} \in \mathcal{N}_1} |n^{-1/2} U_n^L(\beta^{tr}, \hat{\boldsymbol{\alpha}}) - m_2^w(\beta^{tr}, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

Moreover, by strong consistency of $\hat{\boldsymbol{\alpha}}$, it is also true that if $\boldsymbol{\alpha}$ belongs on any neighborhood of $\boldsymbol{\alpha}_0$, say $\mathcal{B}$, then any fixed $\eta^{tr} \in \mathcal{N}_0$, $n^{-1/2} S_n(\eta^{tr}, \boldsymbol{\alpha})$ can be approximated by $m_1(\eta^{tr}, \boldsymbol{\alpha})$. Hence,

$$\sup_{\boldsymbol{\alpha} \in \mathcal{B}, \eta^{tr} \in \mathcal{N}_0} |n^{-1/2} S_n(\eta^{tr}, \boldsymbol{\alpha}) - m_1^w(\eta^{tr}, \boldsymbol{\alpha})| \xrightarrow{p} 0$$

Similarly, for any $\beta^{tr} \in \mathcal{N}_1$ and $\boldsymbol{\alpha} \in \mathcal{B}$,

$$\sup_{\boldsymbol{\alpha} \in \mathcal{B}, \beta^{tr} \in \mathcal{N}_1} |n^{-1/2} U_n^L(\beta^{tr}, \boldsymbol{\alpha}) - m_2^w(\beta^{tr}, \boldsymbol{\alpha})| \xrightarrow{p} 0.$$

Thus $\hat{\eta}^{catr}$ and $\hat{\theta}^{Lcatr}$ are strongly consistent to η_0^{tr} and θ_0^{tr}. Now we want to show strong consistency of $\hat{\theta}^{Pcatr}$. Let $\mathcal{W}$ be a compact set of parameter β^{tr}. From the U-statistics law of large numbers,

$$|n^{-1/2} U_n^P(\beta^{tr}, \boldsymbol{\alpha}_0) - \tilde{\lambda}(\beta, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

for all $\beta^{tr} \in \mathcal{W}$. By decomposing the compact set $\mathcal{W}$ into several finite subsets $\mathcal{W}_1, \ldots, \mathcal{W}_m$ such that $\mathcal{W} \in \cup_{i=1}^m \mathcal{W}_i$, for $(\beta^{tr})^i \in \mathcal{W}_i$,

$$\max_{1 \le i \le m} |n^{-1/2} U_n^P((\beta^{tr})^i, \boldsymbol{\alpha}_0) - \tilde{\lambda}((\beta^{tr})^i, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

Since $w_i(\boldsymbol{\alpha}_0), i = 1 \ldots n$ are bounded, by Appendix of [2],

$$\sup_{||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu} n^{-1/2} |U_n^P(\beta^{tr}, \boldsymbol{\alpha}_0) - U_n^P(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0)| \le \frac{2}{n(n-1)} [\sum_{1 \le i < j \le n} |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \nu)],$$

where

$$\begin{aligned}
L_{ij}^{(1)}(\tilde{\beta}^{tr}, \nu) &= w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) I[\{\beta^{tr} : |\beta^{tr} - \tilde{\beta}^{tr}| \le \nu, \tilde{X}^*_{i(j)}(\beta^{tr}) = \tilde{X}^*_{j(i)}(\beta^{tr})\} \ne \emptyset] \\
L_{ij}^{(2)}(\tilde{\beta}^{tr}, \nu) &= w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) I[\{\beta^{tr} : |\beta^{tr} - \tilde{\beta}^{tr}| \le \nu, \tilde{\delta}^*_{i(j)}(\beta^{tr}) \ne \tilde{\delta}^*_{i(j)}(\tilde{\beta}^{tr})\} \ne \emptyset] \\
K_{ij}(\tilde{\beta}^{tr}, \nu) &= \{L_{ij}^{(1)}(\tilde{\beta}^{tr}, \nu) + L_{ij}^{(2)}(\tilde{\beta}^{tr}, \nu) + L_{ji}^{(2)}(\tilde{\beta}^{tr}, \nu)\}.
\end{aligned}$$

Let $H_{ij}(\tilde{\beta}^{tr}, \nu) = |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \nu)$ and $H(\tilde{\beta}^{tr}, \nu) = \sum_{1 \le i < j \le n} |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \nu)$. By Hoeffding decomposition,

$$H(\tilde{\beta}^{tr}, \nu) - E\{H(\tilde{\beta}^{tr}, \nu)\} = \sum_{i=1}^n B_i(\tilde{\beta}^{tr}, \nu) + \sum_{i<j} B_{ij}(\tilde{\beta}^{tr}, \nu),$$

where

$$\begin{aligned}
B_i(\tilde{\beta}^{tr}, \nu) &= \sum_{j \ne i} [E\{H_{ij}(\tilde{\beta}^{tr}, \nu) | Z_i, e_i(\boldsymbol{\alpha}_0)\} - E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\}] \\
B_{ij}(\tilde{\beta}^{tr}, \nu) &= H_{ij}(\tilde{\beta}^{tr}, \nu) - E\{H_{ij}(\tilde{\beta}^{tr}, \nu) | Z_i, e_i(\boldsymbol{\alpha}_0)\} - E\{H_{ij}(\tilde{\beta}^{tr}, \nu) | Z_j, e_j(\boldsymbol{\alpha}_0)\} \\
&\quad + E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\}.
\end{aligned}$$

To complete the proof of consistency, we need the following lemma.

Lemma 2. There exist constants b_0 and c_0 such that $E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \le b_0 \nu$ and $E\{H(\tilde{\beta}^{tr}, \nu)\} \le c_0 \nu n^2$

The proof of Lemma 2 will be provided later. By Lemma 2 above, $E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \le b_0 \nu$ and $E\{H(\tilde{\beta}^{tr}, \nu)\} \le c_0 \nu n^2$. Note that B_i and B_{ij} are uncorrelated. Hence there exist $v_{10} > 0$ and $v_{20} > 0$ such that

$$Var\{H(\tilde{\beta}^{tr}, \nu)\} = \sum_{i=1}^n Var\{B_i(\tilde{\beta}^{tr}, \nu)\} + \sum_{i<j} Var\{B_{ij}(\tilde{\beta}^{tr}, \nu)\} \le v_{10} n^3 + v_{20} n^2 = O(n^3).$$

Take $\epsilon_0 > 0$ and let $0 < \nu < \epsilon_0/(3b_0)$. Then by Markov inequality,

$$P\{[n(n-1)]^{-1}H(\tilde{\beta}^{tr},\nu) \geq \epsilon\} \leq P\{[n(n-1)]^{-1}[H(\tilde{\beta}^{tr},\nu) - E\{H(\tilde{\beta}^{tr},\nu)]\} \geq \epsilon/3\}$$
$$\leq \frac{9Var\{H(\tilde{\beta}^{tr},\nu)\}}{[n(n-1)]^2\epsilon^2} \to 0.$$

Hence

$$\sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}_0) - U_n^P(\tilde{\beta}^{tr},\boldsymbol{\alpha}_0)| \xrightarrow{p} 0. \tag{2}$$

Note that for any $\nu^* > 0$,

$$\begin{aligned}
&\sup_{||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*,||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}) - U_n^P(\tilde{\beta}^{tr},\boldsymbol{\alpha}_0)| \\
&\leq \sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}) - U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}})| \\
&+ \sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} n^{-1/2}|U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}}) - U_n^P(\beta^{tr},\boldsymbol{\alpha}_0)| \\
&+ \sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}_0) - U_n^P(\tilde{\beta}^{tr},\boldsymbol{\alpha}_0)|.
\end{aligned}$$

For fixed β^{tr}, by Taylor series expansion,

$$n^{-1/2}|U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}}) - U_n^P(\beta^{tr},\boldsymbol{\alpha}_0)| = n^{-1/2}\left|(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\frac{\partial}{\partial\boldsymbol{\alpha}}U_n^P(\beta^{tr},\boldsymbol{\alpha})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right| + o_p(1).$$

Then for fixed β^{tr}, $n^{-1/2}\frac{\partial}{\partial\boldsymbol{\alpha}}U_n^P(\beta^{tr},\boldsymbol{\alpha})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = O(1)$ and $\hat{\boldsymbol{\alpha}} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$, we have

$$\sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} n^{-1/2}|U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}}) - U_n^P(\beta^{tr},\boldsymbol{\alpha}_0)| \xrightarrow{p} 0. \tag{3}$$

By uniqueness of $\boldsymbol{\alpha}_0$ and strong consistency of $\hat{\boldsymbol{\alpha}}$,

$$\sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} |n^{-1/2}U_n^P(\beta^{tr},\boldsymbol{\alpha}) - n^{-1/2}U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}})| \xrightarrow{p} 0. \tag{4}$$

Combining (2), (3) and (4) yields,

$$\sup_{||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*,||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}) - U_n^P(\tilde{\beta}^{tr},\boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

Thus the consistency of $\hat{\theta}^{Pcatr}$ is proved. □

Remark. In this case, the key important element is strong consistency of $\hat{\boldsymbol{\alpha}}$. Given strong consistency of $\hat{\boldsymbol{\alpha}}$, the random function $n^{-1/2}S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}})$ and $n^{-1/2}U_n^L(\beta^{tr}, \hat{\boldsymbol{\alpha}})$ converge to deterministic functions. Moreover, to prove the strong consistency of $\hat{\theta}^{Pcatr}$, we apply arguments from [2] given $\boldsymbol{\alpha}_0$ and use strong consistency of $\hat{\boldsymbol{\alpha}}$.

Proof of Lemma 2 We can use arguments similar to those in [2] and the Appendix of [18]. Note that the set

$$\{||\beta^{tr} - \tilde{\beta}^{tr}|| \leq \nu, \tilde{\delta}^*_{i(j)}(\beta^{tr}) \neq \tilde{\delta}^*_{i(j)}(\tilde{\beta}^{tr})\} \in D_1(\tilde{\beta}^{tr}, \nu) \cup D_2(\tilde{\beta}^{tr}, \nu),$$

where

$$D_1(\tilde{\beta}^{tr}, \nu) = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, X_i - \theta^{tr} Z_i = [D_i + \eta^{tr}(Z_j - Z_i)] - \theta^{tr} Z_j\}$$
$$D_2(\tilde{\beta}^{tr}, \nu) = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, X_i - \theta^{tr} Z_i = [C_i + \eta^{tr}(Z_j - Z_i)] - \theta^{tr} Z_j\}.$$

Then

$$\begin{aligned}
&D_1(\tilde{\beta}^{tr}, \nu) \\
&= \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, X_i + \theta^{tr}(Z_j - Z_i) = D_i + \eta^{tr}(Z_j - Z_i)\} \\
&= \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, \epsilon_i^X + \boldsymbol{\theta}_0^T \boldsymbol{W}_i + \theta^{tr}(Z_j - Z_i) = \epsilon_i^D + \boldsymbol{\eta}_0^T \boldsymbol{W}_i + \eta^{tr}(Z_j - Z_i)\} \\
&= \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, \epsilon_i^X + \boldsymbol{\theta}_0^T \boldsymbol{W}_i + [\tilde{\theta}^{tr} - (\tilde{\theta}^{tr} - \theta^{tr})](Z_j - Z_i) \\
&= \epsilon_i^D + \boldsymbol{\eta}_0^T \boldsymbol{W}_i + [\tilde{\eta}^{tr} - (\tilde{\eta}^{tr} - \eta^{tr})](Z_j - Z_i)\} \\
&\subseteq \{||\epsilon_i^X - \epsilon_i^D + (\boldsymbol{\theta}_0 - \boldsymbol{\eta}_0)^T \boldsymbol{W}_i + (\tilde{\theta}^{tr} - \tilde{\eta}^{tr})(Z_j - Z_i)|| < 2\nu|Z_j - Z_i|\}.
\end{aligned}$$

Thus there exists $d_0 > 0$ such that $w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)P\{D_1(\tilde{\beta}^{tr}, \nu)|e_i(\boldsymbol{\alpha}_0), e_j(\boldsymbol{\alpha}_0), Z_i, Z_j\} \leq 2w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)d_0|Z_j - Z_i|\nu$ by assumption. Similarly,

$$w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)P\{D_2(\tilde{\beta}^{tr}, \nu)|e_i(\boldsymbol{\alpha}_0), e_j(\boldsymbol{\alpha}_0), Z_i, Z_j\} \leq 2w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)|Z_j - Z_i|\nu.$$

Hence, $E\{L_{ij}^{(2)}(\tilde{\beta}^{tr}, \nu)\} \leq 2d_0 E\{w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)|Z_j - Z_i|\}\nu$. Similarly, there exists $d_0^* > 0$ such that $E\{L_{ji}^{(2)}(\tilde{\beta}^{tr}, \nu)\} \leq 2d_0^* E\{w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)|Z_j - Z_i|\}\nu$ and there exists $f_0^* > 0$ and $E\{L_{ij}^{(1)}(\tilde{\beta}^{tr}, \nu)\} \leq 2f_0^* E\{w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)|Z_j - Z_i|\}\nu$. Since $w_i(\boldsymbol{\alpha}_0)$ are bounded, there exists $K_0 > 0$ and $K_1 > 0$ such that $E\{K_{ij}(\tilde{\beta}^{tr}, \nu)\} \leq K_0\nu$ and $E\{K_{ij}^2(\tilde{\beta}^{tr}, \nu)\} \leq K_1\nu$. By the Cauchy-Schwarz inequality,

$$E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \leq \sqrt{E\{K_{ij}^2(\tilde{\beta}^{tr}, \nu)\}E|Z_i - Z_j|^2} \leq K_1\nu\sqrt{E|Z_i - Z_j|^2}.$$

Hence there exists $b_0 > 0$ such that $E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \leq b_0\nu$. Finally, $E\{H(\tilde{\beta}^{tr}, \nu)\} \leq K_1 b_0 n^2$. Thus there exists $c_0 > 0$ such that $E\{H(\tilde{\beta}^{tr}, \nu)\} \leq c_0\nu n^2$. □

Before we state and prove Theorem 2, we first state and prove the following lemma.

Lemma 3. If k_n converges to 0 in probability,

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq k_n} \frac{||\boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) - \boldsymbol{\Lambda}_0 n^{1/2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o_p(1)$$

where

$$\boldsymbol{\Lambda}_0 = \begin{pmatrix} \boldsymbol{L}_1 & 0 & 0 & 0 \\ \boldsymbol{L}_2 & E_1 & 0 & 0 \\ \boldsymbol{L}_3 & E_2 & E_3 & 0 \\ \boldsymbol{L}_4 & E_4 & 0 & E_5 \end{pmatrix}$$

$$\boldsymbol{L}_1 = E\left[\frac{\partial \Psi_1(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right]$$

$$\boldsymbol{L}_2 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\{Z_1 - \bar{z}^{(w,1)}(t;\eta^{tr},\boldsymbol{\alpha})\}dN_{11}^w(t;\eta^{tr},\boldsymbol{\alpha})\right]_{\eta^{tr}=\eta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$\boldsymbol{L}_3 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\{Z_1 - \bar{z}^{(w,2)}(t;\beta^{tr},\boldsymbol{\alpha})\}dN_{21}^w(t;\beta^{tr},\boldsymbol{\alpha})\right]_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$\boldsymbol{L}_4 = \frac{\partial \lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\bigg|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$E_1 = \int_{-\infty}^{\infty} E\left[w_1(\boldsymbol{\alpha}_0) I\{\tilde{D}_1^*(\eta_0^{tr}) \geq t\}\{Z_1 - \bar{z}^{(w,1)}(t)\}\frac{\dot{\lambda}_{10}(t|\boldsymbol{V})}{\lambda_{10}(t|\boldsymbol{V})}g(t)dt\right]$$

$$E_2 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \eta^{tr}}\{Z_1 - \bar{z}^{(w,2)}(t;\beta^{tr},\boldsymbol{\alpha})\}dN_{21}^w(t;\beta^{tr},\boldsymbol{\alpha})\right]_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$E_3 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \theta^{tr}}\{Z_1 - \bar{z}^{(w,2)}(t;\beta^{tr},\boldsymbol{\alpha})\}dN_{21}^w(t;\beta^{tr},\boldsymbol{\alpha})\right]_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$E_4 = \frac{\partial \tilde{\lambda}(\beta^{tr},\boldsymbol{\alpha})}{\partial \eta^{tr}}\bigg|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \qquad E_5 = \frac{\partial \tilde{\lambda}(\beta^{tr},\boldsymbol{\alpha})}{\partial \theta^{tr}}\bigg|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

and $\lambda_{10}(\cdot|\boldsymbol{V})$ is the true common hazard function of $\{D_i^*(\eta_0^{tr})\}_{i=1}^n$ given the true propensity score $e_i(\boldsymbol{\alpha}_0)$ and $\dot{\lambda}_{10}(\cdot|\boldsymbol{V})$ is derivative of $\lambda_{10}(\cdot|\boldsymbol{V})$. If k_n converges to 0 almost surely,

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq k_n} \frac{||\boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) - \boldsymbol{\Lambda}_0 n^{1/2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o(1)$$

Proof Let $\Psi_i(\boldsymbol{\alpha}) = \boldsymbol{H}_i[Z_i - \frac{\exp(\boldsymbol{\alpha}^T\boldsymbol{H}_i)}{1+\exp(\boldsymbol{\alpha}^T\boldsymbol{H}_i)}]$. Then $\boldsymbol{G}_n(\boldsymbol{\alpha}) = n^{-1/2}\sum_{i=1}^n \Psi_i(\boldsymbol{\alpha})$. We will follow the approach of [28] (Chapter 6). Let

$$\begin{pmatrix} \boldsymbol{G}_n(\boldsymbol{\alpha}) \\ S_n(\eta^{tr}, \boldsymbol{\alpha}) \\ U_n^L(\beta^{tr}, \boldsymbol{\alpha}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{G}_n(\boldsymbol{\alpha}_0) \\ S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) \\ U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0) \end{pmatrix} + \begin{pmatrix} \boldsymbol{G}_n(\boldsymbol{\alpha}) - \boldsymbol{G}_n(\boldsymbol{\alpha}_0) \\ S_n(\eta^{tr}, \boldsymbol{\alpha}) - S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) \\ U_n^L(\beta^{tr}, \boldsymbol{\alpha}) - U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0) \end{pmatrix}$$

Clearly,

$$\boldsymbol{G}_n(\boldsymbol{\alpha}) - \boldsymbol{G}_n(\boldsymbol{\alpha}_0) = \dot{\boldsymbol{G}}_n(\boldsymbol{\alpha}_0)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + o_p(n^{1/2}||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0||)$$

where $\dot{\boldsymbol{G}}_n(\boldsymbol{\alpha}_0) = [\partial \boldsymbol{G}_n(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}]_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$ [29]. Let

$$\begin{aligned} \boldsymbol{\gamma}^{sub_1} &= (\boldsymbol{\alpha}^T, \eta^{tr})^T \quad \boldsymbol{\gamma}^{sub_2} = (\boldsymbol{\alpha}^T, \eta^{tr}, \theta^{tr})^T \\ \boldsymbol{\gamma}_0^{sub_1 *} &= (\boldsymbol{\alpha}_0^T, \eta_0^{tr})^T \quad \boldsymbol{\gamma}_0^{sub_2 *} = (\boldsymbol{\alpha}_0^T, \eta_0^{tr}, \theta_0^{tr})^T. \end{aligned}$$

By [26],

$$\begin{aligned} S_n(\eta^{tr}, \boldsymbol{\alpha}) - S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{1/2}\{\boldsymbol{L}_2(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + E_1(\eta^{tr} - \eta_0^{tr})\} + o_p(1 + n^{1/2}||\boldsymbol{\gamma}^{sub_1} - \boldsymbol{\gamma}_0^{sub_1 *}||) \\ U_n^L(\beta^{tr}, \boldsymbol{\alpha}) - U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{1/2}\{\boldsymbol{L}_3(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + E_2(\eta^{tr} - \eta_0^{tr}) + E_3(\theta^{tr} - \theta_0^{tr})\} \\ &\quad + o_p(1 + n^{1/2}||\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2 *}||). \end{aligned}$$

Since $w_i(\boldsymbol{\alpha})$ and $w_j(\boldsymbol{\alpha})$ are bounded by some constant, assumptions N1-N3 in [30] still hold. Hence by Lemma 2 of [30],

$$\sup_{\beta^{tr} \in \mathcal{N}_1, \boldsymbol{\alpha} \in \mathcal{B}} \frac{|U_n^P(\beta^{tr}, \boldsymbol{\alpha}) - U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0) - n^{1/2}\lambda(\beta^{tr}, \boldsymbol{\alpha})|}{1 + n^{1/2}|\lambda(\beta^{tr}, \boldsymbol{\alpha})|} = o_p(1),$$

By Taylor series expansion of $\lambda(\beta^{tr}, \boldsymbol{\alpha})$ at β_0^{tr} and $\boldsymbol{\alpha}_0$,

$$\begin{aligned} \tilde{\lambda}(\beta^{tr}, \boldsymbol{\alpha}) = \tilde{\lambda}(\beta_0^{tr}, \boldsymbol{\alpha}_0) &+ \begin{pmatrix} \left.\frac{\partial \tilde{\lambda}(\beta^{tr}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{\beta^{tr}=\beta_0^{tr}, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \\ \left.\frac{\partial \tilde{\lambda}(\beta^{tr}, \boldsymbol{\alpha})}{\partial \eta^{tr}}\right|_{\beta^{tr}=\beta_0^{tr}, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \\ \left.\frac{\partial \tilde{\lambda}(\beta^{tr}, \boldsymbol{\alpha})}{\partial \theta^{tr}}\right|_{\beta^{tr}=\beta_0^{tr}, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} - \boldsymbol{\alpha}_0 & \eta^{tr} - \eta_0^{tr} & \theta^{tr} - \theta_0^{tr} \end{pmatrix} \\ &+ o(||\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2 *}||). \end{aligned}$$

Extending arguments in the Appendix of [2],

$$\begin{aligned} U_n^P(\beta^{tr}, \boldsymbol{\alpha}^{tr}) = U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0^{tr}) &+ n^{1/2}\{\boldsymbol{L}_4(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + E_4(\eta^{tr} - \eta_0^{tr}) + E_5(\theta^{tr} - \theta_0^{tr})\} \\ &+ o_p(1 + n^{1/2}||\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2 *}||). \end{aligned}$$

Hence,

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq k_n} \frac{||\boldsymbol{Q}_n(\boldsymbol{\gamma}) - \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) - \boldsymbol{\Lambda}_0 n^{1/2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o_p(1),$$

when k_n converges in probability to zero. The second result of Lemma 2 easily follows from the result that the sequence of random variables converges in probability if and only if each subsequence contains a subsequence which converges almost surely. □

Theorem 3. By the conditions (A.1) - (A.9) and by Theorem 2 and Lemma 3, $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has an asymptotic normal distribution with mean 0 and covariance matrix $\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Lambda}_0^{-1}$. where $\boldsymbol{\Lambda}_0$ is a nonsingular matrix, and $\boldsymbol{\Omega}_0$ is the limiting covariance matrix of $\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)$.

Proof By the martingale central limit theorem (Theorem 5.3.5 in [31], pp.227-228) and U-statistic theory,

$$\begin{aligned}
S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{Z_i - \bar{z}^{(w,1)}(t)\}dM_{1i}^{w}(u;\eta_0^{tr},\boldsymbol{\alpha}_0) + o_p(1)\\
U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{Z_i - \bar{z}^{(w,2)}(t)\}dM_{2i}^{w}(u;\beta_0^{tr},\boldsymbol{\alpha}_0) + o_p(1)\\
U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n} 2h_1(Z_i,\beta_0^{tr},\boldsymbol{\alpha}_0) + o_p(1)\\
\boldsymbol{G}_n(\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\Psi_i(\boldsymbol{\alpha}_0),
\end{aligned}$$

where $h(Z_i, Z_j, \boldsymbol{V}_i, \boldsymbol{V}_j, \beta^{tr}, \boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)(Z_i - Z_j)\phi_{ij}(\beta^{tr})$ where $\phi_{ij}(\beta^{tr})$ is defined page 7 and $h_1(z,\beta^{tr},\boldsymbol{\alpha}_0) = E\{h(z,Z_2,\boldsymbol{V}_1,\boldsymbol{V}_2,\beta^{tr},\boldsymbol{\alpha}_0)\}$. Let $\boldsymbol{\tau} = (\eta^{tr},\theta^{tr},\theta^{tr})^T$ and $\boldsymbol{U}_n(\boldsymbol{\tau},\boldsymbol{\alpha}_0) = [S_n^T(\eta^{tr},\boldsymbol{\alpha}_0), \{U_n^L(\beta^{tr},\boldsymbol{\alpha}_0)\}^T, \{U_n^P(\beta^{tr},\boldsymbol{\alpha}_0)\}^T]^T$. By standard asymptotic theory of maximum likelihood estimator and the Cramér-Wold theorem,

$$\begin{pmatrix}\boldsymbol{G}_n(\boldsymbol{\alpha}_0)\\ \boldsymbol{U}_n(\boldsymbol{\tau}_{\mathbf{0}}^{*},\boldsymbol{\alpha}_0)\end{pmatrix} \xrightarrow{d} N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, E\begin{bmatrix}\boldsymbol{v}_1^*\boldsymbol{v}_1^{*T} & \boldsymbol{v}_1^*\boldsymbol{v}_2^{*T}\\ \boldsymbol{v}_2^*\boldsymbol{v}_1^{*T} & \boldsymbol{v}_2\boldsymbol{v}_2^{*T}\end{bmatrix}\right),$$

where

$$\boldsymbol{v}_1^* = \Psi_1(\boldsymbol{\alpha}_0) \quad \boldsymbol{v}_2^* = (v_{21}, v_{22}, v_{23})^T$$

$$v_{21} = \int_{-\infty}^{\infty}\{Z_i - \bar{z}^{(w,1)}(u)\}dM_{1i}^{w}(u;\eta_0^{tr},\boldsymbol{\alpha}_0) \quad v_{22} = \int_{-\infty}^{\infty}\{Z_i - \bar{z}^{(w,2)}(u)\}dM_{2i}^{w}(u;\beta_0^{tr},\boldsymbol{\alpha}_0)$$

$$v_{23} = 2h_1(Z_1,\beta_0^{tr},\boldsymbol{\alpha}_0).$$

By using Lemma 2 above,

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}) = \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) + n^{1/2}\boldsymbol{\Lambda}_0(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + o_p(1 + n^{1/2}||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0||),$$

for $\boldsymbol{\gamma} \in \mathcal{V}$, where $\mathcal{V}$ is any neighborhood of $\boldsymbol{\gamma}_0$. Then

$$\boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}) = \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) + n^{1/2}\boldsymbol{\Lambda}_0(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + o_p(1),$$

and by consistency of $\hat{\boldsymbol{\gamma}}$ and Lemma 3, we get

$$n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = -\boldsymbol{\Lambda}_0^{-1}\boldsymbol{Q}_n(\boldsymbol{\gamma}_0) + o_p(1)$$

Then by Slusky's theorem,

$$n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} N(0, \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Lambda}_0^{-1}),$$

where $\boldsymbol{\Omega}_0$ is

$$\boldsymbol{\Omega}_0 = E\begin{bmatrix} \boldsymbol{v}_1^*\boldsymbol{v}_1^{*T} & \boldsymbol{v}_1^*\boldsymbol{v}_2^{*T} \\ \boldsymbol{v}_2^*\boldsymbol{v}_1^{*T} & \boldsymbol{v}_2^*\boldsymbol{v}_2^{*T} \end{bmatrix}.$$

□

Theorem 4. By the conditions (A.1) - (A.9) and Theorem 3, conditional on observed data, the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}}^* - \hat{\boldsymbol{\gamma}})$ is same as the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$.

Proof To justify the resampling approach in [17], two conditions A(1·1) and A(1·2) in [17] should be verified. The condition A(1·1) follows from Lemma 2 directly. The condition A(1·2) implies that the root of estimating equation should be unique. This condition is also easily satisfied by the assumption. Note that $\boldsymbol{\gamma}^*$ is a solution of

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}) = -n^{-1/2}\sum_{i=1}^{n}\hat{\boldsymbol{v}}_i A_i,$$

From the Theorem 3, we showed that

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}) = \boldsymbol{Q}_n(\boldsymbol{\gamma}_0) + n^{1/2}\boldsymbol{\Lambda}_0(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + o_p(1 + n^{1/2}||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0||).$$

Then

$$\boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}^*) = \boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}) + n^{1/2}\boldsymbol{\Lambda}_0(\hat{\boldsymbol{\gamma}}^* - \hat{\boldsymbol{\gamma}}) + o_p(1)$$
$$n^{1/2}(\hat{\boldsymbol{\gamma}}^* - \hat{\boldsymbol{\gamma}}) = -\boldsymbol{\Lambda}_0^{-1}n^{-1/2}\sum_{i=1}^{n}\hat{\boldsymbol{v}}_i A_i + o_p(1).$$

Since the observed data are independent and identically distributed, given observed data, the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}}^* - \hat{\boldsymbol{\gamma}})$ is normal distribution with zero mean vector and covariance matrix $\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Lambda}_0^{-1}$. Hence the conditional distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}}^* - \hat{\boldsymbol{\gamma}})$ is asymptotically equal to the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$. □

B. Simulation study

In practice, it is difficult to check for the adequacy of the propensity score model. Thus another simulation study we performed was to explore the robustness of the proposed procedure. In this simulation study, it is not assumed that the propensity model relating confounder and treatment variable is true. We generate $\boldsymbol{J} = (J_1, J_2)^T$ from a bivariate normal distribution with mean $(0, 0)^T$ and covariance matrix $\begin{pmatrix} 4 & a \\ a & 1 \end{pmatrix}$, where $a = 0, 1$. In other words, J_1 has normal distribution with mean 0 and variance 4 marginally and J_2 has normal distribution with mean 0 and variance 1 marginally. When $a = 1$, J_1 and J_2 have correlation 0.5. Then we set $V = J_1$ 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 $\rho = 0.25$)

Table 7 and 8 show results for treatment variable from using the entire covariates in the model and from the proposed model. 54 runs, 43 runs, and 96 runs are removed for Case 1, Case 2 and Case 3, respectively when $N = 250$. For $N = 500$, 83 runs, 32 runs and 107 runs are removed for Case 1, Case 2 and Case 3, respectively. However, in the proposed approach, there is no numerical problem for estimating standard errors of estimators. The proposed method has the advantage compared to the full covariates approach in terms of numerical stability.

Numerical results indicate that the proposed method works well. In this case, as for assuming the true propensity model case, three ways of standard error calculation are used. As the simulation using true propensity score, the coverage probability is based on empirical distribution based on resampling runs or bootstrap runs. As for assuming the true propensity model case, $\hat{\eta}^{F,tr}$, $\hat{\theta}^{LF,tr}$ and $\hat{\theta}^{PF,tr}$ are the estimators for Z from utilizing all covariates for the dependent censoring and the event of interest by [2] and [7], respectively. As in the previous simulation scenario, treating estimated propensity score as true results in large standard error of estimators $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$. The

Table 7: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for using all covariates for Z when $N = 250$ and $N = 500$

		$N = 250$				$N = 500$			
		Bias	EMPSD	SEE	CP	Bias	EMPSD	SEE	CP
Case 1	$\hat{\eta}^{F,tr}$	-0.001	0.166	0.153	0.951	-0.009	0.124	0.109	0.945
	$\hat{\theta}^{LF,tr}$	0.003	0.203	0.123	0.78	0.002	0.148	0.118	0.894
	$\hat{\theta}^{PF,tr}$	0.011	0.151	0.151	0.953	-0.003	0.098	0.106	0.966
Case 2	$\hat{\eta}^{F,tr}$	-0.002	0.187	0.153	0.95	-0.012	0.162	0.11	0.949
	$\hat{\theta}^{LF,tr}$	0.025	0.189	0.137	0.84	0.013	0.145	0.123	0.885
	$\hat{\theta}^{PF,tr}$	0.011	0.155	0.154	0.937	0.003	0.102	0.108	0.968
Case 3	$\hat{\eta}^{F,tr}$	-0.037	0.269	0.143	0.941	-0.026	0.194	0.101	0.959
	$\hat{\theta}^{LF,tr}$	-0.005	0.204	0.125	0.787	0.001	0.145	0.116	0.883
	$\hat{\theta}^{PF,tr}$	-0.002	0.144	0.14	0.931	-0.007	0.096	0.098	0.952

Estimators - $\hat{\eta}^{F,tr}$: the estimator of the dependent censoring for Z in the full model; $\hat{\theta}^{LF,tr}$: the estimator by [7] for Z in the full model; $\hat{\theta}^{PF,tr}$: the estimator by [2] for Z in the full model

data bootstrap and new resampling approach reflected variation of estimated propensity score into $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$. Our methodology even works well for the randomized study. One of interesting feature is that more simulation runs are lost in randomized study than those in observational study. Even though our primary goal is to estimate treatment effect in observational study including continuous covariates which have large variations, it also works well for randomized study including continuous covariates with large variance.

Table 9 shows 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.

Table 8: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for proposed estimator when $N = 250$ and $N = 500$

$N = 250$									
		Bias	EMPSD	SSE			CP		
				naive	Bootstrap	Resamp	Naive	Bootstrap	Resamp
Case 1	$\hat{\eta}^{catr}$	0.027	0.183	0.242	0.186	0.186	0.986	0.958	0.958
	$\hat{\theta}^{Lcatr}$	0.033	0.232	0.387	0.227	0.234	0.992	0.94	0.948
	$\hat{\theta}^{Pcatr}$	0.025	0.19	0.373	0.191	0.21	1	0.968	0.974
Case 2	$\hat{\eta}^{catr}$	0.025	0.191	0.242	0.186	0.187	0.98	0.936	0.94
	$\hat{\theta}^{Lcatr}$	0.027	0.243	0.39	0.229	0.238	0.998	0.93	0.94
	$\hat{\theta}^{Pcatr}$	0.017	0.204	0.374	0.194	0.213	1	0.944	0.95
Case 3	$\hat{\eta}^{catr}$	0.009	0.152	0.201	0.154	0.154	0.984	0.952	0.958
	$\hat{\theta}^{Lcatr}$	0.004	0.189	0.313	0.18	0.181	0.998	0.94	0.934
	$\hat{\theta}^{Pcatr}$	0.012	0.161	0.302	0.157	0.157	1	0.946	0.94
$N = 500$									
		Bias	EMPSD	SSE			CP		
				Naive	Bootstrap	Resamp	Naive	Bootstrap	Resamp
Case 1	$\hat{\eta}^{catr}$	0.007	0.133	0.173	0.132	0.131	0.978	0.934	0.93
	$\hat{\theta}^{Lcatr}$	0.02	0.159	0.277	0.16	0.161	1	0.948	0.948
	$\hat{\theta}^{Pcatr}$	0.008	0.118	0.261	0.126	0.131	1	0.952	0.97
Case 2	$\hat{\eta}^{catr}$	0.012	0.134	0.173	0.132	0.132	0.988	0.936	0.94
	$\hat{\theta}^{Lcatr}$	0.022	0.158	0.277	0.162	0.162	1	0.956	0.962
	$\hat{\theta}^{Pcatr}$	0.009	0.122	0.262	0.13	0.134	1	0.966	0.966
Case 3	$\hat{\eta}^{catr}$	-0.0001	0.1	0.141	0.109	0.109	0.994	0.972	0.97
	$\hat{\theta}^{Lcatr}$	-0.002	0.121	0.221	0.128	0.127	0.998	0.96	0.958
	$\hat{\theta}^{Pcatr}$	-0.002	0.108	0.211	0.11	0.11	1	0.958	0.956

Estimators - $\hat{\eta}^{catr}$: the proposed estimator of the dependent censoring ; $\hat{\theta}^{Lcatr}$: the proposed estimator using [7] approach; $\hat{\theta}^{Pcatr}$: the proposed estimator using [2] approach

Table 9: Artificial censoring proportions when not assuming logistic regression model is true

$N = 250$						
	CR_D[a]	CR_X[b]	ACR_{FL}[c]	ACR_{FP}[d]	ACR_{AL}[e]	ACR_{AP}[f]
Case 1	0.088	0.194	0.839	0.143	0.062	0.064
Case 2	0.088	0.215	0.783	0.149	0.062	0.064
Case 3	0.078	0.2	0.887	0.158	0.054	0.053
$N = 500$						
	CR_D[a]	CR_X[b]	ACR_{FL}[c]	ACR_{FP}[d]	ACR_{AL}[e]	ACR_{AP}[f]
Case 1	0.088	0.193	0.877	0.143	0.062	0.064
Case 2	0.087	0.215	0.818	0.147	0.062	0.063
Case 3	0.077	0.2	0.917	0.157	0.053	0.053

[a] the censoring rate subject to the independent censoring

[b] the censoring rate subject to the dependent censoring

[c] the artificial censoring rate from [7] approach considering all covariates

[d] the artificial censoring rate from [2] approach considering all covariates

[e] the artificial censoring rate from proposed method of [7] approach

[f] the artificial censoring rate from the proposed method of [2] approach