Doubly Protected Estimation for Survival Outcomes Utilizing External Controls for Randomized Clinical Trials

Chenyin Gao [1 2] **Shu Yang** [1] **Mingyang Shan** [3] **Wenyu Wendy Ye** [3] **Ilya Lipkovich** [3] **Douglas Faries** [3]

Abstract

Censored survival data are common in clinical trials, but small control groups can pose challenges, particularly in rare diseases or where balanced randomization is impractical. Recent approaches leverage external controls from historical studies or real-world data to strengthen treatment evaluation for survival outcomes. However, using external controls directly may introduce biases due to data heterogeneity. We propose a doubly protected estimator for the treatment-specific restricted mean survival time difference that is more efficient than trial-only estimators and mitigates biases from external data. Our method adjusts for covariate shifts via doubly robust estimation and addresses outcome drift using the DR-Learner for selective borrowing. The approach can incorporate machine learning to approximate survival curves and detect outcome drifts without strict parametric assumptions, borrowing only comparable external controls. Extensive simulation studies and a real-data application evaluating the efficacy of Galcanezumab in mitigating migraine headaches have been conducted to illustrate the effectiveness of our proposed framework.

1. Introduction

Understanding the risk of disease or death and how these risks evolve over time is critical for assisting clinicians in treatment assignment and disease diagnosis. In clinical trials or biomedical studies, evaluating the effectiveness of drugs often suffers from limited sample sizes due to low disease prevalence or restrictive inclusion/exclusion criteria. This issue is exacerbated in survival analysis, as time-to-event or survival endpoints may not always be observed. As a complement to clinical trials, external controls offer a promising avenue to improve statistical inference when recruiting more patients is challenging. However, external controls can differ from clinical trials in many aspects due to differences in the underlying data acquisition and generation mechanisms. Concerns regarding the plausibility of these assumptions have limited their broader deployment. Guidance documents from regulatory agencies, including the recent Food and Drug Administration (FDA) draft guidance on Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products, note several potential issues with the use of external controls, including selection bias, lack of concurrency, differences in the definitions of covariates, treatments, or outcomes, and unmeasured confounding (FDA, 2001; 2019; 2023). Each of these concerns can result in biased treatment effect estimates if external controls are integrated with the trial without further scrutiny.

2. Related Work

Data Integration with Non-survival Outcomes To reliably leverage external data, it is crucial to address these potential issues with the use of external data. One primary concern is the distributional heterogeneity of the baseline disease characteristics between the two studies, leading to the issue of covariate shifts. Likelihood-based frameworks have been explored to mitigate covariate shifts arising from external datasets (Chatterjee et al., 2016; Huang et al., 2016). Other propensity score weighting and matching methods have been proposed to construct new external data that have similar covariate distributions as the trial data (Li et al., 2018). However, these frameworks often rely on the invariance assumption, which posits that the conditional outcome distributions are identical for the trial and external controls. This assumption can be problematic, as the distributions of outcomes may also vary across studies given the baseline covariates, leading to the problem of outcome drift.

In recent years, various adaptive learning methods have been proposed to address the study-specific generating mechanisms for outcomes, aiming to ensure robust estimation even when external controls differ significantly from the

[1]North Carolina State University, Department of Statistics, Raleigh, NC, USA [2]Harvard University, Department of Biostatistics, Boston, MA, USA [3]Eli Lilly & Company, Indianapolis, IN, USA. Correspondence to: Shu Yang <syang24@ncsu.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

trial. The considered analytic frameworks include adaptive information borrowing from diverse populations in linear regression (Li et al., 2022; Yang et al., 2023; Gao & Yang, 2023), generalized linear models (Tian & Feng, 2023; Li et al., 2023a), and nonparametric classification (Cai & Wei, 2021).

Data Integration with Survival Outcomes However, the adaptive learning in survival analysis remains limited due to challenges in incorporating external time-to-event outcomes. Unlike regression settings, where models dynamically borrow information for covariate effects, survival analyses need to address outcome drift in hazards as well. For instance, Liu et al. (2014) and Huang et al. (2016) proposed accommodating outcome heterogeneity by applying a constant factor to the cumulative hazard function. Additionally, Chen et al. (2022) and Wang et al. (2020) developed a propensity score-integrated Bayesian framework for the Kaplan–Meier (KM) estimator, which first selects external controls with similar hazard risks and then down-weights their impacts before incorporating them via the weighted KM estimator.

Nonetheless, these approaches overlook the time-varying nature of the outcome drift for survival analysis. To address this, Chen et al. (2021b) developed an adaptive empirical likelihood estimation that incorporates constraints from external summary-level information to account for time-varying baseline hazard differences. Huang et al. (2023) proposed a federated external control method to estimate hazard ratios in a federated weighted Cox model for time-to-event outcomes. Due to privacy and logistical concerns with data-sharing, these frameworks are designed to utilize external aggregated survival information, which can be restrictive and less efficient.

Furthermore, it is essential to control the information sharing between covariate effects and hazard risks simultaneously. Li et al. (2023b) proposed a transfer learning framework that allows for comparable information borrowing in both covariate effects and baseline hazards through penalized likelihood under Cox models. However, this framework lacks flexibility in modeling time-varying covariate effects. Bellot & van der Schaar (2019) proposed learning the shared representation between two populations via the flexible nonparametric survival trees and correcting distribution mismatches with boosting, aiming to improve prediction performance without providing uncertainty quantification.

Our Contributions Existing integrative methods are limited by the assumption of the Cox model, either on the cause-specific or subdistribution hazard scale, which requires to accurately model the survival curves. In recent years, semiparametric efficient and doubly robust estimators, which leverage the efficient influence function (EIF), including the methodology of solving the EIF-based estimation equation (Gao et al., 2024a; Lee et al., 2024) and the targeted maximum likelihood estimation (Rytgaard et al., 2022; 2023), have gained great popularity to draw inferences about the treatment effects for survival outcomes. Therefore, there is a pressing need for developing a flexible and data-adaptive integrative framework that accounts for outcome drift in time-to-event outcomes, coupled with valid inferential methods, to enhance efficiency and reliability in survival model estimation.

In this paper, we develop a doubly protected estimation method for evaluating treatment effects for survival outcomes. To correct for covariate shifts, we utilize the density ratio of baseline covariates between two datasets in the construction of the doubly robust treatment estimator, motivated by the semi-parametric EIF. Since our framework is developed based on the EIF, it offers an advantage over other non-parametric methods in the construction of confidence intervals.

Next, we recast this influence function into a selection-based integrative framework to address outcome drift, identifying a comparable subset of external data for borrowing. To adjust time-varying hazards for survival outcomes, we propose to detect this subset based on differences in the subject-level restricted mean survival time (RMST) with DR-Learner. By minimizing the bias-variance trade-off, our framework does not require stringent parametric assumptions on the survival curves and allows for the dynamic borrowing of external information with time-varying hazards.

Finally, we establish the asymptotic properties of our proposed data-adaptive integrative estimator for survival outcomes with guaranteed consistency and efficiency improvement, even in the presence of external heterogeneity. Besides, we demonstrate the robustness and reasonable efficiency gains via extensive simulation studies and a real-data application. Our implementation codes will be made publicly available after the acceptance of this manuscript.

3. Methodology

3.1. Notation, Assumptions and Identifications

Following the potential outcomes framework, let $T^{(a)}$ be the potential survival time if a subject received the binary treatment $A = a$. Let $S_a(t)$ and $\lambda_a(t)$ be the corresponding survival and hazard functions, defined as $S_a(t) = \mathbb{P}(T^{(a)} \geq t)$ and $\lambda_a(t) = \lim_{h\to 0} h^{-1}\mathbb{P}(t \leq T^{(a)} \leq t+h)/\mathbb{P}(T^{(a)} \geq t)$. Under the consistency assumption, the observed survival time T is the realization of potential outcome under the actualized treatment, i.e., $T = AT^{(1)} + (1-A)T^{(0)}$. In the presence of censoring, the survival time T is not always observable. Instead, we observe $Y = \min(T, C)$ and $\Delta = \mathbf{1}(T < C)$, where C is the censoring time, and $\mathbf{1}(\cdot)$ is an indicator function. Let $M_a^C(dt \mid X, R = r) = dN_a^C(t) -$

$\mathbf{1}(Y \geq t)\lambda_a^C(t \mid X, R = r)$ be a martingale with $dN_a^C(t) = \mathbf{1}(Y = t, \Delta = 0, A = a)$, and $\lambda_a^C(t \mid X, R = r) = \lim_{h\to 0} h^{-1}P(t \leq C \leq t+h \mid X, A = a, R = r)/P(C \geq t \mid X, A = a, R = r)$.

Suppose we have two data sets: the trial data and the external controls. Let X be the baseline covariates and R be the indicator of the data source, where $R = 1$ if the subject is from the trial data and $R = 0$ for the external controls. For the trial data, we observe $\mathcal{R} = \{V_i = (Y_i, \Delta_i, A_i, X_i, R_i = 1)\}_{i=1}^{N_t}$; for the external controls, only $\mathcal{E} = \{V_i = (Y_i, \Delta_i, A_i = 0, X_i, R_i = 0)\}_{i=N_t+1}^{N_t+N_e}$ are observed since no treatment is assigned. Denote the true distribution for V_i by $\mathbb{P}$, and the empirical measure by $\mathbb{P}_N$ as $\mathbb{P}_N(f) = \sum_i f(V_i)/N$, where $N = N_t + N_e$.

Let $\pi_R(X) = P(R = 1 \mid X)$, $q_R(X) = \pi_R(X)/\{1 - \pi_R(X)\}$ be the density ratio of the baseline covariates, $\pi_A(X) = P(A = 1 \mid X, R = 1)$ be the propensity score for the treatment, and $S_a(t \mid R = 1) = \mathbb{P}(T^{(a)} \geq t \mid R = 1)$ be the treatment-specific survival curves for the trial population. The parameter of our interest θ_τ is the average treatment effect measured by the restricted mean survival time (RMST) difference up to τ, defined by $\theta_\tau = \int_0^\tau \{S_1(t \mid R = 1) - S_0(t \mid R = 1)\}dt$. To identify the difference in RMST, the following assumptions are sufficient.

Assumption 3.1 (Internal validity for the trial data)**.** (i) $T^{(a)} \perp A \mid X, R = 1$ for $a = 0, 1$; and (ii) $0 < \pi_A(X), \pi_R(X) < 1$ in the support of X.

Assumption 3.2 (Informative censoring)**.** $T^{(a)} \perp C \mid X, A = a, R = r$ for $a = 0, 1$ and $r = 0, 1$.

Assumption 3.3 (Comparability for the external data)**.** $S_0(t \mid X) = S_0(t \mid X, R = 0) = S_0(t \mid X, R = 1)$ for any $t < \tau$ in the support of X, where $S_a(t \mid X, R) = \mathbb{P}(T^{(a)} \geq t \mid X, R)$;

Assumption 3.1 holds by the design of trial and is useful to detect the external heterogeneity. Assumption 3.2 is a common censoring at random assumption for survival analysis, which is a special case of the coarsening at random (Tsiatis, 2006). Assumption 3.3 states the external data is comparable to the trial data if there is a rich set of covariates capturing all the outcome predictors that are correlated with the data source indicator R.

However, Assumption 3.3 is prone to violations in practice due to many bias-generating concerns, such as unmeasured confounding, lack of concurrency, and outcome validity. Our proposed framework is two-fold: 1) Under Assumption 3.3 where covariate shift can be present, we develop a semiparametric efficient integrative estimator for the treatment effects evaluation using the combined data sets (Section 3.2); 2) Considering the potential violation of Assumption 3.3 where the outcome drift is allowed, we adapt the efficient estimation into a selective integrative procedure that first detects the biases and only retains a subset of comparable external data for integration (Section 3.3).

3.2. Efficient Integrative Estimation Assuming Population Homogeneity

Under Assumptions 3.1 to 3.3, the average treatment effects θ_τ, or $S_a(t \mid R = 1)$ sufficiently, is identified based on the observed data. The following theorem provides the identification formulas.

Theorem 3.4. *Under Assumptions 3.1 to 3.3, the following identification formulas hold for the treatment-specific survival curves $S_a(t \mid R = 1)$.*

(a) Based on the trial data:

$$S_a(t \mid R = 1) = \frac{1}{\mathbb{P}(R = 1)}\mathbb{E}\{RS_a(t \mid X, R = 1)\} = \frac{1}{\mathbb{P}(R = 1)}\mathbb{E}\left\{\frac{R\mathbf{1}(A = a)\mathbf{\Delta 1}(Y > t)}{P(A = a \mid X)\pi_1^C(Y, X)}\right\},$$

where $\pi_1^C(t, X) = P(C \geq t \mid X, R = 1)$ is the censoring probability for the trial.

(b) Based on the external data:

$$S_0(t \mid R = 1) = \frac{1}{\mathbb{P}(R = 1)}\mathbb{E}\left\{\frac{(1 - R)q_R(X)\mathbf{\Delta 1}(Y > t)}{\pi_0^C(Y, X)}\right\},$$

where $\pi_0^C(t, X) = P(C \geq t \mid X, R = 0)$ is the censoring probability for the external controls.

Theorem 3.4 provides the identification formulas for $S_a(t \mid R = 1)$, which is sufficient to identify the average treatment effect θ_τ among the trial population. In particular, Theorem 3.4(a) identifies $S_a(t \mid R = 1)$ based on the outcome imputation or the inverse probability censoring weighting with the trial data only; Theorem 3.4(b) uses the external controls to identify $S_0(t \mid R = 1)$. As the covariate distribution of the external controls may not be representative of the trial population, the identification formula relies on the density ratio $q_R(X)$ to obtain the survival curves marginalized over the trial population. A detailed proof of Theorem 3.4 is provided in Appendix A.1.

However, the identification formulas provided in Theorem 3.4 could motivate infinitely many estimators for θ_τ under Assumptions 3.1 to 3.3. To construct a more principled estimator, we derive the efficient influence function (EIF) of θ_τ in Theorem 3.5 based on the semiparametric theory (Tsiatis, 2006). The EIF, also known as the canonical gradient (Van der Laan et al., 2011), is a fundamental tool to achieve local semiparametric efficiency for estimation.

Theorem 3.5. *Under Assumptions 3.1 to 3.3,*

(a) the EIF for $S_1(t \mid R=1)$ *is* $\psi_{S_1,\text{eff}}(t,V) = \phi_{S_1,\text{eff}}(t,V) - RS_1(t \mid R=1)/\mathbb{P}(R=1)$*, where*

$$\phi_{S_1,\text{eff}}(t,V) = \frac{RA\mathbf{1}(Y>t)}{\mathbb{P}(R=1)\pi_A(X)\pi_1^C(t,X)} + \int_0^t \frac{RA\cdot dM_1^C(r \mid X)}{\mathbb{P}(R=1)\pi_A(X)\pi_1^C(r,X)} \frac{S_1(t \mid X,R=1)}{S_1(r \mid X,R=1)} + \frac{RS_1(t \mid X,R=1)}{\mathbb{P}(R=1)}\left\{1-\frac{A}{\pi_A(X)}\right\}.$$

(b) the EIF for $S_0(t \mid R=1)$ *is* $\psi_{S_0,\text{eff}}(t,V) = \phi_{S_0,\text{eff}}(t,V) - RS_0(t \mid R=1)/\mathbb{P}(R=1)$*, where*

$$\begin{aligned}\phi_{S_0,\text{eff}}(t,V) &= \frac{R(1-A)}{\mathbb{P}(R=1)}\frac{q_R(X)\mathbf{1}(Y>t)}{\pi_1^C(t,X)D(t,X)} \\ &+ \int_0^t \frac{R(1-A)}{\mathbb{P}(R=1)}\frac{q_R(X)dM_0^C(r \mid X,R=1)}{\pi_1^C(r,X)D(t,X)} \\ &\times \frac{S_0(t \mid X,R=1)}{S_0(r \mid X,R=1)} \\ &+ \int_0^t \frac{(1-R)}{\mathbb{P}(R=1)}\frac{q_R(X)r(t,X)dM_0^C(r \mid X,R=0)}{\pi_0^C(r,X)D(t,X)} \\ &\times \frac{S_0(t \mid X,R=0)}{S_0(r \mid X,R=0)} \\ &+ \frac{(1-R)}{\mathbb{P}(R=1)}\frac{q_R(X)r(t,X)\mathbf{1}(Y>t)}{\pi_0^C(t,X)D(t,X)} \\ &+ \frac{S_0(t \mid X,R=1)}{\mathbb{P}(R=1)}\left\{\frac{Rq_R(X)\{A-\pi_A(X)\}}{D(t,X)} \right. \\ &\left. + \frac{r(t,X)\{R-(1-R)q_R(X)\}}{D(t,X)}\right\},\end{aligned}$$

$D(t,X) = r(t,X) + \{1-\pi_A(X)\}q_R(X)$, $r(t,X) = V_{R1,A0}/V_{R0}$, $V_{R1,A0} = \text{var}\{\mathbf{1}(T>t) \mid R=1, A=0\}$, *and* $V_{R0} = \text{var}\{\mathbf{1}(T>t) \mid R=0\}$.

(c) the EIF for θ_τ *is:*

$$\psi_{\theta_\tau,\text{eff}}(V) = \int_0^\tau \{\psi_{S_1,\text{eff}}(t,V) - \psi_{S_0,\text{eff}}(t,V)\}dt.$$

Theorem 3.5(a) shows the EIF for $S_1(t \mid R=1)$ with the trial data, which is well-studied as the observed-data EIF under monotone coarsening in (Tsiatis, 2006); Theorem 3.5(b) shows the EIF for $S_0(t \mid R=1)$, which is an extension of Gao et al. (2024a) with additional integral terms contributed by the censoring scores of concurrent controls and external data; Theorem 3.5(c) suggests that the EIF for θ_τ is an integral of the EIFs for $S_a(t \mid R=1)$ as the average treatment effect θ_τ is an integral function of the treatment-specific survival curves $S_a(t \mid R=1)$. Furthermore, the proposed integrative framework can be generalized to any estimand that is a function of the survival function $S_a(t)$ (e.g., the mean or median of the survival time), with only a trivial algebraic extension, since the EIFs are derived for $S_a(t)$.

We now give some intuitions behind the EIF $\psi_{S_0,\text{eff}}(t,V)$ for the combined data sets. First, we derive the full-data EIF $\psi^F_{S_0,\text{eff}}(t,V)$ under Assumption 3.3, which involves a part being the weighted-averaged of the EIFs for the trial and external data with weights being inversely proportionately to the variances for the EIFs of the trial and external data, respectively. Next, we project the full-data EIF $\psi^F_{S_0,\text{eff}}(t,V)$ to the coarsened data space induced by the censoring, and gives the observed-data EIF in Theorem 3.5(b); see Theorem 10.4, Tsiatis (2006). A detailed proof of Theorem 3.5 is provided in Appendix A.2.

However, constructing an estimator based on the EIFs first requires approximating the unknown nuisance functions $\pi_R(X)$, $\pi_A(X)$, $S_a(t \mid X)$, and $S^C(t \mid X,R)$. By replacing the true distribution $\mathbb{P}$ with its estimated counterparts and solving the empirical expectation $\mathbb{P}_N\{\widehat{\psi}_{\theta_\tau,\text{eff}}(V)\}$, we have

$$\widehat{\theta}^{\text{acw}}_\tau = N^{-1}\sum_{i\in\mathcal{R}\cup\mathcal{E}}\int_0^\tau \left\{\widehat{\phi}_{S_1,\text{eff}}(t,V_i)dt - \widehat{\phi}_{S_0,\text{eff}}(t,V_i)\right\}dt.$$

Under the conditions in Theorem 3.6, our proposed integrative estimator $\widehat{\theta}^{\text{acw}}_\tau$ is rate doubly robust, asymptotically normal, and locally efficient as established below.

Theorem 3.6. *Denote* $\|f\|_{L_2} = \mathbb{P}\{f(V)^2\}^{1/2}$ *where* $\mathbb{P}$ *is the true distribution. Let* $\widehat{\pi}_R(X)$, $\widehat{\pi}_A(X)$, $\widehat{S}_a(t \mid X,R=1)$, *and* $\widehat{\pi}^C_r(t,X)$ *be general semi-parametric models for* $\pi_R(X)$, $\pi_A(X)$, $S_a(t \mid X,R=1)$, *and* $\pi^C_r(t,X)$*, respectively. Suppose Assumptions 3.1 to 3.3 and the regularity conditions A.1 are satisfied, up to a multiplicative constant, the estimation error* $\widehat{\theta}^{acw}_\tau - \theta_r = \text{err}(\widehat{\theta}^{acw}_\tau, \theta_r)$ *is bounded by*

$$\begin{aligned}\text{err}(\widehat{\theta}^{acw}_\tau,\theta_r) &= \left\{\|\widehat{S}_0(t \mid X,R=1) - S_0(t \mid X,R=1)\|_{L_2} \right. \\ &\left. + \|\widehat{S}_1(t \mid X,R=1) - S_1(t \mid X,R=1)\|_{L_2}\right\} \\ &\times \{\|\widehat{\pi}_A(X) - \pi_A(X)\|_{L_2} + \|\widehat{\pi}_R(X) - \pi_R(X)\|_{L_2} \\ &+ \max_{r<t}\|\widehat{\lambda}^C_0(r \mid X,R) - \lambda^C_0(r \mid X,R)\|_{L_2}\Big\}.\end{aligned}$$

Thus, we have $N^{1/2}(\widehat{\theta}^{acw}_\tau - \theta_\tau) \xrightarrow{d} N(0,\mathbb{V}_\tau)$*, where* $N = N_t + N_e$ *and* $\mathbb{V}_\tau = \mathbb{E}\{\psi^2_{\theta_\tau,\text{eff}}(V)\}$ *is the semi-parametric lower bound for the variance.*

Theorem 3.6 shows that the proposed estimator $\widehat{\theta}_\tau$ can incorporate flexible nonparametric or machine learning methods for estimating the nuisances with the required convergence rates, while maintaining the parametric-rate consistency. The proof of Theorem 3.6 is deferred to Appendix A.3.

3.3. Robust Selective Borrowing with DR-Learner for Outcome Drift Detection

In practice, Assumption 3.3 is often violated, that is, $S_0(t \mid X, R=1) \neq S_0(t \mid X, R=0)$ for some $t \in [0, \tau]$, and $\widehat{\theta}_\tau$ may be biased. Suppose there exists a comparable subset $\mathcal{A} \subseteq \{1, \cdots, N_e\}$ of the external controls such that $S_0(t \mid X_i, R=1) = S_0(t \mid X_i, R=0)$ for any time t and subject $i \in \mathcal{A}$, and we aim to selectively borrow this comparable subset. However, the subset $\mathcal{A}$ is often unknown a priori. In this section, we propose a robust selective borrowing framework to incorporate comparable external controls in estimating the average treatment effect. We introduce a vector of bias parameter $b_0 = (b_{1,0}, \cdots, b_{N_e,0})$, where $b_{i,0} = \int_0^\tau S_0(t \mid X_i, R_i = 1)dt - \int_0^\tau S_0(t \mid X_i, R_i = 0)dt$. To prevent bias in $\widehat{\theta}_\tau$, our goal is to identify the zero-valued subset of the bias parameter and leverage only this subset for the integrative estimation.

One simple estimator to detect the bias is the "plug-in" estimation, defined as $\widehat{b}_i = \widehat{b}(V_i) = \int_0^\tau \widehat{S}_0(t \mid X_i, R_i = 1)dt - \int_0^\tau \widehat{S}_0(t \mid X_i, R_i = 0)dt$. However, the "plug-in" estimator might have large finite-sample biases if one uses flexible models for the conditional survival functions. Heuristically, bias detection is equivalent to estimating the conditional differences in expected control means over two datasets, which parallels the conditional average treatment estimation in the causal inference literature when the study source indicator is perceived as the treatment indicator. To more accurately approximate the biases, the DR-Learner approach (Kennedy Edward, 2020; Kallus & Oprescu, 2023) is utilized to construct the initial pseudo-outcome ξ_i for the bias $b_{i,0}$ by $\xi_i = \xi(V_i) = \int_0^\tau \kappa_0(t, V_i \mid R=1)dt - \int_0^\tau \kappa_0(t, V_i \mid R=0)dt$, where

$$\begin{aligned}
&\kappa_0(t, V_i \mid R=1) = S_0(t \mid X_i, R=1) \\
&+ \int_0^t \frac{R_i(1-A_i)dM_0^C(r \mid X_i, R=1)}{\pi_R(X_i)\{1-\pi_A(X_i)\}\pi_1^C(r, X)} \frac{S_0(t \mid X_i, R=1)}{S_0(r \mid X_i, R=1)} \\
&+ \frac{R_i(1-A_i)}{\pi_R(X_i)\{1-\pi_A(X_i)\}} \\
&\times \left\{ \frac{\mathbf{1}(Y_i > t)}{\pi_1^C(t, X_i)} - S_0(t \mid X_i, R=1) \right\},
\end{aligned}$$

and

$$\begin{aligned}
&\kappa_0(t, V_i \mid R=0) = S_0(t \mid X_i, R=0) \\
&+ \int_0^t \frac{(1-R_i)dM_0^C(r \mid X_i, R=0)}{\{1-\pi_R(X_i)\}\pi_0^C(t, X_i)} \frac{S_0(t \mid X_i, R=0)}{S_0(r \mid X_i, R=0)} \\
&+ \frac{(1-R_i)}{1-\pi_R(X_i)} \left\{ \frac{\mathbf{1}(Y_i > t)}{\pi_0^C(t, X_i)} - S_0(t \mid X_i, R=0) \right\}.
\end{aligned}$$

Lemma 3.7 provides bounds for estimation error for the pseudo-outcomes, which is the key step for bias detection.

Lemma 3.7. *Let $\widehat{\xi}(V)$ be the pseudo-outcome with unknown nuisance functions replaced by their estimated counterparts and $\xi^*(V)$ be its probability limit, the conditional expectation $\|\mathbb{E}\{\xi^*(V) - b_0 \mid X\}\|_{L_2}$ is bounded by*

$$\begin{aligned}
&\sum_{r=0}^{1} \|\widehat{\pi}_R(X) - \pi_R(X)\|_{L_2} \\
&\times \|\widehat{S}_0(t \mid X, R=r) - S_0(t \mid X, R=r)\|_{L_2} \\
&+ \sum_{r=0}^{1} \|\widehat{\pi}_r^C(t, X) - \pi_r^C(t, X)\|_{L_2} \\
&\times \|\widehat{S}_0(t \mid X, R=r) - S_0(t \mid X, R=r)\|_{L_2} \\
&+ \|\widehat{\pi}_A(X) - \pi_A(X)\|_{L_2} \\
&\times \|\widehat{S}_0(t \mid X, R=1) - S_0(t \mid X, R=1)\|_{L_2},
\end{aligned}$$

up to a multiplicative constant.

The results in Lemma 3.7 show that nuisance errors will have smaller impacts on the pseudo-outcomes $\widehat{\xi}$ as its estimation error is bounded by a quadratic form of the nuisance errors. The proof is provided in Appendix A.4.

We now present our adaptive integrative framework in Algorithm 1. The cutoff value τ for computing the RMST is crucial in practice, as the distribution of the tail beyond this point is neglected. Typically, the event rates at this cutoff value should exceed 10% to ensure sufficient data for model development.

In Algorithm 1, Step 1 is a typical strategy to use machine learning techniques for estimating the nuisance functions; The penalty term in Step 2 is chosen to guarantee the selection consistency such that $\mathbb{P}(\tilde{\mathcal{A}} = \mathcal{A}) \to 1$, where $\tilde{\mathcal{A}} = \{i : \tilde{b}_i = 0\}$ is the estimated comparable set. For example, the penalty term $p(\cdot)$ can be the adaptive lasso (Zou, 2006), the smoothly clipped absolute deviation (SCAD) (Fan & Li, 2001), and the minimax concave penalty (MCP) (Zhang, 2010); Finally, Step 3 outputs the adaptive integrative estimator $\widehat{\theta}_\tau^{\text{adapt}}$ with the comparable set $\tilde{\mathcal{A}}$ by

$$\begin{aligned}
\widehat{\theta}_\tau^{\text{adapt}} = N^{-1} \sum_{i \in \mathcal{R} \cup \mathcal{E}} \int_0^\tau \widehat{\phi}_{S_1,\text{eff}}(t, V_i)dt \\
- N^{-1} \sum_{i \in \mathcal{R} \cup \mathcal{E}} \int_0^\tau \widehat{\phi}^{\tilde{\mathcal{A}}}_{S_0,\text{eff}}(t, V_i)dt,
\end{aligned}$$

where the difference between $\phi^{\mathcal{A}}_{S_0,\text{eff}}(t, V)$ in Step 3 and $\phi_{S_0,\text{eff}}(t, V)$ in Theorem 3.5 lies in the focus on leveraging the comparable set $\mathcal{A}$ instead of the whole external controls $\mathcal{E}$ for the integrative analysis.

Theorem 3.8. *Let $\mathbb{V}_\tau^{aipw}$ be the variance of the trial-only estimator. Under the same conditions in Theorem 3.6, we*

Algorithm 1 Doubly Protected Adaptive Integrative Analysis for Survival Outcomes

Input: trial data $\mathcal{R} = \{V_i = (Y_i, \Delta_i, A_i, X_i, R_i = 1)\}_{i=1}^{N_t}$, the external controls $\mathcal{E} = \{V_i = (Y_i, \Delta_i, A_i = 0, X_i, R_i = 0)\}_{i=N_t+1}^{N_t+N_e}$, penalty function $p(\cdot)$, tuning parameter λ_N, and the cutoff value τ.

◁ Preparation

Randomly split the data $\mathcal{R} \cup \mathcal{E}$ into two folds $\mathcal{I}_1$ and $\mathcal{I}_2$.

◁ Step 1

Fit the conditional survival curves $\widehat{S}_a(t \mid X, R = r)$ and $\widehat{S}^C(t \mid X, R = r)$ on $\mathcal{I}_1$ for $r = 0, 1$.

Fit the propensity scores $\widehat{\pi}_R(X)$ and $\widehat{\pi}_A(X)$ on $\mathcal{I}_1$.

Compute the pseudo-outcomes $\widehat{\xi}_i$ by $\kappa_0(t, V_i \mid R = 1)$ and $\kappa_0(t, V_i \mid R = 0)$ on $\mathcal{I}_2$.

◁ Step 2

Refine the pseudo-outcomes by $\tilde{b}_i = \arg\min_{b_i}(\widehat{\xi}_i - b_i)^2 + \lambda_N p(|b_i|)$.

Obtain the comparable set $\tilde{\mathcal{A}} = \{i : \tilde{b}_i = 0\}$.

◁ Step 3

Output the adaptive integrative estimator $\widehat{\theta}_\tau^{\text{adapt}}$.

have $N^{1/2}(\widehat{\theta}_\tau^{adapt} - \theta_\tau) \xrightarrow{d} N(0, \mathbb{V}_\tau^{adapt})$*, where*

$$\mathbb{V}_\tau^{adapt} - \mathbb{V}_\tau^{aipw} = \frac{\pi_R(X) r(t,X) \mathbb{P}(b_0 = 0 \mid X, R = 0)}{\mathbb{P}(R=1)^2 D_{b_0}(t,X)\{1 - \pi_A(X)\}} \times \frac{D^*_{b_0}(t,X)}{D_{b_0}(t,X)} \frac{r(t,X)}{r^*(t,X)} V_{R1,A0},$$

where $D^*_{b_0}(t,X) = r^*(t,X)\mathbb{P}(b_0 = 0 \mid X, R = 0) + \{1 - \pi_A(X)\} q_R(X)$ *and* $r^*(t,X) = V^*_{R1,A0}/V^*_{R0}$*;* V_{R_1,A_0} *are defined in Theorem 3.5, and the modified variance terms* $V^*_{R1,A0}$ *and* V^*_{R0} *are defined in Appendix A.5.*

Theorem 3.8 highlights the benefit of selective incorporating external controls, where the asymptotic variance of $\widehat{\theta}_\tau^{\text{adapt}}$ is strictly smaller than the variance $\mathbb{V}_\tau^{\text{aipw}}$ of the trial-only estimator unless the external study is in extremely poor quality (i.e., $r(t,X) = 0$) or the comparable external subset is empty (i.e., $\mathbb{P}(b_0 = 0 \mid X, R = 0) = 0$). A proof is provided in Appendix A.5.

4. Simulation

In this section, we conduct several simulation studies to evaluate the finite-sample performance of the proposed selective integrative estimator. The sample size for the external controls are fixed at $N_e = 500$, and the parameter of our interest is the difference in RMST with $\tau = 2$. First, we generate $X = (X_{i,1}, \cdots, X_{i,p}) \sim \mathcal{N}(0, I_{p\times p})$ with $p = 3$ for each subject i. Next, we generate the data source indicator R by Bernoulli sampling in Table 1, where $\alpha_{R,0}$ is chosen such that the average of R is around N_t/N. For the trial data, where $R = 1$, the treatment A is generated by Bernoulli sampling:

$$A \mid X, R = 1 \sim \text{Bernoulli}\left\{\frac{\exp(\alpha_{A,0} + 1_p^\intercal X)}{1 + \exp(\alpha_{A,0} + 1_p^\intercal X)}\right\},$$

where $\alpha_{A,0}$ is chosen such that the average of A is around N_1/N_t, and N_1 is the size of treatment group. We consider the following conditional hazard function $\lambda_a(t \mid X, R = 1) = \exp(-.5a - 1_p^\intercal X \cdot 0.2)$, and $\lambda^C(t \mid X, R = 1) = \exp(1_p^\intercal X \cdot 0.1 + \beta_C)$ for the trial. The time-to-event outcomes T and the censoring times C are generated by inverting the survival functions induced by the hazard function $\lambda_a(t \mid X)$ and $\lambda^C(t \mid X)$, respectively. The parameter $\beta_C = 1$ controls the expected censoring time for the trial data where the censoring rates $P(C < T)$ is around 40%; additional simulations where the data is subject to a different intensities of censoring with various values of β_C are available in Appendix B.

Next, to mimic the bias-generating concerns raised by the FDA, namely, selection bias, unmeasured confounding, lack of concurrency, and different covariate effects and time-varying baseline hazards, we consider five simulation settings for the external controls, as summarized in Table 1. Under Setting One, the survival and censoring times for the external controls are generated using the exact same parameters as $\lambda_a(t \mid X)$ with a fixed at 0 (i.e., no treatment). Under Setting Two, the hazard functions for both studies are confounded by an unobserved factor U to maintain the same level of hazard variability across the two datasets. In particular, $U \sim \mathcal{N}(0, 1)$ (zero mean) is included in the hazard function for the trial data, whereas $U + 1 \sim \mathcal{N}(1, 1)$ (non-zero mean) is used for the external controls, which is expected to introduce greater outcome drift. Under Setting Three, the external controls are subject to inconcurrency bias, where δ_i takes values from $\{0, 5\}$ with equal probability $1/2$. Lack of concurrency could occur when the trial data and external control data are collected during different time periods or under varying healthcare settings. Under Setting Four, the external controls have different covariate effects, whereas under Setting Five, they exhibit different baseline time-varying hazards. Such discrepancies may arise if the two populations respond differently to the treatment (or placebo), even when they share the same baseline covariates.

In our evaluation, we compare the proposed selective borrowing estimator $\widehat{\theta}_\tau^{\text{adapt}}$ with the trial-only estimator $\widehat{\theta}_\tau^{\text{aipw}}$ (Tsiatis, 2006), the naive full borrowing estimator $\widehat{\theta}_\tau^{\text{acw}}$, the propensity score-integrated estimator $\widehat{\theta}^{\text{psrwe}}$ (Chen et al., 2020), and the transfer learning Cox regression $\widehat{\theta}_\tau^{\text{TransCox}}$ (Li

Table 1. Summary of considered three simulation settings

Bias	Setting	Details
Covariate shift	Selection bias	$R \mid X \sim \text{Bernoulli}\left\{\frac{\exp(\alpha_{R,0}+1_p^\intercal X)}{1+\exp(\alpha_{R,0}+1_p^\intercal X)}\right\}$, $\lambda_0(t \mid X, R=0) = \exp(-1_p^\intercal X \cdot 0.2)$
Outcome drift	Unmeasured confounder	$R \mid X \sim \text{Bernoulli}\left\{\frac{\exp(\alpha_{R,0}+1_p^\intercal X+U)}{1+\exp(\alpha_{R,0}+1_p^\intercal X+U)}\right\}$, $\lambda_a(t \mid X, U, R) = \exp\{-0.5a - 1_p^\intercal X \cdot 0.2 + 3(U + \mathbf{1}(R=0)\}$
	Lack of concurrency	$R \mid X \sim \text{Bernoulli}\left\{\frac{\exp(\alpha_{R,0}+1_p^\intercal X)}{1+\exp(\alpha_{R,0}+1_p^\intercal X)}\right\}$, $\lambda_a(t \mid X, R) = \exp\{-.5a - 1_p^\intercal X \cdot 0.2 + 3\delta\mathbf{1}(R=0)\}$
	Different covariate effect	$\lambda_a(t \mid X, R=1) = \exp(-.5a - 1_p^\intercal X \cdot 0.2)$ $\lambda_0(t \mid X, R=0) = \exp(-1_p^\intercal X \cdot 0.5)$
	Different baseline hazard	$\lambda_a(t \mid X, R=1) = t\exp(-.5a - 1_p^\intercal X \cdot 0.2)$ $\lambda_0(t \mid X, R=0) = 2t\exp(-1_p^\intercal X \cdot 0.2)$

et al., 2023b). The penalty term $p(\cdot)$ is chosen to be the adaptive lasso (Zou, 2006). The conditional survival curves $S_a(t \mid X)$ and $S^C(t \mid X)$ are modeled by the Cox model, and the propensity scores $\pi_R(X)$ and $\pi_A(X)$ are modeled by the SuperLearner with the logistic regression and random forest as the base learners (Van Der Laan & Rubin, 2006).

Figure 1(Top) presents the bias, standard error (SE), and the square root of the mean squared error (Root-MSE) for each method across three simulation settings, with the size of the concurrent control N_0 ranging from 50 to 400. As expected, the trial-only benchmark estimator $\widehat{\theta}_\tau^{\text{aipw}}$ exhibits small biases across these three settings by design. Under Setting One, where the external controls do not present any outcome heterogeneity, all integrative estimators demonstrate improved Root-MSE relative to the benchmark $\widehat{\theta}_\tau^{\text{aipw}}$. Our proposal $\widehat{\theta}_\tau^{\text{adapt}}$ is less efficient compared to other integrative estimators as it induces extra variability due to bias detection for selective borrowing. However, $\widehat{\theta}_\tau^{\text{acw}}$, $\widehat{\theta}^{\text{psrwe}}$ and $\widehat{\theta}_\tau^{\text{TransCox}}$ can be substantially biased under Settings Two and Three, where the unmeasured confounder or time inconcurrency are present. In particular, our data-adaptive integrative estimator can detect the incompatibility of external controls and selectively borrows the comparable subset, resulting in a similar level of biases, but improved standard errors compared to the benchmark $\widehat{\theta}_\tau^{\text{aipw}}$. The results under Settings Four and Five are presented in Appendix B.

To evaluate the asymptotic properties of our proposed estimators, Figure 1(Bottom) presents the type-I error, the coverage probability, and the power for detecting $\theta_\tau > -0.3$ for the estimators $\widehat{\theta}_\tau^{\text{aipw}}$, $\widehat{\theta}_\tau^{\text{acw}}$, and $\widehat{\theta}_\tau^{\text{adapt}}$. The bootstrap-based variance estimation with bootstrap size 50 is used to construct the 95% Wald confidence intervals for evaluation. Under Setting One, $\widehat{\theta}_\tau^{\text{acw}}$ successfully controls the type-I errors, maintain the nominal coverage rates and achieve the highest powers for detecting treatment effects over a varying size of concurrent controls. However, it has inflated type-I errors and deteriorated coverage probabilities under Setting Two and Three when the outcome drift is present. In contrast, our proposed borrowing estimator effectively controls the type-I error, maintains the satisfactory coverage rates, and achieves improved or comparable power levels compared to the benchmark $\widehat{\theta}_\tau^{\text{aipw}}$, irrespective of the presence of the outcome drift.

5. Real-data Application

This section presents an application of the proposed selective borrowing methodology to evaluate the effectiveness of Galcanezumab (120mg) versus placebo in patients with episodic migraine. The primary trial study is EVOLVE-1, a phase 3 double-blinded trial for patients with episodic migraines that randomized patients 1:1:2 to receive monthly Galcanezumab 120mg, 240mg, or placebo for up to 6 months (Stauffer et al., 2018). In addition, placebo subjects from the REGAIN study were used as external controls to augment the control arm from the EVOLVE-1 study. The REGAIN study is a phase-3 double blinded trial for patients with chronic migraine headaches that randomized patients 1:1:2 to receive monthly galcanezulab 120mg, 240mg, or placebo for up to 3 months, with a subsequent 9-month open label extension follow-up period (Detke et al., 2018).

The primary objective is to assess whether galcanezumab 120mg is superior to placebo in helping patients with episodic migraine achieve a meaningful improvement in migraine headache days (MHD), defined as a 50% reduction in mean MHD per month from baseline. To estimate the difference in time to first meaningful MHD reduction up to 6 months post-baseline, the galcanezumab 120mg and placebo arms from EVOLVE-1 are augmented with

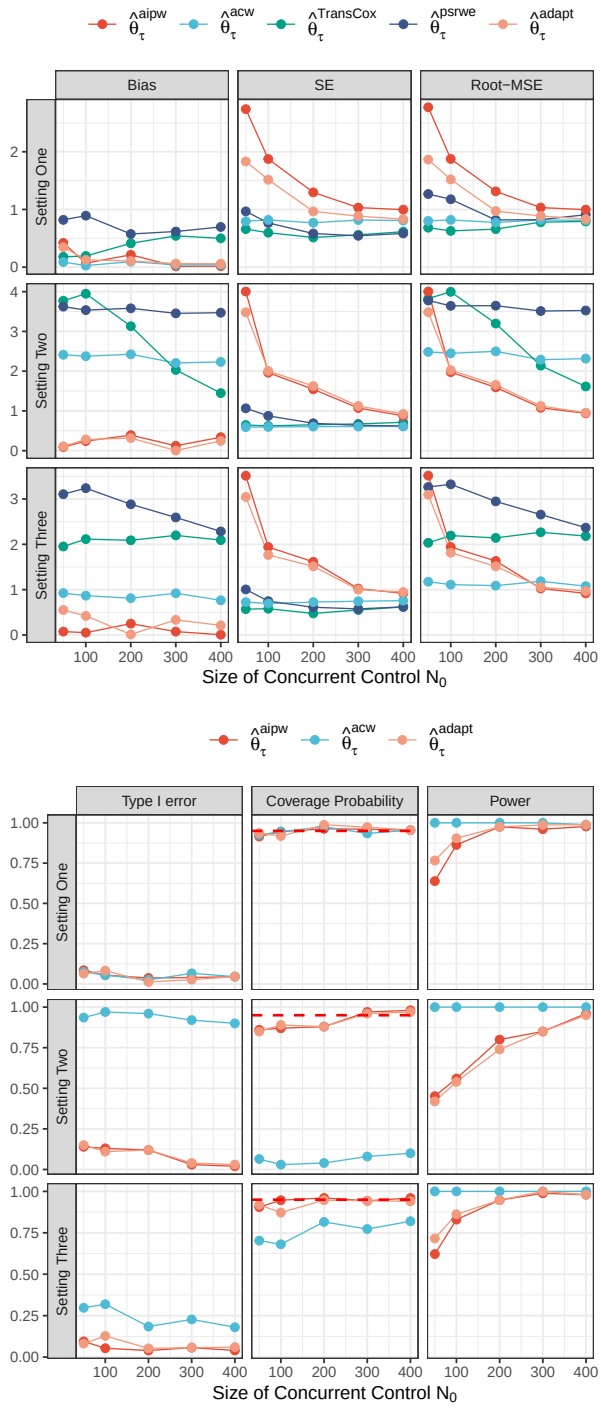

Figure 1. (Top) Point estimation results for RMST over 500 Monte Carlo experiments; (Bottom) Asymptotic results for RMST over 500 Monte Carlo experiments under Settings 1) selection bias only; 2) unmeasured confounding; 3) lack of concurrency.

the placebo arm from the REGAIN study. The treatment effect θ_τ is defined as the RMST difference of the time to first occurrence of 50% MHD reduction up to time $\tau = 6$ months.

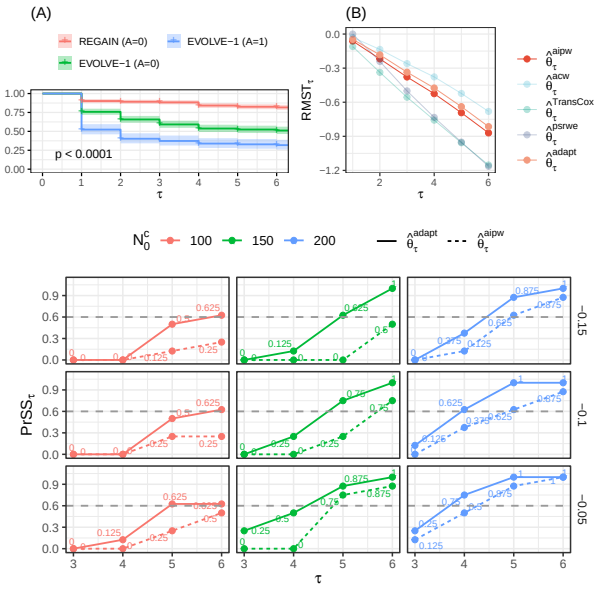

Figure 2. (A) Kaplan-Meier survival curves for the EVOLVE-1 study and the placebo group of the REGAIN study; (B) The estimated treatment effect in RMST for the EVOLVE-1 study; (C) Probability of study success of detecting $\theta_\tau < -0.05, -0.1, -0.15$ using $\widehat{\theta}_\tau^{\text{aipw}}$ and $\widehat{\theta}_\tau^{\text{adapt}}$ over a range of restricted times τ under varying sizes of sub-samples from the placebo group of the REGAIN study.

Figure 2(A) presents the unconditional Kaplan-Meier curves for the placebo group of these two studies. The log-rank test indicates that the time to meaningful MHD improvement is different between the placebo arms of the EVOLVE-1 study and the REGAIN study, suggesting the need to account for outcome drift. Figure 2(B) provides the estimated treatment effect of Galcanezumab (120mg) in terms of RMST as a function of the restricted month τ. All estimators exhibit a trend of decreasing RMST and show significant improvements in mitigating migraines severity, implying a shortened response time for at least a 50% reduction in MHD after being treated with Galcanezumab (120mg). Our proposed estimator $\widehat{\theta}_\tau^{\text{adapt}}$ yields a RMST that is closer to the trial-only estimator $\widehat{\theta}_\tau^{\text{aipw}}$, highlighting its capability to control the bias arising from external heterogeneity.

Next, we benchmark the performances of $\widehat{\theta}_\tau^{\text{adapt}}$ with the trial-only estimator $\widehat{\theta}_\tau^{\text{aipw}}$ to emphasize the efficiency gain of our selective borrowing framework. To accomplish this, we retain the treatment group of the primary EVOLVE-1 study with a size of $N_1 = 212$, and create 50 sub-samples

by randomly selecting N_0^c patients from its placebo group with $N_0^c = 100, 150, 200$. The placebo group of the REGAIN study is then augmented to each selected sub-samples. Figure 2(C) presents the empirical probability of study success (PrSS) over a range of restricted times τ under different sizes of concurrent controls. The empirical PrSS is computed by the proportion of successfully detecting $\theta_\tau < -0.05, -0.1, -0.15$ over the repeated sub-samples. When combined with the placebo group of the REGAIN study, $\widehat{\theta}_\tau^{\text{adapt}}$ enhances the time-to-event analyses and yields a higher PrSS compared to $\widehat{\theta}_\tau^{\text{aipw}}$across all the sizes of sampled concurrent controls.

For example, suppose that we aim to reach PrSS ≥ 0.6 of detecting $\theta_\tau < -0.1$ at month $\tau = 6$, $\widehat{\theta}_\tau^{\text{adapt}}$ only need to recruit 100 patients for the placebo group (solid red line at $\tau = 6$ of the middle panel in Figure 2(C)), however, the trial-only estimator $\widehat{\theta}_\tau^{\text{aipw}}$ needs at least 150 patients for the placebo group (dash green line at $\tau = 6$ of the middle panel in Figure 2(C)). Therefore, our approach could attain similar levels of PrSS with fewer patients and a shortened patient enrollment period by appropriately leveraging the external controls, which could eventually accelerate the drug development for rare diseases where the event rates are typically low and imbalanced trials are often considered.

6. Discussion

In this paper, we introduce a doubly protected borrowing framework that utilizes external controls to enhance treatment estimation for survival outcomes. Unlike most existing approaches, our method is built on semi-parametric efficient estimation coupled with the DR-Learner to detect outcome drift for selective borrowing. This approach effectively incorporates external controls without introducing biases into the integrative treatment evaluation. Moreover, the proposed approach offers a new perspective on the integrative analysis for survival outcomes with a proper method for inference, which could be a valuable contribution to many survival analyses in the machine learning community, such as those involving customer churn (Larivière & Van den Poel, 2004; Gao et al., 2024b), and multi-source domain adaption (Mansour et al., 2008; Shaker & Lawrence, 2023).

Our simulation studies reveal several challenges in controlling type I errors in the presence of unmeasured confounding for small samples, a problem also noted in other methods such as Bayesian dynamic borrowing (Dejardin et al., 2018; Kopp-Schneider et al., 2020). Future work will focus on enhancing type I error control using alternative strategies, such as exact inference. In addition, the current selection criterion focuses on detecting outcome drift based on the differences in RMSTs. However, this may be less efficient when some parts of the conditional survival function are invariant across studies, as our proposal might exclude these valuable external controls that are partially comparable to the trial data. Future research should explore approaches that select comparable information rather than entire comparable subjects. One promising direction involves jointly estimating the average treatment effect and bias functions, as suggested in recent studies (Yang et al., 2020; Wu & Yang, 2023).

Software and Data

Our R codes with illustrative examples are available at `https://github.com/Gaochenyin/SelectiveIntegrative`.

Acknowledgements

We would like to thank the anonymous (meta-)reviewers of ICML 2025 for their helpful comments. This project is supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award U01FD007934 totaling $1,674,013$ over two years funded by FDA/HHS. It is also supported by the National Science Foundation under Award Number SES 2242776, totaling $225,000$ over three years. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS, the National Science Foundation, or the U.S. Government.

Impact Statement

Our proposed method for analyzing censored survival data enhances the evaluation of treatment efficacy in clinical trials with small control groups by incorporating external controls while addressing data heterogeneity to reduce bias. This approach promotes ethical research practices by improving statistical efficiency and minimizing patient exposure to ineffective treatments.

References

Bellot, A. and van der Schaar, M. Boosting transfer learning with survival data from heterogeneous domains. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 57–65. PMLR, 2019.

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. Efficient and adaptive inference in semiparametric models, forthcoming monograph, 1993.

Cai, T. T. and Wei, H. Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *The Annals of Statistics*, 49:100 – 128, 2021.

Chatterjee, N., Chen, Y.-H., Maas, P., and Carroll, R. J.

Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *Journal of the American Statistical Association*, 111(513):107–117, 2016.

Chen, S., Zhang, B., and Ye, T. Minimax rates and adaptivity in combining experimental and observational data. *arXiv preprint arXiv:2109.10522*, 2021a.

Chen, W.-C., Wang, C., Li, H., Lu, N., Tiwari, R., Xu, Y., and Yue, L. Q. Propensity score-integrated composite likelihood approach for augmenting the control arm of a randomized controlled trial by incorporating real-world data. *Journal of biopharmaceutical statistics*, 30(3):508–520, 2020.

Chen, W.-C., Lu, N., Wang, C., Li, H., Song, C., Tiwari, R., Xu, Y., and Yue, L. Q. Propensity score-integrated approach to survival analysis: leveraging external evidence in single-arm studies. *Journal of biopharmaceutical statistics*, 32(3):400–413, 2022.

Chen, Z., Ning, J., Shen, Y., and Qin, J. Combining primary cohort data with external aggregate information without assuming comparability. *Biometrics*, 77(3):1024–1036, 2021b.

Dejardin, D., Delmar, P., Warne, C., Patel, K., van Rosmalen, J., and Lesaffre, E. Use of a historical control group in a noninferiority trial assessing a new antibacterial treatment: A case study and discussion of practical implementation aspects. *Pharmaceutical Statistics*, 17: 169–181, 2018.

Detke, H. C., Goadsby, P. J., Wang, S., Friedman, D. I., Selzler, K. J., and Aurora, S. K. Galcanezumab in chronic migraine: the randomized, double-blind, placebo-controlled regain study. *Neurology*, 91(24):e2211–e2221, 2018.

Fan, J. and Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

FDA. E10 choice of control group and related issues in clinical trials, 2001. Accessed: 2023-02-23.

FDA. Rare diseases: Natural history studies for drug development, 2019. Accessed: 2023-02-23.

FDA. Considerations for the design and conduct of externally controlled trials for drug and biological products guidance for industry, 2023. Accessed: 2023-02-23.

Gao, C. and Yang, S. Pretest estimation in combining probability and non-probability samples. *Electronic Journal of Statistics*, 17(1):1492–1546, 2023.

Gao, C., Yang, S., Shan, M., YE, W., Lipkovich, I., and Faries, D. Improving randomized controlled trial analysis via data-adaptive borrowing. *Biometrika*, pp. asae069, 2024a.

Gao, C., Zhang, Z., and Yang, S. Causal customer churn analysis with low-rank tensor block hazard model. In *International Conference on Machine Learning*, pp. 14920–14953. PMLR, 2024b.

Huang, C., Wei, K., Wang, C., Yu, Y., and Qin, G. Covariate balance-related propensity score weighting in estimating overall hazard ratio with distributed survival data. *BMC medical research methodology*, 23(1):233, 2023.

Huang, C.-Y., Qin, J., and Tsai, H.-T. Efficient estimation of the cox model with auxiliary subgroup survival information. *Journal of the American Statistical Association*, 111(514):787–799, 2016.

Jiang, Z., Yang, S., and Ding, P. Multiply robust estimation of causal effects under principal ignorability. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1423–1445, 2022.

Kallus, N. and Oprescu, M. Robust and agnostic learning of conditional distributional treatment effects. In *International Conference on Artificial Intelligence and Statistics*, pp. 6037–6060. PMLR, 2023.

Kennedy, E. H. Semiparametric theory and empirical processes in causal inference. *Statistical causal inferences and their applications in public health research*, pp. 141–167, 2016.

Kennedy Edward, H. Optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497*, 2020.

Kopp-Schneider, A., Calderazzo, S., and Wiesenfarth, M. Power gains by using external information in clinical trials are typically not possible when requiring strict type i error control. *Biometrical Journal*, 62:361–374, 2020.

Larivière, B. and Van den Poel, D. Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. *Expert Systems with Applications*, 27(2): 277–285, 2004.

Lee, D., Gao, C., Ghosh, S., and Yang, S. Transporting survival of an hiv clinical trial to the external target populations. *Journal of biopharmaceutical statistics*, 34(6): 922–943, 2024.

Li, F., Morgan, K. L., and Zaslavsky, A. M. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400, 2018.

Li, S., Cai, T. T., and Li, H. Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):149–173, 2022.

Li, S., Cai, T., and Duan, R. Targeting underrepresented populations in precision medicine: A federated transfer learning approach. *The Annals of Applied Statistics*, 17 (4):2970–2992, 2023a.

Li, Z., Shen, Y., and Ning, J. Accommodating time-varying heterogeneity in risk estimation under the cox model: A transfer learning approach. *Journal of the American Statistical Association*, 118(544):2276–2287, 2023b.

Liu, D., Zheng, Y., Prentice, R. L., and Hsu, L. Estimating risk with time-to-event data: An application to the women's health initiative. *Journal of the American Statistical Association*, 109(506):514–524, 2014.

Mansour, Y., Mohri, M., and Rostamizadeh, A. Domain adaptation with multiple sources. *Advances in neural information processing systems*, 21, 2008.

Rytgaard, H. C., Gerds, T. A., and van der Laan, M. J. Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes. *The Annals of Statistics*, 50(5):2469–2491, 2022.

Rytgaard, H. C., Eriksson, F., and van der Laan, M. J. Estimation of time-specific intervention effects on continuously distributed time-to-event outcomes by targeted maximum likelihood estimation. *Biometrics*, 79(4):3038–3049, 2023.

Shaker, A. and Lawrence, C. Multi-source survival domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 9752–9762, 2023.

Stauffer, V. L., Dodick, D. W., Zhang, Q., Carter, J. N., Ailani, J., and Conley, R. R. Evaluation of galcanezumab for the prevention of episodic migraine: the evolve-1 randomized clinical trial. *JAMA neurology*, 75(9):1080–1088, 2018.

Tian, Y. and Feng, Y. Transfer learning under high-dimensional generalized linear models. *Journal of the American Statistical Association*, 118(544):2684–2697, 2023.

Tsiatis, A. A. *Semiparametric theory and missing data*. Springer, 2006.

Van Der Laan, M. J. and Rubin, D. Targeted maximum likelihood learning. *The international journal of biostatistics*, 2(1), 2006.

Van der Laan, M. J., Rose, S., et al. *Targeted learning: causal inference for observational and experimental data*, volume 4. Springer, 2011.

Wang, C., Lu, N., Chen, W.-C., Li, H., Tiwari, R., Xu, Y., and Yue, L. Q. Propensity score-integrated composite likelihood approach for incorporating real-world evidence in single-arm clinical studies. *Journal of biopharmaceutical statistics*, 30(3):495–507, 2020.

Wu, L. and Yang, S. Transfer learning of individualized treatment rules from experimental to real-world data. *Journal of Computational and Graphical Statistics*, 32: 1036–1045, 2023.

Yang, S., Zeng, D., and Wang, X. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *arXiv preprint arXiv:2007.12922*, 2020.

Yang, S., Gao, C., Zeng, D., and Wang, X. 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, 2023.

Zeng, D. and Lin, D. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(4):507–564, 2007.

Zhang, C.-H. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38 (2):894 – 942, 2010. doi: 10.1214/09-AOS729. URL https://doi.org/10.1214/09-AOS729.

Zou, H. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.

A. Proofs

A.1. Proof of Theorem 3.4

We only prove the identification formulas for $S_0(t \mid R=1)$, and similar proofs follow for $S_1(t \mid R=1)$. Based on the trial data, we have

$$\begin{aligned} S_0(t \mid R=1) &= \mathbb{E}\{S_0(t \mid X, R=1) \mid R=1\} \\ &= \frac{1}{\mathbb{P}(R=1)}\mathbb{E}\{RS_0(t \mid X, R=1)\} \\ &= \frac{1}{\mathbb{P}(R=1)}\mathbb{E}\{RP(T^{(0)} > t \mid X, R=1)\} \\ &= \frac{1}{\mathbb{P}(R=1)}\mathbb{E}\{RP(T^{(0)} > t \mid X, R=1, A=0, C>T)\} \\ &= \frac{1}{\mathbb{P}(R=1)}\mathbb{E}\left\{\frac{R(1-A)\Delta\mathbf{1}(Y>t)}{\{1-\pi_A(X)\}S^C(Y \mid X, A=0, R=1)}\right\}, \end{aligned} \tag{1}$$

where the fourth equality holds under Assumptions 3.1 and 3.2. Based on the external data, we have

$$\begin{aligned} S_0(t \mid R=1) &= \mathbb{E}\{S_0(t \mid X, R=1) \mid R=1\} \\ &= \frac{1}{\mathbb{P}(R=1)}\mathbb{E}\{RS_0(t \mid X, R=1)\} \\ &= \frac{1}{\mathbb{P}(R=1)}\mathbb{E}\{\pi_R(X)S_0(t \mid X, R=0)\} \\ &= \frac{1}{\mathbb{P}(R=1)}\mathbb{E}\{\pi_R(X)P(T^{(0)} > t \mid X, R=0, A=0, C>T)\} \\ &= \frac{1}{\mathbb{P}(R=1)}\mathbb{E}\left\{\frac{(1-R)q_R(X)\Delta\mathbf{1}(Y>t)}{S^C(Y \mid X, A, R=0)}\right\}, \end{aligned}$$

where the third equality holds under Assumptions 3.3.

A.2. Proof of Theorem 3.5

A.2.1. Preliminaries

We first derive the full-data efficient influence function (EIF) for the treatment-specific survival curves $S_a(t \mid R=1)$ without censoring, i.e., the full data $W_i = (T_i, A_i, X_i, R_i = r)$. We next employ the semi-parametric theory in Bickel et al. (1993) to derive the EIFs. In particular, we consider a one-dimensional parametric submodel $f_\theta(W)$, which contains the true model $f(W)$ at $\theta = 0$, i.e., $f_\theta(W) \mid_{\theta=0} = f(W)$. We use dot to denote the partial derivative with respect to θ, and $s_\theta(\cdot)$ to denote the score function of the submodel. For example, we have

$$\begin{aligned} \dot{\mu}_0 &= \frac{\partial}{\partial\theta}\mathbb{E}_\theta\{\mu_0(X)\} = \int_{\mathcal{X}} \mu_0(X)\frac{\partial f_\theta(X)}{\partial X}dX \\ &= \int_{\mathcal{X}} \mu_0(X)\frac{\dot{f}_\theta(X)}{f_\theta(X)}f_\theta(X)dX = \mathbb{E}\{\mu_0(X)s_\theta(X)\}, \end{aligned}$$

where $s_\theta(X) = \partial \log f_\theta(X)/\partial\theta$. Further, we can factorize the full-data likelihood function as:

$$\begin{aligned} f(W) = f(X)&\mathbb{P}(R=1 \mid X)^R\mathbb{P}(R=0 \mid X)^{1-R} \\ &\times \mathbb{P}(A=1 \mid X, R=1)^{RA}\mathbb{P}(A=0 \mid X, R=1)^{R(1-A)} \\ &\times f(T \mid X, R=1, A=1)^{RA}f(T \mid X, R=1, A=0)^{R(1-A)} \\ &\times f(T \mid X, R=0)^{1-R}, \end{aligned}$$

and the associated score function under the submodel can be decomposed as

$$\begin{aligned} s_\theta(W) = s_\theta(X) &+ \frac{R - \mathbb{P}(R=1 \mid X)}{\mathbb{P}(R=1 \mid X)\{1-\mathbb{P}(R=1 \mid X)\}} \dot{\mathbb{P}}_\theta(R=1 \mid X) \\ &+ \frac{R\{A - \pi_A(X)\}}{\pi_A(X)\{1-\pi_A(X)\}} \dot{\mathbb{P}}_\theta(A=1 \mid X, R=1) \\ &+ RAs_\theta(T \mid X, A=1, R=1) + R(1-A)s_\theta(T \mid X, R=1, A=0) \\ &+ (1-R)s_\theta(T \mid X, R=0), \end{aligned}$$

where $s_\theta(X) = \partial \log f_\theta(X)/\partial\theta$, $s_\theta(T \mid X, R=1, A=a) = \partial \log f_\theta(T \mid X, R=1, A=a)/\partial\theta$ for $a = 0, 1$, and $s_\theta(T \mid X, R=0) = \partial \log f_\theta(T \mid X, R=0)/\partial\theta$. Analogous to our definition $f_\theta(W) \mid_{\theta=0} = f(W)$, we have $s_\theta(\cdot) \mid_{\theta=0} = s(\cdot)$, which is the true score function evaluated at the true parameter under the one-dimensional submodel.

A.2.2. Full-data Efficient Influence Function

From the semiparametric theory, the orthogonal complement of the full-data nuisance tangent space $\Lambda_F^\perp$ equals to

$$\Lambda_F^\perp = H_1 \oplus H_2 \oplus H_3 \oplus H_4, \tag{2}$$

where

$$\begin{aligned} &H_1 = \{\Gamma(X) : \mathbb{E}\{\Gamma(X)\} = 0\}, \\ &H_2 = \{\{R - \mathbb{P}(R=1 \mid X)\}a(X)\}, \quad H_3 = \{R\{A - \pi_A(X)\}b(X)\}, \\ &H_4 = H_{41} \cap H_{42} = \{\Gamma(T, X, R, A) : \mathbb{E}\{\Gamma(T, X, R, A) \mid X, R, A\} = 0\} \\ &\quad \cap \left\{\Gamma(T, X, R, A) : \mathbb{E}\left[\left\{\frac{(1-R)\mathbf{1}(T>t)}{P(R=0 \mid X)} - \frac{R(1-A)\mathbf{1}(T>t)}{P(R=1, A=0 \mid X)}\right\}\Gamma(T, X, R, A) \mid X\right] = 0, t < \tau\right\}, \end{aligned}$$

for any two arbitrary square-integrable measurable functions $a(X)$ and $b(X)$. The tangent space H_{42} is induced by the conditional mean exchangeability in Assumption 3.3, where $S_0(t \mid X, R=1) = S_0(t \mid X, R=0)$ for any $t < \tau$. The EIF for $S_0(t \mid R=1)$, denoted by $\psi^F_{S_0,\text{eff}}(t, W) \in \Lambda_F^\perp$ should satisfy

$$\partial S_0(t \mid R=1)/\partial\theta \mid_{\theta=0} = \mathbb{E}\{\psi^F_{S_0,\text{eff}}(t, W)s(W)\}.$$

Based on our identification formula (1), $S_0(t \mid R=1) = \mathbb{E}\{\pi_R(X)S_0(t \mid X, R=1)\}/\mathbb{P}(R=1)$, which is a ratio for with numerator $N = \mathbb{E}\{\pi_R(X)S_0(t \mid X, R=1)\}$ and denominator $D = \mathbb{P}(R=1)$. Therefore, our strategy is to first derive the EIF for the numerator and denominator, and then combine them to have the final full-data EIF $\psi^F_{S_0,\text{eff}}(t, W)$.

Let N_θ and D_θ denote N and D being evaluated at the submodel $f_\theta(W)$. For the numerator, pathwise derivative is evaluated by the chain rule:

$$\begin{aligned} \frac{\partial N_\theta}{\partial\theta} \mid_{\theta=0} &= \mathbb{E}\{\mathbb{P}(R=1 \mid X)\mathbb{P}(T>t \mid R=1, A=0, X)s(X)\} \\ &\quad + \mathbb{E}\left\{\frac{\partial \mathbb{P}_\theta(R=1 \mid X)}{\partial\theta}\mathbb{P}(T>t \mid R=1, A=0, X)\right\} \mid_{\theta=0} \qquad (3) \\ &\quad + \mathbb{E}\left\{\mathbb{P}(R=1 \mid X)\frac{\partial \mathbb{P}_\theta(T>t \mid R=1, A=0, X)}{\partial\theta}\right\} \mid_{\theta=0}. \qquad (4) \end{aligned}$$

Next, we show the second part of the pathwise derivative (3) is

$$\frac{\partial \mathbb{P}_\theta(R=1 \mid X)}{\partial\theta} = \mathbb{E}[\{R - \mathbb{P}(R=1 \mid X)\}s(A, R \mid X)].$$

However, the pathwise derivative $\partial \mathbb{P}_\theta(T > t \mid X, R=1)/\partial\theta$ in the third part can be derived in different ways under

Assumption 3.3 with the trial data by (5) and the external controls by (6):

$$
\begin{aligned}
&\frac{\partial}{\partial\theta}\mathbb{P}_\theta(T>t \mid R=1,A=0,X)\\
&=\frac{\partial}{\partial\theta}\mathbb{E}_\theta\left\{\mathbf{1}(T>t) \mid R=1,A=0,X\right\}\\
&=\int_t^\infty \mathbf{1}(T>t)\frac{\partial}{\partial\theta}f_\theta(T \mid R=1,A=0,X)dT\\
&=\int_t^\infty \mathbf{1}(T>t)s_\theta(T \mid R=1,A=0,X)f_\theta(T \mid R=1,A=0,X)dT\\
&=\int_t^\infty \frac{R(1-A)f_\theta(T \mid X)\left\{\mathbf{1}(T>t)-S_0(t \mid X,R=1)\right\}}{\mathbb{P}(R=1,A=0 \mid X)}s_\theta(T \mid R,A,X)dT\\
&=\mathbb{E}\left[\frac{R(1-A)\left\{\mathbf{1}(T>t)-S_0(t \mid X,R=1)\right\}s_\theta(T \mid R,A,X)}{\mathbb{P}(R=1,A=0 \mid X)} \mid X\right],
\end{aligned}
\tag{5}
$$

and

$$
\begin{aligned}
&\frac{\partial}{\partial\theta}\mathbb{P}_\theta(T>t \mid R=1,A=0,X)\\
&=\frac{\partial}{\partial\theta}\mathbb{P}_\theta(T>t \mid R=0,A=0,X)\\
&=\frac{\partial}{\partial\theta}\mathbb{E}_\theta\left\{\mathbf{1}(T>t) \mid R=0,A=0,X\right\}\\
&=\int_t^\infty \mathbf{1}(T>t)\frac{\partial}{\partial\theta}f_\theta(T \mid R=0,A=0,X)dT^{(0)}\\
&=\int_t^\infty \mathbf{1}(T>t)s_\theta(T \mid R=0,A=0,X)f_\theta(T \mid R=0,A=0,X)dT\\
&=\int_t^\infty \left\{\mathbf{1}(T>t)-S_0(t \mid X,R=0)\right\}s_\theta(T \mid X,R)\frac{(1-R)f_\theta(T \mid X)}{\mathbb{P}(R=0 \mid X)}dT\\
&=\mathbb{E}\left[\frac{(1-R)\left\{\mathbf{1}(T>t)-S_0(t \mid X,R=0)\right\}s_\theta(T \mid X,R)}{\mathbb{P}(R=0 \mid X)} \mid X\right].
\end{aligned}
\tag{6}
$$

To obtain the efficient influence function of N, we need to find the proper functions C_1 and C_2 of (X,R,A) such that the third part (4) belongs to the tangent space H_4, which satisfies:

$$
\begin{aligned}
&\mathbb{E}\left(\left[C_1(1-R)q(X)\left\{\mathbf{1}(T>t)-S_0(t \mid X,R=1)\right\}+C_2R(1-A)\frac{\mathbf{1}(T>t)-S_0(t \mid X,R=1)}{1-\pi_A(X)}\right]\right.\\
&\left.\times\left\{\frac{(1-R)\mathbf{1}(T>t)}{\mathbb{P}(R=0 \mid X)}-\frac{R(1-A)\mathbf{1}(T>t)}{\mathbb{P}(R=1,A=0 \mid X)}\right\} \mid X\right)=0.
\end{aligned}
\tag{7}
$$

By simple algebra, we can show that

$$
\begin{aligned}
\frac{C_1}{C_2}&=\frac{\mathbb{E}\left[R^2(1-A)^2\{\mathbf{1}(T>t)-S_0(t \mid X,R=1)\}\mathbf{1}(T>t) \mid X\right]}{\{1-\pi_A(X)\}\mathbb{P}(R=1,A=0 \mid X)}\\
&\quad\left(\frac{\mathbb{E}\left[(1-R)^2q_R(X)\{\mathbf{1}(T>t)-S_0(t \mid X,R=1)\}\mathbf{1}(T>t) \mid X\right]}{P(R=0 \mid X)}\right)^{-1}\\
&=\frac{r(t,X)}{\{1-\pi_A(X)\}q_R(X)},
\end{aligned}
$$

where $r(t,X)=\text{var}\left\{\mathbf{1}(T>t) \mid R=1,A=0\right\}/\text{var}\left\{\mathbf{1}(T>t) \mid R=0\right\}$. Plugging the above formulas, we obtain the

EIF for the numerator N as:

$$\begin{aligned}\psi^F_{N,\text{eff}}(t,W) = & RS_0(t \mid X, R=1) \\ & + \frac{(1-R)r(t,X)q_R(X)\left\{\mathbf{1}(T>t) - S_0(t \mid X, R=1)\right\}}{D(t,X)} \\ & + \frac{R(1-A)q_R(X)\left\{\mathbf{1}(T>t) - S_0(t \mid X, R=1)\right\}}{D(t,X)}.\end{aligned}$$

where $D(t,X) = r(t,X) + \{1-\pi_A(X)\}q_R(X)$. For the denominator D_θ, we have $\partial D_\theta/\partial\theta \mid_{\theta=0} = \mathbb{E}\{R\partial\mathbb{P}_\theta(R=1)/\partial\theta\} \mid_{\theta=0}$. By Lemma S1 in Jiang et al. (2022), the full-data EIF of N/D is obtained by:

$$\begin{aligned}\psi^F_{S_0,\text{eff}}(t,W) = & \frac{\psi^F_{N,\text{eff}}(t,W) - RS_0(t \mid R=1)}{\mathbb{P}(R=1)} \\ = & \frac{R(1-A)}{\mathbb{P}(R=1)}\frac{q_R(X)\left\{\mathbf{1}(T>t) - S_0(t \mid X, R=1)\right\}}{D(t,X)} \\ & + \frac{1-R}{\mathbb{P}(R=1)}\frac{r(t,X)q_R(X)\left\{\mathbf{1}(T>t) - S_0(t \mid X, R=1)\right\}}{D(t,X)} \\ & + \frac{R}{\mathbb{P}(R=1)}\{S_0(t \mid X, R=1) - S_0(t \mid R=1)\}, \\ = & \frac{R(1-A)}{\mathbb{P}(R=1)}\frac{q_R(X)\left\{\mathbf{1}(T>t) - S_0(t \mid R=1)\right\}}{D(t,X)} \\ & + \frac{(1-R)q_R(X)}{\mathbb{P}(R=1)}\frac{r(t,X)\left\{\mathbf{1}(T>t) - S_0(t \mid R=1)\right\}}{D(t,X)} \\ & + \frac{S_0(t \mid X, R=1) - S_0(t \mid R=1)}{\mathbb{P}(R=1)}\left\{\frac{R\{A-\pi_A(X)\}q_R(X)}{D(t,X)} + \frac{r(t,X)\{R-(1-R)q_R(X)\}}{D(t,X)}\right\},\end{aligned}$$

which belongs to the tangent space $\Lambda_F^\perp$.

A.2.3. Observed-data Efficient Influence Function

In the presence of censoring, i.e., a special form of monotone coarsening, we observe the data set $V_i = (Y_i, \Delta_i, A_i, X_i, R_i = r)$ instead of W_i. Define the many-to-one linear mapping $\mathcal{K}: \Lambda_\eta^\perp \to \Lambda_F^\perp$ to be $\mathcal{K}(h) = \mathbb{E}\{h(V) \mid W\}$ for any $h \in \Lambda_\eta^\perp$, where $\Lambda_\eta^\perp$ is the orthogonal complement of the observed-data nuisance tangent space by Lemma 7.3, Tsiatis (2006). Let $\psi^F(t,W)$ be a typical element of $\Lambda_F^\perp$, by Theorem 7.2 from Tsiatis (2006), the space $\Lambda_\eta^\perp$ consists of all elements that can be written as

$$\Lambda_\eta^\perp = \mathcal{K}^{-1}(\Lambda_F^\perp) = \frac{\Delta\psi^F(t,W)}{\mathbb{P}(\Delta=1 \mid W)} + \mathcal{K}^{-1}(0) \tag{8}$$

where $\mathcal{K}^{-1}$ as the inverse operator, and $\mathcal{K}^{-1}(0)$ consists any arbitrary functions $L(t,V)$ such that $\mathbb{E}\{L(t,V) \mid W\} = 0$. The first part of (8) is motivated by the inverse censoring weighting of the complete-case estimator where the event time is observed, indicated by $\Delta = 1$. The second part of (8) is referred to as the augmentation space due to censoring. The optimal element (8) with the greatest efficiency improvement is obtained by projecting $\Delta\psi^F(t,W)/\mathbb{P}(\Delta=1 \mid W)$ to the tangent space $\mathcal{K}^{-1}(0)$. By Theorems 9.2 and 10.4, we derive the observed-data EIF for $S_0(t \mid R=1)$ under our monotone

coarsened data:

$$\psi_{S_0,\text{eff}}(t,V) = \frac{R(1-A)\Delta}{\mathbb{P}(R=1)S^C(Y \mid X, R=1)} \frac{q_R(X)\left\{\mathbf{1}(Y>t) - S_0(t \mid R=1)\right\}}{D(t,X)} \tag{9}$$

$$+ \frac{(1-R)\Delta q_R(X)}{\mathbb{P}(R=1)S^C(Y \mid X, R=0)} \frac{r(t,X)\left\{\mathbf{1}(Y>t) - S_0(t \mid R=1)\right\}}{D(t,X)} \tag{10}$$

$$+ \frac{S_0(t \mid X, R=1) - S_0(t \mid R=1)}{\mathbb{P}(R=1)} \left\{ \frac{R\{A-\pi_A(X)\}q_R(X)}{D(t,X)} + \frac{r(t,X)\{R-(1-R)q_R(X)\}}{D(t,X)} \right\} \tag{11}$$

$$+ \int_0^\infty \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=1)} \mathbb{E}\left[\frac{q_R(X)\left\{\mathbf{1}(T>t) - S_0(t \mid R=1)\right\}}{D(t,X)} \mid T>r, X \right] \tag{12}$$

$$+ \int_0^\infty \frac{(1-R)q_R(X)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=0)} \mathbb{E}\left[\frac{r(t,X)\left\{\mathbf{1}(T>t) - S_0(t \mid R=1)\right\}}{D(t,X)} \mid T>r, X \right], \tag{13}$$

where $dM_0^C(r \mid X) = dN_0^C(r) - \mathbf{1}(Y \geq r)\lambda_0^C(r \mid X)$. The last two terms (12) and (13) belong to the augmentation space $\mathcal{K}^{-1}(0)$. Next, we can simplify it to be the EIF in Theorem 3.5. Note that (12) can be expressed by

$$\begin{aligned}
&\int_0^\infty \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=1)} \mathbb{E}\left[\frac{q_R(X)\left\{\mathbf{1}(T>t) - S_0(t \mid R=1)\right\}}{D(t,X)} \mid T>r, X \right] \\
&= \int_0^\infty \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)q_R(X)}{S^C(r \mid X, R=1)D(t,X)} \mathbf{1}(r<t) \left\{ \frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)} - S_0(t \mid R=1) \right\} \\
&\quad + \int_0^\infty \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)q_R(X)}{S^C(r \mid X, R=1)D(t,X)} \mathbf{1}(r \geq t)\left\{1 - S_0(t \mid R=1)\right\} \\
&= \int_0^t \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)q_R(X)}{S^C(r \mid X, R=1)D(t,X)} \frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)} \qquad (14) \\
&\quad + \int_t^\infty \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)q_R(X)}{S^C(r \mid X, R=1)D(t,X)} \qquad (15) \\
&\quad - S_0(t \mid R=1) \int_0^\infty \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)q_R(X)}{S^C(r \mid X, R=1)D(t,X)}. \qquad (16)
\end{aligned}$$

The second term (15) equals to

$$\begin{aligned}
&\int_t^\infty \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)q_R(X)}{S^C(r \mid X, R=1)D(t,X)} \\
&= \frac{R(1-A)q_R(X)}{\mathbb{P}(R=1)D(t,X)} \int_t^\infty \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=1)} \\
&= \frac{R(1-A)q_R(X)}{\mathbb{P}(R=1)D(t,X)} \int_t^\infty \frac{dN_0^C(r) - \mathbf{1}(Y \geq r)\lambda_0^C(r \mid X)}{S^C(r \mid X, R=1)} \\
&= \frac{R(1-A)q_R(X)\mathbf{1}(Y \geq t)}{\mathbb{P}(R=1)D(t,X)} \left\{ \frac{1-\Delta}{S^C(Y \mid X, R=1)} - \int_t^Y \frac{\lambda_0^C(r \mid X)dr}{S^C(r \mid X, R=1)} \right\} \\
&= \frac{R(1-A)q_R(X)\mathbf{1}(Y \geq t)}{\mathbb{P}(R=1)D(t,X)} \left\{ \frac{1}{S^C(t \mid X, R=1)} - \frac{\Delta}{S^C(Y \mid X, R=1)} \right\},
\end{aligned}$$

where the last equality holds as we know that $\lambda_0^C(r \mid X) = -\partial \log S^C(r \mid X, R=1)/\partial r$, and $\int \lambda_0^C(r \mid X)/S^C(r \mid X, R=1) = 1/S^C(r \mid X, R=1)$. For the third term (16), we have

$$\begin{aligned}
&S_0(t \mid R=1) \int_0^\infty \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X)q_R(X)}{S^C(r \mid X, R=1)D(t,X)} \\
&= S_0(t \mid R=1) \frac{R(1-A)q_R(X)}{\mathbb{P}(R=1)D(t,X)} \int_0^\infty \frac{dN_0^C(r) - \mathbf{1}(Y \geq r)\lambda_0^C(r \mid X)}{S^C(r \mid X, R=1)} \\
&= S_0(t \mid R=1) \frac{R(1-A)q_R(X)}{\mathbb{P}(R=1)D(t,X)} \left\{ 1 - \frac{\Delta}{S^C(Y \mid X, R=1)} \right\}.
\end{aligned}$$

Plugging these formulas back with (9) and (12), we obtain

$$
\begin{aligned}
&\frac{R(1-A)\Delta}{\mathbb{P}(R=1)S^C(Y \mid X, R=1)} \frac{q_R(X)\left\{\mathbf{1}(Y>t)-S_0(t \mid R=1)\right\}}{D(t,X)} \\
&+\int_0^t \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X) q_R(X)}{S^C(r \mid X, R=1) D(t,X)} \frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)} \\
&+\frac{R(1-A) q_R(X) \mathbf{1}(Y \geq t)}{\mathbb{P}(R=1) D(t,X)}\left\{\frac{1}{S^C(t \mid X, R=1)}-\frac{\Delta}{S^C(Y \mid X, R=1)}\right\} \\
&-S_0(t \mid R=1) \frac{R(1-A) q_R(X)}{\mathbb{P}(R=1) D(t,X)}\left\{1-\frac{\Delta}{S^C(Y \mid X, R=1)}\right\} \\
&=\frac{R(1-A) q_R(X) \mathbf{1}(Y \geq t)}{\mathbb{P}(R=1) D(t,X) S^C(t \mid X, R=1)}-\frac{R(1-A) q_R(X) S_0(t \mid R=1)}{\mathbb{P}(R=1) D(t,X)} \\
&+\int_0^t \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X) q_R(X)}{S^C(r \mid X, R=1) D(t,X)} \frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)}.
\end{aligned}
$$

Analogous simplification applies to the combination of (10) and (13), which leads to

$$
\begin{aligned}
&\frac{(1-R) q_R(X) r(t,X) \mathbf{1}(Y>t)}{\mathbb{P}(R=1) D(t,X) S^C(t \mid X, R=0)}-\frac{(1-R) q_R(X) r(t,X) S_0(t \mid R=1)}{\mathbb{P}(R=1) D(t,X)} \\
&+\int_0^t \frac{(1-R)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X) q_R(X) r(t,X)}{S^C(r \mid X, R=1) D(t,X)} \frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)}.
\end{aligned}
$$

So the final observed-data EIF now becomes:

$$
\begin{aligned}
\psi_{S_0,\mathrm{eff}}(t,V) = &\frac{R(1-A) q_R(X) \mathbf{1}(Y \geq t)}{\mathbb{P}(R=1) D(t,X) S^C(t \mid X, R=1)} \\
&+\int_0^t \frac{R(1-A)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X) q_R(X)}{S^C(r \mid X, R=1) D(t,X)} \frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)} \\
&+\frac{(1-R) q_R(X) r(t,X) \mathbf{1}(Y>t)}{\mathbb{P}(R=1) D(t,X) S^C(t \mid X, R=0)} \\
&+\int_0^t \frac{(1-R)}{\mathbb{P}(R=1)} \frac{dM_0^C(r \mid X) q_R(X) r(t,X)}{S^C(r \mid X, R=1) D(t,X)} \frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)} \\
&+\frac{S_0(t \mid X, R=1)}{\mathbb{P}(R=1)}\left\{\frac{R\{A-\pi_A(X)\} q_R(X)}{D(t,X)}+\frac{r(t,X)\{R-(1-R) q_R(X)\}}{D(t,X)}\right\} \\
&-\frac{R S_0(t \mid R=1)}{\mathbb{P}(R=1)},
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{R\{A-\pi_A(X)\} q_R(X)+R(1-A) q_R(X)}{\mathbb{P}(R=1) D(t,X)} &= \frac{R\{1-\pi_A(X)\} q_R(X)}{\mathbb{P}(R=1) D(t,X)}, \\
\frac{r(t,X)\{R-(1-R) q_R(X)\}+(1-R) q_R(X) r(t,X)}{\mathbb{P}(R=1) D(t,X)} &= \frac{R r(t,X)}{\mathbb{P}(R=1) D(t,X)}.
\end{aligned}
$$

A.3. Proof of Theorem 3.6

Assumption A.1. Assume (a) $\|\widehat{\pi}_R(X)-\pi_R(X)\|_{L_2}=o_{\mathbb{P}}(1)$, $\|\widehat{\pi}_A(X)-\pi_A(X)\|_{L_2}=o_{\mathbb{P}}(1)$, $\|\widehat{S}_a(t \mid X, R=1)-S_a(t \mid X, R=1)\|_{L_2}=o_{\mathbb{P}}(1)$, and $\|\widehat{S}^C(t \mid X, R)-S^C(t \mid X, R)\|_{L_2}=o_{\mathbb{P}}(1)$; $0<c_1 \leq \pi_R(X), \pi_A(X), S^C(t \mid X, R), S_0(t \mid X, R) \leq c_2<1$ and their estimated counterparts are bounded away from 0 and 1 for some constants c_1 and c_2; (c) $\|\widehat{S}_a(t \mid X, R=1)-S_a(t \mid X, R=1)\|_{L_2} \cdot \|\widehat{\pi}_A(X)-\pi_A(X)\|_{L_2}=o_{\mathbb{P}}(N^{-1/2})$, $\|\widehat{S}_a(t \mid X, R=1)-S_a(t \mid X, R=1)\|_{L_2} \cdot \|\widehat{\pi}_R(X)-\pi_R(X)\|_{L_2}=o_{\mathbb{P}}(N^{-1/2})$, and $\|\widehat{S}_a(t \mid X, R=1)-S_a(t \mid X, R=1)\|_{L_2} \cdot \max_{r<t}\|\widehat{\lambda}_0^C(r \mid X, R)-\lambda_0^C(r \mid X, R)\|_{L_2}=o_{\mathbb{P}}(N^{-1/2})$ for $a \in \{0,1\}$.

Assumption A.1 is analogous to those for double machine learning estimation for average treatment effects (Kennedy, 2016). To investigate the asymptotic properties of $\widehat{\theta}_\tau$, we need to understand the components that constitute $\widehat{\theta}_\tau - \theta_\tau$. Let $\mathbb{P}_N$ be the empirical measure, we have

$$\begin{aligned}
\widehat{\theta}_\tau - \theta_\tau &= \int \widehat{\phi}_{\theta_\tau,\text{eff}}(V) d\mathbb{P}_N - \int \phi_{\theta_\tau,\text{eff}}(V) d\mathbb{P} \\
&= \int \phi_{\theta_\tau,\text{eff}}(V) d\mathbb{P}_N \qquad (17) \\
&+ \int \{\widehat{\phi}_{\theta_\tau,\text{eff}}(V) - \phi_{\theta_\tau,\text{eff}}(V)\} d\mathbb{P} \qquad (18) \\
&+ \int \{\widehat{\phi}_{\theta_\tau,\text{eff}}(V) - \phi_{\theta_\tau,\text{eff}}(V)\} d(\mathbb{P}_N - \mathbb{P}), \qquad (19)
\end{aligned}$$

where the first term (17) is asymptotically normal by the central limit theorem. The third term (19) is the empirical process which is negligible if $\phi_{\theta_\tau,\text{eff}}(V)$ belongs to Donsker classes or the cross-fitting technique is employed. Under the assumptions in Theorem 3.6, and the regularity conditions A.1, we have

$$\widehat{\theta}_\tau = \theta_\tau + \frac{1}{N} \sum_{i \in \mathcal{R} \cup \mathcal{E}} \phi_{\theta_\tau,\text{eff}}(V) + \|\text{Rem}(\widehat{\mathbb{P}}, \mathbb{P})\|_{L_2} + o_{\mathbb{P}}(N^{-1/2}),$$

where $\widehat{\mathbb{P}}$ is the estimated counterpart of the true distribution $\mathbb{P}$, and

$$\|\text{Rem}(\widehat{\mathbb{P}}, \mathbb{P})\|_{L_2}^2 = \int \{\widehat{\phi}_{\theta_\tau,\text{eff}}(V) - \phi_{\theta_\tau,\text{eff}}(V)\}^2 d\mathbb{P}$$

is second-order remainder term (18). By the definition that $\phi_{\theta_\tau,\text{eff}}(V) = \int_0^\tau \{\phi_{S_1,\text{eff}}(t,V) - \phi_{S_0,\text{eff}}(t,V)\} dt$, we will first characterize the remainder term induced by $\phi_{S_0,\text{eff}}(V)$, the remainder term induced by $\phi_{S_1,\text{eff}}(V)$ follows in similar techniques. Combining parts (9), (10), (12), and (13) in $\psi_{S_0,\text{eff}}(V)$, we can show

$$\begin{aligned}
&m_1(t,V) + \left\{ \frac{\Delta}{S^C(Y \mid X, R=1)} - 1 \right\} m_1(t,V) + m_2(t,V) + \left\{ \frac{\Delta}{S^C(Y \mid X, R=0)} - 1 \right\} m_2(t,V) \\
&+ \int_0^\infty \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=1)} \mathbb{E}\left\{ m_1(t,V) \mid T > r, X \right\} + \int_0^\infty \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=0)} \mathbb{E}\left\{ m_2(t,V) \mid T > r, X \right\} \\
&= m_1(t,V) + m_2(t,V) \\
&- \int_0^\infty \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=1)} \left[m_1(t,V) - \mathbb{E}\left\{ m_1(t,V) \mid T > r, X \right\} \right] \\
&- \int_0^\infty \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=0)} \left[m_2(t,V) - \mathbb{E}\left\{ m_2(t,V) \mid T > r, X \right\} \right],
\end{aligned}$$

where

$$\begin{aligned}
m_1(t,V) &= \frac{R(1-A) q_R(X) \left\{ \mathbf{1}(T > t) - S_0(t \mid R=1) \right\}}{\mathbb{P}(R=1) D(t,X)}, \\
m_2(t,V) &= \frac{(1-R) q_R(X) r(t,X) \left\{ \mathbf{1}(T > t) - S_0(t \mid R=1) \right\}}{\mathbb{P}(R=1) D(t,X)}, \\
1 - \frac{\Delta}{S^C(Y \mid X, R=r)} &= \int_0^\infty \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=1)},
\end{aligned}$$

by Lemma 10.4 from Tsiatis (2006); Zeng & Lin (2007). Combine them with (11), the second-order remainder term

$\int\{\widehat{\phi}_{S_0,\text{eff}}(V)-\phi_{S_0,\text{eff}}(V)\}d\mathbb{P}$ becomes

$$\mathbb{P}\left(\frac{\widehat{S}_0(t\mid X,R=1)-S_0(t\mid X,R=1)}{\mathbb{P}(R=1)\widehat{D}(t,X)\{1-\widehat{\pi}_R(X)\}}\left[\pi_R(X)\{\pi_A(X)-\widehat{\pi}_A(X)\}\widehat{\pi}_R(X)+\widehat{r}(t,X)\{\pi_R(X)-\widehat{\pi}_R(X)\}\right]\right)$$
$$+\mathbb{P}\left(\int_0^\infty\frac{d\widehat{M}_0^C(r\mid X)}{\widehat{S}^C(r\mid X,R=1)}\left[\widehat{m}_1(t,V)-\widehat{\mathbb{E}}\left\{\widehat{m}_1(t,V)\mid T>r,X\right\}\right]\right) \tag{20}$$
$$+\mathbb{P}\left(\int_0^\infty\frac{d\widehat{M}_0^C(r\mid X)}{\widehat{S}^C(r\mid X,R=0)}\left[\widehat{m}_2(t,V)-\widehat{\mathbb{E}}\left\{\widehat{m}_2(t,V)\mid T>r,X\right\}\right]\right), \tag{21}$$

since $\mathbb{P}\{m_1(t,V)\}=\mathbb{P}\{m_2(t,V)\}=0$, where

$$\widehat{m}_1(t,V)=\frac{R(1-A)\widehat{q}_R(X)\left\{\mathbf{1}(T>t)-S_0(t\mid R=1)\right\}}{\mathbb{P}(R=1)\widehat{D}(t,X)},$$
$$\widehat{m}_2(t,V)=\frac{(1-R)\widehat{q}_R(X)\widehat{r}(t,X)\left\{\mathbf{1}(T>t)-S_0(t\mid R=1)\right\}}{\mathbb{P}(R=1)\widehat{D}(t,X)}.$$

Next, we compute the expectations of $d\widehat{M}_0^C(r\mid X)$ and $\mathbb{E}\left\{\mathbf{1}(T>t)\mid T>r,X\right\}$ under the true distribution $\mathbb{P}$ by conditioning on $\{T>r,C>r,X\}$:

$$\mathbb{P}\{d\widehat{M}_0^C(r\mid X)\mid T>r,C>r,X\}=\mathbf{1}(Y>r)\{\lambda_0^C(r\mid X)-\widehat{\lambda}_0^C(r\mid X)\}dr,$$
$$\mathbb{P}[\mathbb{E}\{\mathbf{1}(T>t)\mid T>r,C>r,X\}]=\mathbb{P}[\mathbb{E}\{\mathbf{1}(T>t)\mid T>r,X\}],$$

where the first equality holds since $\mathbb{P}\{\mathbf{1}(Y>r,\Delta=0,A=0)\mid T>r,C>r,X\}=\mathbf{1}(Y>r)\lambda_0^C(r\mid X)$, and the second equality holds under Assumption 3.2. By conditioning on $\{T>r,C>r,X\}$ for $r<t$, we show that the expectation (20) is equal to

$$\mathbb{P}\left(\int_0^\infty\frac{\mathbf{1}(Y>r)\pi_R(X)\{1-\pi_A(X)\}\widehat{q}_R(X)}{\mathbb{P}(R=1)\widehat{D}(t,X)\widehat{S}^C(r\mid X,R=1)}\{\lambda_0^C(r\mid X,R=1)-\widehat{\lambda}_0^C(r\mid X,R=1)\}dr\right.$$
$$\left.\times\left[\mathbb{E}\left\{\mathbf{1}(T>t)\mid T>r,X\right\}-\widehat{\mathbb{E}}\left\{\mathbf{1}(T>t)\mid T>r,X\right\}\right]\right).$$

Similar iterated expectation can be applied to (21). Under the regularity conditions A.1, we collect all the terms above and use the Cauchy-Schwarz inequality:

$$\int|\widehat{\phi}_{S_0,\text{eff}}(V)-\phi_{S_0,\text{eff}}(V)|d\mathbb{P}$$
$$\lesssim\left\{\|\widehat{\pi}_A(X)-\pi_A(X)\|_{L_2}+\|\widehat{\pi}_R(X)-\pi_R(X)\|_{L_2}+\max_{r<t}\|\widehat{\lambda}_0^C(r\mid X,R=1)-\lambda_0^C(r\mid X,R=1)\|_{L_2}\right\}$$
$$\times\|\widehat{S}_0(t\mid X,R=1)-S_0(t\mid X,R=1)\|_{L_2},$$

where $\lesssim$ indicates that the inequality holds up to a multiplicative constant. Thus, we can show the second-order remainder term is bounded:

$$\|\text{Rem}(\widehat{\mathbb{P}},\mathbb{P})\|_{L_2}^2=\int\{\widehat{\phi}_{\theta_\tau,\text{eff}}(V)-\phi_{\theta_\tau,\text{eff}}(V)\}^2d\mathbb{P}$$
$$\lesssim\left\{\|\widehat{S}_0(t\mid X,R=1)-S_0(t\mid X,R=1)\|_{L_2}+\|\widehat{S}_1(t\mid X,R=1)-S_1(t\mid X,R=1)\|_{L_2}\right\}$$
$$\times\left\{\|\widehat{\pi}_A(X)-\pi_A(X)\|_{L_2}+\|\widehat{\pi}_R(X)-\pi_R(X)\|_{L_2}+\max_{r<t}\|\widehat{\lambda}_0^C(r\mid X,R)-\lambda_0^C(r\mid X,R)\|_{L_2}\right\},$$

which is negligible under the regularity conditions A.1. Thus, we have $\widehat{\theta}_\tau=\theta_\tau+\sum_{i\in\mathcal{R}\cup\mathcal{E}}\phi_{\theta_\tau,\text{eff}}(V)/N+o_{\mathbb{P}}(N^{-1/2})$, which achieves semiparametric efficiency $\mathbb{V}_\tau=\mathbb{E}\{\psi_{\theta_\tau,\text{eff}}^2(V)\}$.

A.4. Proof of Lemma 3.7

Follow the similar technique in the proof of Theorem 3.5, the observed-data EIF-motivated estimator for $S_0(t \mid X, R=1)$ with the trial data only

$$\begin{aligned}\kappa_0(t, V \mid R=1) = S_0(t \mid X, R=1) \\ + \frac{R(1-A)\Delta\{\mathbf{1}(Y>t) - S_0(t \mid X, R=1)\}}{\pi_R(X)\{1-\pi_A(X)\}S^C(Y \mid X, R=1)} \\ + \int_0^\infty \frac{R(1-A)dM_0^C(r \mid X, R=1)}{\pi_R(X)\{1-\pi_A(X)\}S^C(r \mid X, R=1)}\mathbb{E}\{\mathbf{1}(T>t) - S_0(t \mid X, R=1) \mid T>r\}.\end{aligned}$$

By simple algebra, we obtain the simplified formula in the main paper

$$\begin{aligned}\kappa_0(t, V \mid R=1) &= S_0(t \mid X, R=1) \\ &\quad + \frac{R(1-A)\Delta\{\mathbf{1}(Y>t) - S_0(t \mid X)\}}{\pi_R(X)\{1-\pi_A(X)\}S^C(Y \mid X, R=1)} \\ &\quad + \int_0^t \frac{R(1-A)dM_0^C(r \mid X, R=1)}{\pi_R(X)\{1-\pi_A(X)\}S^C(r \mid X, R=1)}\frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)} \\ &\quad + \frac{R(1-A)\mathbf{1}(Y>t)}{\pi_R(X)\{1-\pi_A(X)\}}\left\{\frac{1}{S^C(t \mid X, R=1)} - \frac{\Delta}{S^C(Y \mid X, R=1)}\right\} \\ &\quad - \frac{R(1-A)S_0(t \mid X)}{\pi_R(X)\{1-\pi_A(X)\}}\left\{1 - \frac{\Delta}{S^C(Y \mid X, R=1)}\right\} \\ &= S_0(t \mid X, R=1) \\ &\quad + \frac{R(1-A)}{\pi_R(X)\{1-\pi_A(X)\}}\left\{\frac{\mathbf{1}(Y>t)}{S^C(t \mid X_i, R=1)} - S_0(t \mid X, R=1)\right\} \\ &\quad + \int_0^t \frac{R(1-A)dM_0^C(r \mid X, R=1)}{\pi_R(X)\{1-\pi_A(X)\}S^C(r \mid X, R=1)}\frac{S_0(t \mid X, R=1)}{S_0(r \mid X, R=1)},\end{aligned}$$

where

$$\begin{aligned}\int_t^\infty \frac{dM_0^C(r \mid X, R=1)}{S^C(r \mid X, R=1)} &= \mathbf{1}(Y>t)\left\{\frac{1}{S^C(t \mid X, R=1)} - \frac{\Delta}{S^C(Y \mid X, R=1)}\right\}, \\ \int_0^\infty \frac{dM_0^C(r \mid X, R=1)}{S^C(r \mid X, R=1)} &= 1 - \frac{\Delta}{S^C(Y \mid X, R=1)},\end{aligned}$$

by our arguments in Theorem 3.6. Following another representation of $\kappa_0(t, V \mid R=1)$, we can show that

$$\begin{aligned}\kappa_0(t, V \mid R=1) &= S_0(t \mid X, R=1) \\ &\quad + \frac{R(1-A)\{\mathbf{1}(T>t) - S_0(t \mid X)\}}{\pi_R(X)\{1-\pi_A(X)\}} \\ &\quad + \left\{\frac{\Delta}{S^C(Y \mid X, R=1)} - 1\right\}\frac{R(1-A)\{\mathbf{1}(T>t) - S_0(t \mid X, R=1)\}}{\pi_R(X)\{1-\pi_A(X)\}} \\ &\quad + \int_0^\infty \frac{R(1-A)dM_0^C(r \mid X, R)}{\pi_R(X)\{1-\pi_A(X)\}S^C(r \mid X, R=1)}\mathbb{E}\{\mathbf{1}(T>t) - S_0(t \mid X) \mid T>r, R=1\} \\ &= S_0(t \mid X, R=1) + \frac{R(1-A)\{\mathbf{1}(T>t) - S_0(t \mid X, R=1)\}}{\pi_R(X)\{1-\pi_A(X)\}} \\ &\quad + \int_0^\infty \frac{R(1-A)dM_0^C(r \mid X, R)}{\pi_R(X)\{1-\pi_A(X)\}S^C(r \mid X, R=1)}\left[\mathbb{E}\{\mathbf{1}(T>t) \mid T>r, R=1\} - \mathbf{1}(T>t)\right],\end{aligned}$$

where $1-\Delta/S^C(Y \mid X, R=r) = \int_0^\infty dM_0^C(r \mid X)/S^C(r \mid X, R=1)$. Let $\kappa_0^*(t, V \mid R=1)$ be the probability limit of $\widehat{\kappa}_0(t, V \mid R=1)$, we can show that

$$
\begin{aligned}
&\mathbb{E}\{\kappa_0^*(t, V \mid R=1) - S_0(t \mid X, R=1) \mid X\} \\
\leq &\{S_0^*(t \mid X, R=1) - S_0(t \mid X, R=1)\} \cdot \left[1 - \frac{\pi_R(X)\{1-\pi_A(X)\}}{\pi_R^*(X)\{1-\pi_A^*(X)\}}\right] \\
&+ \frac{\pi_R(X)\{1-\pi_A(X)\}}{\pi_R^*(X)\{1-\pi_A^*(X)\}} \mathbb{E}\left(\int_0^\infty \frac{\mathbb{E}\{dM_0^{C*}(r \mid X, R=1) \mid C>r, T>r\}}{S^{C*}(r \mid X, R=1)}\right. \\
&\left.\times [\mathbb{E}^*\{\mathbf{1}(T>t) \mid T>r\} - \mathbb{E}\{\mathbf{1}(T>t) \mid T>r\}]\right).
\end{aligned}
$$

Thus, we establish the bound for $\kappa_0^*(t, V \mid R=1)$ by

$$
\begin{aligned}
&\|\mathbb{E}\{\kappa_0^*(t, V \mid R=1) - S_0(t \mid X, R=1) \mid X\}\|_{L_2} \\
&\lesssim \|\pi_R^*(X) - \pi_R(X)\|_{L_2} \cdot \|S_0^*(t \mid X, R=1) - S_0(t \mid X, R=1)\|_{L_2} \\
&+ \|\pi_A^*(X) - \pi_A(X)\|_{L_2} \cdot \|S_0^*(t \mid X, R=1) - S_0(t \mid X, R=1)\|_{L_2} \\
&+ \|\lambda_0^{C*}(t \mid X, R=1) - \lambda_0^C(t \mid X, R=1)\| \cdot \|S_0^*(t \mid X, R=1) - S_0(t \mid X, R=1)\|_{L_2}.
\end{aligned}
$$

Similarly, we can establish the bound for $\kappa_0^*(t, V \mid R=0)$ as

$$
\begin{aligned}
&\|\mathbb{E}\{\kappa_0^*(t, V \mid R=0) - S_0(t \mid X, R=0) \mid X\}\|_{L_2} \\
&\lesssim \|\pi_R^*(X) - \pi_R(X)\|_{L_2} \cdot \|S_0^*(t \mid X, R=0) - S_0(t \mid X, R=0)\|_{L_2} \\
&+ \|\lambda_0^{C*}(t \mid X, R=0) - \lambda_0^C(t \mid X, R=0)\|_{L_2} \cdot \|S_0^*(t \mid X, R=0) - S_0(t \mid X, R=0)\|_{L_2}.
\end{aligned}
$$

Putting the bounds for $\kappa_0^*(t, V \mid R=1)$ and $\kappa_0^*(t, V \mid R=0)$ together, we obtain the desired result:

$$
\begin{aligned}
&\|\mathbb{E}\{\xi(V) - b_0 \mid X\}\|_{L_2} \\
&\lesssim \sum_{r=0}^{1} \|\pi_R^*(X) - \pi_R(X)\|_{L_2} \cdot \|S_0^*(t \mid X, R=r) - S_0(t \mid X, R=r)\|_{L_2} \\
&+ \sum_{r=0}^{1} \|S^{C*}(t \mid X, R=r) - S^C(t \mid X, R=r)\|_{L_2} \cdot \|S_0^*(t \mid X, R=r) - S_0(t \mid X, R=r)\|_{L_2} \\
&+ \|\pi_A^*(X) - \pi_A(X)\|_{L_2} \cdot \|S_0^*(t \mid X, R=1) - S_0(t \mid X, R=1)\|_{L_2},
\end{aligned}
$$

which completes the proof of Lemma 3.7.

A.5. Proof of Theorem 3.8

Under the condition that there exists a subset $\mathcal{A}$of the external controls such that $S(t \mid X_i, R=1) = S(t \mid X_i, R=0)$ for any time t and subject $i \in \mathcal{A}$, the tangent space H_4 in (2) is modified for the updated restricted moment condition

$$
\begin{aligned}
H_4^* = &\{\Gamma(T, X, R, A) : \mathbb{E}\{\Gamma(T, X, R, A) \mid X, R, A\} = 0\} \\
&\cap \left\{\Gamma(Y, X, R, A) : \mathbb{E}\left[\left\{\frac{(1-R)\mathbf{1}(b=0)\mathbf{1}(T>t)}{\mathbb{P}(R=0, b=0 \mid X)} - \frac{R(1-A)\mathbf{1}(T>t)}{\mathbb{P}(R=1, A=0 \mid X)}\right\}\Gamma(Y, X, R, A) \mid X\right] = 0, t<\tau\right\}.
\end{aligned}
$$

Similarly, we find the proper functions C_1^* and C_2^* as in (7) to obtain the full-data EIF for $S_0(t \mid R=1)$:

$$\begin{aligned}
\psi^{F,\mathcal{A}}_{S_0,\text{eff}}(t,W) &= \frac{R(1-A)}{\mathbb{P}(R=1)}\frac{q_R(X)\{\mathbf{1}(T>t)-S_0(t\mid X,R=1)\}}{D_{b_0}(t,X)} \\
&+ \frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)}\frac{r(t,X)q_R(X)\{\mathbf{1}(T>t)-S_0(t\mid X,R=1)\}}{D_{b_0}(t,X)} \\
&+ \frac{R}{\mathbb{P}(R=1)}\{S_0(t\mid X,R=1)-S_0(t\mid R=1)\}, \\
&= \frac{R(1-A)}{\mathbb{P}(R=1)}\frac{q_R(X)\{\mathbf{1}(T>t)-S_0(t\mid R=1)\}}{D_{b_0}(t,X)} \\
&+ \frac{(1-R)\mathbf{1}(b_0=0)q_R(X)}{\mathbb{P}(R=1)}\frac{r(t,X)\{\mathbf{1}(T>t)-S_0(t\mid R=1)\}}{D_{b_0}(t,X)} \\
&+ \frac{R\{A-\pi_A(X)\}q_R(X)}{\mathbb{P}(R=1)D_{b_0}(t,X)}\{S_0(t\mid X,R=1)-S_0(t\mid R=1)\} \\
&+ \frac{r(t,X)\{R\mathbb{P}(b=0\mid X,R=0)-(1-R)\mathbf{1}(b_0=0)q_R(X)\}}{\mathbb{P}(R=1)D_{b_0}(t,X)}\{S_0(t\mid X,R=1)-S_0(t\mid R=1)\},
\end{aligned}$$

where $D_{b_0}(t,X) = r(t,X)P(b=0\mid X,R=0)+\{1-\pi_A(X)\}q_R(X)$. By finding the optimal element of the augmentation space with some algebra similar to Section A.2.3, we obtain the modified observed-data EIF under the updated restricted moment condition

$$\begin{aligned}
\psi^{\mathcal{A}}_{S_0,\text{eff}}(t,V) &= \frac{R(1-A)}{\mathbb{P}(R=1)}\frac{q_R(X)\mathbf{1}(Y>t)}{S^C(t\mid X,R=1)D_{b_0}(t,X)} \\
&+ \frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)}\frac{q_R(X)r(t,X)\mathbf{1}(Y>t)}{S^C(t\mid X,R=0)D_{b_0}(t,X)} \\
&+ \int_0^t \frac{R(1-A)}{\mathbb{P}(R=1)}\frac{q_R(X)dM_0^C(r\mid X)}{S^C(r\mid X,R=1)D_{b_0}(t,X)}\frac{S_0(t\mid X)}{S_0(r\mid X)} \\
&+ \int_0^t \frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)}\frac{q_R(X)r(t,X)dM_0^C(r\mid X)}{S^C(r\mid X,R=0)D_{b_0}(t,X)}\frac{S_0(t\mid X)}{S_0(r\mid X)} \\
&+ \frac{Rq_R(X)\{A-\pi_A(X)\}S_0(t\mid X)}{\mathbb{P}(R=1)D_{b_0}(t,X)} \\
&+ \frac{r(t,X)\{R\mathbb{P}(b=0\mid X,R=0)-(1-R)\mathbf{1}(i\in\mathcal{A})q_R(X)\}S_0(t\mid X)}{\mathbb{P}(R=1)D_{b_0}(t,X)} - \frac{RS_0(t\mid R=1)}{\mathbb{P}(R=1)} \\
&= \phi^{\mathcal{A}}_{S_0,\text{eff}}(t,V) - \frac{RS_0(t\mid R=1)}{\mathbb{P}(R=1)},
\end{aligned}$$

which belongs to the orthogonal complement of the updated observed-data nuisance tangent space $\Lambda_\eta^{*\perp}$. From our previous arguments, we know that the trial-only efficient influence function is $\psi^{\text{rct}}_{S_0,\text{eff}}(t,V)$:

$$\begin{aligned}
\psi^{\text{rct}}_{S_0,\text{eff}}(t,V) &= \frac{R(1-A)\mathbf{1}(Y>t)}{\mathbb{P}(R=1)\{1-\pi_A(X)\}S^C(t\mid X,R=1)} \\
&+ \int_0^t \frac{R(1-A)dM_0^C(r\mid X)}{\mathbb{P}(R=1)\{1-\pi_A(X)\}S^C(r\mid X,R=1)}\frac{S_0(t\mid X)}{S_0(r\mid X)} \\
&+ \frac{R\{A-\pi_A(X)\}S_0(t\mid X,R=1)}{\mathbb{P}(R=1)\{1-\pi_A(X)\}} - \frac{RS_0(t\mid R=1)}{\mathbb{P}(R=1)},
\end{aligned}$$

which leads to the EIF of θ_τ by $\psi^{\text{rct}}_{\theta_\tau,\text{eff}}(V) = \int_0^\tau\{\psi_{S_1,\text{eff}}(t,V)-\psi^{\text{rct}}_{S_0,\text{eff}}(t,V)\}dt$ with the asymptotic variance $\mathbb{V}_\tau^{\text{aipw}}$. Compare the asymptotic variance $\mathbb{V}_\tau^{\text{adapt}}$ to $\mathbb{V}_\tau^{\text{aipw}}$, we have

$$\begin{aligned}
\mathbb{V}_\tau^{\text{aipw}} - \mathbb{V}_\tau^{\text{adapt}} &= \mathbb{E}\{\psi^{\text{rct}}_{\theta_\tau,\text{eff}}(V)\}^2 - \mathbb{E}\{\psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(V)\}^2 \\
&= \mathbb{E}\{\psi^{\text{rct}}_{\theta_\tau,\text{eff}}(V) - \psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(V)\}^2 \\
&+ 2\mathbb{E}[\psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(t,V)\{\psi^{\text{rct}}_{\theta_\tau,\text{eff}}(V) - \psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(V)\}].
\end{aligned} \tag{22}$$

By directional derivative, we have

$$\dot{\theta}_\tau = \mathbb{E}\{\psi^{\text{rct}}_{\theta_\tau,\text{eff}}(V)s(V)\} = \mathbb{E}\{\psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(V)s(V)\},$$

where $s(V)$ is the score function for the observed data, and therefore $\mathbb{E}[\{\psi^{\text{rct}}_{\theta_\tau,\text{eff}}(V) - \psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(V)\}s(V)] = 0$, implying $\psi^{\text{rct}}_{\theta_\tau,\text{eff}}(V) - \psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(V)$ belongs to the updated observed-data nuisance tangent space Λ^*_η. Note that $\psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(t,V) \in \Lambda^{*\perp}_\eta$, we have (22) equals to zero, and

$$\mathbb{V}^{\text{aipw}}_\tau - \mathbb{V}^{\text{adapt}}_\tau = \mathbb{E}\{\psi^{\text{rct}}_{\theta_\tau,\text{eff}}(V) - \psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(V)\}^2,$$

where $\psi^{\text{rct}}_{\theta_\tau,\text{eff}} - \psi^{\mathcal{A}}_{\theta_\tau,\text{eff}}(V) = \int_0^\tau \psi^{\text{rct}}_{S_0,\text{eff}}(t,V)dt - \int_0^\tau \psi^{\mathcal{A}}_{S_0,\text{eff}}(t,V)dt$. Next, we can show

$$\begin{aligned}
&\psi^{\text{rct}}_{S_0,\text{eff}}(t,V) - \psi^{\mathcal{A}}_{S_0,\text{eff}}(t,V)\\
&= \frac{R(1-A)}{\mathbb{P}(R=1)S^C(t \mid X, R=1)}\left\{\frac{1}{1-\pi_A(X)} - \frac{q_R(X)}{D_{b_0}(t,X)}\right\}\mathbf{1}(Y>t)\\
&- \int_0^t \frac{R(1-A)dM^C_0(r \mid X)}{\mathbb{P}(R=1)S^C(t \mid X, R=1)}\frac{S_0(t \mid X)}{S_0(r \mid X)}\left\{\frac{q_R(X)}{D_{b_0}(t,X)} - \frac{1}{1-\pi_A(X)}\right\}\\
&- \frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)}\frac{q_R(X)r(t,X)\mathbf{1}(Y>t)}{S^C(t \mid X, R=0)D_{b_0}(t,X)}\\
&- \int_0^t \frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)}\frac{q_R(X)r(t,X)dM^C_0(r \mid X)}{S^C(r \mid X, R=0)D_{b_0}(t,X)}\frac{S_0(t \mid X)}{S_0(r \mid X)}\\
&+ \frac{R\{A-\pi_A(X)\}S_0(t \mid X, R=1)}{\mathbb{P}(R=1)\{1-\pi_A(X)\}} - \frac{Rq_R(X)\{A-\pi_A(X)\}S_0(t \mid X, R=1)}{\mathbb{P}(R=1)D_{b_0}(t,X)}\\
&- \frac{r(t,X)\{R\mathbb{P}(b=0 \mid X, R=0) - (1-R)\mathbf{1}(b_0=0)q_R(X)\}S_0(t \mid X, R=1)}{\mathbb{P}(R=1)D_{b_0}(t,X)}.
\end{aligned}$$

By some algebra, we have

$$\frac{R\{A-\pi_A(X)\}S_0(t \mid X, R=1)}{\mathbb{P}(R=1)\{1-\pi_A(X)\}} = \frac{R}{\mathbb{P}(R=1)}\left\{1 - \frac{1-A}{1-\pi_A(X)}\right\}S_0(t \mid X, R=1),$$

and

$$\begin{aligned}
&\frac{Rq_R(X)\{A-\pi_A(X)\}S_0(t \mid X, R=1)}{\mathbb{P}(R=1)D_{b_0}(t,X)}\\
&+ \frac{r(t,X)\{R\mathbb{P}(b=0 \mid X, R=0) - (1-R)\mathbf{1}(b_0=0)q_R(X)\}S_0(t \mid X, R=1)}{\mathbb{P}(R=1)D_{b_0}(t,X)}\\
&\frac{1}{\mathbb{P}(R=1)}\left\{R - \frac{R(1-A)q_R(X)}{\mathbb{P}(R=1)D_{b_0}(t,X)} - \frac{r(t,X)(1-R)\mathbf{1}(b_0=0)q_R(X)}{\mathbb{P}(R=1)D_{b_0}(t,X)}\right\}S_0(t \mid X, R=1).
\end{aligned}$$

Plugging these terms back to $\psi^{\text{rct}}_{S_0,\text{eff}}(t,V) - \psi^{\mathcal{A}}_{S_0,\text{eff}}(t,V)$, we have

$$
\begin{aligned}
&\psi^{\text{rct}}_{S_0,\text{eff}}(t,V) - \psi^{\mathcal{A}}_{S_0,\text{eff}}(t,V) \\
&= \frac{R(1-A)}{\mathbb{P}(R=1)S^C(t \mid X, R=1)} \left\{ \frac{1}{1-\pi_A(X)} - \frac{q_R(X)}{D_{b_0}(t,X)} \right\} \mathbf{1}(Y>t) && (23)\\
&- \int_0^t \frac{R(1-A)dM_0^C(r \mid X)}{\mathbb{P}(R=1)S^C(t \mid X, R=1)} \frac{S_0(t \mid X)}{S_0(r \mid X)} \left\{ \frac{q_R(X)}{D_{b_0}(t,X)} - \frac{1}{1-\pi_A(X)} \right\} && (24)\\
&- \frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)} \frac{q_R(X)r(t,X)\mathbf{1}(Y>t)}{S^C(t \mid X, R=0)D_{b_0}(t,X)} && (25)\\
&- \int_0^t \frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)} \frac{q_R(X)r(t,X)dM_0^C(r \mid X)}{S^C(r \mid X, R=0)D_{b_0}(t,X)} \frac{S_0(t \mid X)}{S_0(r \mid X)} && (26)\\
&- \frac{R(1-A)}{\mathbb{P}(R=1)} \left\{ \frac{1}{1-\pi_A(X)} - \frac{q_R(X)}{D_{b_0}(t,X)} \right\} S_0(t \mid X) && (27)\\
&+ \frac{r(t,X)(1-R)\mathbf{1}(b_0=0)q_R(X)}{\mathbb{P}(R=1)D_{b_0}(t,X)} S_0(t \mid X). && (28)
\end{aligned}
$$

Combining (23), (24), and (27) gives us

$$
\begin{aligned}
&\frac{R(1-A)}{\mathbb{P}(R=1)S^C(t \mid X, R=1)} \left\{ \frac{1}{1-\pi_A(X)} - \frac{q_R(X)}{D_{b_0}(t,X)} \right\} \mathbf{1}(Y>t) \\
&- \int_0^t \frac{R(1-A)dM_0^C(r \mid X)}{\mathbb{P}(R=1)S^C(t \mid X, R=1)} \frac{S_0(t \mid X)}{S_0(r \mid X)} \left\{ \frac{q_R(X)}{D_{b_0}(t,X)} - \frac{1}{1-\pi_A(X)} \right\} \\
&- \frac{R(1-A)}{\mathbb{P}(R=1)} \left\{ \frac{1}{1-\pi_A(X)} - \frac{q_R(X)}{D_{b_0}(t,X)} \right\} S_0(t \mid X) \\
=&\frac{R(1-A)}{\mathbb{P}(R=1)} \left\{ \frac{r(t,X)P(b=0 \mid X, R=0)}{\{1-\pi_A(X)\}D_{b_0}(t,X)} \right\} m_1^*(t,V),
\end{aligned}
$$

where

$$
\begin{aligned}
m_1^*(t,V) &= \frac{\mathbf{1}(Y>t)}{S^C(t \mid X, R=1)} + \int_0^\tau \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=1)} \frac{S_0(t \mid X)}{S_0(r \mid X)} - S_0(t \mid X) \\
&= \frac{\Delta\mathbf{1}(Y>t)}{S^C(Y \mid X, R=1)} + \int_0^\infty \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=1)} \frac{S_0(t \mid X)}{S_0(r \mid X)} - S_0(t \mid X).
\end{aligned}
$$

Similarly, (25), (26), and (28) together gives us

$$
\begin{aligned}
&\frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)} \frac{q_R(X)r(t,X)\mathbf{1}(Y>t)}{S^C(t \mid X, R=0)D_{b_0}(t,X)} \\
&+ \int_0^t \frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)} \frac{q_R(X)r(t,X)dM_0^C(r \mid X)}{S^C(r \mid X, R=0)D_{b_0}(t,X)} \frac{S_0(t \mid X)}{S_0(r \mid X)} \\
&- \frac{r(t,X)(1-R)\mathbf{1}(b_0=0)q_R(X)}{\mathbb{P}(R=1)D_{b_0}(t,X)} S_0(t \mid X) \\
=&\frac{(1-R)\mathbf{1}(b_0=0)}{\mathbb{P}(R=1)} \frac{q_R(X)r(t,X)}{D_{b_0}(t,X)} m_2^*(t,V),
\end{aligned}
$$

where

$$
\begin{aligned}
m_2^*(t,V) &= \frac{\mathbf{1}(Y>t)}{S^C(t \mid X, R=0)} + \int_0^t \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=0)} \frac{S_0(t \mid X)}{S_0(r \mid X)} - S_0(t \mid X) \\
&= \frac{\Delta\mathbf{1}(Y>t)}{S^C(Y \mid X, R=0)} + \int_0^\infty \frac{dM_0^C(r \mid X)}{S^C(r \mid X, R=0)} \frac{S_0(t \mid X)}{S_0(r \mid X)} - S_0(t \mid X).
\end{aligned}
$$

Then, we can show that $\mathbb{E}[\{\psi^{\text{rct}}_{S_0,\text{eff}}(t,V) - \psi^{\mathcal{A}}_{S_0,\text{eff}}(t,V)\}^2 \mid X]$ equals to

$$\begin{aligned}
&\frac{\pi_R(X)\{1-\pi_A(X)\}}{\mathbb{P}(R=1)^2}\left\{\frac{r(t,X)\mathbb{P}(b_0=0\mid X,R=0)}{\{1-\pi_A(X)\}D^*(t,X)}\right\}^2 \text{var}\left\{m_1^*(t,V)\mid R=1,A=0\right\}\\
&+\frac{\{1-\pi_R(X)\}\mathbb{P}(b_0=0\mid X,R=0)}{\mathbb{P}(R=1)^2}\left\{\frac{q_R(X)r(t,X)}{D^*(t,X)}\right\}^2 \text{var}\left\{m_2^*(t,V)\mid R=0,b_0=0\right\}\\
&=\frac{\pi_R(X)r(t,X)\mathbb{P}(b_0=0\mid X,R=0)}{\mathbb{P}(R=1)^2 D_{b_0}(t,X)\{1-\pi_A(X)\}}\frac{D^*_{b_0}(t,X)}{D_{b_0}(t,X)}\frac{r(t,X)}{r^*(t,X)}V^*_{R1,A0},
\end{aligned}$$

where

$$\begin{aligned}
r^*(t,X)=\frac{V^*_{R1,A0}}{V^*_{R0}},\quad D^*_{b_0}(t,X)=r^*(t,X)\mathbb{P}(b=0\mid X,R=0)+\{1-\pi_A(X)\}q_R(X),\\
V^*_{R1,A0}=\text{var}\left\{m_1^*(t,V)\mid R=1,A=0\right\},\quad V^*_{R0}=\text{var}\left\{m_2^*(t,V)\mid R=0,b_0=0\right\}.
\end{aligned}$$

Thus, the proof of Theorem 3.8 is completed.

B. Additional Simulations

Additional Bias-generating Settings Figure 3(Left) presents the simulation results under Settings Four and Five. Both $\widehat{\theta}^{\text{adapt}}_\tau$ and $\widehat{\theta}^{\text{TransCox}}_\tau$ account for heterogeneity in covariate effects and the risk associated with varying baseline times in these settings. However, $\widehat{\theta}^{\text{TransCox}}_\tau$ is only valid under the Cox model. For example, when the conditional survival curve $S_a(t\mid X)$ does not follow the Cox model, as in Settings Two and Three of the main paper where the marginalized curves $S_a(t\mid X)$ over U (or δ) result in a model that no longer satisfies the Cox proportional hazards assumption, $\widehat{\theta}^{\text{TransCox}}_\tau$ may exhibit substantial bias, whereas the proposed estimator $\widehat{\theta}^{\text{adapt}}_\tau$ continues to control for bias due to its double robustness and demonstrates improved performance.

Asymptotic Properties of the Proposed Estimator Figure 3(Right) provides more details of our proposed selective integrative estimator, specifically focusing on its average borrowing proportion of the external controls and its relative efficiency. The relative efficiency is measured by the ratio of the width of its confidence intervals to the trial-only estimator. In Setting One, it is reasonable to observe that the selective integrative estimator $\widehat{\theta}^{\text{adapt}}_\tau$ is always more efficient compared to the benchmark as every external control is comparable, and the proportion of borrowing approaches 1 as N_0 increases. Under Setting Two, the proportion of borrowing diminishes to zero as it detects that nearly all the external controls are not comparable when more concurrent controls become available. Under Setting Three, the borrowing proportion approaches 50%, aligning well with the true proportion of comparable external subset in our data-generation process. One side note is that our proposal might be subject to slight relative efficiency inferiority compared to the benchmark in some cases, which is also observed in other literature (Chen et al., 2021a).

Varying Censoring Intensity We conduct additional simulation studies to evaluate our selective borrowing estimator $\widehat{\theta}^{\text{adapt}}_\tau$ under varying censoring rates for the trial. In particular, we vary the value of β_C in the hazard functions $\lambda^C(t\mid X,R=1)$ to represent different censoring levels, with $\beta_C=0$ indicating high censoring (the censoring rate is around 60%) and $\beta_C=-2$ indicating low censoring (the censoring rate is around 20%). The results, presented in Figure 4, demonstrate that our proposed estimator effectively controls external biases across all settings and achieves improved estimation, as shown by smaller Root-MSEs compared to the trial-only estimator.

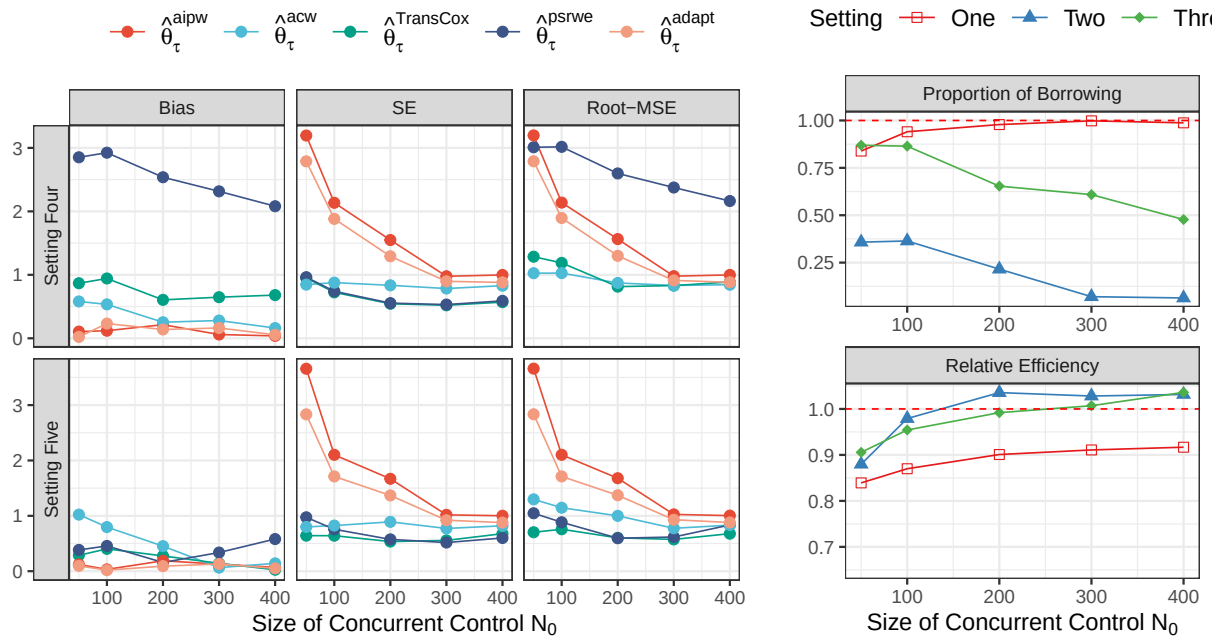

Figure 3. (Left) Point estimation results for RMST over 500 Monte Carlo experiments under Settings 4) different covariate effects, and 5) different baseline time-varying hazards; (Right) Average borrowing proportion of external controls and relative efficiency of the selective integrative estimator $\widehat{\theta}_\tau^{\text{adapt}}$ over 500 Monte Carlo experiments.

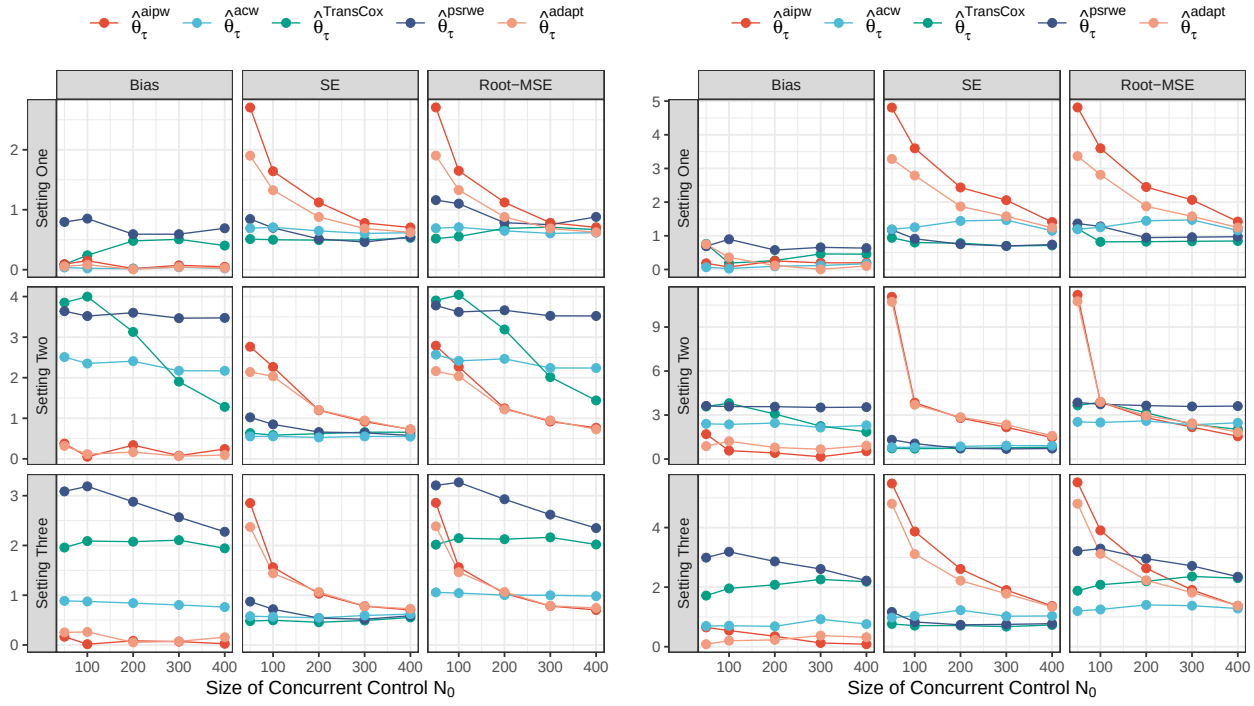

Figure 4. Point estimation results for RMST over 500 Monte Carlo experiments under Settings One, Two and Three when (Left) $\beta_C = -2$ (low censoring rate) and (Right) $\beta_C = 0$ (high censoring rate).