

# Efficient and Robust Transfer Learning of Optimal Individualized Treatment Regimes with Right-Censored Survival Data

**Pan Zhao**

*Statistical Laboratory  
University of Cambridge*

PZ298@CAM.AC.UK

**Julie Josse**

*PreMeDICAL  
Inria Montpellier*

JULIE.JOSSE@INRIA.FR

**Shu Yang**

*Department of Statistics  
North Carolina State University*

SYANG24@NCSSU.EDU

**Editor:** David Sontag

## Abstract

An individualized treatment regime (ITR) is a decision rule that assigns treatments based on patients' characteristics. The value function of an ITR is the expected outcome in a counterfactual world had this ITR been implemented. Recently, there has been increasing interest in combining heterogeneous data sources, such as leveraging the complementary features of randomized controlled trial (RCT) data and a large observational study (OS). Usually, a covariate shift exists between the source and target population, rendering the source-optimal ITR not optimal for the target population. We present an efficient and robust transfer learning framework for estimating the optimal ITR with right-censored survival data that generalizes well to the target population. The value function accommodates a broad class of functionals of survival distributions, including survival probabilities and restrictive mean survival times (RMSTs). We propose a doubly robust estimator of the value function, and the optimal ITR is learned by maximizing the value function within a pre-specified class of ITRs. We establish the cubic rate of convergence for the estimated parameter indexing the optimal ITR, and show that the proposed optimal value estimator is consistent and asymptotically normal even with flexible machine learning methods for nuisance parameter estimation. We evaluate the empirical performance of the proposed method by simulation studies and a real data application of sodium bicarbonate therapy for patients with severe metabolic acidaemia in the intensive care unit (ICU), combining a RCT and an observational study with heterogeneity.

**Keywords:** covariate shift, data integration, policy learning, semiparametric theory, transportability

## 1. Introduction

Data-driven individualized decision making has recently received increasing interest in many fields, such as precision medicine (Kosorok and Laber, 2019; Tsiatis et al., 2019), mobile health (Trella et al., 2022), precision public health (Rasmussen et al., 2020) and econometrics

(Belloni et al., 2017; Athey and Wager, 2021). The goal of optimal ITR estimation is to learn a decision rule that assigns the best treatment among possible options to each patient based on their individual characteristics in order to optimize some functional of the counterfactual outcome distribution in the population of interest, also known as the value function. The optimal ITR is the one with the maximal value function, and the value function of the optimal ITR is the optimal value function.

For completely observed data without censoring, one prevailing line of work in the statistical and biomedical literature uses model-based methods to solve the optimal ITR problem, such as Q-learning (Robins, 2004; Qian and Murphy, 2011; Laber et al., 2014) and A-learning (Murphy, 2003; Schulte et al., 2014; Shi et al., 2018). Alternatively, direct model-free or policy search methods have been proposed recently, including the classification perspective (Zhang et al., 2012a,b; Zhao et al., 2012; Rubin and van der Laan, 2012) and interpretable tree or list-based ITRs (Laber and Zhao, 2015; Zhang et al., 2015, 2018a), among others (Chernozhukov et al., 2019; Colangelo and Lee, 2023). In clinical studies, right-censored survival data are frequently observed as primary outcomes. Recent extensions of optimal ITR with survival data have been established in Goldberg and Kosorok (2012); Cui et al. (2017); Jiang et al. (2017); Bai et al. (2017); Díaz et al. (2018); Zhou et al. (2023). It is worth noting that learning the optimal ITR is closely linked to the estimation of conditional average treatment effects, an area that has seen significant growth recently (Hatt et al., 2022; Guo et al., 2022; Kato et al., 2024; Demirel et al., 2024a).

Researchers have investigated using machine learning algorithms to estimate the optimal ITR from large classes, which cannot be indexed by a finite-dimensional parameter (Luedtke and van der Laan, 2016a,b). One typical instance is that the optimal ITR can be learned from the blip function, which is defined as the additive effect of a blip in treatment on a counterfactual outcome, conditional on baseline covariates (Robins, 2004); and most existing regression or supervised learning methods can be directly applied (Künzel et al., 2019). However, the ITRs learned by machine learning methods can be too complex to inform policy-making and clinical practice; to facilitate the integration of data-driven ITRs into practice, it is crucial that estimated ITRs be interpretable and parsimonious (Zhang et al., 2015).

While policy learning and evaluation under distributional shift have been investigated extensively in the machine learning literature, recently there has been increasing interest in combining heterogeneous data sources, such as leveraging the complementary features of RCT data and a large OS (Rudolph and van der Laan, 2017; Yang et al., 2023a). For example, in biomedical studies and policy research, RCTs are deemed as the gold standard for treatment effects evaluation. However, due to inclusion or exclusion criteria, data availability, and study design, the enrolled participants in RCT who form the source sample may have systematically different characteristics from the target population. Therefore, findings from RCTs cannot be directly extended to the target population of interest (Cole and Stuart, 2010; Dahabreh and Hernán, 2019). On the other hand, OSs are often representative of real-world patient populations, but may be subject to confounding bias. See also Colnet et al. (2024) and Degtiar and Rose (2023) for detailed reviews. Heterogeneity in the populations is of great relevance, and a *covariate shift* usually exists where the covariate distributions differ between the source and target populations; thus, the optimal ITR for the source population is not necessarily optimal for the target population. Zhao et al. (2019)

uses data from a single trial study and proposes a two-stage procedure to derive a robust and parsimonious rule for the target population; Mo et al. (2021) proposes a distributionally robust framework that maximizes the worst-case value function under a set of distributions that are “close” to the training distribution; Kallus (2021) tackles the lack of overlap for different actions in policy learning based on retargeting; Wu and Yang (2022) and Chu et al. (2023) develop a calibration weighting framework that tailors a targeted optimal ITR by leveraging the individual covariate data or summary statistics from a target population; Sahoo et al. (2022) uses distributionally robust optimization and sensitivity analysis tools to learn a decision rule that minimizes the worst-case risk incurred under a family of test distributions. However, these methods focus on continuous or binary outcomes and only consider a single sample for worst-case risk minimization; the extension to right-censored survival outcomes within the data integration context has not been studied.

We have the following contributions. We propose a new transfer learning method of finding an optimal ITR from a restricted ITR class under the super population framework where the source sample is subject to selection bias and the target sample is representative of the target population with a known sampling mechanism. Specifically, in our value search method, the value function accommodates a broad class of functionals of survival distributions, including survival probabilities and RMSTs.

We characterize the efficient influence function (EIF) (Newey, 1990, 1994) of the value function and propose the augmented estimator, which involves models for the survival outcome, propensity score, censoring and sampling processes. The proposed estimator is doubly robust in the sense that it is consistent if either the survival outcome model or the models of the propensity score, censoring, and sampling are correctly specified and is locally efficient when all models are correct. We also consider flexible data-adaptive machine learning algorithms to estimate the nuisance parameters and use the cross-fitting procedure to draw valid inferences under mild regularity conditions and a certain rate of convergence conditions (Semenova and Chernozhukov, 2020; Chernozhukov et al., 2022b).

As we consider a restricted class of ITRs indexed by a Euclidean parameter  $\eta$ , we also establish the cubic convergence rate of  $\hat{\eta}$ , even though its resultant limiting distribution is not standard, and thus very challenging to characterize. Based on this rate of convergence, we show that the proposed estimator for the target value function is consistent and asymptotically normal, even with flexible machine learning methods for nuisance parameter estimation. For causal ITR parameters identified by moment functions that depend on a first step unknown function, the “auto-DML” approach is a promising future direction (Farrell et al., 2021; Chernozhukov et al., 2022a).

Interestingly, when the covariate distributions of the source and target populations are the same, i.e., no covariate shift, the semiparametric efficiency bounds of our method and the standard doubly robust method (Bai et al., 2017) are equal. Moreover, if the true optimal ITR belongs to the restricted class of ITRs, the standard doubly robust method can still learn the optimal ITR despite the covariate shift, but only our method provides valid statistical inference for the value function.

## 2. Statistical Framework

### 2.1 Causal Survival Analysis

Let  $X$  denote the  $p$ -dimensional vector of covariates that belongs to a covariate space  $\mathcal{X} \subset \mathbb{R}^p$ ,  $A \in \mathcal{A} = \{0, 1\}$  denote the binary treatment, and  $T \in \mathbb{R}^+$  denote the *survival time* to the event of interest. In the presence of right censoring, the outcome  $T$  may not be observed. Let  $C \in \mathbb{R}^+$  denote the censoring time and  $\Delta = I\{T \leq C\}$  where  $I\{\cdot\}$  is the indicator function. Let  $U = \min\{T, C\}$  be the observed outcome,  $N(t) = I\{U \leq t, \Delta = 1\}$  the counting process, and  $Y(t) = I\{U \geq t\}$  the at-risk process.

We use the potential outcomes framework (Splawa-Neyman et al., 1990; Rubin, 1974), where for  $a \in \mathcal{A} = \{0, 1\}$ ,  $T(a)$  is the survival time had the subject received treatment  $a$ . The common goal in causal survival analysis is to identify and estimate the counterfactual quantity  $\mathbb{E}[y(T(a))]$  for some deterministic transformation function  $y(\cdot)$ . Such transformations include  $y(T) = \min(T, L)$  for the RMST with some pre-specified maximal time horizon  $L$ , and  $y(T) = I\{T \geq t\}$  for the survival probability at time  $t$ .

Under the standard assumptions (a) consistency:  $T = T(A)$ , (b) positivity:  $Pr(A = a | X) > 0$  for every  $a \in \mathcal{A}$  *almost surely*, (c) unconfoundedness:  $A \perp \{T(1), T(0)\} | X$ , (d) conditionally independent censoring:  $C \perp \{T(1), T(0)\} | \{X, A\}$ , we can nonparametrically identify  $\mathbb{E}[y(T(a))]$  by the outcome regression (OR) formula or the inverse probability weighting (IPW) formula (van der Laan and Robins, 2003).

### 2.2 ITR and Value Function

Without loss of generality, we assume that larger values of  $T$  are more desirable. Typically we aim to identify and estimate an ITR  $d(x) : \mathcal{X} \rightarrow \mathcal{A}$ , which is a mapping from the covariate space  $\mathcal{X}$  to the treatment space  $\mathcal{A} = \{0, 1\}$ , that maximizes the expected outcome in a counterfactual world had this ITR been implemented. Suppose  $\mathcal{D}$  is the class of candidate ITRs of interest, then define the potential outcome  $T(d)$  under any  $d \in \mathcal{D}$  by  $T(d) = d(X)T(1) + (1 - d(X))T(0)$ , and the value function (Manski, 2004) of  $d$  is defined by  $V(d) = \mathbb{E}[y(T(d))]$ . Then by maximizing  $V(d)$  over  $\mathcal{D}$ , the optimal ITR is defined by  $d^{\text{opt}} = \arg \max_{d \in \mathcal{D}} V(d)$ . See Qian and Murphy (2011) for more details.

To estimate the value function, we can use the OR or IPW formulas, and also a doubly robust method (Bai et al., 2017):

$$\begin{aligned}
 V_{DR}(d) = & \mathbb{E} \left[ \frac{I\{A = d(X)\} \Delta y(U)}{Pr(A = d(X) | X) S_C(U | A, X)} \right. \\
 & + \left( 1 - \frac{I\{A = d(X)\}}{Pr(A = d(X) | X)} \right) \mathbb{E}[y(T) | A = d(X), X] \\
 & \left. + \frac{I\{A = d(X)\}}{Pr(A = d(X) | X)} \int_0^\infty \frac{dM_C(u | A, X)}{S_C(u | A, X)} \mathbb{E}[y(T) | T \geq u, A, X] \right], \tag{1}
 \end{aligned}$$

where  $S_C(t | a, x) = Pr(C > t | A = a, X = x)$  is the conditional survival function for the censoring process,  $dM_C(u | A = a, X) = dN_C(u) - Y(u)d\Lambda_C(u | A = a, X)$  is the martingale increment for the censoring process,  $N_C(u) = I\{U \leq u, \Delta = 0\}$  and  $\Lambda_C(u | A = a, X) = -\log(S_C(u | A = a, X))$ . The first term in (1) is the IPW formula, and the augmentation

terms capture additional information from the subjects who do not receive treatment  $d$ , and who receive treatment  $d$  but are censored.

In (clinical) practice, it is usually desirable to consider a class of ITRs indexed by a Euclidean parameter  $\eta = (\eta_1, \dots, \eta_{p+1})^T \in \mathbb{R}^{p+1}$  for feasibility and interpretability, which has become the major trend (Rudin, 2019; Wallace and Moodie, 2015; Chen et al., 2017). Let  $V(\eta) = V(d_\eta)$ . Throughout, we focus on such a class of linear ITRs:

$$\mathcal{D}_\eta = \{d_\eta : d_\eta(X) = I\{\eta^T \tilde{X} \geq 0\}, |\eta_{p+1}| = 1\},$$

where  $\tilde{X} = (1, X^T)^T$ , and for identifiability we assume there exists a continuous covariate whose coefficient has absolute value one (Zhou et al., 2023); without loss of generality, we assume  $|\eta_{p+1}| = 1$ . Therefore, the population parameter  $\eta^*$  indexing the optimal ITR is  $\eta^* = \arg \max_{\eta \in \{\eta \in \mathbb{R}^{p+1} : |\eta_{p+1}| = 1\}} V(\eta)$ , and the optimal value function is  $V(\eta^*)$ .

**Remark 1** *Despite that we focus on linear ITRs, we note that our main identification and semiparametric efficiency results hold for general ITRs. When there is no restrictions on the ITR class  $\mathcal{D}$ , Zhang et al. (2012b) have shown that the optimal ITR, determined by the sign of the conditional average treatment effects (CATEs), equivalently maximizes the value function. That is, the optimal ITR only depends on the CATEs, or the conditional distributions of the outcome given the covariates. But when the ITR class  $\mathcal{D}$  is restricted, the optimal ITR would also hinge on the covariate distribution, so we have to account for the covariate shift. The desirable interpretability and parsimony of linear ITRs can only be achieved at the expense of making suboptimal decisions for some individuals. Maronge et al. (2023) propose a reluctant additive model for interpretable nonlinear ITRs, which addresses the tension between interpretability and accuracy of treatment decisions. To prevent the estimated optimal ITR from being suboptimal or even detrimental to certain disadvantaged subpopulations, Fang et al. (2023) propose new fairness criteria that guarantee the tail performance exceed a prespecified threshold. We leave it for future research to integrate these methods into our transfer learning framework.*

### 2.3 Transfer Learning

If we understand the conditional average treatment effects and, consequently, the optimal individualized treatment rule (ITR) for all values of  $X$ , concerns related to covariate shift become irrelevant. We advocate for the adoption of a transfer learning framework for two primary reasons: 1) As highlighted in the preceding section, simplicity in ITRs is often favored in practical scenarios. However, the efficacy of such an ITR might be compromised by covariate shifts, where there is a discrepancy in population distributions (Sugiyama and Kawanabe, 2012). 2) Our goal extends beyond merely identifying the optimal ITR; we are also keen on estimating and conducting inference on the value function for the target population. This step is crucial for assessing the effectiveness of the ITR before its practical application.

Instead of minimizing the worst-case risk over a single data sample, here we combine the source and target samples. This is a common practice in many clinical settings, for instance generalizing an active-controlled trial’s intention-to-treat effect where only baseline covariates are measured, from historical trial data where complete information is available

(He et al., 2024). Suppose that a source sample of size  $n$  and a target sample of size  $m$  are sampled independently from the target super population with different mechanisms. Let  $I_S$  and  $I_T$  denote the indicator of source and target samples, respectively. Typically, the source data may originate from a randomized controlled trial (RCT), and thus it may be considered as sampled with a selection bias, meaning that its distribution differs from that in the broader target population. Frequently, the target data is uniformly sampled (or with known design weights) from this broader target population, ensuring that their distributions align. However, our approach of integrating two data sources enables us to introduce innovative weighting schemes or designs aimed at enhancing generalization.

More formally, a covariate shift means that  $Pr(I_S = 1 | X) \neq Pr(I_T = 1 | X)$ . In the source sample, independent and identically distributed (i.i.d.) data  $\mathcal{O}_s = \{X_i, A_i, U_i, \Delta_i, I_{S,i} = 1, I_{T,i} = 0\}_{i=1}^n$  are observed from  $n$  subjects; in the target sample, it is common that only the covariates information is available, so i.i.d. data  $\mathcal{O}_t = \{X_i, I_{S,i} = 0, I_{T,i} = 1\}_{i=n+1}^{n+m}$  are observed from  $m$  subjects.

In this framework, we assume that the source and target sampling mechanisms are independent, which holds if two separate studies are conducted independently by different research projects in different locations or in two separate time periods, and the target population is sufficiently large. In the context of combining the RCT and observational study, this framework corresponds to the *non-nested* study design (Dahabreh et al., 2021).

**Remark 2** *We consider the two independent samples from a (target) super population that describes the distribution of all subjects of interest to whom we intend to assign the treatment. We present the identification formulas in Section 3; however, we do not require  $N$  to be fixed and known. Equivalently, it is also possible to assume a pooled population consisting of a source population and a target population, and similar identification formulas can be proposed based on the density ratio of the two populations. Given the sampling mechanism of our data structure, we have that  $n$  and  $m$  scale at the same rate. So in our main asymptotic analysis, we can simply let  $n, m \rightarrow \infty$ . In the Appendix F, we analyze the asymptotic properties of our proposed methods when  $m$  or  $n$  diverges faster.*

### 3. Methodology

#### 3.1 Identification and Semiparametric Efficiency

To identify the causal effects from the observed data, we make the following assumptions.

**Assumption 1** (a)  $T = T(A)$  almost surely. (b)  $Pr(A = a | X, I_S = 1) > 0$  for every  $a$  almost surely. (c)  $A \perp \{T(1), T(0)\} | \{X, I_S = 1\}$ . (d)  $C \perp \{T(1), T(0)\} | \{X, A, I_S = 1\}$ . (e) The transformation  $y(\cdot)$  admits a maximal horizon  $0 < h < \infty$ , such that  $y(t) = y(h)$  for all  $t \geq h$ .  $Pr(C < h | X, A, I_S = 1) < 1$ .

Assumption 1 includes the standard assumptions as we have introduced in Section 2.1. Here we only assume them in the source population. Assumption 1 (a) implies that the observed outcome is the potential outcome under the actual assigned treatment. Assumption 1 (b) states that each subject has a positive probability of receiving both treatments. Assumption 1 (c) requires that all confounding factors are measured so that treatment assignment is as good as random conditionally on  $X$ . Assumption 1 (d) essentially states

that the censoring process is non-informative conditionally on  $X$ . Furthermore, we require additional assumptions for the source and target populations. Assumption 1 (e) says that the transformation is indifferent to survival beyond some maximal time horizon.

**Assumption 2 (Survival mean exchangeability)**  $\mathbb{E}[y(T(a)) | X, I_S = 1] = \mathbb{E}[y(T(a)) | X]$  for every  $a \in \mathcal{A}$ .

**Assumption 3 (Positivity of source inclusion)**  $0 < Pr(I_S = 1 | X) < 1$  almost surely.

**Assumption 4 (Known target design)** The target sample design weight  $e(x) = \pi_T^{-1}(x) = 1/Pr(I_T = 1 | X = x)$  is known by design.

Assumption 2 is similar to the mean exchangeability over trial participation (Dahabreh et al., 2019), and is weaker than the ignorability assumption (Stuart et al., 2011), i.e.,  $I_S \perp \{T(1), T(0)\} | X$ . Assumption 3 states that each subject has a positive probability to be included in the source sample, and implies adequate *overlap* of covariate distributions between the source and target populations. This can be a strong assumption in practice. When Assumption 3 is violated, generalization can only be made to a restricted population without extrapolation. So it would help to first identify the population where we can generalize (e.g. trial eligibility criteria). One can fit a propensity score model and trim the extreme values to ensure positivity is not violated (Yang and Ding, 2018). However, it shifts the target population and complicates the estimation and inference procedure, so we leave it for future work. Another line of research focuses on falsification tests (Hussain et al., 2022, 2023; Demirel et al., 2024b), which may serve as alternatives of Assumption 3. Assumption 4 is commonly assumed in the survey sampling literature; thus the design-weighted target sample is representative of the target population. In an observational study with simple random sampling, we have  $e(x) = 1/Pr(I_T = 1)$ .

Under this framework, we have the following key identity that for any  $g(X)$

$$\mathbb{E} \left[ \frac{I_S}{\pi_S(X)} g(X) \right] = \mathbb{E}[I_T e(X)g(X)] = \mathbb{E}[g(X)], \quad (2)$$

where  $\pi_S(X) = Pr(I_S = 1 | X)$  is the sampling score.

**Proposition 3 (Identification formulas)** Under Assumptions 1 - 4, the value function  $V(d)$  can be identified by the outcome regression formula:

$$V(d) = \mathbb{E}[I_T e(X)\mathbb{E}[y(T) | A = d(X), X, I_S = 1]], \quad (3)$$

and the IPW formula:

$$V(d) = \mathbb{E} \left[ \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} \frac{\Delta y(U)}{S_C(U | A, X)} \right], \quad (4)$$

where  $\pi_d(X) = d(X)\pi_A(X) + (1 - d(X))(1 - \pi_A(X))$  with the propensity score  $\pi_A(X) = Pr(A = 1 | X, I_S = 1)$ , and  $S_C(t | a, x) = Pr(C > t | A = a, X = x, I_S = 1)$ .

Based on the identification formulas (3) and (4), we can construct plug-in estimators for  $V(d)$ , using the sampling score  $\pi_S(X)$  or design weights  $e(X)$  to account for the sampling bias. By the identity (2), the design weights  $I_T e(X)$  in the OR formula (3) with the target sample can also be replaced by the inverse of sampling score  $I_S/\pi_S(X)$  using the source sample. However, these estimators are biased if the posited models are misspecified, and extreme weights from  $\pi_S, \pi_A$  and  $S_C$  usually lead to large variability.

The simple plug-in estimators based on the formulas (3) and (4) are only consistent when the corresponding models are correctly specified, which is often unrealistic in practice. Therefore, we consider a more efficient and robust approach, motivated by the efficient influence function for  $V(d)$ . We use semiparametric efficiency theory to understand the lower bound and statistical difficulty of estimating the target parameter  $V(d)$ . By the EIF, we can construct the optimally efficient estimator, i.e. the one with smallest variance among all regular and asymptotically linear estimators. Such estimators have also been shown to have desirable robustness against model misspecification. We refer to Kennedy (2022) for more details.

**Proposition 4** *Under Assumptions 1 - 4, the efficient influence function of  $V(d)$  is*

$$\begin{aligned} \phi_d = & \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} \frac{\Delta y(U)}{S_C(U | A, X)} - V(d) \\ & + \left( I_T e(X) - \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} \right) \mu(d(X), X) \\ & + \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} \int_0^\infty \frac{dM_C(u | A, X)}{S_C(u | A, X)} Q(u, A, X). \end{aligned} \quad (5)$$

where  $\mu(a, x) = \mathbb{E}[y(T) | A = a, X = x, I_S = 1]$  and  $Q(u, a, x) = \mathbb{E}[y(T) | T \geq u, A = a, X = x, I_S = 1]$ <sup>1</sup>.

The semiparametric EIF guides us in constructing efficient estimators combining the source and target samples. Compared to (1), this EIF captures additional covariates information from the target population via the outcome model and thus removes the sampling bias. An efficient estimation procedure is proposed in the next section, and we show that it enjoys the double robustness property, i.e., it is consistent if either the survival outcome models  $\mu(a, x), Q(u, a, x)$  or the models of propensity score  $\pi_A(x)$ , sampling score  $\pi_S(x)$  and censoring process  $S_C(t | a, x)$  are correct. Moreover, this EIF is Neyman orthogonal in the sense discussed in Chernozhukov et al. (2018). Therefore, a cross-fitting procedure is also proposed, allowing flexible machine learning methods for the nuisance parameters estimation, and  $\sqrt{n}$  rate of convergence can be achieved.

### 3.2 An Efficient and Robust Estimation Procedure

In this section, we focus on estimating the survival function  $S_d(t) = Pr(T(d) > t)$  as the value function under ITR  $d$ . Following the asymptotic linear characterization of survival estimands in Yang et al. (2023b), our results are readily extended to a broad class of

1. Note that  $\mathbb{E}[y(T) | T \geq u, A, X] = -\int_u^\infty y(s) dS(s | A, X) / S(u | A, X)$ . For instance, when  $y(T) = I\{T \geq t\}$ , we have  $\mathbb{E}[y(T) | T \geq u, A, X] = S(t | A, X) / S(u | A, X)$  for  $u \leq t$ .



functionals of survival distributions. For instance, the value function of the RMST under ITR  $d$  is simply  $\int_0^L S_d(t)dt$ .

Based on the EIF (5), we propose an estimator for the survival function

$$\begin{aligned} \hat{S}_d(t) = \frac{1}{n+m} \sum_{i=1}^{n+m} & \left\{ \frac{I_{S,i}}{\hat{\pi}_S(X_i)} \frac{I\{A_i = d(X_i)\}}{\hat{\pi}_d(X_i)} \frac{\Delta_i Y_i(t)}{\hat{S}_C(t|A_i, X_i)} \right. \\ & + \left( I_{T,i} e(X_i) - \frac{I_{S,i}}{\hat{\pi}_S(X_i)} \frac{I\{A_i = d(X_i)\}}{\hat{\pi}_d(X_i)} \right) \hat{S}(t|A = d(X_i), X_i) \\ & \left. + \frac{I_{S,i}}{\hat{\pi}_S(X_i)} \frac{I\{A_i = d(X_i)\}}{\hat{\pi}_d(X_i)} \int_0^\infty \frac{\hat{S}(t|A_i, X_i)}{\hat{S}(u|A_i, X_i)} d\hat{M}_C(u|A_i, X_i) \right\}, \end{aligned} \quad (6)$$

where  $S(t|a, x) = Pr(T > t|A = a, X = x, I_S = 1)$  is the treatment-specific conditional survival function. We posit (semi)parametric models for the nuisance parameters. Let  $\pi_A(X; \theta)$  be the posited propensity score model, for example, using logistic regression  $\text{logit}\{\pi_A(X; \theta)\} = \theta^T \tilde{X}$ , where  $\text{logit}(x) = \log\{x/(1-x)\}$ . We use the Cox proportional hazard model  $\Lambda(t|A = a, X = x) = \Lambda_{0,a}(t) \exp(\beta_a^T x)$  to estimate the survival functions  $S(t|a, x) = \exp\{-\Lambda(t|a, x)\}$  and the cumulative baseline hazard function  $\Lambda_{0,a}(t) = \int_0^t \lambda_{0,a}(u)du$  can be estimated by the Breslow estimator (Breslow, 1972). Similarly, we posit a Cox proportional hazard model for the censoring process  $\Lambda_C(t|A = a, X = x) = \Lambda_{C0,a}(t) \exp(\alpha_a^T x)$ , and the cumulative baseline hazard function  $\Lambda_{C0,a}(t)$  is estimated by the Breslow estimator. The sampling score estimation is discussed in the next section.

Let  $\hat{S}(t; \eta) = \hat{S}_{d_\eta}(t)$  be the estimated value function for the ITR class  $\mathcal{D}_\eta$ , then the optimal ITR is given by  $d_{\hat{\eta}}(x)$ , where  $\hat{\eta} = \arg \max_\eta \hat{S}(t; \eta)$ .

### 3.3 Calibration Weighting

To correct the bias due to the covariate shift between populations, most existing methods directly model the sampling score (Cole and Stuart, 2010), i.e., inverse probability of sampling weighting (IPSW). However, the IPSW method requires the sampling score model to be correctly specified, and it could also be numerically unstable. Alternatively, we introduce the calibration weighting (CW) approach motivated by the identity (2), which is similar to the entropy balancing method (Hainmueller, 2012).

Let  $\mathbf{g}(X)$  be a vector of functions of  $X$  to be calibrated, such as the moments, interactions, and non-linear transformations of  $X$ . Each subject  $i$  in the source sample is assigned a weight  $q_i$  by solving the following optimization task:

$$\min_{q_1, \dots, q_n} \sum_{i=1}^n q_i \log q_i, \quad (7)$$

$$\text{subject to } q_i \geq 0, \sum_{i=1}^n q_i = 1, \sum_{i=1}^n q_i \mathbf{g}(X_i) = \tilde{\mathbf{g}}, \quad (8)$$

where  $\tilde{\mathbf{g}} = \sum_{i=n+1}^{n+m} e(X_i) \mathbf{g}(X_i) / \sum_{i=n+1}^{n+m} e(X_i)$  is a design-weighted estimate of  $\mathbb{E}[\mathbf{g}(X)]$ . The objective function (7) is the negative entropy of the calibration weights, which ensures that the empirical distribution of the weights is not too far away from the uniform, such that

it minimizes the variability due to heterogeneous weights. The final balancing constraint in (8) calibrates the covariate distribution of the weighted source sample to the target population in terms of  $\mathbf{g}(X)$ . By introducing the Lagrange multiplier  $\lambda$ , the minimizer of the optimization task is  $q_i = \exp\{\hat{\lambda}^T \mathbf{g}(X_i)\} / \sum_{i=1}^n \exp\{\hat{\lambda}^T \mathbf{g}(X_i)\}$ , where  $\hat{\lambda}$  solves the estimating equation  $\sum_{i=1}^n \exp\{\lambda^T \mathbf{g}(X_i)\} \{\mathbf{g}(X_i) - \tilde{\mathbf{g}}\} = 0$ . Since we only require specifying  $\mathbf{g}(X)$ , calibration weighting avoids explicitly modeling the sampling score and evades extreme weights.

Moreover, suppose that the sampling score follows a loglinear model  $\pi_S(X; \lambda) = \exp\{\lambda^T \tilde{X}\}$ , Lee et al. (2021, 2022) show that there is a direct correspondence between the calibration weights and the estimated sampling score. We also note that if the fraction  $Pr(I_S = 1)$  is small, the loglinear model is close to the widely used logistic regression model; our simulation studies show the robustness of calibration weights.

**Remark 5** *Other objective functions can also be used for calibration weights estimation. Chu et al. (2023) considers a generic convex distance function  $h(q)$  from the Cressie and Read family of discrepancies (Cressie and Read, 1984). Thus the optimization task is  $\min_{q_1, \dots, q_n} \sum_{i=1}^n h(q_i)$  under the constraints (8), and the correspondence between the sampling score model  $\pi_S$  and the objective function  $h$  has also been established.*

### 3.4 Cross-Fitting

Utilizing the Neyman orthogonality of EIF (5), we consider flexible machine learning methods for estimating the nuisance parameters, where we want to remain agnostic on modeling assumptions for the complex treatment assignment, survival, and censoring processes. There is extensive recent literature on nonparametric methods for heterogeneous treatment effect estimation with survival outcomes. Cui et al. (2023) extends the generalized random forests (Athey et al., 2019) to estimate heterogeneous treatment effects in a survival and observational setting. See Xu et al. (2023) for details and practical considerations. A description of the proposed cross-fitting procedure is given below (Schick, 1986; Chernozhukov et al., 2018). Throughout, we use the subscript  $CF$  to denote the cross-fitted version.

---

**Algorithm 1** Pseudo algorithm for the cross-fitting procedure

---

**Step 1** Randomly split the datasets  $\mathcal{O}_s$  and  $\mathcal{O}_t$  respectively into  $K$ -folds with equal size such that  $\mathcal{O}_s = \cup_{k=1}^K \mathcal{O}_{s,k}$ ,  $\mathcal{O}_t = \cup_{k=1}^K \mathcal{O}_{t,k}$ . For each  $k \in \{1, \dots, K\}$ , let  $\mathcal{O}_{s,k}^c = \mathcal{O}_s \setminus \mathcal{O}_{s,k}$ ,  $\mathcal{O}_{t,k}^c = \mathcal{O}_t \setminus \mathcal{O}_{t,k}$

**Step 2** For each  $k \in \{1, \dots, K\}$ , estimate the nuisance parameters only using data  $\mathcal{O}_{s,k}^c$  and  $\mathcal{O}_{t,k}^c$ ; then obtain an estimate of the value function  $\hat{V}_{CF,k}(\eta)$  using data  $\mathcal{O}_{s,k}$

**Step 3** Aggregate the estimates from  $K$  folds:  $\hat{V}_{CF}(\eta) = \frac{1}{K} \sum_{k=1}^K \hat{V}_{CF,k}(\eta)$

**Step 4** The estimated optimal ITR is indexed by  $\hat{\eta} = \arg \max_{\eta} \hat{V}_{CF}(\eta)$

---

#### 4. Asymptotic Properties

In this section, we present the asymptotic properties of the proposed methods. To establish the asymptotic properties, we require the following assumptions.

**Assumption 5** (i) *The value function  $V(\eta)$  is twice continuously differentiable in a neighborhood of  $\eta^*$ .* (ii) *There exists some constant  $\delta_0 > 0$  such that  $Pr(0 < |\tilde{X}^T \eta| < \delta) = O(\delta)$ , where the big- $O$  term is uniform in  $0 < \delta < \delta_0$ .*

Condition (i) is a standard regularity condition to establish uniform convergence. Similar margin conditions as (ii), which state that  $Pr(0 < |\gamma(X)| < \delta) = O(\delta^\alpha)$ <sup>2</sup>, are often assumed in the literature of classification (Tsybakov, 2004; Audibert and Tsybakov, 2007), reinforcement learning (Farahmand, 2011; Hu et al., 2025) and optimal treatment regimes (Luedtke and van der Laan, 2016a; Luedtke and Chambaz, 2020), to guarantee a fast convergence rate. Note that  $\alpha = 0$  imposes no restriction, which allows  $\gamma(X) = 0$  almost surely, i.e., the challenging setting of exceptional laws where the optimal ITR is not uniquely defined (Robins, 2004; Robins and Rotnitzky, 2014), while the case  $\alpha = 1$  is of particular interest and would hold if  $\gamma(X)$  is absolutely continuous with bounded density.

**Theorem 6** *Under Assumptions 1 - 5 and standard regularity conditions provided in the Supplementary Material, if either the survival outcome model, or the models of the propensity score, the sampling score and the censoring process are correct, we have that as  $n, m \rightarrow \infty$ , (i)  $\hat{S}(t; \eta) \rightarrow S(t; \eta)$  for any  $\eta$  and  $0 < t \leq L$ ; (ii)  $\sqrt{n+m} \left\{ \hat{S}(t; \eta) - S(t; \eta) \right\}$  converges weakly to a mean zero Gaussian process for any  $\eta$ ; (iii)  $(n+m)^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ ; (iv)  $\sqrt{n+m} \left\{ \hat{S}(t; \hat{\eta}) - S(t; \eta^*) \right\} \rightarrow \mathcal{N}(0, \sigma_{t,1}^2)$ , where  $\sigma_{t,1}$  is given in the Supplementary Material.*

Next, to characterize the asymptotic behavior of the estimator with the nonparametric estimation of nuisance parameters, we assume the following consistency and convergence rate conditions of the nonparametric plug-in nuisance estimators.

**Assumption 6** *Assume the following consistency conditions:  $\|\hat{\pi}_A(x) - \pi_A(x)\|_2 = o_p(1)$ ,  $\|\hat{\pi}_S(x) - \pi_S(x)\|_2 = o_p(1)$ , and for  $a = 0, 1$ ,*

$$\sup_{u \leq h} \left\| \hat{S}_C(u | a, X) - S_C(u | a, X) \right\|_2 = o_p(1), \sup_{u \leq h} \left\| \frac{\hat{\lambda}_C(u | a, X)}{\hat{S}_C(u | a, X)} - \frac{\lambda_C(u | a, X)}{S_C(u | a, X)} \right\|_2 = o_p(1),$$

$$\|\hat{\mu}(a, X) - \mu(a, X)\|_2 = o_p(1), \sup_{u \leq h} \left\| \hat{Q}(u, a, X) - Q(u, a, X) \right\| = o_p(1),$$

*and the following rate of convergence conditions:  $\|\hat{\pi}_A(x) - \pi_A(x)\|_2 = o_p(n^{-1/4})$ ,  $\|\hat{\pi}_S(x) - \pi_S(x)\|_2 = o_p(n^{-1/4})$ , and for  $a = 0, 1$ ,*

$$\sup_{u \leq h} \left\| \hat{S}_C(u | a, X) - S_C(u | a, X) \right\|_2 = o_p(n^{-1/4}), \sup_{u \leq h} \left\| \frac{\hat{\lambda}_C(u | a, X)}{\hat{S}_C(u | a, X)} - \frac{\lambda_C(u | a, X)}{S_C(u | a, X)} \right\|_2 = o_p(n^{-1/4}),$$

$$\|\hat{\mu}(a, X) - \mu(a, X)\|_2 = o(n^{-1/4}), \sup_{u \leq h} \left\| \hat{Q}(u, a, X) - Q(u, a, X) \right\|_2 = o(n^{-1/4}).$$

---

2. Let  $\gamma(X) = \mathbb{E}[T | A = 1, X] - \mathbb{E}[T | A = 0, X]$  denote the conditional average treatment effect, then the optimal ITR in an unrestricted class is given by  $d(X) = I\{\gamma(X) > 0\}$ .

The rate conditions in Assumption 6 are generally assumed in the literature (Kennedy, 2022). This rate can be achieved by many existing methods under certain structural assumptions on the nuisance parameters. Note that the nuisance parameters do not necessarily need to be estimated at the same rates  $n^{-1/4}$  for our theorems to hold; it would suffice that the product of rates of any combination of two nuisance parameters is  $n^{-1/2}$ .

**Theorem 7** *Under Assumptions 1 - 6, we have that as  $n, m \rightarrow \infty$ , (i)  $\hat{S}_{CF}(t; \eta) \rightarrow S(t; \eta)$  for any  $\eta$  and  $0 < t \leq L$ ; (ii)  $\sqrt{n+m} \left\{ \hat{S}_{CF}(t; \eta) - S(t; \eta) \right\}$  converges weakly to a mean zero Gaussian process for any  $\eta$ ; (iii)  $(n+m)^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ ; (iv)  $\sqrt{n+m} \left\{ \hat{S}_{CF}(t; \hat{\eta}) - S(t; \eta^*) \right\} \rightarrow \mathcal{N}(0, \sigma_{t,2}^2)$ , where  $\sigma_{t,2}$  is given in the Supplementary Material.*

Besides the survival functions, another common measure of particular interest in survival analysis is the RMST. Let  $V_{RMST}(\eta) = \mathbb{E}[\min(T(d_\eta), L)]$ . We present two corollaries.

**Corollary 8** *Under Assumptions 1 - 5 and standard regularity conditions provided in the Supplementary material, if either the survival outcome model or the models of the propensity score, the censoring and sampling processes are correct, we have that as  $n, m \rightarrow \infty$ , (i)  $\hat{V}_{RMST}(\eta) \rightarrow V_{RMST}(\eta)$  for any  $\eta$ ; (ii)  $(n+m)^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ ; (iii)  $\sqrt{n+m} \left\{ \hat{V}_{RMST}(\hat{\eta}) - V_{RMST}(\eta^*) \right\} \rightarrow \mathcal{N}(0, \sigma_3^2)$ , where  $\sigma_3$  is given in the Supplementary Material.*

**Corollary 9** *Under Assumptions 1 - 6, we have that as  $n, m \rightarrow \infty$ , (i)  $\hat{V}_{RMST,CF}(\eta) \rightarrow V_{RMST}(\eta)$  for any  $\eta$ ; (ii)  $(n+m)^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ ; (iii)  $\sqrt{n+m} \left\{ \hat{V}_{RMST,CF}(\hat{\eta}) - V_{RMST}(\eta^*) \right\} \rightarrow \mathcal{N}(0, \sigma_4^2)$ , where  $\sigma_4$  is given in the Supplementary Material.*

Finally, we show that when the covariate distributions of the source and target populations are the same, the semiparametric efficiency bounds of  $\hat{V}_{DR}(\eta)$  and  $\hat{V}_{CF}(\eta)$  are equal.

**Theorem 10** *Under Assumptions 1 - 6, when the covariate distributions of the source and target populations are the same, both  $\sqrt{n} \left\{ \hat{V}_{DR}(\eta) - V(\eta) \right\}$  and  $\sqrt{n+m} \left\{ \hat{V}_{CF}(\eta) - V(\eta) \right\}$  are asymptotically normal with mean zero and same variance.*

Theorem 10 implies that when there is no covariate shift, our proposed estimator does not lose efficiency in comparison to the original double robust estimator since the augmentation term in EIF (5) from the target population,  $I_T e(X) \mu(d(X), X)$ , is asymptotically equal to this term evaluated on the source population in this case.

Moreover, when the covariate shift exists, we consider the optimal ITR  $d^{\text{opt}}$  without restriction on the ITR class.

**Theorem 11** *Under Assumptions 1 - 6, If  $d^{\text{opt}} \in \mathcal{D}_\eta$ , i.e.,  $d^{\text{opt}} = d_{\eta^*}$ , both the maximizers of  $\hat{V}_{DR}(\eta)$  and  $\hat{V}_{CF}(\eta)$  converge to  $\eta^*$ . However,  $\hat{V}_{DR}(\eta)$  is a biased estimator of  $V(\eta)$ .*

Theorem 11 implies if the true optimal ITR belongs to the restricted ITR class  $\mathcal{D}_\eta$ , standard methods, without accounting for the covariate shift, are still able to recover the optimal ITR but fail to be consistent for the value function, due to the covariate shift. And we can only rely on the proposed method to draw valid inferences.

## 5. Simulation

In this section, we investigate the finite-sample properties of our method through extensive numerical simulations. The R code to replicate all results is available at <https://github.com/panzhaoo/transfer-learning-survival-ITR>. Consider a target population of sample size  $N = 2 \times 10^5$ . The covariates  $(X_1, X_2, X_3)^T$  are generated from a multivariate normal distribution with mean 0, unit variance with  $\text{corr}(X_1, X_3) = 0.2$  and all other pairwise correlations equal to 0, and further truncated below  $-4$  and above  $4$  to satisfy regularity conditions. The target sample is a random sample of size  $m = 8000$  from the target population. The sampling score follows  $\pi_S(X) = \text{expit}(-4.5 - 0.5X_1 - 0.5X_2 - 0.4X_3)$ ; thus the source sampling rate is around 1.6%, and the source sample size around  $n = 3000$ . The treatment assignment mechanism in the source sample follows  $\pi_A(X) = \text{expit}(0.5 + 0.8X_1 - 0.5X_2)$ .

The counterfactual survival times  $T(a)$  are generated according to the hazard functions  $\lambda(t|A=0, X) = \exp(t) \cdot \exp(-2.5 - 1.5X_1 - X_2 - 0.7X_3)$  and  $\lambda(t|A=1, X) = \exp(t) \cdot \exp(-1 - X_1 - 0.9X_2 - X_3 - 2X_2^2 + X_1X_3)$ . The censoring time  $C$  is generated according to the hazard functions  $\lambda_C(t|A=0, X) = 0.04 \exp(t) \cdot \exp(-1.6 + 0.8X_1 - 1.1X_2 - 0.7X_3)$  and  $\lambda_C(t|A=1, X) = 0.04 \exp(t) \cdot \exp(-1.8 - 0.8X_1 - 1.7X_2 - 1.4X_3)$ . The resultant censoring rate is approximately 20%.

We consider the RMST with the maximal time horizon  $L = 4$  as the value function. To evaluate the performance of different estimators for optimal ITRs, we compute the corresponding true value functions and percentages of correct decisions (PCD) for the target population. Specifically, we generate a large sample with size  $\tilde{N} = 1 \times 10^5$  from the target population. The true value function of any ITR  $d(\cdot; \eta)$  is computed by  $V(\eta) = \tilde{N}^{-1} \sum_{i=1}^{\tilde{N}} \min\{d(X_i; \eta)T_i(1) + (1 - d(X_i; \eta))T_i(0), L\}$  and its associated PCD is computed by  $1 - \tilde{N}^{-1} \sum_{i=1}^{\tilde{N}} |d(X_i; \eta^*) - d(X_i; \eta)|$ , where  $\eta^* = \arg \max_{\eta} V(\eta)$ .

We compare the following estimators for the RMST  $\hat{V}(\eta) = \int_0^L \hat{S}(t; \eta) dt$ :

- Naive:  $\hat{S}^{\text{Naive}}(t; \eta) = \frac{1}{n} \sum_{i=1}^n \frac{I\{A_i=d(X_i)\}}{\hat{\pi}_d(X_i)} \frac{\Delta_i Y_i(t)}{\hat{S}_{C(U|A,X)}}$ ; IPW formula (4) without using the sampling score;
- IPSW:  $\hat{S}^{\text{IPSW}}(t; \eta) = \frac{1}{n} \sum_{i=1}^n \frac{I_{S,i}}{\hat{\pi}_S(X_i)} \frac{I\{A_i=d(X_i)\}}{\hat{\pi}_d(X_i)} \frac{\Delta_i Y_i(t)}{\hat{S}_{C(U|A,X)}}$ ; IPW formula (4) where the sampling score is estimated via logistic regression;
- CW-IPW:  $\hat{S}^{\text{CW-IPW}}(t; \eta) = \sum_{i=1}^n q_i \frac{I\{A_i=d(X_i)\}}{\hat{\pi}_d(X_i)} \frac{\Delta_i Y_i(t)}{\hat{S}_{C(U|A,X)}}$  IPW formula (4) where the sampling score is estimated by calibration weighting;
- CW-OR:  $\hat{S}^{\text{CW-OR}}(t; \eta) = \sum_{i=1}^n q_i \hat{S}(t|A = d(X_i), X_i)$ ; OR formula (3) in combination with calibration weights by the identity (2);
- ORt:  $\hat{S}^{\text{ORt}}(t; \eta) = \frac{1}{m} \sum_{i=n+1}^{n+m} \hat{S}(t|A = d(X_i), X_i)$ ; OR formula (3) evaluated on the target sample;
- ACW: augmented estimator (6), where the sampling score is estimated by calibration weighting.

**Remark 12** *Since the estimated value functions are non-convex and non-smooth, multiple local optimal may exist in the optimization task, and many derivatives-based algorithms do*

not work for this challenging setting. Here we utilize the genetic algorithm implemented in the R package *rgenoud* (Mebane Jr. and Sekhon, 2011), which performs well in our numerical experiments. We refer to Mitchell (1998) for algorithmic details.

### 5.1 (Semi)parametric Models

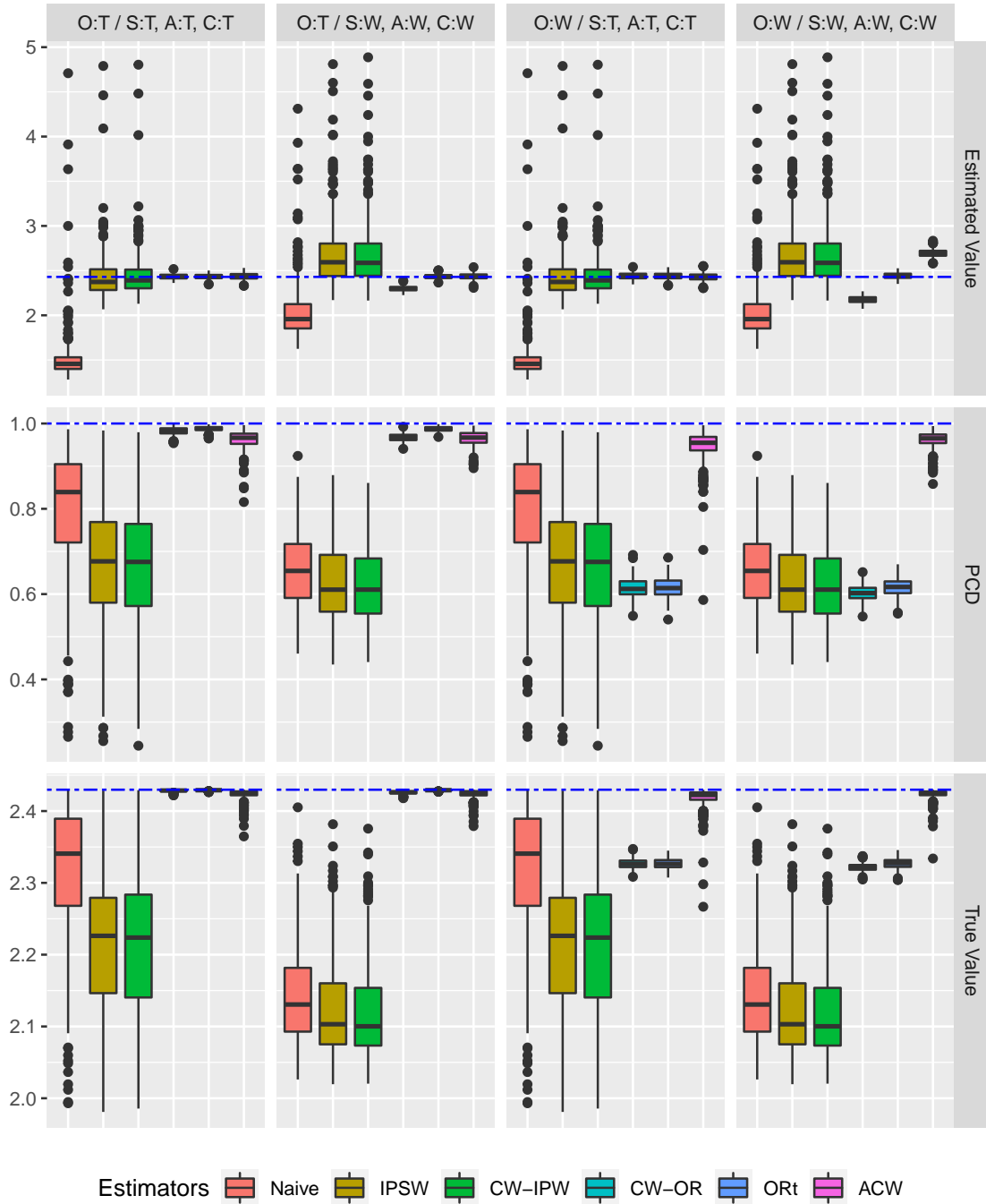
We first consider the setting where the nuisance parameters are estimated by posited (semi)parametric working models as introduced in Section 3.2. To assess the performance of these estimators under model misspecification, we consider four scenarios: (1) all models are correct, (2) only the survival outcome model is correct, (3) only the survival outcome model is wrong, (4) all models are wrong. For the wrong sampling model, the weights are estimated using calibration on  $e^{X_1}$ . The wrong propensity score model is fitted on  $e^{X_3}$ . The wrong Cox models for survival and censoring times are fitted on  $(e^{X_1}, e^{X_2}, e^{X_3})^T$ .

Figure 1 and Table 1 report the simulation results from 350 Monte Carlo replications. Variance is estimated by a bootstrap procedure with  $B = 200$  bootstrap replicates. The proposed ACW estimator is unbiased in scenarios (1) - (3), and the 95% coverage probabilities approximately achieve the nominal level, which shows the double robustness property. In comparison, the bias of the regression based methods is 10 times larger than our proposed method when the outcome model is misspecified, and the bias of the weighting based methods is 100 times larger than our proposed method when the models for sampling, treatment assignment and censoring are wrong.

Table 1: Numerical results under four different model specification scenarios. Bias is the empirical bias of point estimates; SD is the empirical standard deviation of point estimates; SE is the average of bootstrap standard error estimates; CP is the empirical coverage probability of the 95% confidence intervals.

	Bias	SD	SE	CP(%)	Bias	SD	SE	CP(%)
	O:T / S:T, A:T, C:T				O:T / S:W, A:W, C:W			
Naive	-0.8801	0.4595	0.2189	7.43	-0.3528	0.5024	0.4598	37.43
IPSW	0.0185	0.3685	0.2562	87.14	0.3377	0.7144	0.6958	98.29
CW-IPW	0.0378	0.3701	0.2498	88.29	0.3406	0.7144	0.6957	97.71
CW-OR	0.0047	0.0273	0.0286	96.29	-0.1312	0.0269	0.0279	0.57
ORt	0.0041	0.0258	0.0262	95.14	0.0035	0.0258	0.0262	95.71
ACW	0.0070	0.0380	0.0369	94.29	0.0055	0.0316	0.0334	95.43
	O:W / S:T, A:T, C:T				O:W / S:W, A:W, C:W			
Naive	-0.8801	0.4595	0.2207	6.86	-0.3528	0.5024	0.5018	38.57
IPSW	0.0185	0.3685	0.2486	87.71	0.3377	0.7144	0.7586	99.14
CW-IPW	0.0378	0.3701	0.2418	88.86	0.3406	0.7144	0.7570	98.57
CW-OR	0.0103	0.0370	0.0362	92.29	-0.2551	0.0366	0.0391	0.00
ORt	0.0094	0.0365	0.0355	94.00	0.0115	0.0328	0.0355	95.71
ACW	-0.0010	0.0426	0.0419	93.14	0.2644	0.0422	0.0475	0.57

Figure 1: Boxplot of the estimated value, true value and PCD results of estimators under four model specification scenarios. O: survival outcome, S: sampling score, A: propensity score, C: censoring; T: True (correctly specified) model, W: Wrong (misspecified) model.



## 5.2 Flexible Machine Learning Methods

When utilizing flexible ML methods, we construct the cross-fitted ACW estimator as introduced in Section 3.4. The data generation process is the same as above, except that the censoring time  $C$  is generated according to the hazard functions  $\lambda_C(t | A = 0, X) = 0.2 \exp(t) \cdot \exp(-1.6 + 0.8X_1 - 1.1X_2 - 0.7X_3)$  and  $\lambda_C(t | A = 1, X) = 0.2 \exp(t) \cdot \exp(-1.8 - 0.8X_1 - 1.7X_2 - 1.4X_3)$  which leads to an increased censoring rate of approximately 33%, so there are enough observations to get an accurate estimate of the censoring process. The propensity score is estimated by the generalized random forest. The conditional survival and censoring functions are estimated by the random survival forest. The calibration weighting uses calibration on the first- and second-order moments of  $X$ .

First, we study the impact of sample sizes on the performance of the ML methods, and simulation results are given in the Supplementary Material. With a small sample size, the ACW estimator is largely biased, and the bias diminishes as the sample size increases.

Next, we compare the performance of different estimators with target population size  $N = 6 \times 10^5$  and target sample size  $m = 24000$ . Figure 2 shows the simulation results from 200 Monte Carlo replications. The two IPW-based estimators are biased and perform poorly due to the large variability of weights. The two OR-based estimators have comparable performance as the ACW estimator in terms of PCD and true value function but still suffer from the overfitting bias. Only the ACW estimator is consistent and provides valid inferences.

## 6. Real Data Analysis

In this section, to illustrate the proposed method, we study the sodium bicarbonate therapy for patients with severe metabolic acidaemia in the intensive care unit by leveraging the RCT data BICAR-ICU (Jaber et al., 2018) and the observational study (OS) data from Jung et al. (2011). Specifically, we consider the BICAR-ICU data as the source sample and the observational study data as the target sample. The BICAR-ICU is a multi-center, open-label, randomized controlled, phase 3 trial between May 5, 2015, and May 7, 2017, which includes 387 adult patients admitted within 48 hours to the ICU with severe acidaemia. The prospective, multiple-center observational study was conducted over thirteen months in five ICUs, consisting of 193 consecutive patients who presented with severe acidemia within the first 24 hours of their ICU admission. Some heterogeneity exists between the two populations.

Both the RCT and OS datasets contain detailed measurements of ICU patients with severe acidaemia. Motivated by the clinical practice and existing work in the medical literature, we consider ITRs that depend on the following five variables: SEPSIS, AKIN, SOFA, SEX, and AGE. A detailed description of the data preprocessing and variable selection is given in the Supplementary Material. Table 2 summarizes the baseline characteristics of the two datasets. The baseline covariates distribution of the patients in the BICAR-ICU differs from the distribution in the observational study; specifically, the BICAR-ICU patients have higher SOFA scores and the more frequent presence of acute kidney injury and sepsis.

We apply our proposed ACW estimator to learn the optimal ITR for the target population. The calibration weights are estimated based on the means of continuous covariates and the proportions of the binary covariates. The propensity score is estimated using a



Figure 2: Boxplots of the estimated value, true value, and PCD of different estimators using flexible ML methods.

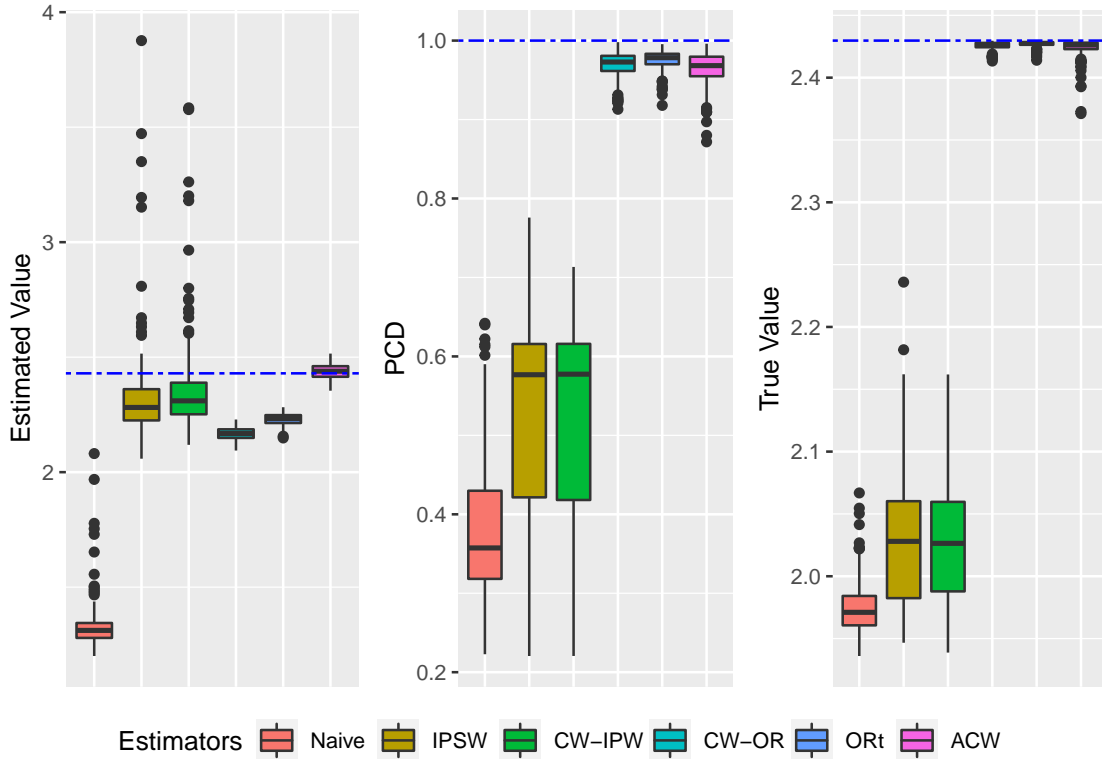


Table 2: Summary of baseline characteristics of the BICAR-ICU trial sample and the OS sample. Mean (standard deviation) for continuous and number (proportion) for the binary covariate.

	SEPSIS	AKIN	SOFA	SEX	AGE
BICAR-ICU ( $n = 387$ )	236 (60.98%)	181 (46.77%)	10.12 (3.72)	237 (61.24%)	63.95 (14.41)
OS ( $m = 193$ )	99(51.30%)	75 (38.86%)	9.10 (4.54)	122 (63.21%)	62.73 (17.49)

logistic regression model, and the Cox proportional hazard model is fitted for the survival outcome with all covariates. The censoring only occurred on the 28th day when the follow-up in ICU ends. We consider the class of linear ITRs that depend on all five variables  $\mathcal{D} = \{I\{\eta_1 + \eta_2\text{SEPSIS} + \eta_3\text{AKIN} + \eta_4\text{SOFA} + \eta_5\text{SEX} + \eta_6\text{AGE} > 0\} : \eta_1, \dots, \eta_6 \in \mathbb{R}, |\eta_6| = 1\}$ , with the aim to maximize the RMST within 28 days in ICU stay. The estimated parameter indexing the optimal ITR is  $\hat{\eta}_{\text{ACW}} = (22.9, -36.1, 87.4, -9.8, 33.7, 1.0)^T$ , which leads to an estimated value function  $\hat{V}(\hat{\eta}_{\text{ACW}}) = 19.52$  days, with confidence interval  $[17.74, 21.30]$  given by 200 bootstraps. In contrast, we also use the standard double robust method to estimate the optimal ITR for the RCT, indexed by  $\hat{\eta}_{\text{DR.RCT}}$  which maximize the value function

$\hat{V}_{\text{DR}}(\eta)$  in (1) with  $y(T) = \min(T, 28)$ . The estimated value function is  $\hat{V}(\hat{\eta}_{\text{DR.RCT}}) = 15.37$  days for the target population.

## 7. Discussion

In this paper, we present an efficient and robust transfer learning framework for estimating optimal ITR with right-censored survival data that generalizes well to the target population. The proposed method can be improved or extended in several directions for future work. Construction and estimation of optimal ITRs for multiple decision points with censored survival data are challenging, taking into account the timing of censoring, events and decision points (Jiang et al., 2017; Hager et al., 2018), e.g., using a reinforcement learning method (Cho et al., 2023). Furthermore, besides the class of ITRs indexed by a Euclidean parameter, it may be possible to consider other classes of ITRs, such as tree or list-based ITRs. The current work focus on value functions in the form  $V(d) = \mathbb{E}[y(T(d))]$  and can also be modified in case of optimizing certain easy-to-interpret quantile criteria, which does not require specifying an outcome regression model and is robust for heavy-tailed distributions (Zhou et al., 2023). And relaxing the restrictive assumptions such as positivity (Yang and Ding, 2018; Jin et al., 2022) and unconfoundedness (Cui and Tchetgen Tchetgen, 2021; Qi et al., 2023) for learning optimal ITRs is also a fruitful direction.

## Acknowledgments

The authors would like to thank the editor, the action editor, and three referees for constructive comments that substantially improved an earlier draft, and thank Maxime Fosset and Boris Jung for their help and support interpreting the BICAR-ICU trial and observational study data. This work is done while Pan Zhao was a PhD student at PreMeDICAL, Inria-Inserm and University of Montpellier. Pan Zhao and Julie Josse were publicly funded through ANR (the French National Research Agency) under the ‘‘Investissements d’avenir’’ programme with the reference ANR-16-IDEX-0006. Yang is partially supported by the USA National Institutes of Health NIA grant 1R01AG066883 and NIEHS grant 1R01ES031651 and the USA National Science Foundation grant SES 2242776.

## Appendix A. Preliminaries

### A.1 Counting Processes for Cox Model

We use the counting process theory of Andersen and Gill (1982) in our theoretical framework to study the large sample properties of Cox model. We state the existing results that are used in our proof.

Let  $X^{\otimes l}$  denote 1 for  $l = 0$ ,  $X$  for  $l = 1$ , and  $XX^T$  for  $l = 2$ . Define

$$U_a^{(l)}(\beta_a, t) = \frac{1}{n_a} \sum_{i=1}^n I\{A_i = a\} X_i^{\otimes l} \exp(\beta_a^T X_i) Y_i(t) \text{ and } u_a^{(l)}(\beta_a, t) = \mathbb{E} \left[ X^{\otimes l} \exp(\beta_a^T X) Y(t) \right],$$

where  $n_a = \sum_{i=1}^n I\{A_i = a\}$ , and define

$$E_a(\beta_a, t) = \frac{U_a^{(1)}(\beta_a, t)}{U_a^{(0)}(\beta_a, t)} \text{ and } e_a(\beta_a, t) = \frac{u_a^{(1)}(\beta_a, t)}{u_a^{(0)}(\beta_a, t)}.$$

The maximum partial likelihood estimator  $\hat{\beta}_a$  for the Cox proportional hazards model solves the estimating equation

$$\mathcal{S}_{a,n}(\beta_a) = \frac{1}{n_a} \sum_{i=1}^n I\{A_i = a\} \int \left\{ X_i - \frac{U_1^{(1)}(\beta_a, u)}{U_1^{(0)}(\beta_a, u)} \right\} dN_i(u) = 0,$$

and the cumulative baseline hazard function  $\hat{\Lambda}_{0,a}$  is estimated by the Breslow estimator:

$$\hat{\Lambda}_{0,a}(t) = \int_0^t \frac{\sum_{i=1}^n I\{A_i = a\} dN_i(u)}{\sum_{i=1}^n I\{A_i = a\} \exp(\hat{\beta}_a^T X_i) Y_i(u)}, a = 0, 1.$$

Under certain regularity conditions (Andersen and Gill, 1982, Conditions A – D),  $\hat{\beta}_a$  and  $\hat{\Lambda}_{0,a}$  converge in probability to the limits  $\beta_a^*$  and  $\Lambda_{0,a}^*$ , respectively; and we have

$$\sqrt{n_a}(\hat{\beta}_a - \beta_a^*) = \Gamma_a^{-1} \frac{1}{\sqrt{n_a}} \sum_{i=1}^n I\{A_i = a\} H_{a,i} + o_p(1),$$

where  $\Gamma_a = \mathbb{E}[-\partial \mathcal{S}_{a,n}(\beta_a^*) / \partial \beta_a^{*T}]$  is the Fisher information matrix of  $\beta_a^*$ ,  $H_{a,i} = \int I\{A_i = a\} \{X_i - e_a(\beta_a^*, u)\} dM_{a,i}(u)$  and  $dM_{a,i}(u) = dN_i(u) - \exp(\beta_a^{*T} X_i) Y_i(u) d\Lambda_{0,a}^*(u)$ . Moreover, let  $S^*(t | a, X) = \exp\{-\Lambda_{0,a}^*(t) \exp(\beta_a^{*T} X)\}$ ; it is shown that  $\sqrt{n_a}\{\hat{S}(t | a, X_i) - S^*(t | a, X_i)\}$  converges uniformly to a mean-zero Gaussian process for all  $X_i$ .

Specifically, we consider the following expansion that we use in our proof of Theorem 6 and Corollary 8,

$$\begin{aligned} \hat{S}(t | a, X_i) - S^*(t | a, X_i) &= -S^*(t | a, X_i) \Lambda_{0,a}^*(t) \exp(\beta_a^{*T} X_i) X_i^T (\hat{\beta}_a - \beta_a^*) \\ &\quad - S^*(t | a, X_i) \exp(\beta_a^{*T} X_i) (\hat{\Lambda}_{0,a}(t) - \Lambda_{0,a}^*(t)), \end{aligned}$$

and furthermore

$$\begin{aligned}
 \hat{\Lambda}_{0,a}(t) - \Lambda_{0,a}^*(t) &= \int_0^t \left\{ \frac{n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dN_i(u)}{U_a^{(0)}(\hat{\beta}_a, u)} - \frac{n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dN_i(u)}{U_a^{(0)}(\beta_a^*, u)} \right\} \\
 &\quad + \int_0^t \left\{ \frac{n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dN_i(u)}{U_a^{(0)}(\beta_a^*, u)} - d\Lambda_{0,a}^*(t) \right\} \\
 &= - \left[ \int_0^t \frac{U_a^{(1)}(\beta_a^*, u)}{\{U_a^{(0)}(\beta_a^*, u)\}^2} \left\{ n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dN_i(u) \right\} \right]^T (\hat{\beta}_a - \beta_a^*) \\
 &\quad + \int_0^t \frac{n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dM_{a,i}(u)}{U_a^{(0)}(\beta_a^*, u)} + o_p(1) \\
 &= - \left\{ \int_0^t e_a(\beta_a^*, u) d\Lambda_{0,a}^*(u) \right\}^T (\hat{\beta}_a - \beta_a^*) \\
 &\quad + \int_0^t \frac{n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dM_{a,i}(u)}{U_a^{(0)}(\beta_a^*, u)} + o_p(1).
 \end{aligned}$$

Combining the above two equations, we obtain

$$\begin{aligned}
 &\hat{S}(t | a, X_i) - S^*(t | a, X_i) \\
 &= \left[ -S^*(t | a, X_i) \Lambda_{0,a}^*(t) \exp(\beta_a^{*T} X_i) X_i^T - \left\{ \int_0^t e_a(\beta_a^*, u) d\Lambda_{0,a}^*(u) \right\}^T \right] (\hat{\beta}_a - \beta_a^*) \\
 &\quad + \int_0^t \frac{n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dM_{a,i}(u)}{U_a^{(0)}(\beta_a^*, u)} + o_p(1).
 \end{aligned}$$

## A.2 Cross-Fitting

To show the high-level idea of cross-fitting, we state the lemma from Kennedy et al. (2020), which is useful in our proof of Theorem 7 and Corollary 9.

**Lemma 13** *Consider two independent samples  $\mathcal{O}_1 = (O_1, \dots, O_n)$  and  $\mathcal{O}_2 = (O_{n+1}, \dots, O_{\tilde{n}})$ , let  $\hat{f}(o)$  be a function estimated from  $\mathcal{O}_2$  and  $\mathbb{P}_n$  the empirical measure over  $\mathcal{O}_1$ , then we have*

$$(\mathbb{P}_n - \mathbb{P})(\hat{f} - f) = O_{\mathbb{P}} \left( \frac{\|\hat{f} - f\|}{\sqrt{n}} \right)$$

**Proof** First note that by conditioning on  $\mathcal{O}_2$  we obtain

$$\mathbb{E} \left\{ \mathbb{P}_n(\hat{f} - f) \mid \mathcal{O}_2 \right\} = \mathbb{E}(\hat{f} - f \mid \mathcal{O}_2) = \mathbb{P}(\hat{f} - f)$$

and the conditional variance is

$$\text{var}\{(\mathbb{P}_n - \mathbb{P})(\hat{f} - f) \mid \mathcal{O}_2\} = \text{var}\{\mathbb{P}_n(\hat{f} - f) \mid \mathcal{O}_2\} = \frac{1}{n} \text{var}(\hat{f} - f \mid \mathcal{O}_2) \leq \|\hat{f} - f\|^2/n$$

therefore by Chebyshev's inequality we have

$$\mathbb{P} \left\{ \frac{|(\mathbb{P}_n - \mathbb{P})(\hat{f} - f)|}{\|\hat{f} - f\|^2/n} \geq t \right\} = \mathbb{E} \left[ \mathbb{P} \left\{ \frac{|(\mathbb{P}_n - \mathbb{P})(\hat{f} - f)|}{\|\hat{f} - f\|^2/n} \geq t \mid \mathcal{O}_2 \right\} \right] \leq \frac{1}{t^2}$$

thus for any  $\epsilon > 0$  we can pick  $t = 1/\sqrt{\epsilon}$  so that the probability above is no more than  $\epsilon$ , which yields the result.  $\blacksquare$

## Appendix B. Proof of Proposition 3

**Proof** We first show the identification by the outcome regression formula.

$$\begin{aligned} V(d) &= \mathbb{E}[\mathbb{E}[y(T(d)) \mid X]] \\ &= \mathbb{E}[d(X)\mathbb{E}[y(T(1)) \mid X] + (1 - d(X))\mathbb{E}[y(T(0)) \mid X]] \\ &= \mathbb{E}[d(X)\mathbb{E}[y(T(1)) \mid X, I_S = 1] + (1 - d(X))\mathbb{E}[y(T(0)) \mid X, I_S = 1]] \\ &= \mathbb{E}[d(X)\mathbb{E}[y(T(1)) \mid A = 1, X, I_S = 1] \\ &\quad + (1 - d(X))\mathbb{E}[y(T(0)) \mid A = 0, X, I_S = 1]] \\ &= \mathbb{E}[d(X)\mathbb{E}[y(T) \mid A = 1, X, I_S = 1] + (1 - d(X))\mathbb{E}[y(T) \mid A = 0, X, I_S = 1]] \\ &= \mathbb{E}[\mathbb{E}[y(T) \mid A = d(X), X, I_S = 1]] \\ &= \mathbb{E}[I_T e(X)\mathbb{E}[y(T) \mid A = d(X), X, I_S = 1]]. \end{aligned}$$

Similarly, we show the identification by the IPW formula.

$$\begin{aligned} V(d) &= \mathbb{E}[\mathbb{E}[y(T) \mid A = d(X), X, I_S = 1]] \\ &= \mathbb{E} \left[ \frac{I_S}{\pi_S(X)} \mathbb{E}[y(T) \mid A = d(X), X, I_S = 1] \right] \\ &= \mathbb{E} \left[ \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} \frac{\Delta y(U)}{S_C(U \mid A, X)} \right], \end{aligned}$$

where the last equation follows from the standard IPTW-IPCW formula (van der Laan and Robins, 2003).  $\blacksquare$

## Appendix C. Proof of Proposition 4

**Proof** While Lee et al. (2022) derived the efficient influence function for the treatment specific survival function, here we derive the EIF for the value function  $V(d) = \mathbb{E}[I_T e(X)\mu(d(X), X)]$ .

First consider the full data  $Z = (X, A, T, I_S, I_T)$ , and we have the factorization as

$$p(Z) = \{p(X)\pi_S(X)p(A \mid X, I_S = 1)p(T \mid A, X, I_S = 1)\}^{I_S} \{p(X)\}^{I_T}.$$

Since  $I_S I_T = 0$ , the score function is  $S(Z) = S(X, A, T, I_S) + I_T S(X)$ . Let  $V_\epsilon(d) = \mathbb{E}_\epsilon[I_T e(X)\mu_\epsilon(d(X), X)]$  denote the parameter of interest evaluated under the law  $p_\epsilon(Z)$ ,

where  $\epsilon$  indexes a regular parametric submodel such that  $p_0(Z)$  is the true data generating law. To establish that  $V(d)$  is pathwise differentiable with EIF  $\phi_d^F$ , we need to show that

$$\left. \frac{\partial}{\partial \epsilon} V_\epsilon(d) \right|_{\epsilon=0} = \mathbb{E}[\phi_d^F S(Z)].$$

First, we compute

$$\left. \frac{\partial}{\partial \epsilon} V_\epsilon(d) \right|_{\epsilon=0} = \mathbb{E}[I_T e(X)\mu(d(X), X)S(X)] + \mathbb{E}\left[\left. \frac{\partial}{\partial \epsilon} \mu_\epsilon(d(X), X) \right|_{\epsilon=0}\right],$$

and further write the first term on the right hand side as

$$\begin{aligned} \mathbb{E}[I_T e(X)\mu(d(X), X)S(X)] &= \mathbb{E}[(I_T e(X)\mu(d(X), X) - V(d))S(X)] \\ &= \mathbb{E}[(I_T e(X)\mu(d(X), X) - V(d))S(Z)], \end{aligned}$$

and the second term as

$$\begin{aligned} &\mathbb{E}\left[\left. \frac{\partial}{\partial \epsilon} \mu_\epsilon(d(X), X) \right|_{\epsilon=0}\right] \\ &= \mathbb{E}[d(X)\mathbb{E}[y(T)S(T | A, X, I_S) | A = 1, X, I_S = 1] \\ &\quad + (1 - d(X))\mathbb{E}[y(T)S(T | A, X, I_S) | A = 0, X, I_S = 1]] \\ &= \mathbb{E}[d(X)\mathbb{E}[(y(T) - \mu(1, X))S(T | A, X, I_S) | A = 1, X, I_S = 1] \\ &\quad + (1 - d(X))\mathbb{E}[(y(T) - \mu(0, X))S(T | A, X, I_S) | A = 0, X, I_S = 1]] \\ &= \mathbb{E}\left[d(X)\mathbb{E}\left[\frac{I_S A}{\pi_S(X)\pi_A(X)}(y(T) - \mu(1, X))S(T | A, X, I_S) \middle| X\right] \right. \\ &\quad \left. + (1 - d(X))\mathbb{E}\left[\frac{I_S(1 - A)}{\pi_S(X)(1 - \pi_A(X))}(y(T) - \mu(0, X))S(T | A, X, I_S) \middle| X\right]\right] \\ &= \mathbb{E}\left[\frac{I_S}{\pi_S(X)}\left(d(X)\frac{A}{\pi_A(X)}(y(T) - \mu(1, X)) \right. \right. \\ &\quad \left. \left. + (1 - d(X))\frac{1 - A}{1 - \pi_A(X)}(y(T) - \mu(0, X))\right)S(T | A, X, I_S)\right] \\ &= \mathbb{E}\left[\frac{I_S}{\pi_S(X)}\frac{I\{A = d(X)\}}{\pi_d(X)}(y(T) - \mu(A, X))S(Z)\right]. \end{aligned}$$

Therefore, the efficient influence function for the full data is

$$\phi_d^F = I_T e(X)\mu(d(X), X) + \frac{I_S}{\pi_S(X)}\frac{I\{A = d(X)\}}{\pi_d(X)}(y(T) - \mu(A, X)) - V(d).$$

Next, we consider the observed data  $O = (X, A, U, \Delta, I_S, I_T)$  due to right censoring. According to Tsiatis (2006, Section 10.4), the EIF based on the observed data is given by

$$\phi_d = \frac{\Delta \phi_d^F}{S_C(U | A, X)} + \int_0^\infty \frac{L(u, A, X)}{S_C(u | A, X)} dM_C(u | A, X),$$

where

$$\begin{aligned} L(u, A, X) &= \mathbb{E}[\phi_d^F | T \geq u, A, X] \\ &= I_T e(X) \mu(d(X), X) + \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} (Q(u, A, X) - \mu(A, X)) - V(d). \end{aligned}$$

Since we have

$$\begin{aligned} \int_0^\infty \frac{dM_C(u | A, X)}{S_C(u | A, X)} &= \int_0^\infty \frac{dN_C(u)}{S_C(u | A, X)} - \int_0^U \frac{d\Lambda_C(u | A, X)}{\exp\{\Lambda_C(u | A, X)\}} \\ &= 1 - \frac{\Delta}{S_C(U | A, X)}, \end{aligned} \quad (9)$$

we conclude that

$$\begin{aligned} \phi_d &= \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} \frac{\Delta y(U)}{S_C(U | A, X)} - V(d) \\ &\quad + \left( I_T e(X) - \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} \right) \mu(d(X), X) \\ &\quad + \frac{I_S}{\pi_S(X)} \frac{I\{A = d(X)\}}{\pi_d(X)} \int_0^\infty \frac{dM_C(u | A, X)}{S_C(u | A, X)} Q(u, A, X). \end{aligned}$$

■

## Appendix D. Proof of Theorem 6 and Corollary 8

### D.1 Double Robustness

**Proof** We start with the proof of the double robustness property. We show that EIF-based estimator is consistent when either the survival outcome model or the models for the sampling score, the propensity score and the censoring process are correctly specified. Under some regularity conditions, the nuisance estimators  $\hat{\mu}(a, x)$ ,  $\hat{Q}(u, a, x)$ ,  $\hat{\pi}_S(x)$ ,  $\hat{\pi}_A(x)$  and  $\hat{S}_C(t | a, x)$  converge in probability to  $\mu^*(a, x)$ ,  $Q^*(u, a, x)$ ,  $\pi_S^*(x)$ ,  $\pi_A^*(x)$  and  $S_C^*(t | a, x)$ , respectively. It suffices to show that  $\mathbb{E}[V^*(d)] = V(d)$ , where

$$\begin{aligned} V^*(d) &= I_T e(X) \mu^*(A = d(X), X) \\ &\quad + \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \left\{ \frac{\Delta y(U)}{S_C^*(U | A, X)} - \mu^*(A, X) \right. \\ &\quad \left. + \int_0^\infty \frac{dM_C^*(u | A, X)}{S_C^*(u | A, X)} Q^*(u, A, X) \right\} \\ &= (I) + (II) + (III). \end{aligned}$$

First, consider the case when the survival outcome model is correct, thus we have

$$(I) = \mathbb{E}[I_T e(X) \mu^*(A = d(X), X)] = V(d)$$

By Equation (9), we obtain

$$(II) + (III) = \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \left\{ y(T) - \mu^*(A, X) - \int_0^\infty \frac{dM_C^*(u | A, X)}{S_C^*(u | A, X)} (y(T) - Q^*(u, A, X)) \right\}.$$

In this case, we have

$$\begin{aligned} & \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} (y(T) - \mu^*(A, X)) \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} (y(T) - \mu^*(A, X)) \mid X \right] \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} (y(T) - \mu^*(A, X)) \mid A, X, I_S = 1 \right] \mid X \right] \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \mathbb{E}[(y(T) - \mu^*(A, X)) \mid A, X, I_S = 1] \mid X \right] \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} (\mathbb{E}[y(T) \mid A, X, I_S = 1] - \mu^*(A, X)) \mid X \right] \right] = 0. \end{aligned}$$

Also define  $d\tilde{M}_C(u | A, X) = d\tilde{N}_C(u) - I\{C \geq u\}d\Lambda_C(u | A, X)$  where  $\tilde{N}_C(u) = I\{C \leq u\}$ , so we have

$$\begin{aligned} & \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \int_0^\infty \frac{dM_C^*(u | A, X)}{S_C^*(u | A, X)} (y(T) - Q^*(u, A, X)) \right] \\ &= \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \int_0^\infty \frac{d\tilde{M}_C(u | A, X)}{S_C^*(u | A, X)} I\{T \geq u\} (y(T) - Q^*(u, A, X)) \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \int_0^\infty \frac{d\tilde{M}_C(u | A, X)}{S_C^*(u | A, X)} I\{T \geq u\} (y(T) - Q^*(u, A, X)) \mid X \right] \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \int_0^\infty \frac{d\tilde{M}_C(u | A, X)}{S_C^*(u | A, X)} I\{T \geq u\} (y(T) - Q^*(u, A, X)) \mid A, X, C, I_S = 1 \right] \mid X \right] \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \int_0^\infty \frac{d\tilde{M}_C(u | A, X)}{S_C^*(u | A, X)} \mathbb{E}[I\{T \geq u\} (y(T) - Q^*(u, A, X)) \mid A, X, C, I_S = 1] \mid X \right] \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \int_0^\infty \frac{d\tilde{M}_C(u | A, X)}{S_C^*(u | A, X)} (\mathbb{E}[I\{T \geq u\}y(T) \mid A, X, I_S = 1] - \mathbb{E}[I\{T \geq u\} \mid A, X, I_S = 1]Q^*(u, A, X)) \mid X \right] \right] = 0. \end{aligned}$$



Next, consider the case when the models for the sampling score, the propensity score and the censoring process are correctly specified. Rearranging the terms of  $V^*(d)$ , we obtain

$$\begin{aligned} V^*(d) &= \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \frac{\Delta y(U)}{S_C^*(U|A, X)} \\ &\quad + \left( I_T e(X) - \frac{I_S}{\pi_S^*(X)} \right) \mu^*(A = d(X), X) \\ &\quad + \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \int_0^\infty \frac{dM_C^*(u|A, X)}{S_C^*(u|A, X)} Q^*(u, A, X) \\ &= (I) + (II) + (III). \end{aligned}$$

In this case, we have

$$\begin{aligned} (I) &= \mathbb{E} \left[ \frac{I_S}{\pi_S^*(X)} \frac{I\{A = d(X)\}}{\pi_d^*(X)} \frac{\Delta y(U)}{S_C^*(U|A, X)} \right] = V(d), \\ (II) &= \mathbb{E} \left[ \left( I_T e(X) - \frac{I_S}{\pi_S^*(X)} \right) \mu^*(A = d(X), X) \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ I_T e(X) - \frac{I_S}{\pi_S^*(X)} \middle| X \right] \mu^*(A = d(X), X) \right] = 0, \end{aligned}$$

and (III) is a stochastic integral with respect to the martingale  $M_C^*(u|A, X)$ , thus equals 0 as well, which completes the double robustness property.  $\blacksquare$

## D.2 Asymptotic Properties

To establish the asymptotic results, we need some regularity conditions such that the nuisance estimators  $\mu(a, x; \hat{\beta}_a, \hat{\Lambda}_{0,a})$ ,  $Q(u, a, x; \hat{\beta}_a, \hat{\Lambda}_{0,a})$ ,  $\pi_S(x; \hat{\lambda})$ ,  $\pi_A(x; \hat{\theta})$  and  $S_C(u|a, x; \hat{\alpha}_a, \hat{\Lambda}_{C0,a})$  converge in probability to  $\mu(a, x; \beta_a^*, \Lambda_{0,a}^*)$ ,  $Q(u, a, x; \beta_a^*, \Lambda_{0,a}^*)$ ,  $\pi_S(x; \lambda^*)$ ,  $\pi_A(x; \theta^*)$  and  $S_C(t|a, x; \alpha_a^*, \Lambda_{C0,a}^*)$ , respectively.

**Condition 1** We assume the following conditions hold:

(C1)  $X$  is bounded almost surely.

(C2) The equation  $\mathbb{E} \left[ \left\{ A - \frac{\exp(\theta^T X)}{1 + \exp(\theta^T X)} \right\} X \right] = 0$  has a unique solution  $\theta^*$ .

(C3) For  $a = 0, 1$ , the equation

$$\mathbb{E} \left[ \int_0^L \left( X_i - \frac{\mathbb{E}[Y_i(u) \exp(\beta_a^T X) X]}{\mathbb{E}[Y_i(u) \exp(\beta_a^T X)]} \right) \times dN_i(u) \right] = 0,$$

has a unique solution  $\beta_a^*$ , where  $L > u$  is a pre-specified time point such that  $\Pr(U_i > L) > 0$ . Moreover, let

$$\Lambda_{0,a}^*(u) = \mathbb{E} \left[ \int_0^u \frac{dN_i(u)}{\mathbb{E}[Y_i(u) \exp(\beta_a^{*T} X_i)]} \right],$$

and assume  $\Lambda_{0,a}^*(L) < \infty$ .

(C4) For  $a = 0, 1$ , the equation

$$\mathbb{E} \left[ \int_0^L \left( X_i - \frac{\mathbb{E}[Y_i(u) \exp(\alpha_a^T X) X]}{\mathbb{E}[Y_i(u) \exp(\alpha_a^T X)]} \right) \times dN_i(u) \right] = 0,$$

has a unique solution  $\alpha_a^*$ . Moreover, let

$$\Lambda_{C0,a}^*(u) = \mathbb{E} \left[ \int_0^u \frac{dN_i(u)}{\mathbb{E}[Y_i(u) \exp(\alpha_a^{*T} X_i)]} \right],$$

and assume  $\Lambda_{C0,a}^*(L) < \infty$ .

(C5) The estimating equation for the sampling score model  $\pi_S(X; \lambda)$  has a unique solution  $\lambda^*$ , and achieves root- $n$  rate of convergence.

Under Condition 1, we have the following asymptotic representations:

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_{\theta_i} + o_p(1), & \sqrt{n}(\hat{\lambda} - \lambda^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_{\lambda_i} + o_p(1), \\ \sqrt{n}(\hat{\beta}_a - \beta_a^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_{\beta_{ai}} + o_p(1), & \sqrt{n}(\hat{\alpha}_a - \alpha_a^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_{\alpha_{ai}} + o_p(1), \quad \text{for } a = 0, 1. \end{aligned}$$

For the ease of notation, we introduce a finite population of size  $N$ , from which the source and target data are independently sampled. Note that we do not assume  $N$  is known and fixed.

We focus on the estimation of survival functions by our proposed method:

$$\begin{aligned} \hat{S}(t; \eta) &= \frac{1}{N} \sum_{i=1}^N \left[ I_{T,i} e(X_i) \hat{S}(t | A = d_\eta(X_i), X_i) \right. \\ &\quad + \frac{I_{S,i} I\{A_i = d_\eta(X_i)\}}{\hat{\pi}_S(X_i) \hat{\pi}_d(X_i)} \left\{ \frac{\Delta_i Y_i(t)}{\hat{S}_C(t | A_i, X_i)} - \hat{S}(t | A_i, X_i) \right. \\ &\quad \left. \left. + \int_0^\infty \frac{\hat{S}(t | A_i, X_i) d\hat{M}_C(u | A_i, X_i)}{\hat{S}(u | A_i, X_i) \hat{S}_C(u | A_i, X_i)} \right\} \right], \end{aligned}$$

and for the ease of notation, define

$$\begin{aligned} \hat{J}(t, a, x) &= \frac{\Delta_i Y_i(t)}{\hat{S}_C(t | a, x)} - \hat{S}(t | a, x) + \int_0^\infty \frac{\hat{S}(t | a, x) d\hat{M}_C(u | a, x)}{\hat{S}(u | a, x) \hat{S}_C(u | a, x)}, \\ J^*(t, a, x) &= \frac{\Delta_i Y_i(t)}{S_C^*(t | a, x)} - S^*(t | a, x) + \int_0^\infty \frac{S^*(t | a, x) dM_C^*(u | a, x)}{S^*(u | a, x) S_C^*(u | a, x)}. \end{aligned}$$

**Proof** Our proof has three main parts as follows.

**PART 1.** By the double robustness property shown in Section D.1, we have, by the strong law of large numbers and uniform consistency, that  $\hat{S}(t; \eta) = S(t; \eta) + o_p(1)$ , which proves (i) of Theorem 6. Moreover, define

$$S_N^*(t; \eta) = \frac{1}{N} \sum_{i=1}^N \left[ I_{T,i} e(X_i) S^*(t | A = d_\eta(X_i), X_i) + \frac{I_{S,i} I\{A_i = d_\eta(X_i)\}}{\pi_S^*(X_i) \pi_d^*(X_i)} J^*(t, A_i, X_i) \right],$$

and by applying the Taylor expansion and the counting processes result in Section A.1, we obtain

$$\begin{aligned} \hat{S}(t; \eta) &= S_n^*(t; \eta) + H_\lambda^T(\hat{\lambda} - \lambda^*) + H_\theta^T(\hat{\theta} - \theta^*) + H_{\beta_0}^T(\hat{\beta}_0 - \beta_0^*) + H_{\beta_1}^T(\hat{\beta}_1 - \beta_1^*) \\ &\quad + H_{\alpha_0}^T(\hat{\alpha}_0 - \alpha_0^*) + H_{\alpha_1}^T(\hat{\alpha}_1 - \alpha_1^*) + R_S + o_p(N^{-1/2}), \end{aligned}$$

where

$$\begin{aligned} H_\lambda &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{\partial \hat{S}(t; \eta)}{\partial \lambda^*}, H_\theta = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{\partial \hat{S}(t; \eta)}{\partial \theta^*}, \\ H_{\beta_a} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left\{ I_{T,i} e(X_i) (-1)^{a+1} G(t, a, X_i) + \frac{I_{S,i} I\{A_i = a\}}{\pi_S^*(X_i) \pi_d^*(X_i)} \left( \int_0^\infty \frac{G(t, a, X_i) dM_C^*(u | a, X_i)}{S^*(u | a, X_i) S_C^*(u | a, X_i)} \right. \right. \\ &\quad \left. \left. - G(t, a, X_i) - \int_0^\infty \frac{G(u, a, X_i) S^*(t | a, X_i) dM_C^*(u | a, X_i)}{S^{*2}(u | a, X_i) S_C^*(u | a, X_i)} \right) \right\}, \\ H_{\alpha_a} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{I_{S,i} I\{A_i = a\}}{\pi_S^*(X_i) \pi_d^*(X_i)} \left\{ \frac{-\Delta_i Y_i(t)}{S_C^*(t | a, X_i)} G_C(t, a, X_i) \right. \\ &\quad \left. - \int_0^\infty \frac{G_C(u, a, X_i) S^*(t | a, X_i) dM_C^*(u | a, X_i)}{S_C^{*2}(u | a, X_i) S^*(u | a, X_i)} + \tilde{G}_C(t, a, X_i) \right\}, \\ R_S &= \frac{1}{N} \sum_{i=1}^N \sum_{a=0,1} \left\{ I_{T,i} e(X_i) (-1)^{a+1} H(t, a, X_i) \right. \\ &\quad + \frac{I_{S,i} I\{A_i = a\}}{\pi_S^*(X_i) \pi_d^*(X_i)} \left( \int_0^\infty \frac{H(t, a, X_i) dM_C^*(u | a, X_i)}{S_C^*(u | a, X_i) S^*(u | a, X_i)} - H(t, a, X_i) \right. \\ &\quad - \int_0^\infty \frac{H(u, a, X_i) S^*(t | a, X_i) dM_C^*(u | a, X_i)}{S_C^*(u | a, X_i) S^{*2}(u | a, X_i)} - \frac{\Delta_i Y_i(t)}{S_C^*(t | a, X_i)} H_C(t, a, X_i) \\ &\quad \left. \left. - \int_0^\infty \frac{H_C(u, a, X_i) S^*(t | a, X_i) dM_C^*(u | a, X_i)}{S_C^{*2}(u | a, X_i) S^*(u | a, X_i)} - \tilde{H}_C(t, a, X_i) \right) \right\} \\ &= \frac{1}{N} \sum_{i=1}^N \phi_{R_S, i}, \end{aligned}$$

with

$$\begin{aligned} G(t, a, x) &= -S^*(t | a, x) \Lambda_{0,a}^*(t) x^T + S^*(t | a, x) \exp(\beta_a^{*T} x) \left\{ \int_0^t e_a(\beta_a^*, u) d\Lambda_{0,a}^*(u) \right\}^T, \\ H(t, a, x) &= -S^*(t | a, x) \exp(\beta_a^{*T} x) \int_0^t \frac{n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dM_{a,i}(u)}{U_a^{(0)}(\beta_a^*, u)}, \\ G_C(t, a, x) &= -S^*(t | a, x) \Lambda_{0,a}^*(t) x^T + S^*(t | a, x) \exp(\beta_a^{*T} x) \left\{ \int_0^t e_a(\beta_a^*, u) d\Lambda_{0,a}^*(u) \right\}^T, \\ H_C(t, a, x) &= -S^*(t | a, x) \exp(\beta_a^{*T} x) \int_0^t \frac{n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dM_{a,i}(u)}{U_a^{(0)}(\beta_a^*, u)}, \end{aligned}$$

$$\tilde{G}_C(t, a, x) = \int_0^{U_i} \frac{S^*(t|a, x) d\Lambda_C^*(u|a, x)}{S_C^*(u|a, x) S^*(u|a, x)} x^T + \left\{ \int_0^t \frac{S^*(t|a, x) e_a(\beta_a^*, u) d\Lambda_{0,a}^*(u)}{S_C^*(u|a, x) S^*(u|a, x)} \right\}^T,$$

$$\tilde{H}_C(t, a, x) = \int_0^t \frac{S^*(t|a, x) n_a^{-1} \sum_{i=1}^n I\{A_i = a\} dM_{a,i}(u)}{S_C^*(u|a, x) S^*(u|a, x) U_a^{(0)}(\beta_a^*, u)}.$$

Thus, we have

$$\sqrt{N} \left\{ \hat{S}(t; \eta) - S(t; \eta) \right\} = \frac{1}{\sqrt{N}} \sum_{i=1}^N (\xi_{1,i}(t; \eta) + \xi_{2,i}(t; \eta)) + o_p(1), \quad (10)$$

where

$$\xi_{1,i}(t; \eta) = S_n^*(t; \eta) - S(t; \eta),$$

$$\xi_{2,i}(t; \eta) = H_\lambda^T \phi_{\lambda^*,i} + H_\theta^T \phi_{\theta^*,i} + \sum_{a=0,1} H_{\beta_a}^T \phi_{\beta_a^*,i} + \sum_{a=0,1} H_{\alpha_a}^T \phi_{\alpha_a^*,i} + H_{\alpha_1}^T + \phi_{Rs,i},$$

and  $\xi_{1,i}(t; \eta), \xi_{2,i}(t; \eta)$  are independent mean-zero processes. Therefore, we obtain that  $\sqrt{N} \left\{ \hat{S}(t; \eta) - S(t; \eta) \right\}$  converges weakly to a mean-zero Gaussian process, which proves (ii) of Theorem 6.

**PART 2.** We show that  $N^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ . Recall that

$$\hat{\eta} = \arg \max_{\eta} \hat{S}(t; \eta) \text{ and } \eta^* = \arg \max_{\eta} S(t; \eta).$$

By Assumption 5 (i),  $S(t; \eta)$  is twice continuously differentiable at a neighborhood of  $\eta^*$ ; in Step 1, we show that  $\hat{S}(t; \eta) = S(t; \eta) + o_p(1), \forall \eta$ ; since  $\hat{\eta}$  maximizes  $\hat{S}(t; \eta)$ , we have that  $\hat{S}(t; \hat{\eta}) \geq \sup_{\eta} \hat{S}(t; \eta)$ , thus by the Argmax theorem, we have  $\hat{\eta} \xrightarrow{P} \eta^*$  as  $N \rightarrow \infty$ .

In order to establish the  $N^{-1/3}$  rate of convergence of  $\hat{\eta}$ , we apply Theorem 14.4 (Rate of convergence) of Kosorok (2008), and need to find the suitable rate that satisfies three conditions below.

**Condition 1** For every  $\eta$  in a neighborhood of  $\eta^*$  such that  $\|\eta - \eta^*\|_2 < \delta$ , by Assumption 5 (i), we apply the second-order Taylor expansion,

$$\begin{aligned} S(t; \eta) - S(t; \eta^*) &= S'(\eta^*) \|\eta - \eta^*\|_2 + \frac{1}{2} S''(\eta^*) \|\eta - \eta^*\|_2^2 + o(\|\eta - \eta^*\|_2^2) \\ &= \frac{1}{2} S''(\eta^*) \|\eta - \eta^*\|_2^2 + o(\|\eta - \eta^*\|_2^2), \end{aligned}$$

and as  $S''(\eta^*) < 0$ , there exists  $c_0 = -\frac{1}{2} S''(\eta^*) > 0$  such that  $S(t; \eta) - S(t; \eta^*) \leq -c_0 \|\eta - \eta^*\|_2^2$ .

**Condition 2** For all  $N$  large enough and sufficiently small  $\delta$ , we consider the centered process  $\hat{S} - S$ , and have that

$$\begin{aligned} & \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{S}(t; \eta) - S(t; \eta) - \left\{ \hat{S}(t; \eta^*) - S(t; \eta^*) \right\} \right| \right] \\ &= \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{S}(t; \eta) - S_n^*(t; \eta) + S_n^*(t; \eta) - S(t; \eta) \right. \right. \\ & \quad \left. \left. - \left\{ \hat{S}(t; \eta^*) - S_n^*(t; \eta^*) + S_n^*(t; \eta^*) - S(t; \eta^*) \right\} \right| \right] \\ &\leq \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{S}(t; \eta) - S_n^*(t; \eta) - \left\{ \hat{S}(t; \eta^*) - S_n^*(t; \eta^*) \right\} \right| \right] \end{aligned} \quad (I)$$

$$+ \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| S_n^*(t; \eta) - S(t; \eta) - \left\{ S_n^*(t; \eta^*) - S(t; \eta^*) \right\} \right| \right], \quad (II)$$

and we bound (I) and (II) respectively as follows.

**Condition 2.1** To bound (II), we need the useful facts that

$$I\{A = d_\eta(X)\} - I\{A = d_{\eta^*}(X)\} = (2A - 1)(d_\eta(X) - d_{\eta^*}(X)),$$

$$S^*(t | d_\eta(X_i), X_i) - S^*(t | d_{\eta^*}(X_i), X_i) = (S^*(t | 1, X_i) - S^*(t | 0, X_i))(d_\eta(X_i) - d_{\eta^*}(X_i)),$$

and obtain

$$\begin{aligned} S_n^*(t; \eta) - S_n^*(t; \eta^*) &= \frac{1}{N} \sum_{i=1}^N (d_\eta(X_i) - d_{\eta^*}(X_i)) \\ &\times \left\{ I_{T,i} e(X_i) (S^*(t | 1, X_i) - S^*(t | 0, X_i)) + \frac{(2A_i - 1)I_{S,i}}{\pi_S^*(X_i)\pi_a^*(X_i)} J^*(t, A_i, X_i) \right\}. \end{aligned}$$

Define a class of functions

$$\begin{aligned} \mathcal{F}_\eta^1 &= \left\{ (d_\eta(x) - d_{\eta^*}(x)) \left( I_T e(x) (S^*(t | 1, x) - S^*(t | 0, x)) + \frac{(2a - 1)I_S}{\pi_a^*(x)\pi_S^*(x)} J^*(t, a, x) \right) : \right. \\ & \quad \left. \|\eta - \eta^*\|_2 < \delta \right\}, \end{aligned}$$

and let  $M_1 = \sup \left| I_T e(x) (S^*(t | 1, x) - S^*(t | 0, x)) + \frac{(2a-1)I_S}{\pi_a^*(x)\pi_S^*(x)} J^*(t, a, x) \right|$ . By Assumption 1, 3 and Condition 1, we have that  $M_1 < \infty$ .

When  $\|\eta - \eta^*\|_2 < \delta$ , by Condition 1 (C1), there exists a constant  $0 < k_0 < \infty$  such that  $|(1, x^T)(\eta - \eta^*)| < k_0\delta$ ; furthermore, we show that  $|d_\eta(x) - d_{\eta^*}(x)| = |I\{(1, x^T)\eta > 0\} - I\{(1, x^T)\eta^* > 0\}| \leq I\{-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta\}$ , by considering the three cases:

- when  $-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta$ , we have  $|d_\eta(x) - d_{\eta^*}(x)| \leq 1 = I\{-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta\}$ ;

- when  $(1, x^T)\eta^* > k_0\delta > 0$ , we have  $(1, x^T)\eta = (1, x^T)(\eta - \eta^*) + (1, x^T)\eta^* > 0$ , so  $|d_\eta(x) - d_{\eta^*}(x)| = 0 = I\{-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta\}$ ;
- when  $(1, x^T)\eta^* < -k_0\delta < 0$ , we have  $(1, x^T)\eta = (1, x^T)(\eta - \eta^*) + (1, x^T)\eta^* < 0$ , so  $|d_\eta(x) - d_{\eta^*}(x)| = 0 = I\{-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta\}$ .

Thus we can define the envelope of  $\mathcal{F}_\eta^1$  as  $F_1 = M_1 I\{-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta\}$ . By Assumption 5 (ii), there exists a constant  $0 < k_1 < \infty$  such that

$$\|F_1\|_{P,2} \leq M_1 \sqrt{\Pr(-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta)} \leq M_1 \sqrt{2k_0 k_1} \delta^{1/2}.$$

By Lemma 9.6 and Lemma 9.9 of Kosorok (2008), we have that  $\mathcal{F}_\eta^1$ , a class of indicator functions, is a Vapnik-Cervonenkis (VC) class with bounded bracketing entropy  $J_{[]}^*(1, \mathcal{F}_\eta^1) < \infty$ .

Since we have the fact that

$$\begin{aligned} \mathbb{G}_N \mathcal{F}_\eta^1 &= N^{-1/2} \sum_{i=1}^N \{\mathcal{F}_\eta^1 - \mathbb{E}[\mathcal{F}_\eta^1]\} \\ &= \sqrt{N} (S_n^*(t; \eta) - S_n^*(t; \eta^*) - \{S(t; \eta) - S(t; \eta^*)\}), \end{aligned}$$

By Theorem 11.2 of Kosorok (2008), we obtain that there exists a constant  $0 < c_1 < \infty$ ,

$$(II) = \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_N \mathcal{F}_\eta^1| \right] \leq c_1 J_{[]}^*(1, \mathcal{F}_\eta^1) \|F_1\|_{P,2} \leq c_1 J_{[]}^*(1, \mathcal{F}_\eta^1) M_1 \sqrt{2k_0 k_1} \delta^{1/2} = \tilde{c}_1 \delta^{1/2},$$

so we conclude that  $(II) \leq \tilde{c}_1 \delta^{1/2}$  where  $\tilde{c}_1 > 0$  is a finite constant.

**Condition 2.2** To bound (I), first we have

$$\begin{aligned} &\hat{S}(t; \eta) - S_n^*(t; \eta) - \{\hat{S}(t; \eta^*) - S_n^*(t; \eta^*)\} = \hat{S}(t; \eta) - \hat{S}(t; \eta^*) - \{S_n^*(t; \eta) - S_n^*(t; \eta^*)\} \\ &= \frac{1}{N} \sum_{i=1}^N (d_\eta(X_i) - d_{\eta^*}(X_i)) \left[ I_{T,i} e(X_i) \{\hat{S}(t|1, X_i) - \hat{S}(t|0, X_i) - (S^*(t|1, X_i) - S^*(t|0, X_i))\} \right. \\ &\quad \left. + \frac{(2A_i - 1)I_{S,i}}{\hat{\pi}_{A_i}(X_i)\hat{\pi}_S(X_i)} \hat{J}(t, A_i, X_i) - \frac{(2A_i - 1)I_{S,i}}{\pi_{A_i}^*(X_i)\pi_S^*(X_i)} J^*(t, A_i, X_i) \right], \end{aligned}$$

and then apply the Taylor expansion and counting processes result in Section A.1,

$$\begin{aligned} &\hat{S}(t; \eta) - S_n^*(t; \eta) - \{\hat{S}(t; \eta^*) - S_n^*(t; \eta^*)\} \\ &= \frac{1}{N} \sum_{i=1}^N (d_\eta(X_i) - d_{\eta^*}(X_i)) \times \left\{ D_\lambda(\hat{\lambda} - \lambda^*) + D_\theta(\hat{\theta} - \theta^*) + D_{\beta_0}(\hat{\beta}_0 - \beta_0^*) \right. \\ &\quad \left. + D_{\beta_1}(\hat{\beta}_1 - \beta_1^*) + D_{\alpha_0}(\hat{\alpha}_0 - \alpha_0^*) + D_{\alpha_1}(\hat{\alpha}_1 - \alpha_1^*) + R_{S,i} \right\} + o_p(N^{-1/2}), \end{aligned} \quad (11)$$

where

$$D_\lambda = -\frac{(2A_i - 1)I_{S,i}}{\pi_{A_i}^*(X_i)\pi_S^{*2}(X_i)} J^*(t, A_i, X_i) \left( \frac{\partial \pi_S^*(X_i)}{\partial \lambda} \right)^T,$$

$$D_\theta = -\frac{I_{S,i}}{\pi_{A_i}^*(X_i)\pi_S^*(X_i)} J^*(t, A_i, X_i) \left( \frac{\partial \pi_A^*(X_i)}{\partial \theta} \right)^T,$$

$$D_{\beta_a} = I_{T,i} e(X_i) (-1)^{a+1} G(t, a, X_i) + \frac{(2A_i - 1)I\{A_i = a\}I_{S,i}}{\pi_{A_i}^*(X_i)\pi_S^*(X_i)} \left( \int_0^\infty \frac{G(t, a, X_i) dM_C^*(u | a, X_i)}{S_C^*(u | a, X_i) S^*(u | a, X_i)} \right. \\ \left. - G(t, a, X_i) - \int_0^\infty \frac{G(u, a, X_i) S^*(t | a, X_i) dM_C^*(u | a, X_i)}{S_C^*(u | a, X_i) S^{*2}(u | a, X_i)} \right),$$

$$D_{\alpha_a} = \frac{(2A_i - 1)I\{A_i = a\}I_{S,i}}{\pi_{A_i}^*(X_i)\pi_S^*(X_i)} \left\{ -\frac{\Delta_i Y_i(t)}{S_C^*(t | a, X_i)} G_C(t, a, X_i) \right. \\ \left. - \int_0^\infty \frac{G_C(u, a, X_i) S^*(t | a, X_i) dM_C^*(u | a, X_i)}{S_C^{*2}(u | a, X_i) S^*(u | a, X_i)} + \tilde{G}_C(t, a, X_i) \right\},$$

$$R_{S,i} = \sum_{a=0,1} \left[ I_{T,i} e(X_i) (-1)^{a+1} H(t, a, X_i) + \frac{(2A_i - 1)I\{A_i = a\}I_{S,i}}{\pi_{A_i}^*(X_i)\pi_S^*(X_i)} \left( \int_0^\infty \frac{H(t, a, X_i) dM_C^*(u | a, X_i)}{S_C^*(u | a, X_i) S^*(u | a, X_i)} \right. \right. \\ \left. - H(t, a, X_i) - \int_0^\infty \frac{H(u, a, X_i) S^*(t | a, X_i) dM_C^*(u | a, X_i)}{S_C^*(u | a, X_i) S^{*2}(u | a, X_i)} \right. \\ \left. - \frac{\Delta_i Y_i(t)}{S_C^*(t | a, X_i)} H_C(t, a, X_i) - \int_0^\infty \frac{H_C(u, a, X_i) S^*(t | a, X_i) dM_C^*(u | a, X_i)}{S_C^{*2}(u | a, X_i) S^*(u | a, X_i)} - \tilde{H}_C(t, a, X_i) \right)].$$

Similarly, we define the following classes of functions:

$$\mathcal{F}_\eta^2 = \left\{ (d_\eta(x) - d_{\eta^*}(x)) \frac{(2a-1)I_{S,i}}{\pi_a^*(x)\pi_S^{*2}(x)} J^*(t, a, x) \left( \frac{\partial \pi_S^*(x)}{\partial \lambda} \right)^T : \|\eta - \eta^*\|_2 < \delta \right\},$$

$$\mathcal{F}_\eta^3 = \left\{ (d_\eta(x) - d_{\eta^*}(x)) \frac{-I_{S,i}}{\pi_a^*(x)\pi_S^*(x)} J^*(t, a, x) \left( \frac{\partial \pi_A^*(x)}{\partial \theta} \right)^T : \|\eta - \eta^*\|_2 < \delta \right\},$$

$$\mathcal{F}_\eta^4 = \left\{ (d_\eta(x) - d_{\eta^*}(x)) \left[ I_T e(x) (-1)^{a+1} G(t, a, x) + \frac{(2a-1)I_S}{\pi_a^*(x)\pi_S^*(x)} \right. \right. \\ \times \left( \int_0^\infty \frac{G(t, a, x) dM_C^*(u | a, x)}{S_C^*(u | a, x) S^*(u | a, x)} - G(t, a, x) \right. \\ \left. \left. - \int_0^\infty \frac{G(u, a, x) S^*(t | a, x) dM_C^*(u | a, x)}{S_C^*(u | a, x) S^{*2}(u | a, x)} \right) \right] : \|\eta - \eta^*\|_2 < \delta \right\},$$

$$\mathcal{F}_\eta^5 = \left\{ (d_\eta(x) - d_{\eta^*}(x)) \left[ I_T e(x) (-1)^{a+1} G(t, a, x) + \frac{(2a-1)I_S}{\pi_a^*(x)\pi_S^*(x)} \right. \right. \\ \times \left( \int_0^\infty \frac{G(t, a, x) dM_C^*(u | a, x)}{S_C^*(u | a, x) S^*(u | a, x)} - G(t, a, x) \right. \\ \left. \left. - \int_0^\infty \frac{G(u, a, x) S^*(t | a, x) dM_C^*(u | a, x)}{S_C^*(u | a, x) S^{*2}(u | a, x)} \right) \right] : \|\eta - \eta^*\|_2 < \delta \right\},$$

$$\mathcal{F}_\eta^6 = \left\{ (d_\eta(x) - d_{\eta^*}(x)) \left[ \frac{(2a-1)I_S}{\pi_a^*(x)\pi_S^*(x)} \left\{ -\frac{\Delta Y(t)}{S_C^*(t|a,x)} G_C(t,a,x) - \int_0^\infty \frac{G_C(u,a,x)S^*(t|a,x)dM_C^*(u|a,x)}{S_C^{*2}(u|a,x)S^*(u|a,x)} + \tilde{G}_C(t,a,x) \right\} \right] : \|\eta - \eta^*\|_2 < \delta \right\},$$

$$\mathcal{F}_\eta^7 = \left\{ (d_\eta(x) - d_{\eta^*}(x)) \left[ \frac{(2a-1)I_S}{\pi_a^*(x)\pi_S^*(x)} \left\{ -\frac{\Delta Y(t)}{S_C^*(t|a,x)} G_C(t,a,x) - \int_0^\infty \frac{G_C(u,a,x)S^*(t|a,x)dM_C^*(u|a,x)}{S_C^{*2}(u|a,x)S^*(u|a,x)} + \tilde{G}_C(t,a,x) \right\} \right] : \|\eta - \eta^*\|_2 < \delta \right\},$$

$$\begin{aligned} \mathcal{F}_\eta^8 = & \left\{ (d_\eta(x) - d_{\eta^*}(x)) \left[ \sum_{a=0,1} \left[ I_T e(x)^{a+1} H(t,a,x) + \frac{(2a-1)I_S}{\pi_a^*(x)\pi_S^*(x)} \right. \right. \right. \\ & \times \left( \int_0^\infty \frac{H(t,a,x)dM_C^*(u|a,x)}{S_C^*(u|a,x)S^*(u|a,x)} - H(t,a,x) \right. \\ & - \int_0^\infty \frac{H(u,a,x)S^*(t|a,x)dM_C^*(u|a,x)}{S_C^*(u|a,x)S^{*2}(u|a,x)} - \frac{\Delta Y(t)}{S_C^*(t|a,x)} H_C(t,a,x) \\ & \left. \left. \left. - \int_0^\infty \frac{H_C(u,a,x)S^*(t|a,x)dM_C^*(u|a,x)}{S_C^{*2}(u|a,x)S^*(u|a,x)} - \tilde{H}_C(t,a,x) \right) \right] \right] : \|\eta - \eta^*\|_2 < \delta \right\}. \end{aligned}$$

Let

$$M_2 = \sup \left| \frac{(2a-1)}{\pi_a^*(x)} J^*(t,a,x) \left( \frac{\partial \pi_S^*(x)}{\partial \lambda} \right)^T \right|,$$

where  $M_2 \in \mathbb{R}^+$  and the supremum is taken over all the coordinates; and  $M_3, \dots, M_8$  are defined accordingly for  $\mathcal{F}_\eta^3, \dots, \mathcal{F}_\eta^8$ . By Assumption 1, 3 and Condition 1, we have that  $M_2, \dots, M_8 < \infty$ .

Using the same technique as in **Condition 2.1**, we define the envelop of  $\mathcal{F}_\eta^j$  as  $F_j = M_j I \{-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta\}$  for  $j = 2, \dots, 8$ , and obtain that

$$\|F_j\|_{P,2} \leq \tilde{M}_j \delta^{1/2} < \infty, \quad j = 2, \dots, 8,$$

where  $\tilde{M}_2, \dots, \tilde{M}_8$  are some finite constants, and that  $\mathcal{F}_\eta^j$  is a VC class with bounded bracketing entropy  $J_{[]}^*(1, \mathcal{F}_\eta^j) < \infty$ , for  $j = 2, \dots, 8$ . By Theorem 11.2 of Kosorok (2008), we obtain

$$\mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_N \mathcal{F}_\eta^j| \right] \leq c_j J_{[]}^*(1, \mathcal{F}_\eta^j) \|F_j\|_{P,2}, \quad j = 2, \dots, 8,$$

where  $c_2, \dots, c_8$  are some finite constants. That is, we have

$$\mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_N \mathcal{F}_\eta^8| \right] \leq \tilde{c}_8 \delta^{1/2},$$



and furthermore by Theorem 2.14.5 of van der Vaart and Wellner (1996), we obtain

$$\begin{aligned}
 \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} \|\mathbb{G}_n \mathcal{F}_\eta^j\|_2^2 \right] \right\}^{1/2} &\leq l_j \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_n \mathcal{F}_\eta^j| \right] + \|F_j\|_{P,2} \right\} \\
 &\leq l_j \{c_j J_{\square}^*(1, \mathcal{F}_\eta^j) + 1\} \|F_j\|_{P,2} \\
 &\leq \tilde{c}_j \delta^{1/2}, \quad j = 2, \dots, 7,
 \end{aligned}$$

where  $l_2, \dots, l_7$  and  $\tilde{c}_2, \dots, \tilde{c}_7$  are some finite constants.

By Equation (11), we have that

$$\begin{aligned}
 (I) &= \mathbb{E} \left[ N^{1/2} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{S}(t; \eta) - S_N^*(t; \eta) - \{\hat{S}(t; \eta^*) - S_N^*(t; \eta^*)\} \right| \right] \\
 &\leq \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} \left\{ |\mathbb{G}_n \mathcal{F}_\eta^2(\hat{\lambda} - \lambda^*)| + |\mathbb{G}_n \mathcal{F}_\eta^3(\hat{\theta} - \theta^*)| + |\mathbb{G}_n \mathcal{F}_\eta^4(\hat{\beta}_0 - \beta_0^*)| + |\mathbb{G}_n \mathcal{F}_\eta^5(\hat{\beta}_1 - \beta_1^*)| \right. \right. \\
 &\quad \left. \left. + |\mathbb{G}_n \mathcal{F}_\eta^6(\hat{\alpha}_0 - \alpha_0^*)| + |\mathbb{G}_n \mathcal{F}_\eta^7(\hat{\alpha}_1 - \alpha_1^*)| + |\mathbb{G}_n \mathcal{F}_\eta^8 \right\} + o_p(1) \right] \\
 &\leq N^{-1/2} \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_n \mathcal{F}_\eta^2 \cdot N^{1/2}(\hat{\lambda} - \lambda^*)| \right] + \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_n \mathcal{F}_\eta^3 \cdot N^{1/2}(\hat{\theta} - \theta^*)| \right] \right. \\
 &\quad \left. + \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_n \mathcal{F}_\eta^4 \cdot N^{1/2}(\hat{\beta}_0 - \beta_0^*)| \right] + \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_n \mathcal{F}_\eta^5 \cdot N^{1/2}(\hat{\beta}_1 - \beta_1^*)| \right] \right. \\
 &\quad \left. + \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_n \mathcal{F}_\eta^6 \cdot N^{1/2}(\hat{\alpha}_0 - \alpha_0^*)| \right] + \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_n \mathcal{F}_\eta^7 \cdot N^{1/2}(\hat{\alpha}_1 - \alpha_1^*)| \right] \right\} \\
 &\quad + \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_N \mathcal{F}_\eta^8| \right] + o_p(1),
 \end{aligned}$$

and then by the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
 (I) &\leq N^{-1/2} \left\{ \mathbb{E}[N \|\hat{\lambda} - \lambda^*\|_2^2] \right\}^{1/2} \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} \|\mathbb{G}_N \mathcal{F}_\eta^2\|_2^2 \right] \right\}^{1/2} \\
 &+ N^{-1/2} \left\{ \mathbb{E}[N \|\hat{\theta} - \theta^*\|_2^2] \right\}^{1/2} \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} \|\mathbb{G}_N \mathcal{F}_\eta^3\|_2^2 \right] \right\}^{1/2} \\
 &+ N^{-1/2} \left\{ \mathbb{E}[N \|\hat{\beta}_0 - \beta_0^*\|_2^2] \right\}^{1/2} \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} \|\mathbb{G}_N \mathcal{F}_\eta^4\|_2^2 \right] \right\}^{1/2} \\
 &+ N^{-1/2} \left\{ \mathbb{E}[N \|\hat{\beta}_1 - \beta_1^*\|_2^2] \right\}^{1/2} \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} \|\mathbb{G}_N \mathcal{F}_\eta^5\|_2^2 \right] \right\}^{1/2} \\
 &+ N^{-1/2} \left\{ \mathbb{E}[N \|\hat{\alpha}_0 - \alpha_0^*\|_2^2] \right\}^{1/2} \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} \|\mathbb{G}_N \mathcal{F}_\eta^6\|_2^2 \right] \right\}^{1/2} \\
 &+ N^{-1/2} \left\{ \mathbb{E}[N \|\hat{\alpha}_1 - \alpha_1^*\|_2^2] \right\}^{1/2} \left\{ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} \|\mathbb{G}_N \mathcal{F}_\eta^7\|_2^2 \right] \right\}^{1/2} \\
 &+ \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_N \mathcal{F}_\eta^8| \right].
 \end{aligned}$$

Let  $M_\lambda = \left\{ \mathbb{E}[N \|\hat{\lambda} - \lambda^*\|_2^2] \right\}^{1/2}$ , and  $M_\theta, M_{\beta_0}, M_{\beta_1}, M_{\alpha_0}, M_{\alpha_1}$  are defined accordingly. By Condition 1, we have that  $M_\lambda, M_\theta, M_{\beta_0}, M_{\beta_1}, M_{\alpha_0}, M_{\alpha_1} < \infty$ , and therefore

$$(I) \leq N^{-1/2} (M_\lambda \tilde{c}_2 + M_\theta \tilde{c}_3 + M_{\beta_0} \tilde{c}_4 + M_{\beta_1} \tilde{c}_5 + M_{\alpha_0} \tilde{c}_6 + M_{\alpha_1} \tilde{c}_7) \delta^{1/2} + \tilde{c}_8 \delta^{1/2}.$$

In summary, we obtain that, let  $N \rightarrow \infty$ , the centered process satisfies

$$\begin{aligned}
 &\mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{S}(t; \eta) - S(t; \eta) - \{ \hat{S}(t; \eta^*) - S(t; \eta^*) \} \right| \right] \\
 &\leq (I) + (II) \leq (\tilde{c}_1 + \tilde{c}_8) \delta^{1/2}.
 \end{aligned} \tag{12}$$

Let  $\phi_N(\delta) = \delta^{1/2}$  and  $\alpha = \frac{3}{2} < 2$ , thus we have  $\frac{\phi_N(\delta)}{\delta^\alpha} = \delta^{-1}$  is decreasing, and  $\alpha$  does not depend on  $N$ . That is, the second condition holds.

**Condition 3** By the facts that  $\hat{\eta} \xrightarrow{P} \eta^*$  as  $N \rightarrow \infty$ , and that  $\hat{S}(t; \hat{\eta}) \geq \sup_\eta \hat{S}(t; \eta)$ , we choose  $r_N = N^{1/3}$  such that  $r_N^2 \phi_N(r_N^{-1}) = N^{2/3} \phi_N(N^{-1/3}) = N^{1/2}$ . The third condition holds.

In the end, the three conditions are satisfied with  $r_N = N^{1/3}$ ; thus we conclude that  $N^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ , which completes the proof of (iii) of Theorem 6.

**PART 3.** We characterize the asymptotic distribution of  $\hat{S}(t; \hat{\eta})$ . Since we have

$$\sqrt{N} \{ \hat{S}(t; \hat{\eta}) - S(t; \eta^*) \} = \sqrt{N} \{ \hat{S}(t; \hat{\eta}) - \hat{S}(t; \eta^*) \} + \sqrt{N} \{ \hat{S}(t; \eta^*) - S(t; \eta^*) \},$$

we study the two terms in two steps.

**Step 3.1** To establish  $\sqrt{N}\{\hat{S}(t; \hat{\eta}) - \hat{S}(t; \eta^*)\} = o_p(1)$ , it suffices to show that  $\sqrt{N}\{S(t; \hat{\eta}) - S(t; \eta^*)\} = o_p(1)$  and  $\sqrt{N}(\hat{S}(t; \hat{\eta}) - \hat{S}(t; \eta^*) - \{S(t; \hat{\eta}) - S(t; \eta^*)\}) = o_p(1)$ .

First, as  $N^{1/3}\|\hat{\eta} - \eta^*\|_2 = O_p(1)$ , we take the second-order Taylor expansion

$$\begin{aligned} \sqrt{N}\{S(t; \hat{\eta}) - S(t; \eta^*)\} &= \sqrt{N} \left\{ S'(\eta^*)\|\hat{\eta} - \eta^*\|_2 + \frac{1}{2}S''(\eta^*)\|\hat{\eta} - \eta^*\|_2^2 + o_p(\|\hat{\eta} - \eta^*\|_2^2) \right\} \\ &= \sqrt{N} \left\{ \frac{1}{2}S''(\eta^*)\|\hat{\eta} - \eta^*\|_2^2 + o_p(\|\hat{\eta} - \eta^*\|_2^2) \right\} \\ &= \sqrt{N} \left\{ \frac{1}{2}S''(\eta^*)O_p(N^{-2/3}) + o_p(N^{-2/3}) \right\} = o_p(1). \end{aligned}$$

Next, we follow the result (12) obtained in **PART 2**. As  $N^{1/3}\|\hat{\eta} - \eta^*\|_2 = O_p(1)$ , there exists  $\tilde{\delta} = c_9N^{-1/3}$ , where  $c_9 < \infty$  is a finite constant, such that  $\|\hat{\eta} - \eta^*\|_2 \leq \tilde{\delta}$ . Therefore we have

$$\begin{aligned} &\sqrt{N}(\hat{S}(t; \hat{\eta}) - \hat{S}(t; \eta^*) - \{S(t; \hat{\eta}) - S(t; \eta^*)\}) \\ &\leq \mathbb{E} \left[ \sqrt{N} \sup_{\|\hat{\eta} - \eta^*\|_2 < \tilde{\delta}} \left| \hat{S}(t; \hat{\eta}) - S(t; \hat{\eta}) - \{\hat{S}(t; \eta^*) - S(t; \eta^*)\} \right| \right] \\ &\leq (\tilde{c}_1 + \tilde{c}_8)\tilde{\delta}^{1/2} = (\tilde{c}_1 + \tilde{c}_8)\sqrt{c_9}N^{-1/6} = o_p(1), \end{aligned}$$

which yields the result.

**Step 3.2** To derive the asymptotic distribution of  $\sqrt{n}\{\hat{S}(t; \eta^*) - S(t; \eta^*)\}$ , we follow the result (10) obtained in **PART 1** and have that

$$\sqrt{N} \left\{ \hat{S}(t; \eta^*) - S(t; \eta^*) \right\} \xrightarrow{D} \mathcal{N}(0, \sigma_{t,1}^2),$$

where  $\sigma_{t,1}^2 = \mathbb{E}[(\xi_{1,i}(t; \eta^*) + \xi_{2,i}(t; \eta^*))^2]$ . Therefore we obtain in the end

$$\begin{aligned} \sqrt{N}\{\hat{S}(t; \hat{\eta}) - S(t; \eta^*)\} &= \sqrt{N}\{\hat{S}(t; \hat{\eta}) - \hat{S}(t; \eta^*)\} + \sqrt{N}\{\hat{S}(t; \eta^*) - S(t; \eta^*)\} \\ &= o_p(1) + \sqrt{N}\{\hat{S}(t; \eta^*) - S(t; \eta^*)\} \\ &\xrightarrow{D} \mathcal{N}(0, \sigma_{t,1}^2), \end{aligned}$$

which completes the proof.

For Corollary 8 where we consider RMST, the proof can follow the same steps as before, and is thus omitted here.  $\blacksquare$

## Appendix E. Proof of Theorem 7 and Corollary 9

**Proof** Our proof has three main parts below.

**PART 1.** Recall that the cross-fitting technique, at a high level as exemplified in Lemma 13, uses sample splitting to avoid bias due to over-fitting. For simplicity, consider that the datasets  $\mathcal{O}_s$  and  $\mathcal{O}_t$  are randomly split into 2 folds with equal size respectively such that  $\mathcal{O}_s = \mathcal{O}_{s,1} \cup \mathcal{O}_{s,2}$ ,  $\mathcal{O}_t = \mathcal{O}_{t,1} \cup \mathcal{O}_{t,2}$ . The extension to  $K$ -folds as described in Algorithm

1 is straightforward. Here the subscript  $CF$  is omitted to simplify the notation. Define  $\mathcal{I}_1 = \mathcal{O}_{s,1} \cup \mathcal{O}_{t,1}$ ,  $\mathcal{I}_2 = \mathcal{O}_{s,2} \cup \mathcal{O}_{t,2}$ , and  $N_1 = |\mathcal{I}_1|$ ,  $N_2 = |\mathcal{I}_2|$ . The cross-fitted estimator for the value function under the ITR  $d_\eta$  is

$$\hat{V}(\eta) = \frac{N_1}{N} \hat{V}^{\mathcal{I}_1}(\eta) + \frac{N_2}{N} \hat{V}^{\mathcal{I}_2}(\eta),$$

where

$$\begin{aligned} \hat{V}^{\mathcal{I}_1}(\eta) = & \frac{1}{N_1} \sum_{\mathcal{I}_1} \left\{ I_{T,i} e(X_i) \hat{\mu}(d_\eta(X_i), X_i) + \frac{I_{S,i}}{\hat{\pi}_S(X_i)} \frac{I\{A_i = d_\eta(X_i)\}}{\hat{\pi}_d(X_i)} \right. \\ & \left. \times \left( \frac{\Delta_i y(U_i)}{\hat{S}_C(U_i | A_i, X_i)} - \hat{\mu}(A_i, X_i) + \int_0^\infty \frac{d\hat{M}_C(u | A_i, X_i)}{\hat{S}_C(u | A_i, X_i)} \hat{Q}(u, A_i, X_i) \right) \right\}, \end{aligned}$$

and the nuisance parameters are estimated from  $\mathcal{I}_2$ .  $\hat{V}^{\mathcal{I}_2}(\eta)$  is defined accordingly.

In this step, we show that

$$\hat{V}(\eta) - V_N(\eta) = o_p(N^{-1/2}),$$

and essentially it suffices to prove that

$$\hat{V}^{\mathcal{I}_1}(\eta) - V_N^{\mathcal{I}_1}(\eta) = o_p(N^{-1/2}),$$

where

$$\begin{aligned} V_N(\eta) = & \frac{1}{N} \sum_{i=1}^N \left\{ I_{T,i} e(X_i) \mu(d_\eta(X_i), X_i) + \frac{I_{S,i}}{\pi_S(X_i)} \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} \right. \\ & \left. \times \left( \frac{\Delta_i y(U_i)}{S_C(U_i | A_i, X_i)} - \mu(A_i, X_i) + \int_0^\infty \frac{dM_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} Q(u, A_i, X_i) \right) \right\}, \end{aligned}$$

and  $V_N^{\mathcal{I}_1}(\eta)$  is defined accordingly.

First, we have the following decomposition

$$\begin{aligned}
 & \hat{V}^{\mathcal{I}_1}(\eta) - V_N^{\mathcal{I}_1}(\eta) \\
 &= \frac{1}{N_1} \sum_{\mathcal{I}_1} \left\{ I_{T,i} e(X_i) (\hat{\mu}(d_\eta(X_i), X_i) - \mu(d_\eta(X_i), X_i)) \right. \\
 &+ I_{S,i} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} K(A_i, X_i) \\
 &+ \frac{I_{S,i} I\{A_i = d_\eta(X_i)\}}{\pi_S(X_i)} \left( \frac{1}{\pi_d(X_i)} - \frac{1}{\hat{\pi}_d(X_i)} \right) K(A_i, X_i) \\
 &+ \frac{I_{S,i}}{\pi_S(X_i)} \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} (\hat{K}(A_i, X_i) - K(A_i, X_i)) \\
 &+ I_{S,i} I\{A_i = d_\eta(X_i)\} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) \left( \frac{1}{\pi_d(X_i)} - \frac{1}{\hat{\pi}_d(X_i)} \right) K(A_i, X_i) \\
 &+ \frac{I_{S,i} I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) (\hat{K}(A_i, X_i) - K(A_i, X_i)) \\
 &+ \frac{I_{S,i} I\{A_i = d_\eta(X_i)\}}{\pi_S(X_i)} \left( \frac{1}{\pi_d(X_i)} - \frac{1}{\hat{\pi}_d(X_i)} \right) (\hat{K}(A_i, X_i) - K(A_i, X_i)) \\
 &\left. + I_{S,i} I\{A_i = d_\eta(X_i)\} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) \left( \frac{1}{\pi_d(X_i)} - \frac{1}{\hat{\pi}_d(X_i)} \right) (\hat{K}(A_i, X_i) - K(A_i, X_i)) \right\}, \tag{13}
 \end{aligned}$$

where

$$\begin{aligned}
 \hat{K}(A_i, X_i) &= \frac{\Delta_i y(U_i)}{\hat{S}_C(U_i | A_i, X_i)} - \hat{\mu}(A_i, X_i) + \int_0^\infty \frac{d\hat{M}_C(u | A_i, X_i)}{\hat{S}_C(u | A_i, X_i)} \hat{Q}(u, A_i, X_i), \\
 K(A_i, X_i) &= \frac{\Delta_i y(U_i)}{S_C(U_i | A_i, X_i)} - \mu(A_i, X_i) + \int_0^\infty \frac{dM_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} Q(u, A_i, X_i).
 \end{aligned}$$

In summary, the decomposition (13) consists of two types of terms: four mean-zero terms and four product terms. For the mean-zero terms, we utilize the method introduced in Section A.2; since

$$\mathbb{E}[I_{T,i} e(X_i) (\hat{\mu}(d_\eta(X_i), X_i) - \mu(d_\eta(X_i), X_i))] = 0,$$

by applying Lemma 13, we obtain

$$\frac{1}{N_1} \sum_{\mathcal{I}_1} I_{T,i} e(X_i) (\hat{\mu}(d_\eta(X_i), X_i) - \mu(d_\eta(X_i), X_i)) = o_p(N^{-1/2}).$$

Similarly we have

$$\mathbb{E} \left[ I_{S,i} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} K(A_i, X_i) \right] = 0,$$

so we obtain

$$\begin{aligned}
 & \mathbb{E} \left[ \left( \frac{1}{N_1} \sum_{\mathcal{I}_1} I_{S,i} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} K(A_i, X_i) \right)^2 \right] \\
 &= \mathbb{E} \left[ \mathbb{E} \left[ \left( \frac{1}{N_1} \sum_{\mathcal{I}_1} I_{S,i} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} K(A_i, X_i) \right)^2 \middle| \mathcal{I}_2 \right] \right] \\
 &= \mathbb{E} \left[ \text{var} \left[ \frac{1}{N_1} \sum_{\mathcal{I}_1} I_{S,i} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} K(A_i, X_i) \middle| \mathcal{I}_2 \right] \right] \\
 &= \frac{1}{N_1} \mathbb{E} \left[ \text{var} \left[ I_{S,i} \left( \frac{1}{\pi_S(X_i)} - \frac{1}{\hat{\pi}_S(X_i)} \right) \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} K(A_i, X_i) \middle| \mathcal{I}_2 \right] \right] \\
 &\leq \frac{O_p(1)}{N_1} = o_p\left(\frac{1}{N}\right).
 \end{aligned}$$

We also have

$$\mathbb{E} \left[ \frac{I_{S,i} I\{A_i = d_\eta(X_i)\}}{\pi_S(X_i)} \left( \frac{1}{\pi_d(X_i)} - \frac{1}{\hat{\pi}_d(X_i)} \right) K(A_i, X_i) \right] = 0,$$

$$\mathbb{E} \left[ \frac{I_{S,i}}{\pi_S(X_i)} \frac{I\{A_i = d_\eta(X_i)\}}{\pi_d(X_i)} (\hat{K}(A_i, X_i) - K(A_i, X_i)) \right] = 0,$$

and using the same technique, we conclude that these two mean-zero terms are  $o_p(N^{-1/2})$  as well.

The product terms can be handled simply by the Cauchy-Schwarz inequality and the rate of convergence conditions in Assumption 6. Additionally we have the decomposition

as follows

$$\begin{aligned}
 & \frac{1}{N_1} \sum_{\mathcal{I}_1} (\hat{K}(A_i, X_i) - K(A_i, X_i)) \\
 &= \frac{1}{N_1} \sum_{\mathcal{I}_1} \left\{ -(\hat{\mu}(A_i, X_i) - \mu(A_i, X_i)) + \frac{1 - \Delta_i}{S_C(U_i | A_i, X_i)} (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i)) \right. \\
 & \quad - \int_0^{U_i} \frac{\lambda_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i)) du \\
 & \quad + (1 - \Delta_i) \left( \frac{1}{\hat{S}_C(U_i | A_i, X_i)} - \frac{1}{S_C(U_i | A_i, X_i)} \right) Q(U_i | A_i, X_i) \\
 & \quad + \left( \frac{1}{\hat{S}_C(U_i | A_i, X_i)} - \frac{1}{S_C(U_i | A_i, X_i)} \right) \Delta_i y(U_i) \\
 & \quad - \int_0^{U_i} \left( \frac{\hat{\lambda}_C(u | A_i, X_i)}{\hat{S}_C(u | A_i, X_i)} - \frac{\lambda_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} \right) Q(U_i | A_i, X_i) du \\
 & \quad + (1 - \Delta_i) \left( \frac{1}{\hat{S}_C(U_i | A_i, X_i)} - \frac{1}{S_C(U_i | A_i, X_i)} \right) (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i)) \\
 & \quad \left. - \int_0^{U_i} \left( \frac{\hat{\lambda}_C(u | A_i, X_i)}{\hat{S}_C(u | A_i, X_i)} - \frac{\lambda_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} \right) (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i)) du, \right.
 \end{aligned}$$

and similarly we have three mean-zero terms which are  $o_p(N^{-1/2})$  by the same technique in Section A.2 and the facts that

$$\mathbb{E}[\hat{\mu}(A_i, X_i) - \mu(A_i, X_i)] = 0,$$

$$\begin{aligned}
 & \mathbb{E} \left[ \frac{1 - \Delta_i}{S_C(U_i | A_i, X_i)} (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i)) \right. \\
 & \quad \left. - \int_0^{U_i} \frac{\lambda_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} (\hat{Q}(u | A_i, X_i) - Q(u | A_i, X_i)) du \right] = 0,
 \end{aligned}$$

$$\begin{aligned}
 & \mathbb{E} \left[ (1 - \Delta_i) \left( \frac{1}{\hat{S}_C(U_i | A_i, X_i)} - \frac{1}{S_C(U_i | A_i, X_i)} \right) Q(U_i | A_i, X_i) \right. \\
 & \quad + \left( \frac{1}{\hat{S}_C(U_i | A_i, X_i)} - \frac{1}{S_C(U_i | A_i, X_i)} \right) \Delta_i y(U_i) \\
 & \quad \left. - \int_0^{U_i} \left( \frac{\hat{\lambda}_C(u | A_i, X_i)}{\hat{S}_C(u | A_i, X_i)} - \frac{\lambda_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} \right) Q(U_i | A_i, X_i) du \right] = 0,
 \end{aligned}$$

and we can bound the two product terms as well

$$\begin{aligned}
 & \frac{1}{N_1} \sum_{\mathcal{I}_1} \left[ (1 - \Delta_i) \left( \frac{1}{\hat{S}_C(U_i | A_i, X_i)} - \frac{1}{S_C(U_i | A_i, X_i)} \right) (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i)) \right. \\
 & \quad \left. - \int_0^{U_i} \left( \frac{\hat{\lambda}_C(u | A_i, X_i)}{\hat{S}_C(u | A_i, X_i)} - \frac{\lambda_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} \right) (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i)) du \right] \\
 & \leq \left[ \frac{1}{N_1} \sum_{\mathcal{I}_1} (1 - \Delta_i) \left( \frac{1}{\hat{S}_C(U_i | A_i, X_i)} - \frac{1}{S_C(U_i | A_i, X_i)} \right)^2 \right]^{1/2} \\
 & \quad \times \left[ \frac{1}{N_1} \sum_{\mathcal{I}_1} (1 - \Delta_i) (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i))^2 \right]^{1/2} \\
 & \quad - \int_0^{U_i} \left[ \frac{1}{N_1} \sum_{\mathcal{I}_1} \left( \frac{\hat{\lambda}_C(u | A_i, X_i)}{\hat{S}_C(u | A_i, X_i)} - \frac{\lambda_C(u | A_i, X_i)}{S_C(u | A_i, X_i)} \right)^2 \right]^{1/2} \\
 & \quad \times \left[ \frac{1}{N_1} \sum_{\mathcal{I}_1} (\hat{Q}(U_i | A_i, X_i) - Q(U_i | A_i, X_i))^2 \right]^{1/2} du \\
 & = o_p(N^{-1/2}),
 \end{aligned}$$

which proves that  $\frac{1}{N_1} \sum_{\mathcal{I}_1} (\hat{K}(A_i, X_i) - K(A_i, X_i)) = o_p(N^{-1/2})$ .

Therefore, we conclude that the four product terms in (13) are  $o_p(N^{-1/2})$  as well, which completes the proof of (i) in Theorem 7.

**PART 2:** We show that  $N^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ .

By Assumption 5 (i),  $V(\eta)$  is twice continuously differentiable at a neighborhood of  $\eta^*$ ; in **PART 1**, we show that  $\hat{V}(\eta) = V(\eta) + o_p(1), \forall \eta$ ; since  $\hat{\eta}$  maximizes  $\hat{V}(\eta)$ , we have that  $\hat{V}(\hat{\eta}) \geq \sup_{\eta} \hat{V}(\eta)$ , thus by the Argmax theorem, we have  $\hat{\eta} \xrightarrow{P} \eta^*$  as  $N \rightarrow \infty$ .

In order to establish the  $N^{-1/3}$  rate of convergence of  $\hat{\eta}$ , we apply Theorem 14.4 (Rate of convergence) of Kosorok (2008), and need to find the suitable rate that satisfies three conditions below.

**Condition 1** For every  $\eta$  in a neighborhood of  $\eta^*$  such that  $\|\eta - \eta^*\|_2 < \delta$ , by Assumption 5 (i), we apply the second-order Taylor expansion,

$$\begin{aligned}
 V(\eta) - V(\eta^*) &= V'(\eta^*) \|\eta - \eta^*\|_2 + \frac{1}{2} V''(\eta^*) \|\eta - \eta^*\|_2^2 + o(\|\eta - \eta^*\|_2^2) \\
 &= \frac{1}{2} V''(\eta^*) \|\eta - \eta^*\|_2^2 + o(\|\eta - \eta^*\|_2^2),
 \end{aligned}$$

and as  $V''(\eta^*) < 0$ , there exists  $c_{10} = -\frac{1}{2} V''(\eta^*) > 0$  such that  $V(\eta) - V(\eta^*) \leq -c_{10} \|\eta - \eta^*\|_2^2$ .



**Condition 2** For all  $N$  large enough and sufficiently small  $\delta$ , we consider the centered process  $\hat{V} - V$ , and have that

$$\begin{aligned}
 & \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{V}(\eta) - V(\eta) - \{\hat{V}(\eta^*) - V(\eta^*)\} \right| \right] \\
 &= \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{V}(\eta) - V_n(\eta) + V_n(\eta) - V(\eta) - \{\hat{V}(\eta^*) - V_n(\eta^*) + V_n(\eta^*) - V(\eta^*)\} \right| \right] \\
 &\leq \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{V}(\eta) - V_n(\eta) - \{\hat{V}(\eta^*) - V_n(\eta^*)\} \right| \right] \tag{I} \\
 &+ \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| V_n(\eta) - V(\eta) - \{V_n(\eta^*) - V(\eta^*)\} \right| \right] \tag{II}
 \end{aligned}$$

It follows from the result in **PART 1** that (I) =  $o_p(1)$ . To bound (II), we have

$$\begin{aligned}
 & V_n(\eta) - V_n(\eta^*) \\
 &= \frac{1}{N} \sum_{i=1}^N (d_\eta(X_i) - d_{\eta^*}(X_i)) \times \left( I_{T,i} e(X_i) (\mu(1, X_i) - \mu(0, X_i)) + \frac{(2A_i - 1)I_{S,i}}{\pi_{A_i}(X_i)\pi_S(X_i)} K(A_i, X_i) \right).
 \end{aligned}$$

Define a class of functions

$$\mathcal{F}_\eta^9 = \left\{ (d_\eta(x) - d_{\eta^*}(x)) \times \left( I_T e(x) (\mu(1, x) - \mu(0, x)) + \frac{(2a - 1)I_S}{\pi_a(x)\pi_S(x)} K(a, x) \right) : \|\eta - \eta^*\|_2 < \delta \right\},$$

and let  $M_9 = \sup \left| I_T e(x) (\mu(1, x) - \mu(0, x)) + \frac{(2a-1)I_S}{\pi_a(x)\pi_S(x)} K(a, x) \right|$ . By Assumption 1, 3 and Condition 1, we have that  $M_9 < \infty$ . Using the same technique as in Section D.2 **Condition 2.1**, we define the envelop of  $\mathcal{F}_\eta^9$  as  $F_9 = M_9 I\{-k_0\delta \leq (1, x^T)\eta^* \leq k_0\delta\}$ , and obtain that  $\|F_9\|_{P,2} \leq \tilde{M}_9\delta^{1/2} < \infty$ , where  $\tilde{M}_9$  is a finite constant, and that  $\mathcal{F}_\eta^9$  is a VC class with bounded entropy  $J_{[]}^*(1, \mathcal{F}_\eta^9) < \infty$ . By Theorem 11.2 of Kosorok (2008), we obtain

$$\mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_N \mathcal{F}_\eta^9| \right] \leq \tilde{c}_9 \delta^{1/2},$$

where  $\tilde{c}_9$  is a finite constant. Therefore, we obtain

$$\begin{aligned}
 (II) &= \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| V_N(\eta) - V(\eta) - \{V_N(\eta^*) - V(\eta^*)\} \right| \right] \\
 &= \mathbb{E} \left[ \sup_{\|\eta - \eta^*\|_2 < \delta} |\mathbb{G}_n \mathcal{F}_\eta^9| \right] \leq \tilde{c}_9 \delta^{1/2}.
 \end{aligned}$$

In summary, we obtain that the centered process satisfies

$$\begin{aligned}
 & \mathbb{E} \left[ \sqrt{N} \sup_{\|\eta - \eta^*\|_2 < \delta} \left| \hat{S}(t; \eta) - S(t; \eta) - \{\hat{S}(t; \eta^*) - S(t; \eta^*)\} \right| \right] \\
 &\leq (I) + (II) \leq \tilde{c}_9 \delta^{1/2}.
 \end{aligned} \tag{14}$$

Let  $\phi_N(\delta) = \delta^{1/2}$  and  $\alpha = \frac{3}{2} < 2$ , thus we have  $\frac{\phi_n(\delta)}{\delta^\alpha} = \delta^{-1}$  is decreasing, and  $\alpha$  does not depend on  $N$ . That is, the second condition holds.

**Condition 3** By the facts that  $\hat{\eta} \xrightarrow{P} \eta^*$  as  $N \rightarrow \infty$ , and that  $\hat{S}(t; \hat{\eta}) \geq \sup_{\eta} \hat{S}(t; \eta)$ , we choose  $r_N = N^{1/3}$  such that  $r_N^2 \phi_N(r_N^{-1}) = N^{2/3} \phi_N(N^{-1/3}) = N^{1/2}$ . The third condition holds.

In the end, the three conditions are satisfied with  $r_N = N^{1/3}$ ; thus we conclude that  $N^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ , which completes the proof of (ii) in Theorem 7.

**PART 3:** We characterize the asymptotic distribution of  $\hat{V}(\hat{\eta})$ . Since we have

$$\sqrt{N}\{\hat{V}(\hat{\eta}) - V(\eta^*)\} = \sqrt{N}\{\hat{V}(\hat{\eta}) - \hat{V}(\eta^*)\} + \sqrt{N}\{\hat{V}(\eta^*) - V(\eta^*)\},$$

we study the two terms in two steps.

**Step 3.1** To establish  $\sqrt{N}\{\hat{V}(\hat{\eta}) - \hat{V}(\eta^*)\} = o_p(1)$ , it suffices to show that  $\sqrt{N}\{V(\hat{\eta}) - V(\eta^*)\} = o_p(1)$  and  $\sqrt{N}(\hat{V}(\hat{\eta}) - \hat{V}(\eta^*) - \{V(\hat{\eta}) - V(\eta^*)\}) = o_p(1)$ .

First, as  $N^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ , we take the second-order Taylor expansion

$$\begin{aligned} \sqrt{N}\{V(\hat{\eta}) - V(\eta^*)\} &= \sqrt{N} \left\{ V'(\eta^*) \|\hat{\eta} - \eta^*\|_2 + \frac{1}{2} V''(\eta^*) \|\hat{\eta} - \eta^*\|_2^2 + o_p(\|\hat{\eta} - \eta^*\|_2^2) \right\} \\ &= \sqrt{N} \left\{ \frac{1}{2} V''(\eta^*) \|\hat{\eta} - \eta^*\|_2^2 + o_p(\|\hat{\eta} - \eta^*\|_2^2) \right\} \\ &= \sqrt{N} \left\{ \frac{1}{2} V''(\eta^*) O_p(N^{-2/3}) + o_p(N^{-2/3}) \right\} = o_p(1). \end{aligned}$$

Next, we follow the result (14) obtained in **PART 2**. As  $N^{1/3} \|\hat{\eta} - \eta^*\|_2 = O_p(1)$ , there exists  $\tilde{\delta}_2 = c_{11} N^{-1/3}$ , where  $c_{11} < \infty$  is a finite constant, such that  $\|\hat{\eta} - \eta^*\|_2 \leq \tilde{\delta}_2$ . Therefore we have

$$\begin{aligned} &\sqrt{N}(\hat{V}(\hat{\eta}) - \hat{V}(\eta^*) - \{V(\hat{\eta}) - V(\eta^*)\}) \\ &\leq \mathbb{E} \left[ \sqrt{N} \sup_{\|\hat{\eta} - \eta^*\|_2 < \tilde{\delta}_2} \left| \hat{V}(\hat{\eta}) - V(\hat{\eta}) - \{\hat{V}(\eta^*) - V(\eta^*)\} \right| \right] \\ &\leq \tilde{c}_9 \tilde{\delta}_2^{1/2} = \tilde{c}_9 \sqrt{c_{11}} N^{-1/6} = o_p(1), \end{aligned}$$

which yields the result.

**Step 3.2** To derive the asymptotic distribution of  $\sqrt{N}\{\hat{V}(\eta^*) - V(\eta^*)\}$ , we follow the result obtained in **PART 1** that  $\hat{V}(\eta^*) = V_N(\eta^*) + o_p(N^{-1/2})$ , and thus

$$\sqrt{N} \left\{ \hat{V}(\eta^*) - V(\eta^*) \right\} \xrightarrow{D} \mathcal{N}(0, \sigma_2^2),$$

where  $\sigma_2^2 = \mathbb{E}[\phi_{d_{\eta^*}}^2]$  is the semiparametric efficiency bound.

Therefore we obtain in the end

$$\begin{aligned} \sqrt{N}\{\hat{V}(\hat{\eta}) - v(\eta^*)\} &= \sqrt{N}\{\hat{V}(\hat{\eta}) - \hat{V}(\eta^*)\} + \sqrt{N}\{\hat{V}(\eta^*) - V(\eta^*)\} \\ &= o_p(1) + \sqrt{N}\{\hat{V}(\eta^*) - V(\eta^*)\} \\ &\xrightarrow{D} \mathcal{N}(0, \sigma_2^2), \end{aligned}$$

which completes the proof of Theorem 7 and Corollary 9. ■

## Appendix F. Analysis of divergent sample sizes

In this section, we analyze the asymptotic properties of the ACW estimator

$$\begin{aligned} \hat{V}(d) = & \frac{1}{m} \sum_{i=n+1}^{n+m} \hat{\mu}(A = d(X_i), X_i) \\ & + \frac{1}{n} \sum_{i=1}^n \frac{I\{A_i = d(X_i)\}}{\hat{\pi}_S(X_i)\hat{\pi}_d(X_i)} \left\{ \frac{\Delta_i y(U_i)}{\hat{S}_C(U_i | A_i, X_i)} - \hat{\mu}(A_i, X_i) + \int_0^\infty \frac{d\hat{M}_C(u | A_i, X_i)}{\hat{S}_C(u | A_i, X_i)} \hat{Q}(u, A_i, X_i) \right\}, \end{aligned}$$

when the sample sizes  $n, m$  diverge.

Note that we solve the following estimation equations to obtain the calibration weights:

$$\begin{aligned} \frac{1}{m} \sum_{i=n+1}^{n+m} \{g(X_i) - \mu_g\} &= 0, \\ \frac{1}{n} \sum_{i=1}^n \exp\{\lambda^T g(X_i)\} \{g(X_i) - \mu_g\} &= 0. \end{aligned}$$

When  $m$  diverges faster than  $n$ , we characterize  $\sqrt{n}\{\hat{V}(d) - V(d)\}$ . First we have that  $\mu_g = E[g(X)]$ , and  $\hat{\lambda}$  is the solution to the estimation equation  $n^{-1} \sum_{i=1}^n \exp\{\lambda^T g(X_i)\} \{g(X_i) - \mu_g\} = 0$ . Thus we obtain that

$$\begin{aligned} \sqrt{n}\{\hat{V}(d) - V(d)\} &= \frac{1}{m} \sum_{i=n+1}^{n+m} \sqrt{n}\{\hat{\mu}(A = d(X_i), X_i) - V(d)\} \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(O_i)q_i, \end{aligned}$$

where  $\phi(O_i)$  is the influence function. Finally, we conclude that  $\sqrt{n}\{\hat{V}(d) - V(d)\} \rightarrow \mathcal{N}(0, \sigma_1^2)$ , where  $\sigma_1^2 = E[\phi^2(O)q]$ .

When  $n$  diverges faster than  $m$ , we characterize  $\sqrt{m}\{\hat{V}(d) - V(d)\}$ . First we have that  $\hat{\lambda}$  is the solution to the estimation equation

$$\frac{1}{m} \sum_{i=n+1}^{n+m} \{g(X_i)M_{S,1}(\lambda) - M_{S,2}(\lambda)\} = 0,$$

where  $M_{S,1}(\lambda) = E[I_S \exp\{\lambda^T g(X)\}]$ , and  $M_{S,2}(\lambda) = E[I_S \exp\{\lambda^T g(X)\}g(X)]$ .

Thus we obtain that

$$\begin{aligned} \sqrt{m}\{\hat{V}(d) - V(d)\} &= \frac{1}{\sqrt{m}} \sum_{i=n+1}^{n+m} \{\hat{\mu}(A = d(X_i), X_i) - V(d)\} \\ &+ \frac{1}{n} \sum_{i=1}^n \sqrt{m}\phi(O_i)q_i, \end{aligned}$$

Finally, we conclude that  $\sqrt{m}\{\hat{V}(d) - V(d)\} \rightarrow \mathcal{N}(0, \sigma_2^2)$ , where  $\sigma_2^2 = E[(\mu(d(X), X) - V(d))^2]$ .

**Appendix G. Proof of Theorem 10 and Theorem 11**

**Proof** When the source and target populations have the same distributions, both  $\hat{V}_{DR}(\eta)$  and  $\hat{V}_{CF}(\eta)$  converge to  $V(\eta)$ . The asymptotic variance of  $\hat{V}_{DR}(\eta)$  is

$$\begin{aligned}\sigma_{DR}^2 &= \mathbb{E} \left[ \frac{I_S}{\mathbb{P}(I_S = 1)} \left( \mu(d(X), X) + \frac{I\{A = d(X)\}}{\pi_d(X)} K(A, X) - V(\eta) \right)^2 \right] \\ &= \mathbb{E} \left[ \frac{I_S}{\mathbb{P}(I_S = 1)} \left( \mu^2(d(X), X) + \frac{I\{A = d(X)\}}{\pi_d^2(X)} K^2(A, X) - V^2(\eta) \right. \right. \\ &\quad \left. \left. + \frac{2I\{A = d(X)\}}{\pi_d(X)} K(A, X)\mu(d(X), X) - 2\mu(d(X), X)V(\eta) \right. \right. \\ &\quad \left. \left. - \frac{2I\{A = d(X)\}}{\pi_d(X)} K(A, X)V(\eta) \right) \right],\end{aligned}$$

while the asymptotic variance of  $\hat{V}_{CF}(\eta)$  is

$$\begin{aligned}\sigma_{CF}^2 &= \mathbb{E} \left[ \left( I_T e(X)\mu(d(X), X) + \frac{I_S I\{A = d(X)\}}{\pi_S(X)\pi_d(X)} K(A, X) - V(\eta) \right)^2 \right] \\ &= \mathbb{E} \left[ \left( I_T e^2(X)\mu^2(d(X), X) + \frac{I_S I\{A = d(X)\}}{\pi_S^2(X)\pi_d^2(X)} K^2(A, X) - V^2(\eta) \right. \right. \\ &\quad \left. \left. - 2I_T e^2(X)\mu(d(X), X)V(\eta) - \frac{2I_S I\{A = d(X)\}}{\pi_S(X)\pi_d(X)} K(A, X)V(\eta) \right) \right],\end{aligned}$$

where

$$K(A, X) = \frac{\Delta y(U)}{S_C(U|A, X)} - \mu(A, X) + \int_0^\infty \frac{dM_C(u|A, X)}{S_C(u|A, X)} Q(u, A, X).$$

Since we have that

$$\mathbb{E} \left[ \frac{I_S}{\mathbb{P}(I_S = 1)} \frac{2I\{A = d(X)\}}{\pi_d(X)} K(A, X)\mu(d(X), X) \right] = 0,$$

and for

$$B \in \left\{ \mu^2(d(X), X), \frac{I\{A = d(X)\}}{\pi_d^2(X)} K^2(A, X), \mu(d(X), X)V(\eta), \frac{I\{A = d(X)\}}{\pi_d^2(X)} K(A, X)V(\eta) \right\},$$

we also have that

$$\mathbb{E} \left[ \frac{I_S}{\mathbb{P}(I_S = 1)} B \right] = \mathbb{E}[I_T e(X)B] = \mathbb{E} \left[ \frac{I_S}{\pi_S(X)} B \right],$$

we conclude that  $\sigma_{DR}^2 = \sigma_{CF}^2$ .

By the law of iterated expectations, the value function  $V_d = \mathbb{E}[y(T(d))] = \mathbb{E}_X[\mathbb{E}[y(T(d)) | X]]$ . When there is no restriction on the class of ITRs, the true optimal ITR is

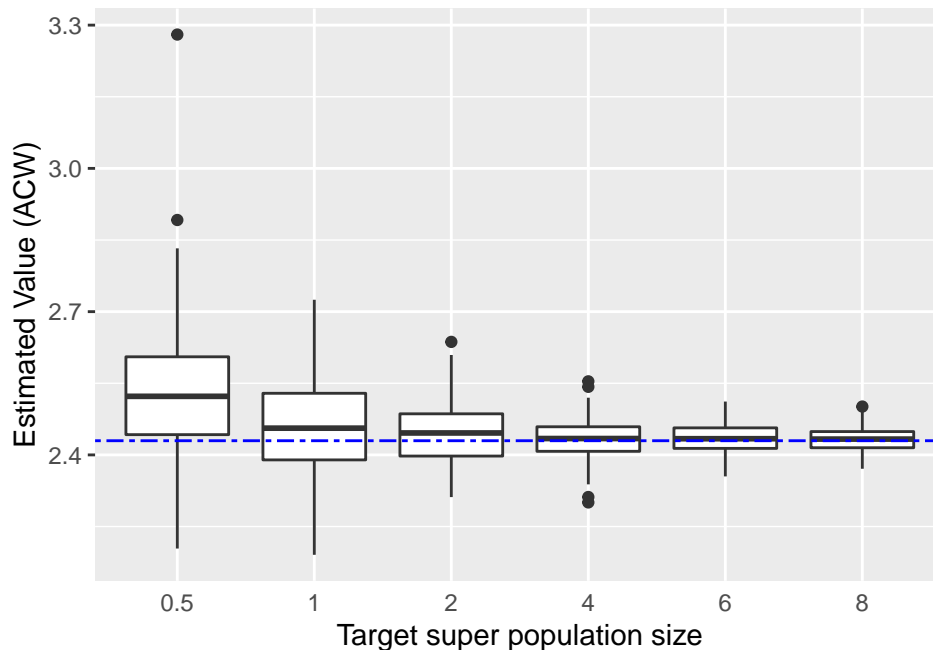
$$\begin{aligned}d^{**}(X) &= \arg \max_d V_d = \arg \max_d \mathbb{E}_X[\mathbb{E}[y(T(d)) | X]] \\ &= I\{\mathbb{E}[y(T(1)) | X] > \mathbb{E}[y(T(0)) | X]\}.\end{aligned}$$

That is, the optimal ITR does not depend on the covariate distributions, but only the bilp function which is the same in both the source and target populations by Assumption 2. Thus both the maximizers of  $\hat{V}_{DR}(\eta)$  and  $\hat{V}_{CF}(\eta)$  converge to the true population parameter  $\eta^{**}$ . However,  $\hat{V}_{DR}(\eta)$  is biased since the expectation  $\mathbb{E}_X$  is taken with respect to the source population. ■

### Appendix H. Additional simulations

We first investigate the performance of the cross-fitted ACW estimator with different sample sizes  $(N, m) = (5 \times 10^4, 2000), (1 \times 10^5, 4000), (2 \times 10^5, 8000), (4 \times 10^5, 16000), (6 \times 10^5, 24000), (8 \times 10^5, 32000)$ . Figure 3 and Table 3 report the results from 200 Monte Carlo replications. The variance is computed using the EIF.

Figure 3: Boxplot of estimated value by ACW estimator with different sample sizes.



### Appendix I. Details of real data analysis

There are around 0.5% and 1.6% missing values in the RCT and OS data, respectively. We use the `mice` function in the R package `mice` (van Buuren and Groothuis-Oudshoorn, 2011) to impute the missing values.

Motivated by the clinical practice and existing work in the medical literature, we consider ITRs that depend on the following five variables:

Table 3: Numeric results of the ACW estimator. Bias is the empirical bias of point estimates; SD is the empirical standard deviation of point estimates; SE is the average of standard error estimates; CP is the empirical coverage probability of the 95% Wald confidence intervals.

$n; m(\times 10^3)$	$\sim 780; 2$	$\sim 1560; 4$	$\sim 3120; 8$	$\sim 6240; 16$	$\sim 9360; 24$	$\sim 12480; 32$
Bias	0.1041	0.0253	0.0134	0.0046	0.0031	0.0030
SD	0.1394	0.0985	0.0635	0.0419	0.0317	0.0267
SE	0.1611	0.0942	0.0627	0.0417	0.0330	0.0284
CP(%)	97.5	93.5	96.0	94.5	97.5	97.0

- AGE, SEX and Sequential Organ Failure Assessment (SOFA) score: these three baseline variables are well related to mortality in ICUs, so we consider them as important risk factors.
- Acute Kidney Injury Network (AKIN) score: Jaber et al. (2018) observed that the infusion of sodium bicarbonate improved survival outcomes and mortality rate in critically ill patients with severe metabolic acidemia and acute kidney injury. In the observational data, the AKIN score was not recorded, so we computed the score using serum creatinine measurement (Závada et al., 2010).
- SEPSIS: we consider the presence of sepsis as a risk factor because it is the main condition associated with severe acidemia at the arrival in ICU. The effect of sodium bicarbonate infusion on patients with acidemia and acute kidney injury was also observed in septic patients (Zhang et al., 2018b).

## References

- Per Kragh Andersen and Richard D. Gill. Cox’s Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, 10(4):1100–1120, 1982.
- Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.
- Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019.
- Jean-Yves Audibert and Alexandre B. Tsybakov. Fast learning rates for plug-in classifiers. *The Annals of statistics*, 35(2):608–633, 2007.
- Xiaofei Bai, Anastasios A. Tsiatis, Wenbin Lu, and Rui Song. Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. *Lifetime data analysis*, 23(4):585–604, 2017.
- Alexandre Belloni, Victor Chernozhukov, Ivan Fernández-Val, and Christian Hansen. Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298, 2017.

- Norman E. Breslow. Discussion on professor cox’s paper. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):216–217, 1972.
- Shuai Chen, Lu Tian, Tianxi Cai, and Menggang Yu. A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics*, 73(4):1199–1209, 2017.
- Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.
- Victor Chernozhukov, Mert Demirer, Greg Lewis, and Vasilis Syrgkanis. Semi-parametric efficient policy learning with continuous actions. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- Victor Chernozhukov, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K. Newey, and James M. Robins. Locally robust semiparametric estimation. *Econometrica*, 90(4):1501–1535, 2022a.
- Victor Chernozhukov, Whitney K. Newey, and Rahul Singh. Automatic debiased machine learning of causal and structural effects. *Econometrica*, 90(3):967–1027, 2022b.
- Hunyoung Cho, Shannon T. Holloway, David J. Couper, and Michael R. Kosorok. Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring. *Biometrika*, 110(2):395–410, 2023.
- Jianing Chu, Wenbin Lu, and Shu Yang. Targeted optimal treatment regime learning using summary statistics. *Biometrika*, 110(4):913–931, 2023.
- Kyle Colangelo and Ying-Ying Lee. Double debiased machine learning nonparametric inference with continuous treatments, 2023. URL <https://arxiv.org/abs/2004.03036>.
- Stephen R. Cole and Elizabeth A. Stuart. Generalizing evidence from randomized clinical trials to target populations: the actg 320 trial. *American Journal of Epidemiology*, 172(1):107–115, 2010.
- Bénédicte Colnet, Imke Mayer, Guanhua Chen, Awa Dieng, Ruohong Li, Gaël Varoquaux, Jean-Philippe Vert, Julie Josse, and Shu Yang. Causal inference methods for combining randomized trials and observational studies: A review. *Statistical Science*, 39(1):165–191, 2024.
- Noel Cressie and Timothy R.C. Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):440–464, 1984.
- Yifan Cui and Eric Tchetgen Tchetgen. A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 116(533):162–173, 2021.
- Yifan Cui, Ruoqing Zhu, and Michael Kosorok. Tree based weighted learning for estimating individualized treatment rules with censored data. *Electronic Journal of Statistics*, 11(2):3927–3953, 2017.

- Yifan Cui, Michael R. Kosorok, Erik Sverdrup, Stefan Wager, and Ruoqing Zhu. Estimating heterogeneous treatment effects with right-censored data via causal survival forests. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(2):179–211, 2023.
- Issa J. Dahabreh and Miguel A. Hernán. Extending inferences from a randomized trial to a target population. *European Journal of Epidemiology*, 34(8):719–722, 2019.
- Issa J. Dahabreh, Sarah E. Robertson, Eric J. Tchetgen, Elizabeth A. Stuart, and Miguel A. Hernán. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*, 75(2):685–694, 2019.
- Issa J Dahabreh, Sebastien J-P A Haneuse, James M Robins, Sarah E Robertson, Ashley L Buchanan, Elizabeth A Stuart, and Miguel A Hernán. Study designs for extending causal inferences from a randomized trial to a target population. *American Journal of Epidemiology*, 190(8):1632–1642, 2021.
- Irina Degtiar and Sherri Rose. A review of generalizability and transportability. *Annual Review of Statistics and Its Application*, 10:501–524, 2023.
- Ilker Demirel, Ahmed Alaa, Anthony Philippakis, and David Sontag. Prediction-powered generalization of causal inferences. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 10385–10408, 2024a.
- Ilker Demirel, Edward De Brouwer, Zeshan M Hussain, Michael Oberst, Anthony A Philippakis, and David Sontag. Benchmarking observational studies with experimental data under right-censoring. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238, pages 4285–4293. PMLR, 2024b.
- Iván Díaz, Oleksandr Savenkov, and Karla Ballman. Targeted learning ensembles for optimal individualized treatment rules with time-to-event outcomes. *Biometrika*, 105(3):723–738, 2018.
- Ethan X. Fang, Zhaoran Wang, and Lan Wang. Fairness-oriented learning for optimal individualized treatment rules. *Journal of the American Statistical Association*, 118(543):1733–1746, 2023.
- Amir-massoud Farahmand. Action-gap phenomenon in reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 24, 2011.
- Max H. Farrell, Tengyuan Liang, and Sanjog Misra. Deep learning for individual heterogeneity: An automatic inference framework, 2021. URL <https://arxiv.org/abs/2010.14694>.
- Yair Goldberg and Michael R. Kosorok. Q-learning with censored data. *The Annals of Statistics*, 40(1):529–560, 2012.
- Wenshuo Guo, Serena Lutong Wang, Peng Ding, Yixin Wang, and Michael Jordan. Multi-source causal inference using control variates under outcome selection bias. *Transactions on Machine Learning Research*, 2022.



- Rebecca Hager, Anastasios A. Tsiatis, and Marie Davidian. Optimal two-stage dynamic treatment regimes from a classification perspective with censored survival data. *Biometrics*, 74(4):1180–1192, 2018.
- Jens Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46, 2012.
- Tobias Hatt, Jeroen Berrevoets, Alicia Curth, Stefan Feuerriegel, and Mihaela van der Schaar. Combining observational and randomized data for estimating heterogeneous treatment effects, 2022. URL <https://arxiv.org/abs/2202.12891>.
- Qijia He, Fei Gao, Oliver Dukes, Sinead Delany-Moretlwe, and Bo Zhang. Generalizing the intention-to-treat effect of an active control from historical placebo-controlled trials: A case study of the efficacy of daily oral tdf/ftc in the hptn 084 study. *Journal of the American Statistical Association*, 119(548):2478–2492, 2024.
- Yichun Hu, Nathan Kallus, and Masatoshi Uehara. Fast rates for the regret of offline reinforcement learning. *Mathematics of Operations Research*, 50(1):633–655, 2025.
- Zeshan Hussain, Michael Oberst, Ming-Chieh Shih, and David Sontag. Falsification before extrapolation in causal effect estimation. In *Advances in Neural Information Processing Systems*, 2022.
- Zeshan Hussain, Ming-Chieh Shih, Michael Oberst, Ilker Demirel, and David Sontag. Falsification of internal and external validity in observational studies via conditional moment restrictions. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206, pages 5869–5898. PMLR, 2023.
- Samir Jaber, Catherine Paugam, Emmanuel Futier, Jean-Yves Lefrant, Sigismond Lasocki, Thomas Lescot, Julien Pottecher, Alexandre Demoule, Martine Ferrandiere, Karim Asehnoune, et al. Sodium bicarbonate therapy for patients with severe metabolic acidaemia in the intensive care unit (bicar-icu): a multicentre, open-label, randomised controlled, phase 3 trial. *The Lancet*, 392(10141):31–40, 2018.
- Runchao Jiang, Wenbin Lu, Rui Song, and Marie Davidian. On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1165–1185, 2017.
- Ying Jin, Zhimei Ren, Zhuoran Yang, and Zhaoran Wang. Policy learning “without” overlap: Pessimism and generalized empirical bernstein’s inequality. *arXiv preprint arXiv:2212.09900*, 2022.
- Boris Jung, Thomas Rimmel, Charlotte Le Goff, Gérald Chanques, Philippe Corne, Olivier Jonquet, Laurent Muller, Jean-Yves Lefrant, Christophe Guervilly, Laurent Papazian, et al. Severe metabolic or mixed acidemia on intensive care unit admission: incidence, prognosis and administration of buffer therapy. a prospective, multiple-center study. *Critical Care*, 15(5):1–9, 2011.

- Nathan Kallus. More efficient policy learning via optimal retargeting. *Journal of the American Statistical Association*, 116(534):646–658, 2021.
- Masahiro Kato, Akihiro Oga, Wataru Komatsubara, and Ryo Inokuchi. Active adaptive experimental design for treatment effect estimation with covariate choice. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.
- Edward H. Kennedy. Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*, 2022.
- Edward H. Kennedy, Sivaraman Balakrishnan, and Max G’Sell. Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, 48(4):2008–2030, 2020.
- Michael R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, 2008.
- Michael R. Kosorok and Eric B. Laber. Precision medicine. *Annual Review of Statistics and Its Application*, 6:263–286, 2019.
- Sören R. Künnel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019.
- Eric B. Laber and Ying-Qi Zhao. Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514, 2015.
- Eric B. Laber, Kristin A. Linn, and Leonard A. Stefanski. Interactive model building for q-learning. *Biometrika*, 101(4):831–847, 2014.
- Dasom Lee, Shu Yang, Lin Dong, Xiaofei Wang, Donglin Zeng, and Jianwen Cai. Improving trial generalizability using observational studies. *Biometrics*, 79(2):1213–1225, 2021.
- Dasom Lee, Shu Yang, and Xiaofei Wang. Doubly robust estimators for generalizing treatment effects on survival outcomes from randomized controlled trials to a target population. *Journal of Causal Inference*, 10(1):415–440, 2022.
- Alex Luedtke and Antoine Chambaz. Performance guarantees for policy learning. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):2162–2188, 2020.
- Alexander R. Luedtke and Mark J. van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of Statistics*, 44(2):713–742, 2016a.
- Alexander R. Luedtke and Mark J. van der Laan. Super-learning of an optimal dynamic treatment rule. *The International Journal of Biostatistics*, 12(1):305–332, 2016b.
- Charles F. Manski. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246, 2004.

- Jacob M. Maronge, Jared D. Huling, and Guanhua Chen. A reluctant additive model framework for interpretable nonlinear individualized treatment rules. *The Annals of Applied Statistics*, 17(4):3384–3402, 2023.
- Walter R. Mebane Jr. and Jasjeet S. Sekhon. Genetic optimization using derivatives: The rgenoud package for r. *Journal of Statistical Software*, 42(11):1–26, 2011.
- Melanie Mitchell. *An Introduction to Genetic Algorithms*. The MIT Press, 1998.
- Weibin Mo, Zhengling Qi, and Yufeng Liu. Learning optimal distributionally robust individualized treatment rules. *Journal of the American Statistical Association*, 116(534):659–674, 2021.
- Susan A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.
- Whitney Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):1349–1382, 1994.
- Whitney K. Newey. Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2):99–135, 1990.
- Zhengling Qi, Rui Miao, and Xiaoke Zhang. Proximal learning for individualized treatment regimes under unmeasured confounding. *Journal of the American Statistical Association*, 119(546):915–928, 2023.
- Min Qian and Susan A. Murphy. Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180–1210, 2011.
- Sonja A. Rasmussen, Muin J. Khoury, and Carlos del Rio. Precision public health as a key tool in the covid-19 response. *JAMA*, 324(10):933–934, 09 2020.
- James Robins and Andrea Rotnitzky. Discussion of “Dynamic treatment regimes: Technical challenges and applications”. *Electronic Journal of Statistics*, 8(1):1273–1289, 2014.
- James M. Robins. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, pages 189–326. Springer New York, 2004.
- Daniel B. Rubin and Mark J. van der Laan. Statistical issues and limitations in personalized medicine research with clinical trials. *The International Journal of Biostatistics*, 8(1):18, 2012.
- Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.
- Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1:206–215, 2019.

- Kara E. Rudolph and Mark J. van der Laan. Robust estimation of encouragement design intervention effects transported across sites. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(5):1509–1525, 2017.
- Roshni Sahoo, Lihua Lei, and Stefan Wager. Learning from a biased sample. *arXiv preprint arXiv:2209.01754*, 2022.
- Anton Schick. On Asymptotically Efficient Estimation in Semiparametric Models. *The Annals of Statistics*, 14(3):1139–1151, 1986.
- Phillip J. Schulte, Anastasios A. Tsiatis, Eric B. Laber, and Marie Davidian. **Q**- and **A**-Learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Statistical Science*, 29(4):640–661, 2014.
- Vira Semenova and Victor Chernozhukov. Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24(2):264–289, 2020.
- Chengchun Shi, Ailin Fan, Rui Song, and Wenbin Lu. High-dimensional **A**-learning for optimal dynamic treatment regimes. *The Annals of Statistics*, 46(3):925 – 957, 2018.
- Jerzy Splawa-Neyman, D. M. Dabrowska, and T. P. Speed. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472, 1990. doi: 10.1214/ss/1177012031.
- Elizabeth A. Stuart, Stephen R. Cole, Catherine P. Bradshaw, and Philip J. Leaf. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2):369–386, 2011.
- Masashi Sugiyama and Motoaki Kawanabe. *Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation*. MIT Press, 2012.
- Anna L. Trella, Kelly W. Zhang, Inbal Nahum-Shani, Vivek Shetty, Finale Doshi-Velez, and Susan A. Murphy. Designing reinforcement learning algorithms for digital interventions: Pre-implementation guidelines. *Algorithms*, 15(8), 2022.
- Anastasios A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, New York, 2006.
- Anastasios A. Tsiatis, Marie Davidian, Shannon T. Holloway, and Eric B. Laber. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. Chapman & Hall, 2019.
- Alexander B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135–166, 2004.
- Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011.
- Mark J. van der Laan and James M. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York, 2003.

- Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York, 1996.
- Michael P. Wallace and Erica E.M. Moodie. Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71(3):636–644, 2015.
- Lili Wu and Shu Yang. Transfer learning of individualized treatment rules from experimental to real-world data. *Journal of Computation and Graphical Statistics*, 32(3):1036–1045, 2022.
- Yizhe Xu, Nikolaos Ignatiadis, Erik Sverdrup, Scott Fleming, Stefan Wager, and Nigam Shah. *Treatment Heterogeneity with Survival Outcomes*, chapter 21. Handbook of Matching and Weighting Adjustments for Causal Inference. Chapman and Hall/CRC, Boca Raton, FL, 2023.
- Shu Yang and Peng Ding. Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika*, 105(2):1–7, 2018.
- Shu Yang, Chenyin Gao, Donglin Zeng, and Xiaofei Wang. Elastic integrative analysis of randomised trial and real-world data for treatment heterogeneity estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):575–596, 2023a.
- Shu Yang, Yilong Zhang, Guanghan Frank Liu, and Qian Guan. Smim: A unified framework of survival sensitivity analysis using multiple imputation and martingale. *Biometrics*, 79(1):230–240, 2023b.
- Jakub Závada, Eric Hoste, Rodrigo Cartin-Ceba, Paolo Calzavacca, Ognjen Gajic, Gilles Clermont, Rinaldo Bellomo, John A Kellum, and AKI6 investigators. A comparison of three methods to estimate baseline creatinine for rifle classification. *Nephrology Dialysis Transplantation*, 25(12):3911–3918, 2010.
- Baqun Zhang, Anastasios A. Tsiatis, Marie Davidian, Min Zhang, and Eric Laber. Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114, 2012a.
- Baqun Zhang, Anastasios A. Tsiatis, Eric B. Laber, and Marie Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012b.
- Yichi Zhang, Eric B. Laber, Anastasios Tsiatis, and Marie Davidian. Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71(4):895–904, 2015.
- Yichi Zhang, Eric B. Laber, Marie Davidian, and Anastasios A. Tsiatis. Interpretable dynamic treatment regimes. *Journal of the American Statistical Association*, 113(524):1541–1549, 2018a.
- Zhongheng Zhang, Carlie Zhu, Lei Mo, and Yucai Hong. Effectiveness of sodium bicarbonate infusion on mortality in septic patients with metabolic acidosis. *Intensive Care Medicine*, 44(11):1888–1895, 2018b.

Ying-Qi Zhao, Donglin Zeng, Catherine M. Tangen, and Michael L. Leblanc. Robustifying trial-derived optimal treatment rules for a target population. *Electronic Journal of Statistics*, 13(1):1717–1743, 2019.

Yingqi Zhao, Donglin Zeng, A. John Rush, and Michael R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012.

Yu Zhou, Lan Wang, Rui Song, and Tuoyi Zhao. Transformation-invariant learning of optimal individualized decision rules with time-to-event outcomes. *Journal of the American Statistical Association*, 118(544):2632–2644, 2023.