ADAPTIVE EXPERIMENTATION FOR CENSORED SURVIVAL OUTCOMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Adaptive experimentation enables efficient estimation of causal effects, but existing methods are not designed for survival data with censoring, where event times are only partially observed (e.g., overall survival in cancer trials but with dropout). In this paper, we develop a novel framework for adaptive experimentation to estimate causal effects under right censoring. For this, we derive the semiparametric efficiency bound for the average survival effect curve as a function of the treatment allocation policy and thereby obtain a closed-form efficiency-optimal allocation policy. The policy generalizes classical Neyman allocation to survival settings by prioritizing patient strata where both event and censoring dynamics induce high uncertainty. Building on this, we propose the *Adaptive Survival Estimator (ASE)*, an adaptive framework that learns the allocation policy and estimates the average survival effect curve sequentially. Our framework has three main benefits: (i) it accommodates arbitrary machine learning models for nuisance estimation; (ii) it is guided by a closed-form efficiency-optimal allocation policy; and (iii) it admits strong theoretical guarantees, including asymptotic normality via a martingale central limit theorem. We demonstrate our framework across various numerical experiments to show consistent efficiency gains over uniform randomization and censoring-agnostic baselines.

1 INTRODUCTION

Randomized controlled trials (RCTs) are the gold standard for causal inference (Imbens and Rubin, 2015; Wager, 2024), but RCTs have important limitations. By design, RCTs rely on a static rule to allocate patients to treatment arms that is fixed before data collection, and, as a result, RCTs ignore information that becomes available during the trial and may allocate patients inefficiently. This is particularly problematic in survival trials, where patient dropout and administrative censoring are common; this causes event times to be only partially observed, which further complicates estimation.

An alternative to RCTs is ***adaptive experimentation***, which updates how patients are assigned to treatment arms as data are collected (Robertson et al., 2023; van der Laan, 2008; Zhang and van der Laan, 2025). Here, outcomes from previous patients are used to inform how future patients are assigned to treatments. The idea of adaptive randomization is particularly relevant in settings where population sizes are limited by costly recruitment, ethical concerns, or rare diseases (Berry, 2011; Guidance, 2010; van Amsterdam et al., 2026).

Prominent examples are response-adaptive randomization (RAR) and the covariate-adjusted variant (CRAR), which adjust the treatment assignment probabilities based on observed outcomes and patient characteristics (e.g., Chow and Chang, 2008; Hu and Rosenberger, 2006; Rosenberger et al., 2001; Thompson, 1933; Zhang et al., 2007). These methods are primarily designed to improve patient welfare by allocating more patients to better-performing treatments during the trial. A separate line of work studies adaptive designs that aim to minimize asymptotic variance of the ATE estimator, thus leading to an allocation policy based on outcome variability, such as Neyman allocation (e.g., Hahn et al., 2011; Kato et al., 2024; Cook et al., 2024; Dai et al., 2023). However, both paradigms are developed for *fully observed outcomes*, and are **not** directly applicable to survival data with censoring, where event times are only *partially observed* (e.g., overall survival time in cancer trials (Klein and Moeschberger, 2003; Wiegrebe et al., 2024)).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

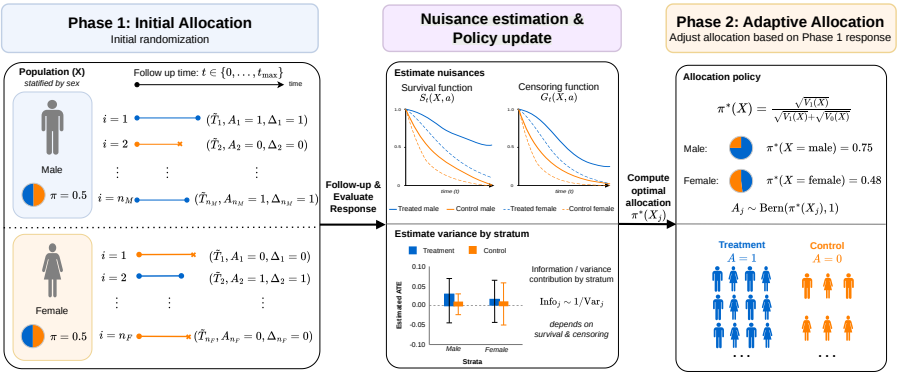


Figure 1: **Our semi-parametric framework for adaptive experimentation with censored survival outcomes.**

Motivating example. GBM AGILE (Cloughesy et al., 2022) is an adaptive clinical trial for glioblastoma (an aggressive form of brain cancer) that compares different treatments based on overall survival time. In cancer care, outcomes are time-to-event and often censored, as patients may not experience the event during the study or may drop out (e.g., due to changes in therapies or loss to follow-up) (Klein and Moeschberger, 2003).

However, in such settings, adaptive trials may update treatment allocation based on observed survival outcomes, while not explicitly accounting for how censoring affects information accumulation. As a result, treatment assignment may overlook censoring-induced uncertainty, leading to inefficient allocation. For example, if patients in one treatment group drop out more frequently, more observations are censored in that group, making it harder to precisely estimate the overall treatment effect.

Existing methods for adaptive experimentation (e.g., Hahn et al., 2011; Kato et al., 2024; Thompson, 1933; van der Laan, 2008) have two main **shortcomings** in survival settings: (1) *Optimal allocation policy*: Standard methods for adaptive experimentation account for uncertainty in observed outcomes, but survival data introduce *additional* uncertainty due to *censoring*. $\Rightarrow$ As a result, an efficient allocation policy should prioritize an arm not only based on variability in outcomes, but also based on the level of censoring, which may vary across treatments and patient subgroups. (2) *Estimation*: Existing methods are designed for settings with fully observed outcomes. In survival settings, outcomes are only partially observed, which makes inference more challenging. $\Rightarrow$ Therefore, it requires an estimator for the average survival effect curve that remains valid as data accumulate across rounds under adaptive allocation.

In this paper, we address the above challenges and propose *a novel adaptive experimentation framework for censored survival outcomes* (see Figure 1):

(1) **Optimal allocation under censoring.** We derive the semiparametric efficiency bound for the average survival effect curve as a function of the treatment allocation policy. This allows us to characterize the variance structure induced by both the survival process and censoring. We thereby obtain a *closed-form*, efficiency-optimal allocation policy that is tailored to survival settings.

(2) **Adaptive Survival Estimator (ASE).** We propose the ***A**daptive **S**urvival **E**stimator (ASE)*, a sequential procedure that learns the optimal allocation policy and estimates the average survival effect curve. Specifically, ASE uses flexible machine learning models to estimate nuisance components, then updates the probability of assigning newly enrolled patients to different treatment arms, and offers valid inference for the average survival effect curve.

Our **contributions**[1] are: (1) We develop a novel framework for adaptive experimentation with censored survival outcomes. Thereby, we generalize the Neyman allocation by deriving an allocation policy that minimizes the trace of the semiparametric efficiency bound for the average survival effect curve. (2) We introduce the *Adaptive Survival Estimator (ASE)*, a new sequential approach for updating treatment allocation policy and estimating the average survival effect curve. We provide strong theoretical guarantees, including asymptotic normality of ASE, and time-uniform asymptotic

[1]Code available at: `https://anonymous.4open.science/r/adaptive_survival-7371`

confidence sequences for valid inference at arbitrary stopping times. (3) We evaluate our ASE across various numerical experiments and observe consistent efficiency gains from our adaptive allocation policy over uniform randomization and censoring-agnostic baselines.

2 Related work[2]

Response-adaptive randomization: Response-adaptive randomization methods are primarily designed to improve patient outcomes during a trial by allocating more patients to treatment arms that appear more effective. • *Response-adaptive randomization (RAR)* traces back to the foundational contribution of Thompson (1933) with Bayesian extensions that drive allocation via posterior estimates of treatment effects (Berry et al., 2010; Robertson et al., 2023). Frequentist formulations update the probabilities of assigning patients to treatment arms as data accumulate during a trial, so that the allocation to treatment can adapt to observed outcomes (Chow and Chang, 2008; Hu and Rosenberger, 2006). • *Covariate-adjusted response-adaptive randomization (CRAR)* extends RAR by incorporating baseline covariates into the allocation rule, which thus yields individualized assignment probabilities that can be used to improve patient welfare (Mukherjee et al., 2025; Rosenberger et al., 2001; Rosenberger and Sverdlov, 2008; Zhang et al., 2007). ⇒ *However, these methods are primarily developed for welfare optimization rather than for improving estimation efficiency.*

Adaptive experimentation for efficient ATE estimation: Another line of work studies adaptive designs that aim to minimize the variance of treatment effect estimation (van der Laan, 2008). In these settings, optimal allocation is given by *Neyman allocation*, which assigns more samples to treatment arms with higher outcome variability. Existing methods can be grouped by the experiment design: (i) in two-staged designs, one can asymptotically attain the semiparametric efficiency bound (Hahn et al., 2011), while (ii) fully sequential designs such as A2IPW achieve variance-optimal Neyman allocation (Cook et al., 2024; Dai et al., 2023; Kato et al., 2024). There are also several extensions (see Appendix B). ⇒ *However, these methods assume* ***fully observed outcomes*** *and apply Neyman allocation based only on outcome variance but ignore the censoring-induced uncertainty.*

Semiparametric estimation of survival treatment effects: We build on semiparametric efficiency theory for treatment effects with time-to-event outcomes, but our focus is different: we leverage the concepts to improve *adaptive* experimental design rather than estimation under a *fixed* treatment assignment. This literature stream provides efficient estimators under censoring (Cai and van der Laan, 2020; Mao et al., 2018; Rubin and van der Laan, 2007; Schaubel and Wei, 2011; Westling et al., 2024; van der Laan and Robins, 2003). There are also several extensions (see Appendix B). ⇒ *However, this line of work focuses on estimation under a fixed treatment assignment mechanism and does not address adaptive design. While it characterizes the semiparametric efficiency bound, it does not leverage it to guide treatment allocation. In contrast, our work uses the efficient influence function to derive sequential allocation policies that are optimal under censoring.*

Research gap: In sum, no existing work jointly derives an efficiency-optimal allocation policy including a valid estimator for survival outcomes under censoring in adaptive experiments (see Table 2 in the Appendix).

3 Problem setup

Setting. We consider the problem of estimating the average treatment effect (ATE) from time-to-event data Cui et al. (2023); Curth and van der Schaar (2021); Frauen et al. (2025).[3] That is, we consider a population $(X, A, T, C) \sim \mathbb{P}$, where $X \in \mathcal{X} \subseteq \mathbb{R}^p$ are observed covariates, $A \in \mathcal{A} = \{0, 1\}$ is a binary treatment, $T \in \mathcal{T}$ is the event time of interest (e.g., death of the patients), and $C \in \mathcal{T}$ is the censoring time (e.g., dropout time). We further use $\Delta = \mathbf{1}(T \leq C)$ as the event indicator[4], and $\tilde{T} = \min\{T, C\}$. Let $\mathcal{O} = (X, A, \tilde{T}, \Delta)$ denote the observed population. Throughout, we assume a discrete-time setting $(T, C) \in \mathcal{T} = \{0, \dots, t_{\max}\}$ and all units share the same discrete time horizon across experiment rounds.

[2] We provide an extended literature review in Appendix B.

[3] We deal with the problem of right censoring, which is common in survival analysis settings.

[4] By defining $\Delta = \mathbf{1}(T \leq C)$, we allow for ties between T and C, which is reflected in the definition of G_{t-1}. We extend our results to the no-ties case, i.e., $\Delta = \mathbf{1}(T < C)$, in Appendix G.

The experiment proceeds over $R \in \mathbb{N}$ rounds. At each round r, a new patient with covariates X_r is i.i.d. drawn from $\mathcal{X}$. The experimenter observes X_r and selects a treatment $A_r \sim \pi_r(\cdot \mid X_r, \mathcal{H}_{r-1})$, where π_r is an **adaptive treatment allocation policy** that depends on the current covariates X_r and past observations

$$\mathcal{H}_{r-1} = \left\{(x_1, a_1, \tilde{t}_1, \delta_1), \ldots, (x_{r-1}, a_{r-1}, \tilde{t}_{r-1}, \delta_{r-1})\right\}. \tag{1}$$

Following the treatment assignment a_r, one observes the outcome $\tilde{t}_r$ and the event indicator δ_r. The full observation at round r is thus $(x_r, a_r, \tilde{t}_r, \delta_r)$. After R rounds, the experimenter estimates the ATE from accumulated data $\mathcal{H}_R = \{(x_r, a_r, \tilde{t}_r, \delta_r)\}_{r=1}^R$. As a result, we allow the treatment assignment policy to evolve over rounds based on accumulated data.

Key definitions: For $t \in \mathcal{T}$, we define the *(conditional) survival and censoring functions* $S_t(x,a) = \mathbb{P}(T > t \mid X = x, A = a)$, $G_t(x,a) = \mathbb{P}(C > t \mid X = x, A = a)$, with the convention $S_{-1}(x,a) = G_{-1}(x,a) \equiv 1$. We also define the observed *discrete-time hazards* $\lambda_t^S(x,a) = \mathbb{P}(\widetilde{T} = t, \Delta = 1 \mid \widetilde{T} \geq t, X = x, A = a)$, and $\lambda_t^G(x,a) = \mathbb{P}(\widetilde{T} = t, \Delta = 0 \mid \widetilde{T} \geq t, X = x, A = a)$.

Estimand: We use the potential outcomes framework (Rubin, 1974) to formalize our causal inference task. Let $T(a) \in \mathcal{T}$ denote the potential event time under treatment $a \in \{0,1\}$. For a fixed $t \in \mathcal{T}$, we are interested in the *average survival effect curve*

$$\tau_t = \mathbb{E}[\tau_t(X)], \text{ where } \tau_t(x) = \mathbb{P}(T(1) > t \mid X = x) - \mathbb{P}(T(0) > t \mid X = x). \tag{2}$$

We define the nuisance vector $\eta_t(X) := \left(\lambda_i^S(X,1), \lambda_i^S(X,0), \lambda_{i-1}^G(X,1), \lambda_{i-1}^G(X,0)\right)_{i=0}^t$.

Identifiability: We impose the following standard assumptions to ensure the identifiability of τ_t.

Assumption 3.1 (Standard causal inference assumptions)**.** For all $a \in \{0,1\}$ and $x \in \mathcal{X}$: (i) *consistency*: $T(a) = T$ whenever $A = a$; (ii) *treatment overlap*: $0 < \pi(x) < 1$ whenever $\mathbb{P}(X = x) > 0$; (iii) *ignorability*: $A \perp (T(0), T(1)) \mid X = x$.

Assumption 3.2 (Survival-specific assumptions)**.** For all $a \in \{0,1\}$ and $x \in \mathcal{X}$ with $\mathbb{P}(X = x, A = a) > 0$: (i) *censoring overlap*: $G_{t-1}(x,a) > 0$; (ii) *survival overlap*: $S_{t-1}(x,a) > 0$; (iii) *non-informative censoring*: $T \perp C \mid X = x, A = a$.

Assumption 3.1 is standard in causal inference (Imbens, 2004; Robins, 2000; van der Laan and Rubin, 2006): (i) consistency ensures that a patient's observed outcome under treatment a equals their potential outcome $T(a)$ and rules out interference between patients; (ii) treatment overlap means that every patient has a positive probability of receiving each treatment, which thereby ensures sufficient support in the data; and (iii) ignorability means there are no unobserved confounders that can bias our estimation. Assumption 3.2 is commonly imposed for survival analysis (van der Laan and Robins, 2003; Cui et al., 2023) and ensures that we have sufficient uncensored and surviving individuals for each covariate value, and censoring mechanism is independent of a patient's survival time given covariates and treatment. Under Assumptions 3.1 and 3.2, one can identify τ_t via (van der Laan and Robins, 2003):

$$\tau_t = \mathbb{E}\left[S_t(X,1) - S_t(X,0)\right] = \mathbb{E}\left[\prod_{i=0}^{t}(1 - \lambda_i^S(X,1)) - \prod_{i=0}^{t}(1 - \lambda_i^S(X,0))\right]. \tag{3}$$

Efficient influence function (EIF) in the non-adaptive setting. Cui et al. (2023) have derived the non-centered EIF ϕ_t for the ATE estimator in Eq. 4 in the *non-adaptive* setting, where the treatment assignment policy is *fixed* over time:

$$\phi_t(\mathcal{O}; \tau_t, \eta_t) = S_t(X,1) - S_t(X,0) - \frac{(A - \pi(X))\,\xi(\mathcal{O}, \eta_t)\,S_t(X,A)}{\pi(X)(1 - \pi(X))}, \tag{4}$$

where $\xi(\mathcal{O}, \eta_t) = \sum_{i=0}^t \frac{\mathbf{1}(\widetilde{T}=i, \Delta=1) - \mathbf{1}(\widetilde{T} \geq i)\,\lambda_i^S(X,A)}{S_i(X,A)\,G_{i-1}(X,A)}$ and $G_{t-1}(X,A) = \prod_{i=0}^{t-1} 1 - \frac{\lambda_i^G(x,a)}{1 - \lambda_i^S(x,a)}$.

4 Semiparametric efficiency bound and optimal treatment allocation

To guide optimal experiment design in survival settings, we first derive the semiparametric efficiency bound for ATE estimation. This then characterizes the variance-minimizing allocation strategy and motivates our ASE method later (see § 5).

Semiparametric efficiency bound: We consider a setting where the experimenter has direct control over π and can update the allocation policy to improve estimation efficiency. Hence, we are interested in the problem of *designing* π to minimize the asymptotic variance of the ATE estimator.

Theorem 4.1 (Semiparametric efficiency bound). *Under Assumptions 3.1 and 3.2, the semiparametric efficiency bound for estimating* $\tau = (\tau_0, \ldots, \tau_{t_{\max}})^\top$ *is* $\Sigma_{\text{eff}}(\pi) \in \mathbb{R}^{(t_{\max}+1)\times(t_{\max}+1)}$,

$$\Sigma_{\text{eff}}(\pi) = \mathbb{E}\left[\frac{\Sigma_1(X)}{\pi(X)} + \frac{\Sigma_0(X)}{1-\pi(X)}\right] + \mathbb{E}[b(X)b(X)^\top], \tag{5}$$

where $\Sigma_a(X) = \text{Var}(\boldsymbol{S}(X,a) \odot \boldsymbol{\xi}(\mathcal{O}; \eta_{t_{\max}}) \mid X, A=a)$, *and* $b(X) := \boldsymbol{S}(X,1) - \boldsymbol{S}(X,0) - \boldsymbol{\tau}$.

Proof. See Appendix F.1, we provide a detailed summary of notation in Appendix A. □

Illustrative example

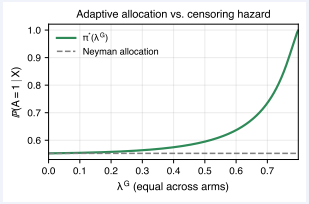


Figure 2: Censoring aware A-Optimal policy $\pi^\star(X)$ vs. censoring hazard.

Consider the case where both arms share the same censoring hazard (i.e., $\lambda_t^G(x,0) = \lambda_t^G(x,1) =: \lambda^G$); yet where the event dynamics differ (i.e., $\lambda_t^S(x,0) \neq \lambda_t^S(x,1)$). One might expect that equal censoring renders $\pi^\star$ insensitive to λ^G, and that one would recover the classical Neyman allocation. However, Figure 2 refutes this: at $\lambda^G = 0$, $\pi^\star$ coincides with the Neyman allocation (dashed), but diverges monotonically as λ^G grows. The key mechanism is the following: equal λ^G does not imply equal G_{t-1} across arms, since the conditional censoring hazard $G_{t-1}(x,a) = \prod_{i=0}^{t-1} 1 - \lambda_i^G/(1-\lambda_i^S(x,a))$ depends on arm-specific event dynamics, so the arm with higher event hazard experiences a larger increase in the censoring variance even under equal λ^G. We provide an additional example for intuition in Appendix C.

⇒ *This demonstrates that censoring affects the optimal allocation through both* $\lambda_t^G(X,a)$ *and* $\lambda_t^S(X,a)$*; and that the standard Neyman allocation is suboptimal whenever* $\lambda_t^G(X,a) > 0$.

Optimal treatment allocation. Given the efficiency bound $\Sigma_{\text{eff}}(\pi)$ from Theorem 4.1, we now derive the treatment allocation policy $\pi^\star$ that minimizes it. Since $\Sigma_{\text{eff}}(\pi)$ is a matrix, an optimal $\pi^\star$ should minimize a scalar summary of $\Sigma_{\text{eff}}(\pi)$ over all feasible allocation policies $\Pi_\alpha = \{\pi : \alpha \leq \pi(x) \leq 1-\alpha \text{ for all } x\}$.

Definition 4.2. *A treatment allocation policy* π^* *is A-optimal if* $\pi^* \in \arg\min \text{tr}(\Sigma_{\text{eff}}(\pi))$, *where* tr *is the trace operator. That is, an A-optimal policy minimizes the sum of the variance of the discrete horizon* t *estimators.*

In the following, we adopt the trace criterion $\text{tr}(\Sigma_{\text{eff}}(\pi))$, corresponding to *A-optimality*[5] in the classical theory of optimal experimental design (Atkinson et al., 2023), which calculates the sum of component-wise asymptotic variances across horizons. Since the trace separates additively over t, it reduces to independently minimizing the marginal variance at each time horizon. The following result explicitly states the A-optimal treatment allocation policy for our task.

Proposition 4.3 (A-optimal treatment allocation policy). *Assume* $\alpha \in (0, 1/2)$ *and* $V_0(X) + V_1(X) > 0$ *for all* $x \in \mathcal{X}$. *Then, the allocation policy minimizing* $\text{tr}(\Sigma_{\text{eff}}(\pi))$ *subject to* $\pi \in \Pi_\alpha$ *is given by*

$$\pi^\star(X) = \text{clip}_\alpha\left(\frac{\sqrt{V_1(X)}}{\sqrt{V_1(X)} + \sqrt{V_0(X)}}\right), \quad \textit{where } V_a(X) = \sum_{t=0}^{t_{\max}} v_{t,a}(X) = \text{tr}(\Sigma_a(X)), \tag{6}$$

where $v_{t,a}(X)$ *are the* t*-th diagonal entries of* $\Sigma_a(X)$.

Proof. See Appendix F.2. □

Interpretation: Proposition 4.3 implies that the optimal policy $\pi^\star(X)$ allocates more probability mass to the arm a with higher $V_a(X)$, where

$$V_a(x) = \sum_{t=0}^{t_{\max}} v_{t,a}(X) = \sum_{t=0}^{t_{\max}} \text{Var}[S_t(X,a)\,\xi(\mathcal{O}, \eta_t) \mid X, A=a] = \sum_{t=0}^{t_{\max}} S_t(x,a)^2 \sum_{i=0}^{t} \frac{\lambda_i^S(x,a)}{S_i(x,a)\,G_{i-1}(x,a)}. \tag{7}$$

Unlike standard adaptive ATE estimation, $V_a(X)$ jointly captures two sources of uncertainty. (i) *survival uncertainty*: arms where the variance of the conditional survival probability (i.e.,

[5] In contrast to A-optimality, alternatives such as D-optimality (i.e., $\det \Sigma_{\text{eff}}$, which corresponds to confidence-ellipsoid volume) or E-optimality (i.e., $\lambda_{\max}(\Sigma_{\text{eff}})$ account for inter-horizon covariance structure but do not admit a closed-form solution; discussion in Appendix D.

$\mathrm{Var}[S_t(X,a) \mid X, A=a]$) is large contribute more to $V_a(X)$; and (ii) *censoring uncertainty*: since $\xi(\mathcal{O},\eta_t)$ involves the inverse censoring weight $1/G_{i-1}(X,a)$, arms with larger censoring inflate $V_a(X)$, thus causing $\pi^\star$ to compensate by allocating more patients to that arm.

Neyman allocation as a special case: In the special case of no censoring in the trial (i.e., $G_{t-1}(x,a)\equiv 1$, $\widetilde{T}=T$ with $\Delta\equiv 1$), one can directly compute $V_a(X)=\sum_{t=0}^{t_{\max}}\mathrm{Var}[S_t(X,a)\mid X,A=a]=\sum_{t=0}^{t_{\max}}\mathrm{Var}\big(\mathbf{1}\{T>t\}\big|X,A=a\big)$, so that the trace-optimal allocation in Proposition 4.3 reduces to

$$\pi^\star(x) \;=\; \mathrm{clip}_\alpha\!\left(\frac{\sqrt{\mathrm{tr}(\mathrm{Var}(\mathbf{1}\{T>t\}\mid X=x,A=1))}}{\sqrt{\mathrm{tr}(\mathrm{Var}(\mathbf{1}\{T>t\}\mid X=x,A=1))}+\sqrt{\mathrm{tr}(\mathrm{Var}(\mathbf{1}\{T>t\}\mid X=x,A=0))}}\right), \tag{8}$$

which exactly matches the classical Neyman allocation for minimizing the variance of a difference-in-means estimator with a binary outcome $\mathbf{1}\{T>t\}$ (Neyman, 1934). However, unlike standard Neyman allocation, which targets outcome variance alone, our policy accounts for a second source of uncertainty: $v_{t,a}(X)$ inflates whenever censoring is large in arm a, which directly forces the policy to compensate by allocating more units there.

5 The Adaptive Survival Estimator (ASE)

We now propose the **Adaptive Survival Estimator (ASE)**, a sequential framework for learning the optimal allocation policy and estimating the average survival effect curve. Our goal is to minimize the semiparametric efficiency bound from Theorem 4.1. We proceed as follows: (1) We introduce the *adaptive treatment allocation policy* in Section 5.1 and the adaptive version of *average survival effect curve estimation* in Section 5.2. We provide an extension to batch settings in Appendix I.

5.1 Adaptive treatment allocation policy

Our framework operates in two steps. In the first R_0 rounds, units are assigned under a fixed policy $\pi_{\mathrm{init}}(X)$ (e.g., Bernoulli$(1/2)$ randomization) to accumulate sufficient history for stable estimation of the survival and censoring hazards λ_t^S, λ_t^G. From round R_0+1 onward, the policy switches to the data-driven policy $\pi_r(X\mid\mathcal{H}_{r-1})$ below, which sequentially approximates the A-optimal allocation of Proposition 4.3. Concretely, we obtain a direct estimate $\widetilde{\pi}_r(X\mid\mathcal{H}_{r-1})$ by

$$\widetilde{\pi}_r(X\mid\mathcal{H}_{r-1}) := \frac{\sqrt{\hat{V}_{1,r-1}(X)}}{\sqrt{\hat{V}_{1,r-1}(X)}+\sqrt{\hat{V}_{0,r-1}(X)}}, \tag{9}$$

where $V_{a,r-1}(X)$ is the trace of the conditional covariance matrix based on data in $\mathcal{H}_{r-1}$. We then apply a truncation step to $\widetilde{\pi}_r(X\mid\mathcal{H}_{r-1})$ to obtain the assignment policy $\pi_r(X\mid\mathcal{H}_{r-1})$ at round r.

Closed-form $V_{a,r-1}(X)$ estimation: We now present the closed-form estimation of $V_{a,r-1}(X)$ by combining Theorem 4.1 with A-optimal policy:

$$\hat{V}_{a,r-1}(X)=\sum_{t=0}^{t_{\max}}\hat{S}_{t,r-1}(x,a)^2\sum_{i=0}^{t}\frac{\hat{\lambda}^S_{i,r-1}(x,a)}{\hat{S}_{i,r-1}(x,a)\,\hat{G}_{i-1,r-1}(x,a)}, \tag{10}$$

where $\hat{S}_{i,r-1}(x,a)$ and $\hat{G}_{i-1,r-1}(x,a)$ are the survival- and censoring-hazard nuisances already required by the EIF in Eq. 4, but now estimated sequentially from the accumulated history $\mathcal{H}_{r-1}$. Specifically, $\widetilde{\pi}_r(x)$ is a deterministic function of $\eta_{t,r-1}(X)$ alone.

Remark 5.1. *These conclusions highlight the difference to existing work (Oprescu et al., 2025), which typically requires a second-stage regression to estimate the variance target governing the optimal allocation. In contrast, Eq. 10 expresses $\hat{V}_{a,r-1}(x)$ as a closed-form function of the hazard nuisances $\hat{\eta}_{r-1}$, which are already used inside the EIF construction at round $r-1$. This yields two practical advantages: (i) no auxiliary second-stage learner for the variance is needed; (ii) the non-negativity of $\hat{V}_{a,r-1}$ holds by construction.*

Truncation for treatment allocation policy: Extreme censoring or small within-stratum sample sizes can drive $\hat{V}_{a,r-1}(X)$ toward zero, thus causing the direct estimates of policy $\widetilde{\pi}_r(X\mid\mathcal{H}_{r-1})$ to collapse to the boundary of $[0,1]$. Following the literature (Cook et al., 2024; Dai et al., 2023; Oprescu et al., 2025), we clip $\widetilde{\pi}_r$ to a shrinking interval, defining the final assignment policy as

$$\pi_r(X\mid\mathcal{H}_{r-1}) := \min\left\{1-\frac{1}{k_r}, \max\left\{\frac{1}{k_r}, \widetilde{\pi}_r(X\mid\mathcal{H}_{r-1})\right\}\right\}, \tag{11}$$

where $k_r\in[2,\infty)$ is a non-decreasing sequence with $k_r\to\infty$ as $r\to\infty$. This keeps all assignment probabilities strictly interior, controls the growth of the censoring-weighted scores in the EIF, and is necessary for the theoretical guarantees in Section 6.

5.2 SEQUENTIALLY CROSS-FITTED ADAPTIVE ATE ESTIMATION

Why a custom cross-fitting approach is needed: The reason is the sequential dependence: at round r, the estimated nuisances $\hat{\eta}_{t,r-1}$ are functions of the same accumulated history $\mathcal{H}_{r-1}$. The standard central limit theorems (CLTs) therefore do not apply. Our approach resolves this via a near-martingale decomposition: we isolate an oracle term that satisfies a martingale CLT and a nuisance remainder whose bias vanishes at rate $o_p(R^{-1/2})$, which is guaranteed once each hazard estimator achieves the $o_p(r^{-1/4})$ rate imposed in Theorem 6.3.

Estimation via sequential cross-fitting: To mitigate finite-sample bias and ensure the nuisance remainder is of order $o_p(R^{-1/2})$, we apply the sequential cross-fitting scheme of Waudby-Smith et al. (2024) with $K = 2$ folds. For each unit $r \in \mathbb{N}$, define the fold index $J_r := r \bmod 2 \in \{0, 1\}$, and split $\mathcal{H}_{r-1}$ into two temporal folds, i.e., $\mathcal{H}_{r-1} = \mathcal{H}_{r-1}^{(0)} \cup \mathcal{H}_{r-1}^{(1)}$,

$$\mathcal{H}_{r-1}^{(j)} := \big\{(X_{\tilde{r}}, A_{\tilde{r}}, \widetilde{T}_{\tilde{r}}, \Delta_{\tilde{r}}) : \tilde{r} \in [1, 2, ..., r-1],\ J_{\tilde{r}} = j\big\}, \qquad j \in \{0, 1\}. \tag{12}$$

Given the cross-fitted nuisances $\hat{\eta}_{t,r-1}^{(-J_r)}$ and the adaptive policy $\pi_r(X \mid \mathcal{H}_{r-1})$, we define the cross-fitted EIF pseudo-outcome at horizon t by

$$\hat{\phi}_{t,r}^{\mathrm{cf}} := \hat{S}_{t,r-1}^{(-J_r)}(X_r, 1) - \hat{S}_{t,r-1}^{(-J_r)}(X_r, 0) - \frac{\{A_r - \pi_r(X_r)\}\hat{\xi}\big(\mathcal{H}_r, \hat{\eta}_{t,r-1}^{(-J_r)}\big)}{\pi_r(X_r)\{1 - \pi_r(X_r)\}} \hat{S}_{t,r-1}^{(-J_r)}(X_r, A_r), \tag{13}$$

where $\hat{\xi}(\mathcal{H}_r, \hat{\eta}_{t,r-1}^{(-J_r)})$ denotes the sequentially estimated version of $\xi(\mathcal{H}_r, \eta_t)$ in Eq. 4, with all nuisance components replaced by $\hat{\eta}_{t,r-1}^{(-J_r)}$. We now yield our ASE: our aim is to sequentially estimate the ATE from the non-centered efficient influence function in Eq. 4 with sequentially updated nuisance functions. In ASE, the estimator at horizon t is defined as

$$\hat{\tau}_{t,R}^{\mathrm{ASE}} = \frac{1}{R} \sum_{r=1}^{R} \hat{\phi}^{\mathrm{cf}}(\mathcal{H}_r; \pi_r, \hat{\eta}_{t,r}^{\mathrm{cf}}), \tag{14}$$

where $\hat{\eta}_{t,r-1}^{\mathrm{cf}} = \{\hat{\lambda}_{i,r-1}^{S,\mathrm{cf}}(X, 0), \hat{\lambda}_{i,r-1}^{S,\mathrm{cf}}(X, 1), \hat{\lambda}_{i,r-1}^{G,\mathrm{cf}}(X, 0), \hat{\lambda}_{i,r-1}^{G,\mathrm{cf}}(X, 1)\}_{i=0}^{t}$ denotes the estimates of the nuisance functions at round r with sequential cross-fitting, which are constructed solely from the past data $\mathcal{H}_{r-1}$. The treatment assignment policy $\pi_r(X \mid \mathcal{H}_{r-1})$, defined by the experimenter based on the estimated optimal policy from Section 5.1, is treated as known and does not require further estimation from data. The full procedure is summarized in Algorithm 1.

6 THEORETICAL GUARANTEES

This section provides several theoretical guarantees for the ASE. We establish asymptotic normality and an A-optimal semiparametric efficiency bound in Section 6.1, and robustness of the ASE under partial nuisance misspecification in Section 6.2. Both results extend naturally to batch settings where policy updates occur at the end of each batch rather than every round, since the near-martingale decomposition depends only on L_2-consistency of the nuisance estimates and *not* on the update frequency. In Appendix H, we additionally derive asymptotically-valid time-uniform confidence sequences for the ASE, thereby enabling valid inference at arbitrary data-dependent stopping times.

6.1 ASYMPTOTIC NORMALITY OF THE ASE

To derive our asymptotic guarantees[6], we impose the following uniform overlap assumption on the survival and censoring hazards. This condition is common in semiparametric survival analysis (Cui et al., 2023; van der Laan and Robins, 2003) and ensures that the inverse probability of censoring weights in the Eq. 4 remain bounded.

Assumption 6.1 (Uniform overlap up to horizon t)**.** For each fixed $t \in \{0, \ldots, t_{\max}\}$, there exist constants $\underline{c}_t^S, \underline{c}_t^G > 0$ such that, for all $a \in \{0, 1\}$, all x, and all $i \le t$, we have $1 - \lambda_i^S(x, a) \ge \underline{c}_t^S$ and $1 - \lambda_i^G(x, a)/(1 - \lambda_i^S(x, a)) \ge \underline{c}_t^G$.

Remark 6.2 (Necessity of stepwise overlap). *Assumption 6.1 imposes bounds at every intermediate step $i \le t$ rather than only at the endpoint, since S_t and G_{t-1} are cumulative products: a single step with $\lambda_i^S \to 1$ drives the $\xi(\mathcal{H}_r; \eta)$ in the EIF to become unbounded. Hence, this constitutes the discrete-time survival analogue of the positivity condition standard in semiparametric survival analysis (Cui et al., 2023; van der Laan and Robins, 2003; Westling et al., 2024). See Appendix F.3 for details.*

[6]Throughout, estimators use sequential cross-fitting of Section 5.2; we suppress the superscript $^{\mathrm{cf}}$ for brevity.

Theorem 6.3 (Asymptotic normality of the adaptive survival estimator)**.** *Suppose Assumptions 3.1, 3.2, and Assumption 6.1 hold, and let $k_r \in [2, \infty)$ be a non-decreasing truncation sequence satisfying $k_r \to \infty$ and $k_r = o(r^{1/4})$ as $r \to \infty$. Suppose there exists a non-adaptive policy $\pi(X) \in [\varepsilon, 1-\varepsilon]$ for some $\varepsilon > 0$ such that the adaptive assignment policy $\pi_r(X \mid \mathcal{H}_{r-1})$ are L_2-consistent relative to truncation schedule, i.e., $k_r\|\pi_r - \pi\|_2 = o_p(1)$. Furthermore, assume that, for each fixed t and treatment a, we have $\max_{0\le i\le t_{\max}} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2 = o_p(r^{-1/4})$ and $\max_{0\le i\le t_{\max}-1} \|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \lambda^G_i(a,\cdot)\|_2 = o_p(r^{-1/4})$. Then, the adaptive estimator is asymptotically normal:*

$$\sqrt{R}\left(\hat{\tau}^{\mathrm{ASE}}_{t,R} - \tau_t\right) \to^d \mathcal{N}(0, V_{\mathrm{eff},t}(\pi)), \tag{15}$$

where $V_{\mathrm{eff},t}(\pi) = \mathbb{E}\left[\frac{v_{t,1}(X)}{\pi(X)} + \frac{v_{t,0}(X)}{1-\pi(X)}\right] + b_t$ denotes the semiparametric efficiency bound for τ_t under policy π and $b_t = \mathbb{E}[(S_t(X,1) - S_t(X,0) - \tau_t)^2]$. In particular, if we have $\pi = \pi^\star$, then $\hat{\tau}^{\mathrm{ASE}}_{t,R}$ attains the A-optimal semiparametric efficiency bound.

Proof. See Appendix F.4. □

The proof decomposes $\sqrt{R}(\hat{\tau}^{\mathrm{ASE}}_t - \tau_t) = \frac{1}{\sqrt{R}}\sum_{r=1}^R z_{t,r} + \frac{1}{\sqrt{R}}\sum_{r=1}^R m_{t,r}$ into an oracle term and a nuisance remainder. The oracle term $z_{t,r} = \phi_{t,r} - \tau_t$ forms a martingale difference sequence with respect to $\mathcal{H}_{r-1}$, since the adaptive policy π_r is determined by $\mathcal{H}_{r-1}$ and the inverse censoring augmentation satisfies $\mathbb{E}[\xi(\mathcal{H}_r, \eta_t) \mid A_r, X_r, \mathcal{H}_{r-1}] = 0$. The martingale central limit theorem (CLT) then yields asymptotic normality. The remainder $m_{t,r} = \hat{\phi}_{t,r} - \phi_{t,r}$ is controlled via the cause-specific structure of $G_{t-1}(X,a)$: the bias term vanishes at rate $o_p(r^{-1/2})$ under $o_p(r^{-1/4})$ hazard consistency, and the martingale fluctuation term is negligible by L_2-boundedness of the cross-fitted scores, thus requiring no Donsker conditions.

Importantly, the truncation does *not* interfere with the optimal allocation in the limit. The reason is that, since $\pi^\star(X) \in [\alpha, 1-\alpha]$ and $1/k_r \to 0$, the clipping interval eventually contains $\pi^\star(X)$.

Proposition 6.4 (Policy convergence under hazard consistency)**.** *Suppose Assumptions 3.1, 3.2, and 6.1 hold, and that the hazard estimators satisfy the rate conditions of Theorem 6.3. Assume further that $\pi^\star(X) \in [\alpha, 1-\alpha]$ almost surely for some $\alpha > 0$. Then, the direct variance estimates satisfy $\|\hat{V}_{a,r-1} - V_a\|_2 = o_p(r^{-1/4})$ for each $a \in \{0,1\}$. Consequently, for any non-decreasing truncation sequence $k_r = o(r^{-1/4})$, we have $k_r\|\pi_r - \pi^\star\|_2 = o_p(1)$.*

Proof. See Appendix F.5. □

6.2 Consistency guarantees under partial nuisance misspecification

As shown in Theorem 6.3, the convergence rate of ASE is primarily governed by the estimation error of $\hat{\lambda}^S_t(X,\cdot)$ and $\hat{\lambda}^G_t(X,\cdot)$. This reflects a *robustness* property formalized in the following proposition and corollary: misspecification of the censoring hazard $\hat{\lambda}^G_t(X,\cdot)$ alone does *not* affect consistency, provided the event hazard $\hat{\lambda}^S_t(X,\cdot)$ is estimated consistently.

Proposition 6.5 (Convergence under censoring-hazard misspecification)**.** *Suppose Assumptions 3.1, 3.2, and 6.1 hold, and suppose that there exists a non-adaptive policy $\pi(X) \in [\varepsilon, 1-\varepsilon]$ such that $k_r\|\pi_r - \pi\|_2 = o_p(1)$. Let $\tilde{\lambda}^G_i(a,\cdot)$ denote a possibly misspecified limit of the censoring-hazard estimators, and suppose that $\|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \tilde{\lambda}^G_i(a,\cdot)\|_2 = o_p(1)$ for all $i < t$ and $a \in \{0,1\}$. Then, for each $t \le t_{\max}$,*

$$|\hat{\tau}^{\mathrm{ASE}}_{t,R} - \tau_t| = O_p(R^{-1/2}) \tag{16}$$
$$+ O_p\left(\frac{1}{R}\sum_{r=1,a\in\{0,1\}}^{R}\left[\left(\sum_{i=0}^{t}\|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2\right)^2 + \sum_{i=0}^{t}\|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2 \sum_{i=0}^{t-1}\|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \lambda^G_i(a,\cdot)\|_2\right]\right).$$

Corollary 6.6 (Consistency under censoring-hazard misspecification)**.** *Under the conditions of Proposition 6.5, if the event-hazard estimators are L_2-consistent for all $i \le t$ and $a \in \{0,1\}$, then $\hat{\tau}^{\mathrm{ASE}}_{t,R}$ is consistent for τ_t, even when the censoring-hazard estimators converge to a possibly misspecified limit.*

For the no-ties case, we can even provide a double-robustness result (see Appendix G, where $\Delta = \mathbf{1}(T < C)$ excludes simultaneous event and censoring times. In this setting, the censoring survival function simplifies to $G_{t-1}(x,a) = \prod_{i=0}^{t-1}(1 - \lambda^G_i(x,a))$, which is independent of the event hazard λ^S. As a result, the ASE is doubly-robust (i.e., to misspecification of both λ^S and λ^G), and the remainder vanishes whenever either nuisance is consistently estimated, regardless of the other.

7 EXPERIMENTAL RESULTS

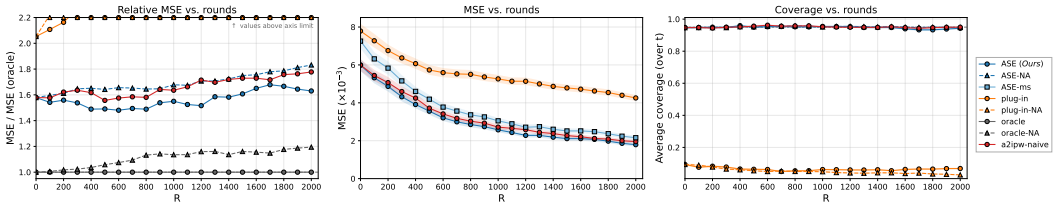


Figure 3: **Synthetic experiments:** (a) *(left)*: Relative MSE with respect to Oracle; ASE achieves the lowest error among baselines. (b) *(middle)*: MSE across rounds; ASE converges fastest while ASE-MS remains consistent under nuisance misspecification. (c) *(right)*: Empirical coverage of nominal 95% CIs; ASE achieves coverage while plug-in and a2ipw-NAïve deteriorate as R grows.

Baselines: We compare our estimator (**ASE**) against several baselines: (1) its non-adaptive counterpart (**ASE-NA**), which assigns the treatment uniformly at random; (2) the plug-in estimator from Eq. 3 (**plug-in**); and (3) its non-adaptive version (**plug-in-NA**); (4) the **A2IPW-NAïve** estimator from Kato et al. (2024) as the Neyman-allocation which ignores censoring in allocation policy; we also report two oracle baselines for comparison: (5) a fully use oracle-efficient estimator that uses the oracle nuisance function (**Oracle**) and (6) a non-adaptive version (**Oracle-NA**). To assess the benefit of robustness, we also evaluate a misspecified version of ASE, denoted as **ASE-MS**, in which the $\lambda_t^G(X, a)$ estimator is deliberately misspecified.

Implementation details: We estimate the nuisance functions using LightGBM, trained with standard gradient-boosted tree objectives for classification (see Appendix K for details). **Synthetic data:** We follow the common practice in adaptive experimentation (Oprescu et al., 2025; Zhang and van der Laan, 2025) and evaluate the effectiveness of our framework on both synthetic and semi-synthetic datasets to benchmark against ground-truth results. We provide the estimated average survival effect curves for synthetic and semi-synthetic data in Appendix J.2. In our experiments, we aim to evaluate three research questions:

(a) *How much do we benefit from adaptive allocation?* ⇒ **Effectiveness of A-optimal treatment allocation**: For this, we report the relative MSE with respect to the Oracle estimator in Fig. 3 (a). Our ASE achieves a relative MSE of approximately $1.05\times$ of the oracle at $R = 2000$, which nearly matches oracle performance despite using estimated nuisances. In contrast, the non-adaptive baselines (ASE-NA and plug-in-NA) stabilize at $1.35\times$ and $1.40\times$ of the oracle MSE, respectively, maintaining a persistent efficiency gap. The a2ipw-NAïve method falls in an intermediate position at roughly 1.20–$1.25\times$, which confirms that our censoring-aware A-optimal allocation (and not just adaptive assignment alone) is the key driver of the efficiency gains over uniform randomization.

(b) *Does ASE improve estimation accuracy over rounds?* ⇒ **Estimation accuracy**: We thus plot the MSE against R in Fig. 3 (b). First, ASE, ASE-MS, plug-in, and a2ipw-NAïve converge to the true τ. Out of them, ASE achieves the lowest error and thus performs best. The results further highlight the importance of our robustness property: ASE-MS remains consistent and even achieves a similar speed of convergence as ASE, despite one nuisance component being misspecified.

(c) *Is ASE inferentially valid?* ⇒ **Inferential validity**: We here report empirical coverage of nominal 95% confidence intervals (CIs) in Fig. 3 (c). Consistent with our theoretical guarantee, our ASE and ASE-NA achieve nominal coverage. In contrast, the other methods fail to do so. For example, the plug-in, plug-in-NA, and a2ipw-NAïve perform poorly as R grows, which is due to finite-sample bias and censoring bias.

Semi-synthetic data: We also evaluate ASE on a semi-synthetic dataset based on the Twins dataset as in Frauen et al. (2025); Louizos et al. (2017) to demonstrate the applicability to medical data. The results are in Appendix J.1. **Results:** Our findings are consistent with the synthetic setting: (i) adaptive treatment allocation policy improves estimation efficiency, (ii) ASE achieves superior coverage and consistency, and (iii) robustness holds under partial nuisance misspecification.

Conclusion: Our work highlights the potential broader impact of adaptive experimentation to improve clinical trials with survival data, especially in settings where efficient evidence generation is critical. **Limitations:** Future work could be extended to multi-arm trials and demonstrate its practical value in real clinical settings with limited and heterogeneous patient populations.

AI USE STATEMENT

The research questions and core methodological ideas were developed by the authors. Generative AI tools were used as research assistants for literature retrieval, feedback on research and experimental design, implementation and debugging, synthetic-data and figure preparation, checking and refining mathematical claims and proofs, interpreting results, and drafting and editing parts of the manuscript. All AI-assisted outputs were reviewed and, where applicable, verified against the implementation, experiments, literature, or mathematical derivations; the authors made all final decisions and take responsibility for the content. LLM use as part of the experimental data-generation pipelines is described separately in the corresponding experimental appendices.

REFERENCES

Ahthony Atkinson, Alexander Donev, and Randall Tobias. *Optimum experimental designs, with SAS*. Number 34 in Oxford Statistical Science Series. Oxford University Press, 2023.

Ad Barker, Cc Sigman, Gj Kelloff, Nm Hylton, Da Berry, and Lj Esserman. I-SPY 2: An adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clinical Pharmacology & Therapeutics*, 86(1):97–100, 2009.

Donald A. Berry. Adaptive clinical trials: The promise and the caution. *Journal of Clinical Oncology*, 29(6):606–609, 2011.

Scott M. Berry, Bradley P. Carlin, J. Jack Lee, and Peter Muller. *Bayesian Adaptive Methods for Clinical Trials*. CRC Press, 2010.

Weixin Cai and Mark J. van der Laan. One-step targeted maximum likelihood estimation for time-to-event outcomes. *Biometrics*, 76(3):722–733, 2020.

Emmanuel Candès, Lihua Lei, and Zhimei Ren. Conformalized survival analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(1):24–45, 2023.

Fangyi Chen, Shu Ge, Jian Qian, and Christopher Harshaw. Sigmoid-FTRL: Design-based adaptive neyman allocation for AIPW estimators. *arXiv preprint*, arXiv:2511.19905, 2025.

Shein-Chung Chow and Mark Chang. Adaptive design methods in clinical trials – a review. *Orphanet Journal of Rare Diseases*, 3:11, 2008.

Timothy Francis Cloughesy, Brian Michael Alexander, Donald A. Berry, Howard Colman, John Frederick De Groot, Benjamin M. Ellingson, Gary B. Gordon, Mustafa Khasraw, Andrew B. Lassman, Eudocia Quant Lee, Michael Lim, Ingo K. Mellinghoff, Apoorva Nelli, James R. Perry, Erik P. Sulman, Kirk Tanner, Michael Weller, Patrick Y. Wen, W. K. Alfred Yung, and GBM AGILE Investigators. GBM AGILE: A global, phase 2/3 adaptive platform trial to evaluate multiple regimens in newly diagnosed and recurrent glioblastoma. *Journal of Clinical Oncology*, 40: TPS2078–TPS2078, 2022.

Thomas Cook, Alan Mishler, and Aaditya Ramdas. Semiparametric efficient inference in adaptive experiments. In *Proceedings of the Third Conference on Causal Learning and Reasoning*, pages 1033–1064. PMLR, 2024.

Yifan Cui, Michael R Kosorok, Erik Sverdrup, Stefan Wager, and Ruoqing Zhu. Estimating heterogeneous treatment effects with right-censored data via causal survival forests. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(2):179–211, 2023.

Alicia Curth and Mihaela van der Schaar. Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms. In *AISTATS*, 2021.

Jessica Dai, Paula Gradu, and Christopher Harshaw. CLIP-OGD: An experimental design for adaptive neyman allocation in sequential experiments. In *NeurIPS*, 2023.

Abhinandan Dalal, Patrick Blöbaum, Shiva Kasiviswanathan, and Aaditya Ramdas. Anytime-valid inference for double/debiased machine learning of causal parameters. *arXiv preprint*, arXiv:2408.09598, 2024.

Hen Davidov, Shai Feldman, Gil Shamai, Ron Kimmel, and Yaniv Romano. Conformalized survival analysis for general right-censored data. In *ICLR*, 2025.

Aryeh Dvoretzky. Asymptotic normality for sums of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, volume 6.2, pages 513–536. University of California Press, 1972.

Dennis Frauen, Maresa Schröder, Konstantin Hess, and Stefan Feuerriegel. Orthogonal survival learners for estimating heterogeneous treatment effects from time-to-event data. In *NeurIPS*, 2025.

Zijun Gao and Trevor Hastie. Estimating heterogeneous treatment effects for general responses. *Biometrics*, 81(4):ujaf162, 2025.

Yu Gui, Rohan Hore, Zhimei Ren, and Rina Foygel Barber. Conformalized survival analysis with adaptive cut-offs. *Biometrika*, 111(2):459–477, 2024.

FDA Guidance. Adaptive design clinical trials for drugs and biologics. *Biotechnol Law Rep*, 29(2): 173, 2010.

Jinyong Hahn, Keisuke Hirano, and Dean Karlan. Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, 29(1):96–108, 2011.

Nicholas C. Henderson, Thomas A. Louis, Gary L. Rosner, and Ravi Varadhan. Individualized treatment effects with censored data via fully nonparametric Bayesian accelerated failure time models. *Biostatistics*, 21(1):50–68, 2020.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17, 2020.

Feifang Hu and William F. Rosenberger. *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley & Sons, 2006.

Liangyuan Hu, Jiayi Ji, and Fan Li. Estimating heterogeneous survival treatment effect in observational data using machine learning. *Statistics in Medicine*, 40(21):4691–4713, 2021.

Guido W. Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86(1):4–29, 2004.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press, 2015.

Masahiro Kato, Akihiro Oga, Wataru Komatsubara, and Ryo Inokuchi. Active adaptive experimental design for treatment effect estimation with covariate choices. In *ICML*, 2024.

Jared L. Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Medical Research Methodology*, 18(1):24, 2018.

Emilie Kaufmann and Aurélien Garivier. Learning the distribution with largest mean: two bandit frameworks. *ESAIM: Proceedings and Surveys*, 60:114–131, 2017.

Edward S. Kim, Roy S. Herbst, Ignacio I. Wistuba, J. Jack Lee, George R. Blumenschein, Anne Tsao, David J. Stewart, Marshall E. Hicks, Jeremy Erasmus, Sanjay Gupta, Christine M. Alden, Suyu Liu, Ximing Tang, Fadlo R. Khuri, Hai T. Tran, Bruce E. Johnson, John V. Heymach, Li Mao, Frank Fossella, Merrill S. Kies, Vassiliki Papadimitrakopoulou, Suzanne E. Davis, Scott M. Lippman, and Waun K. Hong. The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discovery*, 1(1):44–53, 2011.

John P. Klein and Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer, 2003.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 1 edition, 2020.

Jeonghwan Lee and Cong Ma. Off-policy estimation with adaptively collected data: the power of online learning. In *NeurIPS*, 2024.

Jiachun Li, David Simchi-Levi, and Yunxiao Zhao. Optimal adaptive experimental design for estimating treatment effect. *arXiv preprint*, arXiv:2410.05552, 2024.

Michael Lindon and Nathan Kallus. Anytime-valid inference under outcome delay: A design-based approach. *arXiv preprint*, arXiv:2603.25971, 2026.

Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In *NeurIPS*, 2017.

Huzhang Mao, Liang Li, Wei Yang, and Yu Shen. On the propensity score weighting analysis with survival outcome: Estimands, estimation, and inference. *Statistics in Medicine*, 37(26):3745–3763, 2018.

Ayon Mukherjee, Sayantee Jana, and Stephen Coad. Covariate-adjusted response-adaptive designs for semiparametric survival models. *Statistical Methods in Medical Research*, 34(9):1697–1723, 2025.

Ojash Neopane, Aaditya Ramdas, and Aarti Singh. Logarithmic neyman regret for adaptive estimation of the average treatment effect. In *AISTATS*, 2025a.

Ojash Neopane, Aaditya Ramdas, and Aarti Singh. Optimistic algorithms for adaptive estimation of the average treatment effect. In *ICML*, 2025b.

Jerzy Neyman. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4): 558–625, 1934.

Georgy Noarov, Riccardo Fogliato, Martin Andres Bertran, and Aaron Roth. Stronger neyman regret guarantees for adaptive experimental design. In *ICML*, 2025.

Miruna Oprescu, Brian M. Cho, and Nathan Kallus. Efficient adaptive experimentation with noncompliance. In *NeurIPS*, 2025.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *arXiv preprint*, arXiv:2210.01948, 2022.

David S. Robertson, Kim May Lee, Boryana C. López-Kolkovska, and Sofía S. Villar. Response-adaptive randomization in clinical trials: From myths to practical considerations. *Statistical Science*, 38(2), 2023.

James M Robins. Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the American Statistical Association*, volume 1999, pages 6–10. Indianapolis, IN, 2000.

W. F. Rosenberger, A. N. Vidyashankar, and D. K. Agarwal. Covariate-adjusted response-adaptive designs for binary response. *Journal of Biopharmaceutical Statistics*, 11(4):227–236, 2001.

William F. Rosenberger and Oleksandr Sverdlov. Handling covariates in the design of clinical trials. *Statistical Science*, 23(3):404–419, 2008.

Daniel Rubin and Mark J. van der Laan. A doubly robust censoring unbiased transformation. *The International Journal of Biostatistics*, 3(1):Article 4, 2007.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688, 1974.

Douglas E. Schaubel and Guanghui Wei. Double inverse-weighted estimation of cumulative treatment effects under nonproportional hazards and dependent censoring. *Biometrics*, 67(1):29–38, 2011.

Stefan Schrod, Andreas Schäfer, Stefan Solbrig, Robert Lohmayer, Wolfram Gronwald, Peter J. Oefner, Tim Beißbarth, Rainer Spang, Helena U. Zacharias, and Michael Altenbuchinger. BITES: balanced individual treatment effect for survival data. *Bioinformatics*, 38(Supplement_1):i60–i67, 2022.

Sami Tabib and Denis Larocque. Non-parametric individual treatment effect estimation for survival data with random forests. *Bioinformatics (Oxford, England)*, 36(2):629–636, 2020.

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

Wouter AC van Amsterdam, Michael Oberst, Jean Feng, Jenna Wiens, Shengpu Tang, Shalmali Joshi, Rajesh Ranganath, Mark Sendak, Uri Shalit, Julia E Vogt, et al. Clinical trials for continuously monitored and updated AI systems. *Nature medicine*, pages 1–3, 2026.

Mark J. van der Laan. The construction and analysis of adaptive group sequential designs. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 2008.

Mark J. van der Laan and James M. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. Springer Series in Statistics. Springer, New York, NY, 2003.

Mark J. van der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1), 2006.

Stefan Wager. Causal inference: A statistical learning approach. *Preprint*, 2024.

Ian Waudby-Smith, David Arbour, Ritwik Sinha, Edward H. Kennedy, and Aaditya Ramdas. Time-uniform central limit theory and asymptotic confidence sequences. *The Annals of Statistics*, 52(6): 2613–2640, 2024.

Ted Westling, Alex Luedtke, Peter B. Gilbert, and Marco Carone. Inference for treatment-specific survival curves using machine learning. *Journal of the American Statistical Association*, 119(546): 1541–1553, 2024.

Simon Wiegrebe, Philipp Kopper, Raphael Sonabend, Bernd Bischl, and Andreas Bender. Deep learning for survival analysis: A review. *Artificial Intelligence Review*, 57(3):65, 2024.

Shenbo Xu, Raluca Cobzaru, Stan N. Finkelstein, Roy E. Welsch, Kenney Ng, and Zach Shahn. Estimating heterogeneous treatment effects on survival outcomes using counterfactual censoring unbiased transformations. *arXiv preprint*, arXiv:2401.11263, 2024.

Yizhe Xu, Nikolaos Ignatiadis, Erik Sverdrup, Scott Fleming, Stefan Wager, and Nigam Shah. Treatment heterogeneity for survival outcomes. *arXiv preprint*, arXiv:2207.07758, 2022.

Kelly Zhang, Lucas Janson, and Susan Murphy. Statistical inference with m-estimators on adaptively collected data. In *NeurIPS*, 2021.

Li-Xin Zhang, Feifang Hu, Siu Hung Cheung, and Wai Sum Chan. Asymptotic properties of covariate-adjusted response-adaptive designs. *The Annals of Statistics*, 35(3):1166–1182, 2007.

Weijia Zhang, Thuc Duy Le, Lin Liu, Zhi-Hua Zhou, and Jiuyong Li. Mining heterogeneous causal effects for personalized cancer treatment. *Bioinformatics*, 33(15):2372–2378, 2017.

Wenxin Zhang and Mark.J van der Laan. Efficient statistical estimation for sequential adaptive experiments with implications for adaptive designs. *arXiv preprint*, arXiv:2508.09135, 2025.

A NOTATION

Table 1: Notation

Symbol	Description
X	Observed covariates, $X \in \mathcal{X} \subseteq \mathbb{R}^p$
A	Binary treatment, $A \in \{0, 1\}$
T	Event time of interest (e.g., overall survival time (OS), disease-free survival time (DFS))
C	Censoring time (e.g., dropout)
$\widetilde{T}$	Observed time, $\widetilde{T} = \min\{T, C\}$
Δ	Event indicator, $\Delta = \mathbf{1}(T \le C)$
$\mathcal{O}$	Observed population, $\mathcal{O} = (X, A, \widetilde{T}, \Delta)$
$\mathcal{H}_r$	Observed accumulated data at round r, $\mathcal{H}_r = \{(x_1, a_1, \tilde{t}_1, \delta_1), \ldots, (x_r, a_r, \tilde{t}_r, \delta_r)\}$
$\boldsymbol{\tau}$, τ_t, $\tau_t(x)$	Average survival effect curve vector $\tau = (\tau_0, \ldots, \tau_{t_{\max}})^\top$ and its t-th entry $\tau_t = \mathbb{E}[S_t(X, 1) - S_t(X, 0)]$ Conditional ATE at horizon t: $\tau_t(x) = \mathbb{P}(T(1) > t \mid X = x) - \mathbb{P}(T(0) > t \mid X = x)$
$\hat{\tau}_{t,R}^{\mathrm{ASE}}$	ASE estimator at horizon t after R rounds
$\boldsymbol{S}(x, a)$, $S_t(x, a)$	Survival function vector $\boldsymbol{S}(x, a) = (S_0(x, a), \ldots, S_{t_{\max}}(x, a))^\top$ and its t-th entry $S_t(x, a) = \mathbb{P}(T > t \mid X = x, A = a)$
$\hat{\boldsymbol{S}}_r(x, a)$, $\hat{S}_{t,r}(x, a)$, $\hat{S}_{t,r}^{(-J_r)}(x, a)$	Estimated survival function vector, its t-th entry, and its cross-fitted version using fold $\mathcal{H}_{r-1}^{(-J_r)}$
$\boldsymbol{G}(x, a)$, $G_t(x, a)$	Censoring survival function vector $\boldsymbol{G}(x, a) = (G_0(x, a), \ldots, G_{t_{\max}}(x, a))^\top$ and its t-th entry $G_t(x, a) = \mathbb{P}(C > t \mid X = x, A = a)$
$\hat{\boldsymbol{G}}_r(x, a)$, $\hat{G}_{t,r}(x, a)$, $\hat{G}_{t,r}^{(-J_r)}(x, a)$	Estimated censoring survival function vector, its t-th entry, and its cross-fitted version using fold $\mathcal{H}_{r-1}^{(-J_r)}$
$\boldsymbol{\lambda}^S(x, a)$, $\lambda_t^S(x, a)$	Event hazard vector $\boldsymbol{\lambda}^S(x, a) = (\lambda_0^S(x, a), \ldots, \lambda_{t_{\max}}^S(x, a))^\top$ and its t-th entry $\lambda_t^S(x, a) = \mathbb{P}(\widetilde{T} = t, \Delta = 1 \mid \widetilde{T} \ge t, X = x, A = a)$
$\hat{\boldsymbol{\lambda}}_r^S(x, a)$, $\hat{\lambda}_{t,r}^S(x, a)$, $\hat{\lambda}_{t,r}^{S,(-J_r)}(x, a)$	Estimated event hazard vector, its t-th entry, and its cross-fitted version using fold $\mathcal{H}_{r-1}^{(-J_r)}$
$\boldsymbol{\lambda}^G(x, a)$, $\lambda_t^G(x, a)$	Censoring hazard vector $\boldsymbol{\lambda}^G(x, a) = (\lambda_0^G(x, a), \ldots, \lambda_{t_{\max}}^G(x, a))^\top$ and its t-th entry $\lambda_t^G(x, a) = \mathbb{P}(\widetilde{T} = t, \Delta = 0 \mid \widetilde{T} \ge t, X = x, A = a)$
$\hat{\boldsymbol{\lambda}}_r^G(x, a)$, $\hat{\lambda}_{t,r}^G(x, a)$, $\hat{\lambda}_{t,r}^{G,(-J_r)}(x, a)$	Estimated censoring hazard vector, its t-th entry, and its cross-fitted version using fold $\mathcal{H}_{r-1}^{(-J_r)}$
η_t, $\hat{\eta}_{t,r-1}$, $\hat{\eta}_{t,r-1}^{(-J_r)}$	Oracle nuisance vector $\eta_t = \{\lambda_i^S(\cdot, \cdot), \lambda_i^G(\cdot, \cdot)\}_{i=0}^t$ and its estimate from $\mathcal{H}_{r-1}$: $\hat{\eta}_{t,r-1} = \{\hat{\lambda}_{i,r-1}^S(\cdot, \cdot), \hat{\lambda}_{i,r-1}^G(\cdot, \cdot)\}_{i=0}^t$ and its cross-fitted version trained on fold $\mathcal{H}_{r-1}^{(J_r)}$
$\boldsymbol{\xi}(\mathcal{O}, \eta_{t_{\max}})$	IPCW augmentation vector: $\boldsymbol{\xi}(\mathcal{O}, \eta_{t_{\max}}) = (\xi(\mathcal{O}, \eta_0), \ldots, \xi(\mathcal{O}, \eta_{t_{\max}}))^\top$
$\xi(\mathcal{O}, \eta_t)$, $\xi(\mathcal{H}_r, \eta_t)$	IPCW augmentation term in the EIF from observed population and accumulated data, respectively
$\hat{\xi}(\mathcal{O}, \eta_t)$, $\hat{\xi}(\mathcal{H}_r, \eta_t)$, $\hat{\xi}(\mathcal{H}_r, \eta_t^{\mathrm{cf}})$	Estimated IPCW augmentation term in the EIF from observed population and accumulated data, and with sequential cross-fitting, respectively.
$\boldsymbol{\phi}(\cdot; \tau, \eta_{t_{\max}})$, $\phi_t(\cdot; \tau_t, \eta_t)$	EIF vector $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{t_{\max}})^\top$ and its t-th entry (non-centered EIF at horizon t, see Eq. 4)
$\phi_{t,r}$, $\hat{\phi}_{t,r}$, $\hat{\phi}_{t,r}^{\mathrm{cf}}$	Denotes non-centered EIF with oracle nuisances $\phi_{t,r} = \phi(\cdot; \tau_t, \pi_r, \eta_t)$ non-centered EIF with estimated nuisances $\hat{\phi}_{t,r} = \phi(\cdot; \tau_t, \pi_r, \hat{\eta}_{t,r})$ $\hat{\phi}_{t,r}^{\mathrm{cf}}$ denotes the cross-fitted nuisance at round r
$\pi(x)$	Treatment assignment probability: $\mathbb{P}(A = 1 \mid X = x)$
$\pi^\star(x)$	A-optimal treatment allocation policy (Proposition 4.3)
$\pi_r(x \mid \mathcal{H}_{r-1})$	Adaptive policy at round r
$\widetilde{\pi}_r(x \mid \mathcal{H}_{r-1})$	Plug-in (pre-truncation) policy at round r
Π_α	Feasible policy class: $\{\pi : \alpha \le \pi(x) \le 1 - \alpha\}$
R, R_0	Total number of rounds and burn-in period (initial exploration)
k_r	Truncation parameter at round r (ensures $\pi_r \in [1/k_r, 1 - 1/k_r]$)
J_r	Sequential cross-fitting fold index, $J_r := r \bmod 2$
$\mathcal{H}_{r-1}^{(j)}$	Temporal cross-fitting fold $j \in \{0, 1\}$
$\Sigma_{\mathrm{eff}}(\pi)$, $\Sigma_a(X)$	Semiparametric efficiency bound matrix $\Sigma_{\mathrm{eff}}(\pi) = \mathbb{E}\big[(\boldsymbol{\phi}(\mathcal{O}; \tau, \eta_{t_{\max}}) - \tau)(\boldsymbol{\phi}(\mathcal{O}; \tau, \eta_{t_{\max}}) - \tau)^\top\big] \in \mathbb{R}^{(t_{\max}+1)\times(t_{\max}+1)}$ $\Sigma_a(X) = \mathrm{Var}(\boldsymbol{S}(X, a) \odot \boldsymbol{\xi}(\mathcal{O}; \boldsymbol{\eta}) \mid X, A = a) \in \mathbb{R}^{(t_{\max}+1)\times(t_{\max}+1)}$
$v_{t,a}(x)$, $V_a(x)$	t-th component of $\Sigma_a(X)$, Variance target $V_a(x) = \sum_{t=0}^{t_{\max}} v_{t,a}(x) = \mathrm{tr}(\Sigma_a(x))$
$V_{\mathrm{eff},t}(\pi)$	and its t-th diagonal entry $\Sigma_{\mathrm{eff}}(\pi) = [\Sigma_{\mathrm{eff}}(\pi)]_{tt}$
$\|\cdot\|_2$	L_2 norm: $\|g\|_2 := \mathbb{E}[g(X)^2]^{1/2}$
$o_p(1)$, $O_p(\cdot)$	Standard probabilistic asymptotic order notation

B EXTENDED RELATED WORK

Literature stream	Allocation objective	Censored outcome	Efficiency bound	Model-agnostic
RAR / CARA designs (Hu and Rosenberger, 2006; Rosenberger et al., 2001; Thompson, 1933)	welfare optimization	✗	✗	✗[†]
CARA designs for survival (Mukherjee et al., 2025)	welfare optimization	✓	✗	✗[†]
Adaptive ATE estimation (two-stage) (Hahn et al., 2011)	estimation of outcome variability	✗	✓	✓
Adaptive ATE estimation (multi-round) (Cook et al., 2024; Dai et al., 2023; Kato et al., 2024)	estimation of outcome variability	✗	✓	✓
Ours	estimation of survival & censoring variability	✓	✓	✓

[†] RAR/CRAR rely on parametric response models; Mukherjee et al. (2025) is Cox proportional hazards (PH) model specific.

Table 2: **Overview of adaptive randomization literature streams.**

Estimation of treatment effects with censored outcomes. A related line of work extends causal inference to survival settings, focusing on estimating treatment effects such as the CATE under censoring (Hu et al., 2021). Methods here divide broadly into (i) model-based learners, including tree-based (Cui et al., 2023; Henderson et al., 2020; Tabib and Larocque, 2020; Zhang et al., 2017) and neural-network-based approaches (Curth and van der Schaar, 2021; Katzman et al., 2018; Schrod et al., 2022), and (ii) (orthogonal) meta-learners specifically designed for censored time-to-event data (Frauen et al., 2025; Gao and Hastie, 2025; van der Laan and Robins, 2003; Xu et al., 2024; 2022). While these methods provide principled estimation under censoring, they operate under a *fixed* treatment assignment mechanism and focus on point estimation, typically requiring censoring-independence assumptions that limit applicability. A complementary line develops inferential methods via conformal inference (Candès et al., 2023; Davidov et al., 2025; Gui et al., 2024), thereby targeting prediction intervals for individual treatment effects rather than population-level causal estimands. $\Rightarrow$ Neither stream addresses adaptive experimental design, and neither leverages semiparametric efficiency bounds to guide treatment allocation under censoring.

Bayesian response-adaptive randomization. In parallel, Bayesian RAR (BRAR) and Thompson sampling (Kaufmann and Garivier, 2017; Lattimore and Szepesvári, 2020) have seen substantial uptake in both the machine learning and clinical trial communities (Barker et al., 2009; Kim et al., 2011), driven by empirical gains in system performance and patient welfare. However, BRAR methods are primarily designed to maximize reward or accelerate identification of superior treatments, rather than to minimize the asymptotic variance of a causal estimand. As a result, they do not target the semiparametric efficiency bound, and their allocation rules do not account for the additional variance inflation induced by censoring. Our framework is complementary in spirit but distinct in objective: we derive allocation policy that are provably optimal with respect to the efficiency bound for the average survival effect curve.

CRAR with survival outcomes: The closest work to ours is Mukherjee et al. (2025), who propose CRAR designs for survival data under a proportional hazards assumption, using a sequentially estimated Cox model to drive allocation. However, their approach is driven by model-based estimates rather than optimizing allocation for statistical efficiency, and therefore does not target the semiparametric efficiency bound.

Adaptive allocation for efficient ATE estimation. There are several extensions that add principled policy truncation with anytime-valid inference (Cook et al., 2024), and establish sublinear or logarithmic Neyman regret via clipped and optimistic algorithms (Chen et al., 2025; Dai et al., 2023; Neopane et al., 2025a;b; Noarov et al., 2025). Further work covers covariate-adaptive and off-policy settings (Lee and Ma, 2024; Li et al., 2024) and develops TMLE-based guarantees for sequential experiments (Zhang and van der Laan, 2025).

Confidence sequences and anytime-valid inference. Unlike fixed-time confidence intervals, a confidence sequence (CS) provides a time-uniform coverage guarantee, thereby enabling valid inference at any data-dependent stopping time (Howard et al., 2020). Martingale methods and empirical Bernstein inequalities have been central to recent constructions of CSs for influence-function-based estimators (Cook et al., 2024; Waudby-Smith et al., 2024), and Lindon and Kallus (2026) extends this to the delayed outcome regime. Our approach builds on these tools to construct an asymptotic CS for the ASE estimator, where the key challenge is simultaneously accommodating adaptive allocation, right censoring, and the almost-sure consistency requirements that time-uniform inference imposes on the nuisance remainder.

C ADDITIONAL ILLUSTRATIVE EXAMPLE

Consider an example with *arm-dependent censoring*, the two treatment arms experience different rates of dropout, as when treatment improves retention ($G_t(X,1) > G_t(X,0)$) or induces side effects leading to dropout ($G_t(X,1) < G_t(X,0)$). Suppose the censoring survival functions have time-invariant ratio, $G_t(x,0)/G_t(x,1) = g(x)$ for all t, while event dynamics $S_t(x,a) = s_{t,a}$ may differ across arms. Proposition 4.3 yields the closed-form optimum $\pi^\star(x) = \mathrm{clip}_\alpha\left(\frac{\sqrt{K_1}}{\sqrt{K_0/g_t(x)}+\sqrt{K_1}}\right)$, where $K_a(x)$ captures arm-a event dynamics and is censoring-independent up to a common factor that cancels in $\pi^\star$ (proof as following). Figure 4 plots $\pi^\star$ against difference value of $g(x) = G_t(X,0)/G_t(X,1)$, with closed form $\pi^\star(x) = \sqrt{g(x)}/(\sqrt{\kappa}+\sqrt{g(x)})$ where $\kappa := K_0/K_1$ captures the event-information ratio across arms. Under heavier censoring in one arm, allocating more units to that arm compensates for the censoring variance inflation, preserving estimator efficiency.

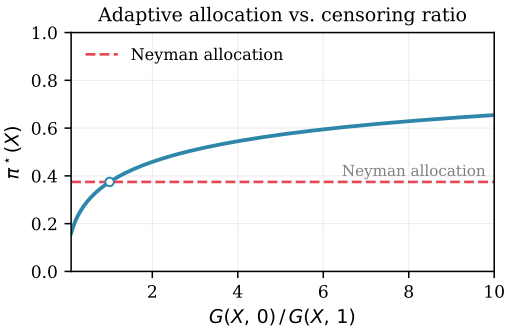


Figure 4: Optimal allocation $\pi^\star$ under arm-dependent censoring ratio $g(x) = G_t(x,0)/G_t(x,1)$: $\pi^\star$ departs from uniform allocation $1/2$ and Neyman-allocation as the censoring asymmetry increases.

Proof. We derive the closed-form expression for $\pi^\star$ claimed in the motivating illustration above. Assume that the event dynamics may differ across arms and that the censoring survival functions have a time-invariant ratio: $S_t(x,a) = s_{t,a}$, $G_{i-1}(x,0)/G_{i-1}(x,1) = g(x)$ for all i. First, we write out the expression from Eq. 6 as

$$v_{t,a}(x) \;=\; S_t(x,a)^2 \sum_{i=0}^{t} \frac{\lambda_i^S(x,a)}{S_i(x,a)\, G_{i-1}(x,a)} \;=\; \frac{1}{g(x)^{\mathbf{1}\{a=0\}}} \cdot s_{t,a}^2 \sum_{i=0}^{t} \frac{\lambda_{i,a}}{s_{i,a}\, G_{i-1}(x,1)}, \tag{17}$$

where we pulled the time-constant factor $g(x)$ out of the sum (for $a = 0$) using the ratio assumption $G_{i-1}(x,0) = g(x)\, G_{i-1}(x,1)$. Then, we prove that

$$s_{t,a}^2 \sum_{i=0}^{t} \frac{\lambda_{i,a}}{s_{i,a}} \;=\; s_{t,a}(1 - s_{t,a}). \tag{18}$$

The key observation is that $\lambda_{i,a}/s_{i,a}$ is a first-order difference of $1/s_{i,a}$. Using $s_{i,a} = s_{i-1,a}(1-\lambda_{i,a})$, we have $s_{i-1,a} - s_{i,a} = s_{i-1,a}\lambda_{i,a}$, so

$$\frac{1}{s_{i,a}} - \frac{1}{s_{i-1,a}} \;=\; \frac{s_{i-1,a} - s_{i,a}}{s_{i,a}\, s_{i-1,a}} \;=\; \frac{\lambda_{i,a}}{s_{i,a}} \qquad (i \geq 1). \tag{19}$$

Then, we have

$$\sum_{i=0}^{t} \frac{\lambda_{i,a}}{s_{i,a}} \;=\; \frac{1}{s_{t,a}} - \frac{1}{s_{-1,a}} \;=\; \frac{1 - s_{t,a}}{s_{t,a}}, \tag{20}$$

and multiplying by $s_{t,a}^2$ yields Eq. 18. By absorbing the shared time-varying factor $G_{i-1}(x,1)$ into a common weighting that cancels in the final ratio, we write

$$V_a(x) \;=\; \sum_{t=0}^{t_{\max}} v_{t,a}(x) \;=:\; \frac{K_a(x)}{g(x)^{\mathbf{1}\{a=0\}}}, \qquad K_a(x) := \sum_{t=0}^{t_{\max}} s_{t,a}^2 \sum_{i=0}^{t} \frac{\lambda_{i,a}}{s_{i,a}\, G_{i-1}(x,1)}, \tag{21}$$

so that $V_1(x) = K_1(x)$ and $V_0(x) = K_0(x)/g(x)$. Substituting into Proposition 4.3, we arrive at

$$\frac{\sqrt{V_1(x)}}{\sqrt{V_1(x)} + \sqrt{V_0(x)}} \;=\; \frac{\sqrt{K_1}}{\sqrt{K_1} + \sqrt{K_0/g(x)}}. \tag{22}$$

Applying the overlap constraint, we yield

$$\pi^\star(x) \;=\; \mathrm{clip}_\alpha\left(\frac{\sqrt{K_1(x)}}{\sqrt{K_0(x)/g(x)} + \sqrt{K_1(x)}}\right). \tag{23}$$

□

D ADDITIONAL THEORETICAL RESULTS FOR OPTIMAL DATA ACQUISITION

In the main text, we focus on *A-optimal* allocation, i.e., minimizing $\mathrm{tr}(\Sigma_{\mathrm{eff}}(\pi))$, which targets small marginal variances across horizons and admits the closed-form solution in Proposition 4.3. Alternative classical criteria from optimal experimental design are *D-optimality* and *E-optimality*, which target the joint uncertainty of the full survival effect curve vector $\tau = (\tau_0, \ldots, \tau_{t_{\max}})^\top$ (Atkinson et al., 2023), or control the worst-case linear combination of survival effect curve contrasts. We extend our derivations to D-optimality and E-optimality below.

D.1 D-OPTIMALITY

Definition D.1 (D-optimal allocation policy)**.** *A feasible allocation policy π^D is* D-optimal *if*

$$\pi^D \in \arg\min_{\pi \in \Pi_\alpha} \det(\Sigma_{\mathrm{eff}}(\pi)) \quad \textit{equivalently} \quad \pi^D \in \arg\min_{\pi \in \Pi_\alpha} \log\det(\Sigma_{\mathrm{eff}}(\pi)), \tag{24}$$

where $\Pi_\alpha = \{\pi : \alpha \le \pi(x) \le 1 - \alpha \textit{ for all } x\}$.

Interpretation: The confidence ellipsoid for τ has volume proportional to $\det(\Sigma_{\mathrm{eff}}(\pi))^{1/2}$ (up to constants independent of π). Hence, D-optimal allocation tends to yield small *joint* uncertainty across all time horizons simultaneously, which is desirable when the full survival effect curve is the object of interest rather than any individual τ_t.

Theorem D.2 (D-optimal allocation policy)**.** *Assume $\Pi_\alpha \neq \emptyset$ and that Assumptions 3.1 and 3.2 hold. Then, any D-optimal policy π^D satisfies, for a.e. x,*

$$\pi^D(x) = \mathrm{clip}_\alpha\left(\frac{\sqrt{\mathrm{tr}(\Sigma_{\mathrm{eff}}(\pi^D)^{-1}\Sigma_1(x))}}{\sqrt{\mathrm{tr}(\Sigma_{\mathrm{eff}}(\pi^D)^{-1}\Sigma_1(x))} + \sqrt{\mathrm{tr}(\Sigma_{\mathrm{eff}}(\pi^D)^{-1}\Sigma_0(x))}}\right). \tag{25}$$

The characterization is a fixed-point equation since $\Sigma_{\mathrm{eff}}(\pi^D)$ depends on π^D.

Remark D.3. The form of π^D parallels the A-optimal rule in Proposition 4.3, but with $V_a(X) = \mathrm{tr}(\Sigma_a(X))$ replaced by $\mathrm{tr}(\Sigma_{\mathrm{eff}}(\pi^D)^{-1}\Sigma_a(x))$. This difference has a clean calculus origin: differentiating $\mathrm{tr}(\Sigma_{\mathrm{eff}}(\pi))$ w.r.t. $\pi(x)$ yields $\mathrm{tr}(\Sigma_a(x))$, whereas differentiating $\log\det(\Sigma_{\mathrm{eff}}(\pi))$ via the identity $\frac{\mathrm{d}}{\mathrm{d}A}\log\det(A) = \mathrm{tr}(A^{-1}dA)$ yields $\mathrm{tr}(\Sigma_{\mathrm{eff}}^{-1}\Sigma_a(x))$. The $\Sigma_{\mathrm{eff}}^{-1}$ weighting introduces a global dependence on the full covariance structure: directions already well-estimated are upweighted, redirecting allocation toward reducing joint uncertainty across horizons. This is precisely what prevents π^D from admitting a closed-form expression, since $\Sigma_{\mathrm{eff}}(\pi^D)$ depends on π^D itself, making the characterization a fixed-point equation.

Computation via fixed-point iteration: Unlike the A-optimal rule in Proposition 4.3, the D-optimal policy depends on $\Sigma_{\mathrm{eff}}(\pi^D)^{-1}$ and has therefore no closed-form solution. A simple solver is a fixed-point iteration:

1. Initialize $\pi^{(0)}$ (e.g., the A-optimal policy from Proposition 4.3).
2. For $k = 0, 1, 2, \ldots$:
   (a) Estimate $\Sigma_{\mathrm{eff}}(\pi^{(k)})$ using the accumulated data $\mathcal{H}_{r-1}$ or the population formula if available.
   (b) Update

$$\tilde{\pi}^{(k+1)}(x) = \mathrm{clip}_\alpha\left(\frac{\sqrt{\mathrm{tr}\big(\Sigma_{\mathrm{eff}}(\pi^{(k)})^{-1}\Sigma_1(x)\big)}}{\sqrt{\mathrm{tr}\big(\Sigma_{\mathrm{eff}}(\pi^{(k)})^{-1}\Sigma_1(x)\big)} + \sqrt{\mathrm{tr}\big(\Sigma_{\mathrm{eff}}(\pi^{(k)})^{-1}\Sigma_0(x)\big)}}\right). \tag{26}$$

   (c) Set $\pi^{(k+1)} \leftarrow \tilde{\pi}^{(k+1)}$ and stop when the iterates stabilize.

D.2 E-OPTIMALITY

E-optimality minimizes the largest eigenvalue of $\Sigma_{\mathrm{eff}}(\pi)$ and thereby controls the worst-case variance over all unit-norm linear contrasts $v^\top \tau$.

Definition D.4 (E-optimal allocation policy)**.** *A feasible allocation policy π^E is* E-optimal *if*

$$\pi^E \in \arg\min_{\pi \in \Pi_\alpha} \lambda_{\max}(\Sigma_{\text{eff}}(\pi)), \tag{27}$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue.

Interpretation: By the variational characterization $\lambda_{\max}(\Sigma_{\text{eff}}(\pi)) = \max_{\|v\|=1} v^\top \Sigma_{\text{eff}}(\pi) v$, the E-optimal design minimizes the variance of the worst-case linear combination of survival effect curve contrasts $v^\top \tau$. This is the natural criterion when the experimenter lacks a prior preference over individual horizons τ_t and wishes to guard against the worst-case linear contrast.

Theorem D.5 (E-optimal allocation policy)**.** *Under the same conditions as Theorem D.2, any E-optimal policy π^E satisfies, for a.e. x,*

$$\pi^E(x) = \text{clip}_\alpha\left(\frac{\sqrt{(v^\star)^\top \Sigma_1(x)\, v^\star}}{\sqrt{(v^\star)^\top \Sigma_1(x)\, v^\star} + \sqrt{(v^\star)^\top \Sigma_0(x)\, v^\star}}\right), \tag{28}$$

where $v^\star = v^\star(\pi^E) \in \mathbb{R}^{t_{\max}+1}$ is the unit eigenvector corresponding to $\lambda_{\max}(\Sigma_{\text{eff}}(\pi^E))$. The characterization is a fixed-point equation since $v^\star$ depends on π^E through $\Sigma_{\text{eff}}(\pi^E)$.

By the envelope theorem, at an interior optimum the first-order condition w.r.t. $\pi(x)$ yields $(v^\star)^\top[\partial\Sigma_{\text{eff}}(\pi)/\partial\pi(x)]v^\star = 0$, which gives $-(v^\star)^\top\Sigma_1(x)v^\star/\pi(x)^2 + (v^\star)^\top\Sigma_0(x)v^\star/(1-\pi(x))^2 = 0$. Solving for $\pi(x)$ and applying the overlap constraint yields the stated expression.

Remark D.6. The quantity $(v^\star)^\top\Sigma_a(x)v^\star = \text{Var}\big(\sum_t v_t^\star S_t(X,a)\,\xi_t(\mathcal{O},\eta_t) \mid X=x, A=a\big)$ is the conditional variance of the worst-case linear contrast $v^{\star\top}\phi(\mathcal{O};\tau,\eta)$ in arm a at covariate value x. Thus, the E-optimal policy has the same Neyman form as the A- and D-optimal policies, but with the scalar $V_a(x)$ replaced by the worst-case directional variance $(v^\star)^\top\Sigma_a(x)v^\star$.

Computation via fixed-point iteration: The E-optimal policy can be computed analogously to the D-optimal case:

1. Initialize $\pi^{(0)}$ (e.g., the A-optimal policy from Proposition 4.3).
2. For $k = 0, 1, 2, \ldots$:
   (a) Estimate $\Sigma_{\text{eff}}(\pi^{(k)})$ from accumulated data $\mathcal{H}_{r-1}$.
   (b) Compute the leading eigenvector $v^{(k)} = \arg\max_{\|v\|=1} v^\top \Sigma_{\text{eff}}(\pi^{(k)})v$.
   (c) Update

$$\tilde{\pi}^{(k+1)}(x) = \text{clip}_\alpha\left(\frac{\sqrt{(v^{(k)})^\top \Sigma_1(x)\, v^{(k)}}}{\sqrt{(v^{(k)})^\top \Sigma_1(x)\, v^{(k)}} + \sqrt{(v^{(k)})^\top \Sigma_0(x)\, v^{(k)}}}\right). \tag{29}$$

   (d) Set $\pi^{(k+1)} \leftarrow \tilde{\pi}^{(k+1)}$ and stop when the iterates stabilize.

D.3 Comparison of A-, D-, and E-optimality

All three criteria share the same Neyman allocation structure $\pi^\star(x) = \sqrt{Q_1(x)}/(\sqrt{Q_1(x)} + \sqrt{Q_0(x)})$ for an arm-specific scalar $Q_a(x)$, but differ in how they aggregate the cross-horizon covariance matrix $\Sigma_a(x)$:

1. A-optimality uses $Q_a(x) = \text{tr}(\Sigma_a(x)) = V_a(x)$ (sum of marginal variances, closed-form);
2. D-optimality uses $Q_a(x) = \text{tr}(\Sigma_{\text{eff}}^{-1}\Sigma_a(x))$ (precision-weighted trace, fixed-point);
3. E-optimality uses $Q_a(x) = (v^\star)^\top\Sigma_a(x)v^\star$ (worst-case directional variance, fixed-point).

When $t_{\max} = 0$, $v^\star = 1$ and, therefore, all three criteria coincide. The A-optimal criterion is the only one that separates additively over t, which is why it yields a closed-form solution; D- and E-optimal criteria couple all horizons jointly through Σ_{eff} and its eigenvector and therefore require iterative computation.

E ALGORITHM

Algorithm 1 Adaptive Survival Estimator

Input: Burn-in period R_0; initial policy $\pi_{\text{init}}(X)$; truncation schedule $(k_r)_{r\geq 1}$; regression learners for $\lambda_t^S(X,a)$, $\lambda_t^G(X,a)$

for $r \leftarrow 1$ **to** R **do**
  Observe covariates X_r
  if $r \leq R_0$ **then**
    Assign $A_r \sim \text{Bern}(\pi_{\text{init}}(X_r))$
  else
    Estimate nuisance functions $\hat{\lambda}_{t,r-1}^{S,\text{cf}}(X,a)$, $\hat{\lambda}_{t,r-1}^{G,\text{cf}}(X,a)$ from $\mathcal{H}_{r-1}$ using cross-fitting
    Compute $\widehat{S}_{t,r-1}^{\text{cf}}(X,a) = \prod_{i=0}^{t}(1-\widehat{\lambda}_{i,r-1}^{S}(X,a))$
    Compute $\widehat{G}_{t-1,r-1}^{\text{cf}}(X,a) = \prod_{i=0}^{t-1}\Big(1-\widehat{\lambda}_{i,r-1}^{G}(X,a)/[1-\widehat{\lambda}_{i,r-1}^{S}(X,a)]\Big)$
    Compute $\hat{V}_{a,r-1}(X_r)$ for $a \in \{0,1\}$ using Eq. 10
    Compute plug-in assignment probability:
    $\widetilde{\pi}_r(X_r \mid \mathcal{H}_{r-1}) = \sqrt{\hat{V}_{1,r-1}(X_r)}/(\sqrt{\hat{V}_{0,r-1}(X_r)} + \sqrt{\hat{V}_{1,r-1}(X_r)})$
    Apply truncation:
    $\pi_r(X_r \mid \mathcal{H}_{r-1}) \leftarrow \min\Big\{1-\frac{1}{k_r}, \max\Big\{\frac{1}{k_r}, \widetilde{\pi}_r(X_r \mid \mathcal{H}_{r-1})\Big\}\Big\}$
    Assign $A_r \sim \text{Bern}(\pi_r(X_r \mid \mathcal{H}_{r-1}))$
  end if
  Observe $\widetilde{T}_r$, Δ_r
  Compute cross-fitted EIF pseudo-outcome $\hat{\boldsymbol{\phi}}_r^{\text{cf}} = \big(\hat{\phi}_{0,r}^{\text{cf}}, \ldots, \hat{\phi}_{t_{\max},r}^{\text{cf}}\big)^\top$ using Eq. 13 with nuisances $\hat{\eta}_{t,r-1}^{(-J_r)}$
end for
return $\hat{\tau}_R^{\text{ASE}} = \dfrac{1}{R}\displaystyle\sum_{r=1}^{R}\hat{\boldsymbol{\phi}}_r^{\text{cf}}$

F PROOFS

F.1 PROOF OF THEOREM 4.1

The semiparametric efficiency bound for the vector parameter $\tau = (\tau_0, \ldots, \tau_{t_{\max}})$ under a fixed allocation π is the covariance of the efficient influence function vector, i.e.,

$$\Sigma_{\text{eff}}(\pi) = \text{Cov}\big(\boldsymbol{\phi}(\mathcal{O}; \tau, \eta_{t_{\max}})\big) = \mathbb{E}\Big[\big(\boldsymbol{\phi}(\mathcal{O}; \tau, \eta_{t_{\max}}) - \boldsymbol{\tau}\big)\big(\boldsymbol{\phi}(\mathcal{O}; \tau, \eta_{t_{\max}}) - \boldsymbol{\tau}\big)^\top\Big], \tag{30}$$

So first, we decompose it at a discrete time t. To ease notation, we write $U_t = S_t(X, A)\,\xi(\mathcal{O}, \eta_t)$, so that the centered EIF scores from Eq. 4 are given by

$$\phi_t(\mathcal{O}; \tau_t, \eta_t) = S_t(X, 1) - S_t(X, 0) - \underbrace{\frac{A - \pi(X)}{\pi(X)(1 - \pi(X))}\, U_t}_{\Lambda_t} - \tau_t, \tag{31}$$

We show $\mathbb{E}[\xi(\mathcal{O}, \eta_t) \mid X, A] = 0$ first. Under Assumption 3.2, we have $S_i(X, A) > 0$ and $G_{i-1}(X, A) > 0$, so each denominator is strictly positive and the summand is well defined. It therefore suffices to show that the numerator has a conditional mean zero for every i:

$$\mathbb{E}\Big[\mathbf{1}\{\widetilde{T} = i, \Delta = 1\} - \mathbf{1}\{\widetilde{T} \geq i\}\,\lambda_i^S(X, A)\,\Big|\,X, A\Big] \;=\; 0. \tag{32}$$

By the definition of the discrete-time event hazard, we yield

$$\lambda_i^S(X, A) \;=\; \mathbb{P}\big(\widetilde{T} = i,\, \Delta = 1 \,\big|\, \widetilde{T} \geq i,\, X,\, A\big). \tag{33}$$

Since $\{\widetilde{T} = i, \Delta = 1\} \subseteq \{\widetilde{T} \geq i\}$, the event $\{\widetilde{T} = i, \Delta = 1\}$ implies $\mathbf{1}\{\widetilde{T} \geq i\} = 1$, and therefore

$$\begin{aligned}
\mathbb{E}\Big[\mathbf{1}\{\widetilde{T} = i, \Delta = 1\} \,\Big|\, X, A\Big] &= \mathbb{P}\big(\widetilde{T} = i,\, \Delta = 1 \,\big|\, X, A\big) \\
&= \mathbb{P}\big(\widetilde{T} \geq i \,\big|\, X, A\big) \cdot \mathbb{P}\big(\widetilde{T} = i,\, \Delta = 1 \,\big|\, \widetilde{T} \geq i,\, X, A\big) \\
&= \;\mathbb{E}\Big[\mathbf{1}\{\widetilde{T} \geq i\} \,\Big|\, X, A\Big] \cdot \lambda_i^S(X, A).
\end{aligned} \tag{34}$$

Rearranging yields Eq. 32. Summing over $i = 0, \ldots, t$ and dividing each term by the $\sigma(X, A)$-measurable positive factor $S_i(X, A)\,G_{i-1}(X, A)$, we yield

$$\mathbb{E}[\xi(\mathcal{O}, \eta_t) \mid X, A] \;=\; \sum_{i=0}^{t} \frac{\mathbb{E}\Big[\mathbf{1}\{\widetilde{T} = i, \Delta = 1\} - \mathbf{1}\{\widetilde{T} \geq i\}\,\lambda_i^S(X, A) \,\Big|\, X, A\Big]}{S_i(X, A)\,G_{i-1}(X, A)} \;=\; 0. \tag{35}$$

Then, we have $\mathbb{E}[U_t \mid X, A] = 0$ and therefore

$$\mathbb{E}[\Lambda_t \mid X] = \mathbb{E}\big[\mathbb{E}[\Lambda_t \mid X, A] \,\big|\, X\big] = 0. \tag{36}$$

Thus, we can expand $V_{\text{eff},t}(\pi)$ as

$$\text{Var}(\phi_t) = \mathbb{E}[(\phi_t - \tau_t)^2] = \mathbb{E}[(S_t(X, 1) - S_t(X, 0) - \tau_t)^2] + \mathbb{E}[\Lambda_t^2]. \tag{37}$$

For the second term, the identity $\mathbb{E}[\Lambda_t^2] = \mathbb{E}[\mathbb{E}[\Lambda_t^2 \mid X]]$ can be rewritten via

$$\begin{aligned}
&\mathbb{E}[\Lambda_t^2 \mid X] \\
=&\mathbb{E}\left[\left(\frac{A - \pi(X)}{\pi(X)(1 - \pi(X))}\right)^2 U_t^2 \mid X\right] \\
=&\pi(X)\mathbb{E}\left[\left(\frac{1}{\pi(X)}\right)^2 U_t^2 \mid X, A = 1\right] + (1 - \pi(X))\mathbb{E}\left[\left(\frac{1}{1 - \pi(X)}\right)^2 U_t^2 \mid X, A = 0\right] \\
=&\frac{1}{\pi(X)}\text{Var}[S_t(X, 1)\xi(\mathcal{O}, \eta_t) \mid X, A = 1] + \frac{1}{1 - \pi(X)}\text{Var}[S_t(X, 0)\xi(\mathcal{O}, \eta_t) \mid X, A = 0].
\end{aligned} \tag{38}$$

Taking the outer expectation, we have

$$\mathrm{Var}(\phi_t) = \mathbb{E}\left[\frac{v_{t,1}(X)}{\pi(X)} + \frac{v_{t,0}(X)}{1-\pi(X)}\right] + \mathbb{E}[b_t(X)^2], \tag{39}$$

where $v_{t,a}(X) := \mathbb{E}[U_t^2 \mid X, A = a]$, $b_t(X) = S_t(X,1) - S_t(X,0) - \tau_t$, and the second term is $\pi(X)$ independent. □

F.2 Proof of Proposition 4.3

Since C in Theorem 4.1 does not depend on π, we have

$$\begin{aligned}
\pi^\star &= \arg\min_\pi \mathbb{E}\left[\frac{V_1(X)}{\pi(X)} + \frac{V_0(X)}{1-\pi(X)}\right] \\
&= \arg\min_\pi \mathbb{E}\left[\mathbb{E}\left[\frac{V_1(X)}{\pi(X)} + \frac{V_0(X)}{1-\pi(X)} \,\middle|\, X\right]\right] \\
\Rightarrow \quad \pi^\star(x) &= \arg\min_{p\in[\alpha,1-\alpha]} \left(\frac{V_1(x)}{p} + \frac{V_0(x)}{1-p}\right),
\end{aligned} \tag{40}$$

where $V_a(x) := \sum_{t=0}^{t_{\max}} v_{t,a}(x)$. The first-order condition gives

$$\frac{V_0(x)}{(1-p)^2} - \frac{V_1(x)}{p^2} = 0, \tag{41}$$

which yields $p = \sqrt{V_1(x)}/(\sqrt{V_1(x)} + \sqrt{V_0(x)})$. Applying the overlap constraint, we yield

$$\pi^\star(x) = \mathrm{clip}_\alpha\left(\frac{\sqrt{V_1(x)}}{\sqrt{V_1(x)} + \sqrt{V_0(x)}}\right). \tag{42}$$

F.3 Remark on necessity of stepwise overlap

Remark F.1 (Necessity of stepwise overlap). Assumption 6.1 imposes lower bounds on $1-\lambda_i^S(x,a)$ and $1-\lambda_i^G(x,a)/(1-\lambda_i^S(x,a))$ at every intermediate step $i \le t$, rather than only at the endpoint t. This is necessary: since $G_{t-1}(x,a) = \prod_{i=0}^{t-1}\big(1-\lambda_i^G(x,a)/(1-\lambda_i^S(x,a))\big)$ and $S_i(x,a) = \prod_{j=0}^{i}(1-\lambda_j^S(x,a))$ are both cumulative products, a single intermediate step at which $\lambda_i^S(x,a) \to 1$ drives $S_i(x,a) \to 0$ and hence $G_{i-1}(x,a)^{-1} \to \infty$, which the inverse probability of censoring weights in the EIF to become unbounded regardless of the behavior at other time steps. Assumption 6.1 rules this out uniformly across $i \le t$ and thus ensures that $S_i(x,a) \ge (c_t^S)^{t+1}$ and $G_{i-1}(x,a) \ge (c_t^G)^{t+1}$ for all $i \le t$. Hence, this constitutes the discrete-time survival analogue of the positivity condition standard in semiparametric survival analysis (Cui et al., 2023; van der Laan and Robins, 2003; Westling et al., 2024).

F.4 Proof of Theorem 6.3

The asymptotic argument relies on the martingale central limit theorem under a Lindeberg-type condition. We use a streamlined version of the martingale difference sequence (MDS) central limit theorem originally due to Dvoretzky (1972), as presented in Theorem 2 of Zhang et al. (2021).

To begin our proof, we define the centered EIF score at horizon t as $\psi_{t,r} := \phi(X_r, A_r, \tilde{T}_r, \Delta_r; \tau_t, \pi_r, \hat{\eta}_{t,r})$, where ϕ is given in Eq. 4. Let $\eta_t = \{\lambda_i^S(X,0), \lambda_i^S(X,1), \lambda_i^G(X,0), \lambda_i^G(X,1)\}_{i=0}^t$ denote the oracle nuisance functions. We can decompose $\psi_{t,r}$ as

$$\begin{aligned}\psi_{t,r} =& \phi_{t,r} - \tau_t + \hat{\phi}_{t,r} - \phi_{t,r} \qquad (43)\\ =& \underbrace{\phi(X_r, A_r, \tilde{T}_r, \Delta_r; \tau_t, \pi_r, \eta_t) - \tau_t}_{z_{t,r}} \\ &+ \underbrace{\phi(X_r, A_r, \tilde{T}_r, \Delta_r; \tau_t, \pi_r, \hat{\eta}_{t,r}) - \phi(X_r, A_r, \tilde{T}_r, \Delta_r; \tau_t, \pi_r, \eta_t)}_{m_{t,r}}, \qquad (44)\end{aligned}$$

such that

$$\sqrt{R}\big(\hat{\tau}_{t,R}^{\mathrm{ASE}} - \tau_t\big) = \frac{1}{\sqrt{R}}\sum_{r=1}^{R} z_{t,r} + \frac{1}{\sqrt{R}}\sum_{r=1}^{R} m_{t,r}. \qquad (45)$$

Our proof proceeds in three steps: (1) we prove that, for each horizon t, $\{z_{t,r}\}_{r=1}^R$ forms a martingale difference sequence, (2) we show that $\{z_{t,r}\}_{r=1}^R$ satisfy conditional variance convergence and Lindeberg condition, so we have the $\frac{1}{\sqrt{R}}\sum_{r=1}^R z_{t,r} \xrightarrow{d} \mathcal{N}(0, V_{\text{eff},t}(\pi))$, and (3) we show that $\frac{1}{\sqrt{R}}\sum_{r=1}^R m_{t,r} = o_p(1)$.

F.4.1 Martingale difference sequence

Before we start to the first steps, we state the martingale central limit theorem first.

Theorem F.2 (Martingale CLT, adapted from Theorem 2 in Zhang et al. (2021))**.** *Let $\{z_{t,r}, \mathcal{H}_r\}_{r=1}^R$ be a real-valued sequence, and let $\bar{z}_{t,r} = \frac{1}{R}\sum_{r=1}^R z_{t,r}$ such that:*

1. *(Martingale difference sequence) $\{z_{t,r}\}_{r=1}^R$ is a martingale difference sequence; that is $\mathbb{E}[z_{t,r} \mid \mathcal{H}_{r-1}] = 0$ for all $r \in [1, R]$.*
2. *(Conditional variance convergence) There exists a constant $V_{\text{eff},t}(\pi) > 0$ such that*
$$\frac{1}{R}\sum_{r=1}^{R} \mathbb{E}[z_{t,r}^2 \mid \mathcal{H}_{r-1}] \xrightarrow{p} V_{\text{eff},t}(\pi), \qquad (46)$$
3. *(Lindeberg condition) For every $\varepsilon > 0$,*
$$\frac{1}{R}\sum_{r=1}^{R} \mathbb{E}\left[z_{t,r}^2 \mathbb{1}\{|z_{t,r}| > \varepsilon\sqrt{R}\} \mid \mathcal{H}_{r-1}\right] \xrightarrow{p} 0. \qquad (47)$$

Then, $\sqrt{R}\bar{z}_{t,r} \xrightarrow{d} \mathcal{N}(0, V_{\mathit{eff},t}(\pi))$.

The first step is that we need to prove $\mathbb{E}[z_{t,r} \mid \mathcal{H}_{r-1}] = 0$:

$$
\begin{aligned}
&\mathbb{E}[z_{t,r} \mid \mathcal{H}_{r-1}] \\
=&\mathbb{E}[S_t(X_r,1) - S_t(X_r,0) - \frac{A_r - \pi_r(X_r)}{\pi_r(X_r)\{1-\pi_r(X_r)\}}\xi(\mathcal{H}_r,\eta_t)S_t(X_r,A_r) \mid \mathcal{H}_{r-1}] - \tau_t \\
=&\tau_t - \tau_t - \mathbb{E}\left[\mathbb{E}\left[\frac{A_r - \pi_r(X_r)}{\pi_r(X_r)\{1-\pi_r(X_r)\}}\xi(\mathcal{H}_r,\eta)S_t(X_r,A_r) \mid X_r, \mathcal{H}_{r-1}\right] \mid \mathcal{H}_{r-1}\right] \\
=&-\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[\frac{A_r - \pi_r(X_r)}{\pi_r(X_r)\{1-\pi_r(X_r)\}}\xi(\mathcal{H}_r,\eta)S_t(X_r,A_r) \mid A_r, X_r, \mathcal{H}_{r-1}\right] \mid X_r, \mathcal{H}_{r-1}\right] \mid \mathcal{H}_{r-1}\right] \\
=&0. \qquad (48)
\end{aligned}
$$

We further yield

$$
\begin{aligned}
&\mathbb{E}\left[\xi(\mathcal{H}_r,\eta_t) \mid A_r, X_r, \mathcal{H}_{r-1}\right] \qquad (49)\\
=&\mathbb{E}\left[\sum_{i=0}^{t}\frac{\mathbb{1}(\tilde{T}=i)\mathbb{1}(\Delta=1) - \mathbb{1}(\tilde{T}\geq i)\lambda_i^s(X_r,A_r)}{S_i(X_r,A_r)G_{i-1}(X_r,A_r)} \mid A_r, X_r, \mathcal{H}_{r-1}\right] \qquad (50)\\
=&\sum_{i=0}^{t}\mathbb{E}\left[\frac{\mathbb{1}(\tilde{T}=i)\mathbb{1}(\Delta=1) - \mathbb{1}(\tilde{T}\geq i)\lambda_i^s(X_r,A_r)}{S_i(X_r,A_r)G_{i-1}(X_r,A_r)} \mid A_r, X_r, \mathcal{H}_{r-1}\right] \qquad (51)\\
=&\sum_{i=0}^{t}\frac{\mathbb{P}\left(\tilde{T}=i,\Delta=1 \mid X_r,A_r,\mathcal{H}_{r-1}\right) - \mathbb{P}\left(\tilde{T}\geq i \mid X_r,A_r,\mathcal{H}_{r-1}\right)\mathbb{P}\left(\tilde{T}=i,\Delta=1 \mid \tilde{T}>i,X_r,A_r\right)}{S_i(X_r,A_r)G_{i-1}(X_r,A_r)} \qquad (52)\\
=&\sum_{i=0}^{t}\frac{\mathbb{P}\left(\tilde{T}=i,\Delta=1 \mid X_r,A_r,\mathcal{H}_{r-1}\right) - \mathbb{P}\left(\tilde{T}=i,\Delta=1 \mid X_r,A_r,\mathcal{H}_{r-1}\right)}{S_i(X_r,A_r)G_{i-1}(X_r,A_r)} \qquad (53)\\
=&0. \qquad (54)
\end{aligned}
$$

Thus, for each fixed t, $\{z_{t,r}, \mathcal{H}_r\}_{r\geq 1}$ is a martingale difference sequence, since the assignment probability π_r is $\mathcal{H}_{r-1}$-measurable and the oracle augmentation term satisfies $\mathbb{E}[\xi(\mathcal{H}_r,\eta_t) \mid A_r, X_r, \mathcal{H}_{r-1}] = 0$.

F.4.2 Conditional variance convergence and Lindeberg condition

We first need to show that $\mathbb{E}[z_{t,r}^2 \mid \mathcal{H}_{r-1}] - V_{\mathrm{eff},t}(\pi) \xrightarrow{p} 0$. Let $b_t(X) := S_t(X,1) - S_t(X,0) - \tau_t$. Using $\mathbb{E}[\xi_t(\mathcal{H}_r,\eta_t) \mid X_r, A_r, \mathcal{H}_{r-1}] = 0$, we have

$$
\mathbb{E}[z_{t,r}^2 \mid \mathcal{H}_{r-1}] = \mathbb{E}\left[b_t(X_r)^2 + \frac{v_{t,1}(X_r)}{\pi_r(X_r)} + \frac{v_{t,0}(X_r)}{1-\pi_r(X_r)} \;\middle|\; \mathcal{H}_{r-1}\right]. \qquad (55)
$$

Therefore, we have

$$
\begin{aligned}
&\left|\mathbb{E}[z_{t,r}^2 \mid \mathcal{H}_{r-1}] - V_{\mathrm{eff},t}(\pi)\right| \\
&\leq \mathbb{E}\left[v_{t,1}(X_r)\left|\frac{1}{\pi_r(X_r)} - \frac{1}{\pi(X_r)}\right| + v_{t,0}(X_r)\left|\frac{1}{1-\pi_r(X_r)} - \frac{1}{1-\pi(X_r)}\right| \;\middle|\; \mathcal{H}_{r-1}\right] \\
\leq& 4\frac{k_r}{\varepsilon}\left(\frac{t+1}{\underline{s}_t\,\underline{g}_t}\right)^2 \|\pi_r - \pi\|_2 = o_p(1), \qquad (56)
\end{aligned}
$$

According to Assumption 6.1, we have $1-\lambda_i^S(x,a) \geq \underline{c}_t^S$ $1-\lambda_i^G(x,a)/(1-\lambda_i^S(x,a)) \geq \underline{c}_t^G$, so we have $S_i(x,a) = \prod_{j=0}^{i} 1-\lambda_j^S(x,a) \geq (\underline{c}_t^S)^{i+1} \geq (\underline{c}_t^S)^{t+1} = \underline{s}_t$ and $G_{i-1}(x,a) = \prod_{j=0}^{i-1} 1-$

$\lambda_j^G(x,a)/(1-\lambda_j^S(x,a)) \geq (\underline{c}_t^G)^i \geq (\underline{c}_t^G)^{t+1} = \underline{g}_t$. We note $\left|\sum_{i=0}^t \frac{S_i(X,a)}{S_i(X,a)G_{i-1}(X,a)}(\mathbf{1}(\widetilde{T}=i,\Delta=1)-\mathbf{1}(\widetilde{T}\geq i)\,\lambda_i^S(X,A))\right| \leq \sum_{i=0}^t \frac{S_i(X,a)}{S_i(X,a)G_{i-1}(X,a)} \leq \sum_{i=0}^t \frac{1}{\underline{s}_t\underline{g}_t}$. Therefore, $v_{t,a}(X) = \mathrm{Var}(\phi_t \mid X, A=a) \leq \mathbb{E}[\phi_t^2 \mid X, A=a] \leq \left(\frac{t+1}{\underline{s}_t\,\underline{g}_t}\right)^2$ where the last line we use (i) the boundedness assumption from Assumption 6.1 and thus $\|v_{t,1}(X)\|_\infty \leq \left(\frac{t+1}{\underline{s}_t\,\underline{g}_t}\right)^2$, (ii) $\pi(X_r), 1-\pi(X_r) > \varepsilon$ from Theorem 6.3 statement, (iii) $\pi(X_r), 1-\pi(X_r) > 1/k_r$ by construction, and the L_1-norm is bounded by the L_2 norm. Thus, setting $\sigma^2 := V_{\text{eff},t}(\pi)$, we have that the term converges in probability to σ^2, i.e., $E[z_r^2 \mid \mathcal{H}_{r-1}] \xrightarrow{p} \sigma^2$, where σ^2 is finite. And we further have that $\frac{1}{R}\sum_{r=1}^R \mathbb{E}[z_r^2 \mid \mathcal{H}_{r-1}] \xrightarrow{p} \sigma^2$.

Now, we prove the **Lindeberg condition**. For every $\varepsilon > 0$, by the truncation rule, $1/k_r \leq \pi_r(X_r) \leq 1 - 1/k_r$. Hence,

$$\left|\frac{A_r - \pi_r(X_r)}{\pi_r(X_r)\{1-\pi_r(X_r)\}}\right| \leq k_r. \tag{57}$$

Moreover, by Assumption 6.1, the survival and censoring weights in the oracle EIF are uniformly bounded for each fixed t. Therefore, there exists a constant $C_t < \infty$ such that $|z_{t,r}| \leq C_t k_r$. Since $k_r = o(r^{1/4})$, we also have $k_r = o(\sqrt{r})$. If k_r is nondecreasing, then

$$\max_{1\leq r\leq R} |z_{t,r}| \leq C_t k_R = o(\sqrt{R}). \tag{58}$$

Thus, for every $\varepsilon > 0$, eventually $|z_{t,r}| \leq \varepsilon\sqrt{R}$ for all $r \leq R$, and hence $\mathbb{1}\{|z_{t,r}| > \varepsilon\sqrt{R}\} = 0$ for all $r \leq R$. Consequently,

$$\frac{1}{R}\sum_{r=1}^R \mathbb{E}\left[z_{t,r}^2 \mathbb{1}\{|z_{t,r}| > \varepsilon\sqrt{R}\} \mid \mathcal{H}_{r-1}\right] = 0, \tag{59}$$

which proves the Lindeberg condition. Thus, by Theorem F.2 we proved that $\frac{1}{\sqrt{R}}\sum_{r=1}^R z_{t,r} \xrightarrow{d} \mathcal{N}(0, V_{\text{eff},t}(\pi))$.

F.4.3 The remainder term

For a fixed horizon t, recall that

$$m_{t,r} := \phi_t(\mathcal{H}_r; \pi_r, \hat{\eta}_{t,r-1}) - \phi_t(\mathcal{H}_r; \pi_r, \eta_t). \tag{60}$$

We aim to show

$$\frac{1}{\sqrt{R}}\sum_{r=1}^R m_{t,r} = o_p(1). \tag{61}$$

We decompose

$$\begin{aligned}\frac{1}{\sqrt{R}}\sum_{r=1}^R m_{t,r} &= \sqrt{R}\left(\frac{1}{R}\sum_{r=1}^R \mathbb{E}[m_{t,r} \mid \mathcal{H}_{r-1}]\right) + \frac{1}{\sqrt{R}}\sum_{r=1}^R \Big(m_{t,r} - \mathbb{E}[m_{t,r} \mid \mathcal{H}_{r-1}]\Big) \\ &:= \Delta_t^A + \Delta_t^B.\end{aligned} \tag{62}$$

We show separately that $\Delta_t^A = o_p(1)$ and $\Delta_t^B = o_p(1)$.

Step 1: Bias term Δ_t^A. Let

$$\Delta_{t,r}^A := \mathbb{E}[m_{t,r} \mid \mathcal{H}_{r-1}] = \mathbb{E}[\phi_t(\mathcal{H}_r; \pi_r, \hat{\eta}_{t,r-1}) - \phi_t(\mathcal{H}_r; \pi_r, \eta_t) \mid \mathcal{H}_{r-1}]\,. \tag{63}$$

Using the same expansion as in the proof of the robustness survival estimator, one obtains

$$\Delta_{t,r}^A = \mathbb{E}[\mathcal{R}_{t,r,1}(X_r) - \mathcal{R}_{t,r,0}(X_r) \mid \mathcal{H}_{r-1}]\,, \tag{64}$$

where, for $a \in \{0,1\}$,

$$\mathcal{R}_{t,r,a}(x) = \sum_{i=0}^{t} \frac{\hat{S}_{t,r-1}(x,a)S_{i-1}(x,a)}{\hat{S}_{i,r-1}(x,a)\hat{G}_{i-1,r-1}(x,a)} \{\hat{\lambda}^S_{i,r-1}(x,a) - \lambda^S_i(x,a)\}\{G_{i-1}(x,a) - \hat{G}_{i-1,r-1}(x,a)\}. \tag{65}$$

According to assumptions in Theorem 6.3, for each arm a, we have $\max_{0\le i\le t_{\max}} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2 = o_p(r^{-1/4})$ and $\max_{0\le i\le t_{\max}-1} \|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \lambda^G_i(a,\cdot)\|_2 = o_p(r^{-1/4})$. This implies

$$\left(\sum_{i=0}^{t} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2\right)^2 = o_p(r^{-1/2}), \tag{66}$$

$$\left(\sum_{i=0}^{t} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2\right)\left(\sum_{i=0}^{t-1} \|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \lambda^G_i(a,\cdot)\|_2\right) = o_p(r^{-1/2}). \tag{67}$$

Then, the prefactor in Eq.65 is uniformly bounded by a constant $C_t < \infty$ depending only on $t, \underline{s}_t, \underline{g}_t$. Hence

$$|\Delta^A_{t,r}| \le C_t \sum_{a\in\{0,1\}} \left[\left(\sum_{i=0}^{t} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2\right)^2 + \left(\sum_{i=0}^{t} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2\right)\left(\sum_{i=0}^{t-1} \|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \lambda^G_i(a,\cdot)\|_2\right)\right]. \tag{68}$$

By the assumed product-rate condition, the right-hand side is $o_p(r^{-1/2})$. Therefore, we have

$$\Delta^A_{t,r} = o_p(r^{-1/2}), \qquad \text{and hence} \qquad \Delta^A_t = \sqrt{R}\left(\frac{1}{R}\sum_{r=1}^{R} \Delta^A_{t,r}\right) = o_p(1). \tag{69}$$

Step 2: Martingale fluctuation term Δ^B_t. Define $U_{t,r} := m_{t,r} - \mathbb{E}[m_{t,r} \mid \mathcal{H}_{r-1}]$. Then $\{U_{t,r}, \mathcal{H}_r\}_{r\ge1}$ is a martingale difference sequence, and $\Delta^B_t = \frac{1}{\sqrt{R}}\sum_{r=1}^{R} U_{t,r}$. It therefore suffices to show that

$$\mathrm{Var}(\Delta^B_t) = o(1). \tag{70}$$

Since the cross-terms vanish for a martingale difference sequence, we have

$$\mathrm{Var}(\Delta^B_t) = \frac{1}{R}\sum_{r=1}^{R} \mathbb{E}[U^2_{t,r}] \le \frac{1}{R}\sum_{r=1}^{R} \mathbb{E}\left[m^2_{t,r}\right]. \tag{71}$$

We now bound $m^2_{t,r}$. We write

$$\hat{\xi}_{t,r} := \xi_t(\mathcal{H}_r, \hat{\eta}_{t,r-1}), \qquad \xi_{t,r} := \xi_t(\mathcal{H}_r, \eta_t), \qquad W_r := \frac{A_r - \pi_r(X_r)}{\pi_r(X_r)\{1-\pi_r(X_r)\}}, \tag{72}$$

and

$$\Delta^S_{t,r}(a) := \hat{S}_{t,r-1}(X_r, a) - S_t(X_r, a), \qquad a \in \{0,1\}. \tag{73}$$

Then, we yield

$$m_{t,r} = \Delta^S_{t,r}(1) - \Delta^S_{t,r}(0) - W_r\Big[(\hat{\xi}_{t,r} - \xi_{t,r})\hat{S}_{t,r-1}(X_r, A_r) + \xi_{t,r}\,\Delta^S_{t,r}(A_r)\Big]. \tag{74}$$

By $(u_1 + \cdots + u_4)^2 \le 4(u_1^2 + \cdots + u_4^2)$, we have

$$m^2_{t,r} \le C_t k_r^2\Big(|\Delta^S_{t,r}(1)|^2 + |\Delta^S_{t,r}(0)|^2 + |\hat{\xi}_{t,r} - \xi_{t,r}|^2\Big), \tag{75}$$

where $C_t < \infty$ depends only on t and the overlap constants $\underline{s}_t, \underline{g}_t$.

It remains to bound the two ingredients on the right-hand side.

Step 2a: Control of $\Delta^S_{t,r}(a)$. Since

$$S_t(x,a) = \prod_{i=0}^{t}(1-\lambda_i^S(x,a)), \qquad \hat{S}_{t,r-1}(x,a) = \prod_{i=0}^{t}(1-\hat{\lambda}^S_{i,r-1}(x,a)), \tag{76}$$

the product-difference identity gives

$$|\Delta^S_{t,r}(a)| \le \sum_{i=0}^{t} |\hat{\lambda}^S_{i,r-1}(X_r,a) - \lambda_i^S(X_r,a)|. \tag{77}$$

Hence, by the Cauchy–Schwarz inequality, we yield

$$\mathbb{E}\big[|\Delta^S_{t,r}(a)|^2 \mid \mathcal{H}_{r-1}\big] \le (t+1)\sum_{i=0}^{t}\mathbb{E}\Big[\big(\hat{\lambda}^S_{i,r-1}(X_r,a) - \lambda_i^S(X_r,a)\big)^2 \,\Big|\, \mathcal{H}_{r-1}\Big]. \tag{78}$$

Step 2b: Control of $\hat{\xi}_{t,r} - \xi_{t,r}$. Using the definition of ξ_t and a first-order decomposition of the difference between

$$\frac{N_{t,i} - Y_{t,i}\hat{\lambda}^S_{i,r-1}}{\hat{S}_{i,r-1}\hat{G}_{i-1,r-1}} \qquad \text{and} \qquad \frac{N_{t,i} - Y_{t,i}\lambda_i^S}{S_i G_{i-1}}, \tag{79}$$

together with the uniform lower bounds $\underline{s}_t, \underline{g}_t$, we obtain

$$|\hat{\xi}_{t,r} - \xi_{t,r}|^2 \le C_t \sum_{a\in\{0,1\}} \left\{ \sum_{i=0}^{t}\big(\hat{\lambda}^S_{i,r-1}(X_r,a) - \lambda_i^S(X_r,a)\big)^2 + \sum_{i=0}^{t-1}\big(\hat{\lambda}^G_{i,r-1}(X_r,a) - \lambda_i^G(X_r,a)\big)^2 \right\}, \tag{80}$$

for some constant $C_t < \infty$.

Combining Eq. 75, Eq. 78, and Eq. 80, we get

$$\mathbb{E}[m^2_{t,r} \mid \mathcal{H}_{r-1}] \le C_t k_r^2 \sum_{a\in\{0,1\}} \left\{ \sum_{i=0}^{t} \|\hat{\lambda}^S_{i,r-1}(\cdot,a) - \lambda_i^S(\cdot,a)\|_2^2 + \sum_{i=0}^{t-1} \|\hat{\lambda}^G_{i,r-1}(\cdot,a) - \lambda_i^G(\cdot,a)\|_2^2 \right\}. \tag{81}$$

Under the stated L_2 rate conditions and the boundedness implied by Assumption 6.1, the expectation of the right-hand side of Eq. 81 is $o(1)$.

Conclusion. Since both $\Delta^A_t = o_p(1)$ and $\Delta^B_t = o_p(1)$, we conclude that $\frac{1}{\sqrt{R}}\sum_{r=1}^{R} m_{t,r} = o_p(1)$. This proves that the nuisance remainder is asymptotically negligible.

If the limiting policy satisfies $\pi = \pi^\star$, then the limiting variance in the above display becomes $V_{\text{eff},t}(\pi^\star)$. By Theorem 4.1, this is the semiparametric efficiency bound under the A-optimal allocation.

F.5 Proof of Proposition 6.4

By Eq. 10, $V_a(x) = \sum_{t=0}^{t_{\max}} S_t(x,a)^2 \sum_{i=0}^{t} \lambda_i^S(x,a)/(S_i(x,a)\, G_{i-1}(x,a))$ is a Lipschitz function of $(\boldsymbol{\lambda}^S(X,a), \boldsymbol{\lambda}^G(X,a))$ on the region where $S_i(x,a)$ and $G_{i-1}(x,a)$ are bounded away from zero; Assumption 6.1 guarantees precisely this uniform lower bound. Hence, the $o_p(r^{-1/4})$ hazard consistency implies $\|\hat{V}_{a,r-1} - V_a\|_2 = o_p(r^{-1/4})$ for each $a \in \{0,1\}$ by Lipschitz composition. Furthermore, Assumption 6.1 implies $S_t(x,a) \ge (\underline{c}_t^S)^{t+1} > 0$ and $G_{i-1}(x,a) \ge \prod_j (1-\lambda_j^G/(1-\lambda_j^S)) \cdot (1-\lambda_j^S) > 0$ uniformly, so that the denominator $\sqrt{V_0(x)} + \sqrt{V_1(x)}$ in the plug-in policy is bounded away from zero. The map $(V_0, V_1) \mapsto \sqrt{V_1}/(\sqrt{V_0} + \sqrt{V_1})$ is therefore Lipschitz, giving $\|\widetilde{\pi}_r - \pi^\star\|_2 = O_p(r^{-1/4})$ by Lipschitz composition. Since $\pi^\star(X) \in [\alpha, 1-\alpha]$ and $k_r \to \infty$, we have $1/k_r < \alpha$ for all sufficiently large r, so $\pi^\star(X) \in [1/k_r, 1-1/k_r]$. Because clipping onto $[1/k_r, 1-1/k_r]$ is the Euclidean projection and is non-expansive with respect to any point inside the interval, we have $|\pi_r(x) - \pi^\star(x)| \le |\widetilde{\pi}_r(x) - \pi^\star(x)|$ pointwise. Hence, we arrive at $\|\pi_r - \pi^\star\|_2 \le \|\widetilde{\pi}_r - \pi^\star\|_2 = O_p(r^{-1/4})$. Since $k_r = o(r^{1/4})$, it follows that $k_r\|\pi_r - \pi^\star\|_2 = o_p(1)$. □

F.6 PROOF OF COROLLARY 6.6

The claim is an immediate consequence of Theorem 6.3 and Eq. 66, which yields

$$\left|\hat{\tau}_{t,R}^{\mathrm{ASE}} - \tau_t\right| = O_p(R^{-1/2}) \tag{82}$$
$$+ O_p\left((\sum_{i=0}^{t} \|\hat{\lambda}_{i,r-1}^{S}(a,\cdot) - \lambda_i^S(a,\cdot)\|_2)^2 + \sum_{i=0}^{t} \|\hat{\lambda}_{i,r-1}^{S}(a,\cdot) - \lambda_i^S(a,\cdot)\|_2 \sum_{i=0}^{t-1} \|\hat{\lambda}_{i,r-1}^{G}(a,\cdot) - \lambda_i^G(a,\cdot)\|_2\right).$$

Since $\left(\sum_{i=0}^{t} \|\hat{\lambda}_{i,r-1}^{S}(a,\cdot) - \lambda_i^S(a,\cdot)\|_2\right)^2 = o_p(r^{-1/2})$ and $\sum_{i=0}^{t} \|\hat{\lambda}_{i,r-1}^{S}(a,\cdot) - \lambda_i^S(a,\cdot)\|_2 \sum_{i=0}^{t-1} \|\hat{\lambda}_{i,r-1}^{G}(a,\cdot) - \lambda_i^G(a,\cdot)\|_2 = o_p(r^{-1/2})$, so $\hat{\tau}_{t,R}^{\mathrm{ASE}} \xrightarrow{p} \tau_t$. The absence of any standalone term $O_p(\|\hat{\lambda}_t^G - \lambda_t^G\|_2)$ shows that perturbations in λ^G enter only through interaction terms, which is exactly the robustness condition we aimed to show. □

G EXTENSIONS TO NO TIES

In the no-ties setting, we refine the observed data structure by recording separate event and censoring indicators. Specifically, each observation is $\mathcal{O} = (X, A, \widetilde{T}, \Delta^S, \Delta^G)$, where $\widetilde{T} = \min(T, C)$, $\Delta^S = \mathbf{1}(T = \widetilde{T})$, $\Delta^G = \mathbf{1}(C = \widetilde{T})$, so that $\Delta^S + \Delta^G = 1$ almost surely. In this case, the event and censoring hazards are defined separately by

$$\lambda_t^S(x, a) = \mathbb{P}(T = t \mid T \geq t, X = x, A = a), \tag{83}$$

$$\lambda_t^G(x, a) = \mathbb{P}(C = t \mid C \geq t, X = x, A = a), \tag{84}$$

with corresponding survival functions

$$S_t(x, a) = \prod_{i=0}^{t} (1 - \lambda_i^S(x, a)), \qquad G_{t-1}(x, a) = \prod_{i=0}^{t-1} (1 - \lambda_i^G(x, a)). \tag{85}$$

For a possibly misspecified nuisance limit $\tilde{\eta}_t = (\tilde{\lambda}_i^S(\cdot, \cdot), \tilde{\lambda}_i^G(\cdot, \cdot))_{i=0}^t$, we analogously define

$$\tilde{S}_t(x, a) = \prod_{i=0}^{t} (1 - \tilde{\lambda}_i^S(x, a)), \qquad \tilde{G}_t(x, a) = \prod_{i=0}^{t-1} (1 - \tilde{\lambda}_i^G(x, a)). \tag{86}$$

In the no-ties setting, the censoring survival function G_t depends only on the censoring hazard λ^G, so the event and censoring nuisances form two variation-independent blocks. As a result, the adaptive survival estimator satisfies a literal two-block double robustness property.

Theorem G.1 (Convergence rate of the adaptive survival estimator under no ties)**.** *Suppose Assumptions 3.1 and 3.2 hold, and that there exists a non-adaptive policy $\pi(X) \in [\varepsilon, 1 - \varepsilon]$ for some $\varepsilon > 0$ such that*

$$k_r \|\pi_r - \pi\|_2 = o_p(1). \tag{87}$$

Let $\tilde{\eta}_t$ denote a possibly misspecified limit of the cross-fitted nuisance estimators, and suppose

$$k_r \|\hat{f}_{t,r-1} - \tilde{f}_t\|_2 = o_p(1) \tag{88}$$

componentwise for $\tilde{f}_t \in \tilde{\eta}_t$. Then, for each horizon $t \leq t_{\max}$,

$$\left|\hat{\tau}_t^{\text{ad-cf}} - \tau_t\right| = O_p(R^{-1/2}) + O_p\Big(\|\tilde{S}_t - S_t\|_2 \, \|\tilde{G}_t - G_t\|_2\Big) + \sum_{i=0}^{t} O_p\Big(\|\tilde{\lambda}_i^S - \lambda_i^S\|_2 \, \|\tilde{G}_{i-1} - G_{i-1}\|_2\Big). \tag{89}$$

Corollary G.2 (Double robustness under no ties)**.** *Under the conditions of Theorem G.1, if either*

$$\|\hat{\lambda}_{t,r-1}^S - \lambda_t^S\|_2 = o_p(1) \qquad \textit{for all } t, \tag{90}$$

or

$$\|\hat{\lambda}_{t,r-1}^G - \lambda_t^G\|_2 = o_p(1) \qquad \textit{for all } t, \tag{91}$$

then

$$\hat{\tau}_{R,t}^{\text{ad-cf}} \xrightarrow{p} \tau_t \qquad \textit{for each } t \leq t_{\max}. \tag{92}$$

Proof. The claim follows directly from Theorem G.1. If either the event hazard estimators or the censoring hazard estimators are L_2-consistent, then the corresponding product terms in the remainder are $o_p(1)$, and hence $\left|\hat{\tau}_t^{\text{ad-cf}} - \tau_t\right| = o_p(1)$. Therefore, $\hat{\tau}_{R,t}^{\text{ad-cf}} \xrightarrow{p} \tau_t$ for each, $t \leq t_{\max}$. □

Why the tied-case remainder is not of pure product form:

We explain the structural reason why the second-order remainder in the tied setting, stated in Theorem 6.3, contains both a pure event-hazard quadratic term and an event–censoring cross-interaction term, in contrast to the clean two-block product structure of the no-ties case (Appendix G).

The source of changes: In both settings, the key remainder after first-order cancellation involves, for each time horizon t and arm a, a sum of the form

$$\mathrm{Rem}_t := \sum_{i=0}^{t} \big(\hat{\lambda}_i^S - \lambda_i^S\big) \cdot \big(\hat{G}_{i-1} - G_{i-1}\big). \tag{93}$$

This cross-product arises because the EIF augmentation term at each time i is weighted by $1/(S_i \cdot G_{i-1})$: when the event hazard estimator deviates from truth, the error in $\hat{G}_{i-1}$ multiplies it. The two settings differ precisely in how G_{i-1} depends on the nuisance parameters.

Tied setting: In the tied setting, the censoring survival function is constructed via a cause-specific factorization, i.e.,

$$G_{i-1}(x,a) = \prod_{k=0}^{i-1} \left(1 - \frac{\lambda_k^G(x,a)}{1 - \lambda_k^S(x,a)}\right), \tag{94}$$

because at each discrete time k the conditional probability of censoring given survival to time k is $\lambda_k^G/(1-\lambda_k^S)$, not λ_k^G alone. Let $\delta_k^S := \hat{\lambda}_k^S - \lambda_k^S$ and $\delta_k^G := \hat{\lambda}_k^G - \lambda_k^G$. A first-order Taylor expansion of $\hat{G}_{i-1} - G_{i-1}$ gives

$$\hat{G}_{i-1} - G_{i-1} \approx -G_{i-1} \sum_{k=0}^{i-1} \frac{1}{1 - \lambda_k^G/(1-\lambda_k^S)} \left[\frac{\delta_k^G}{1-\lambda_k^S} + \frac{\lambda_k^G\, \delta_k^S}{(1-\lambda_k^S)^2}\right]. \tag{95}$$

The critical observation is the second term inside the brackets: because λ_k^G appears in the numerator of the cause-specific hazard, a perturbation δ_k^S in the event hazard also perturbs G_{i-1}. Substituting into Eq. 93 and bounding in L_2, the remainder can be rewritten as

$$\|\mathrm{Rem}_t\| \lesssim \underbrace{\sum_{i=0}^{t} \|\delta_i^S\|_2 \sum_{k=0}^{i-1} \frac{\lambda_k^G}{(1-\lambda_k^S)^2} \|\delta_k^S\|_2}_{\text{pure } \lambda^S \text{ quadratic}} + \underbrace{\sum_{i=0}^{t} \|\delta_i^S\|_2 \sum_{k=0}^{i-1} \frac{\|\delta_k^G\|_2}{1-\lambda_k^S}}_{\lambda^S\text{–}\lambda^G \text{ cross-interaction}}. \tag{96}$$

Since $t \le t_{\max}$ is fixed, summing over i and k contributes only a finite multiplicative constant, so, at the level of asymptotic rates, this gives

$$\|\mathrm{Rem}_t\| = O_p\left(\Big(\sum_{i=0}^{t} \|\hat{\lambda}_i^S - \lambda_i^S\|_2\Big)^2 + \Big(\sum_{i=0}^{t} \|\hat{\lambda}_i^S - \lambda_i^S\|_2\Big)\Big(\sum_{i=0}^{t-1} \|\hat{\lambda}_i^G - \lambda_i^G\|_2\Big)\right). \tag{97}$$

This is precisely the remainder structure appearing in Theorem 6.3. Note in particular that there is no standalone term of the form $O_p(\sum_i \|\delta_i^G\|_2^2)$: a perturbation in λ^G alone, without any event-hazard error, does not appear in Rem_t at all, which is the robustness of the estimating equation with respect to the censoring nuisance.

No-ties setting (contrast): In the no-ties setting, the censoring survival function factorizes independently of the event hazard:

$$G_{i-1}(x,a) = \prod_{k=0}^{i-1} (1 - \lambda_k^G(x,a)). \tag{98}$$

The first-order expansion then gives

$$\hat{G}_{i-1} - G_{i-1} \approx -G_{i-1} \sum_{k=0}^{i-1} \frac{\delta_k^G}{1-\lambda_k^G}, \tag{99}$$

which involves only δ^G and contains no δ^S term. Consequently, Rem_t involves only the cross-product $\delta^S \cdot \delta^G$, yielding the clean remainder structure of Theorem G.1:

$$\|\mathrm{Rem}_t\| = O_p\left(\sum_{i=0}^{t} \|\tilde{\lambda}_i^S - \lambda_i^S\|_2 \, \|\tilde{G}_{i-1} - G_{i-1}\|_2\right). \tag{100}$$

Because the two nuisance blocks are now variation-independent, the remainder vanishes whenever either block is L_2-consistent, giving the literal two-block double robustness of Corollary G.2. In the tied setting, this property fails: even if λ^G is estimated perfectly, a nonzero δ^S still induces a nonzero $\hat{G}_{i-1} - G_{i-1}$ through the cause-specific factorization, so the pure λ^S quadratic term remains.

H ASYMPTOTIC CONFIDENCE SEQUENCES

The fixed-time intervals in Section 6 guarantee $(1-\alpha)$ coverage *solely* at a pre-specified sample size R. In practice, however, analysts often *peek* at interim results and may stop the study early once a decision rule is met (Ramdas et al., 2022); this behavior invalidates fixed-time intervals. To remain valid under such data-dependent stopping, one needs a *confidence sequence (CS)*, i.e., a collection of intervals $[L_{t,r}, U_{t,r}]_{r\geq 1}$ at horizon t satisfying the time-uniform guarantee $\mathbb{P}(\forall r \in \mathbb{N}^+ : \tau_t \in [L_{t,r}, U_{t,r}]) \geq 1-\alpha$.

Constructing *non-asymptotic*, anytime-valid CSs can be difficult when the target estimand contains estimated nuisance functions. Fortunately, for the adaptive survival estimator, the nuisance-induced remainder is asymptotically negligible under the almost-sure rate conditions below, so an *asymptotic CS* (valid after a finite burn-in phase) remains both tractable and practically useful.

Definition H.1 (Asymptotic time-uniform coverage (Dalal et al., 2024))**.** *A sequence of random intervals $\widetilde{C}_{t,r} = [\widetilde{L}_{t,r}, \widetilde{U}_{t,r}]_{r\geq 1}$ is an asymptotic time-uniform $(1-\alpha)$ confidence sequence (AsympCS) for a parameter τ_t at horizon t if the following two conditions hold:*

1. ***Asymptotic confidence sequence:*** *there exists an exact (potentially unknown) $(1-\alpha)$ confidence sequence $C^\star_{t,r} = [L^\star_{t,r}, U^\star_{t,r}]_{r\geq 1}$ such that $\widetilde{L}_{t,r}/L^\star_{t,r} \to 1$ and $\widetilde{U}_{t,r}/U^\star_{t,r} \to 1$ almost surely.*

2. ***Asymptotic time-uniform coverage:*** $\lim_{R_0\to\infty} \mathbb{P}(\forall r \geq R_0 : \tau_t \in \widetilde{C}_{t,r}) \geq 1-\alpha$.

Definition H.1 can be read as follows: if one waits to "peek" until the sample size is sufficiently large ($R \geq R_0$ for some burn-in R_0), the band then covers the true parameter at *every* later time with probability approaching $1-\alpha$. In practice, rare coverage failures occur almost exclusively during this short initial window; later, the intervals tighten rapidly and deliver appreciable power gains over fully non-asymptotic sequences (Cook et al., 2024; Waudby-Smith et al., 2024).

Building on the methodologies of Waudby-Smith et al. (2024) and Cook et al. (2024), we now present the corresponding asymptotic confidence sequence for our estimator.

Theorem H.2 (AsympCS for adaptive survival estimator)**.** *Suppose the conditions of Theorem 6.3 hold. For the time-uniform result, strengthen the policy and nuisance convergence conditions to almost-sure convergence as follows. Let the truncation sequence satisfy $k_r = O(r^\kappa)$ for some $\kappa < 1/4$, and suppose $k_r\|\pi_r - \pi\|_2 = o_{a.s.}(1)$. Furthermore, for each fixed t and treatment a, suppose*

$$\max_{0\leq i\leq t_{\max}} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2 = o_{a.s.}\left(\left(\frac{\log r}{r}\right)^{1/4}\right), \tag{101}$$

$$\max_{0\leq i\leq t_{\max}-1} \|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \lambda^G_i(a,\cdot)\|_2 = o_{a.s.}\left(\left(\frac{\log r}{r}\right)^{1/4}\right). \tag{102}$$

We additionally evaluate the estimated EIF after clipping the nuisance hazards so that, for every $i \leq t$, $a \in \{0,1\}$, and x,

$$1 - \hat{\lambda}^S_{i,r-1}(x,a) \geq \frac{c^S_t}{2}, \qquad 1 - \frac{\hat{\lambda}^G_{i,r-1}(x,a)}{1-\hat{\lambda}^S_{i,r-1}(x,a)} \geq \frac{c^G_t}{2}. \tag{103}$$

Let

$$\hat{V}_{t,R} := \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\phi}_{t,r} - \hat{\tau}^{\mathrm{ASE}}_{t,R}\right)^2 \tag{104}$$

denote the estimated variance process, and fix a user-specified $\rho > 0$. Then the interval

$$\widetilde{C}^{\mathrm{AsympCS}}_R := \left(\hat{\tau}^{\mathrm{ASE}}_{t,R} \pm \sqrt{\frac{2(R\hat{V}_{t,R}\rho^2+1)}{R^2\rho^2}\log\left(\frac{\sqrt{R\hat{V}_{t,R}\rho^2+1}}{\alpha}\right)}\right) \tag{105}$$

is an asymptotic $(1-\alpha)$ *confidence sequence for* τ_t*. The width is approximately minimized at*

$$\rho^\star = \sqrt{\frac{-2\log\alpha + \log(-2\log\alpha + 1)}{R}}. \tag{106}$$

The clipping in Eq. 103 is only used to keep the estimated survival and censoring denominators uniformly away from zero. It can be implemented by projecting the fitted event hazard and the fitted conditional censoring hazard onto the corresponding interior intervals. Since the true hazards satisfy Assumption 6.1, this projection does not worsen the stated L_2 convergence rates. The treatment allocation itself continues to use the truncation rule in Eq. 11; the additional polynomial condition $k_r = O(r^\kappa)$ with $\kappa < 1/4$ is used below to obtain almost-sure control of the centered nuisance-remainder martingale.

Remark H.3 (Difference in convergence rates for fixed-time and anytime-valid inference.). For fixed-time inference, the second-order nuisance remainder is required to be $o_p(r^{-1/2})$, which is implied by pointwise nuisance rates $o_p(r^{-1/4})$. For the AsympCS, the asymptotic linearity remainder only needs to be $o_{a.s.}(\sqrt{\log r/r})$, which is implied by pointwise nuisance rates $o_{a.s.}((\log r/r)^{1/4})$. Thus, the AsympCS condition allows a slightly slower rate up to a logarithmic factor, but requires almost-sure rather than in-probability convergence.

For the proof, we use the following martingale specialization of the asymptotic confidence-sequence result in Waudby-Smith et al. (2024); see also the adaptive-experiment construction of Cook et al. (2024).

Lemma H.4 (Asymptotic CS for an asymptotically linear martingale estimator, adapted from Waudby-Smith et al. (2024))**.** *Let* $\{z_r, \mathcal{H}_r\}_{r\geq 1}$ *be a martingale difference sequence and suppose an estimator* $\hat{\theta}_R$ *satisfies*

$$\hat{\theta}_R - \theta = \frac{1}{R}\sum_{r=1}^{R} z_r + o_{a.s.}\left(\sqrt{\frac{\log R}{R}}\right). \tag{107}$$

Suppose further that, for some $V \in (0,\infty)$*,*

$$\frac{1}{R}\sum_{r=1}^{R}\mathbb{E}[z_r^2 \mid \mathcal{H}_{r-1}] \longrightarrow V \qquad a.s., \tag{108}$$

the martingale increments satisfy the corresponding Lindeberg condition, and an empirical variance estimator $\hat{V}_R$ *satisfies* $\hat{V}_R \to V$ *almost surely. Then the normal-mixture boundary in Eq. 105, with* $\hat{\theta}_R$ *and* $\hat{V}_R$ *substituted for* $\hat{\tau}^{\mathrm{ASE}}_{t,R}$ *and* $\hat{V}_{t,R}$*, forms an asymptotic* $(1-\alpha)$ *confidence sequence for* θ*.*

The main difference from the fixed-time proof in Appendix F.4 is the nuisance term. For fixed-time asymptotic normality, it is enough to show $\frac{1}{\sqrt{R}}\sum_{r=1}^{R} m_{t,r} = o_p(1)$. For the AsympCS, we instead require the stronger almost-sure, time-uniform scale $\sum_{r=1}^{R} m_{t,r} = o_{a.s.}(\sqrt{R\log R})$. Importantly, this cannot be obtained by bounding the realized $m_{t,r}$ pointwise by L_2 nuisance errors, since the first-order nuisance terms cancel only after taking conditional expectation. We therefore separate the conditional bias from the centered martingale fluctuation.

Lemma H.5 (Time-uniform control of the nuisance remainder)**.** *For a fixed horizon* t*, define*

$$m_{t,r} := \hat{\phi}_{t,r} - \phi_{t,r}, \qquad b_{t,r} := \mathbb{E}[m_{t,r} \mid \mathcal{H}_{r-1}], \qquad u_{t,r} := m_{t,r} - b_{t,r}. \tag{109}$$

Under the conditions of Theorem H.2, $\{u_{t,r}, \mathcal{H}_r\}_{r\geq 1}$ *is a martingale difference sequence and*

$$\sum_{r=1}^{R} b_{t,r} = o_{a.s.}(\sqrt{R\log R}), \tag{110}$$

$$\sum_{r=1}^{R} u_{t,r} = o_{a.s.}(\sqrt{R\log R}). \tag{111}$$

Consequently,

$$\frac{1}{R}\sum_{r=1}^{R} m_{t,r} = o_{a.s.}\left(\sqrt{\frac{\log R}{R}}\right). \tag{112}$$

Proof. Throughout the proof, the sequentially cross-fitted nuisance estimates are constructed from $\mathcal{H}_{r-1}$ and are therefore $\mathcal{H}_{r-1}$-measurable. We suppress the cross-fitting superscript for readability. For each arm a, define

$$e^S_{r,a} := \sum_{i=0}^{t} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2, \qquad e^G_{r,a} := \sum_{i=0}^{t-1} \|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \lambda^G_i(a,\cdot)\|_2. \tag{113}$$

Since $t \le t_{\max}$ is fixed, Eqs. 101–102 imply

$$e^S_{r,a} = o_{a.s.}\left(\left(\frac{\log r}{r}\right)^{1/4}\right), \qquad e^G_{r,a} = o_{a.s.}\left(\left(\frac{\log r}{r}\right)^{1/4}\right). \tag{114}$$

Step 1: Conditional bias. The first-order nuisance terms are not required to vanish pathwise. Instead, their cancellation occurs after conditioning on the history and averaging over the current observation. To make this explicit, consider the arm-specific non-centered score

$$\hat{\phi}^a_{t,r} = \hat{S}_{t,r-1}(X_r,a) - \frac{\mathbf{1}(A_r=a)}{\pi_{r,a}(X_r)}\hat{S}_{t,r-1}(X_r,a)\sum_{i=0}^{t}\frac{\mathbf{1}(\widetilde{T}_r = i, \Delta_r = 1) - \mathbf{1}(\widetilde{T}_r \ge i)\hat{\lambda}^S_{i,r-1}(X_r,a)}{\hat{S}_{i,r-1}(X_r,a)\hat{G}_{i-1,r-1}(X_r,a)}, \tag{115}$$

where $\pi_{r,1}(x) = \pi_r(x)$ and $\pi_{r,0}(x) = 1 - \pi_r(x)$. Conditional on X_r and $\mathcal{H}_{r-1}$, the treatment-assignment factor integrates out because $\mathbb{P}(A_r = a \mid X_r, \mathcal{H}_{r-1}) = \pi_{r,a}(X_r)$. The same algebra as in Eq. 64 therefore gives

$$b_{t,r} = \mathbb{E}[\mathcal{R}_{t,r,1}(X_r) - \mathcal{R}_{t,r,0}(X_r) \mid \mathcal{H}_{r-1}], \tag{116}$$

where

$$\mathcal{R}_{t,r,a}(x) = \sum_{i=0}^{t}\frac{\hat{S}_{t,r-1}(x,a)S_{i-1}(x,a)}{\hat{S}_{i,r-1}(x,a)\hat{G}_{i-1,r-1}(x,a)}\{\hat{\lambda}^S_{i,r-1}(x,a) - \lambda^S_i(x,a)\}\{G_{i-1}(x,a) - \hat{G}_{i-1,r-1}(x,a)\}. \tag{117}$$

Thus, the first-order event-hazard perturbation has already canceled after conditional expectation; only its interaction with the error in $\hat{G}_{i-1,r-1}$ remains.

Under Assumption 6.1 and the nuisance clipping in Eq. 103, all survival and censoring denominators in Eq. 117 are uniformly bounded away from zero. Moreover, because

$$G_{i-1}(x,a) = \prod_{j=0}^{i-1}\left(1 - \frac{\lambda^G_j(x,a)}{1-\lambda^S_j(x,a)}\right), \tag{118}$$

the map $(\lambda^S, \lambda^G) \mapsto G_{i-1}$ is Lipschitz on this clipped region. Hence, for a constant $C_t < \infty$ depending only on t and the overlap constants,

$$\|\hat{G}_{i-1,r-1}(a,\cdot) - G_{i-1}(a,\cdot)\|_2 \le C_t\left(e^S_{r,a} + e^G_{r,a}\right). \tag{119}$$

Applying Cauchy–Schwarz to Eq. 117 therefore yields

$$|b_{t,r}| \le C_t \sum_{a\in\{0,1\}}\left[(e^S_{r,a})^2 + e^S_{r,a}e^G_{r,a}\right]. \tag{120}$$

By Eq. 114,

$$|b_{t,r}| = o_{a.s.}\left(\sqrt{\frac{\log r}{r}}\right). \tag{121}$$

Since

$$\sum_{r=1}^{R}\sqrt{\frac{\log r}{r}} = O(\sqrt{R\log R}), \tag{122}$$

Eq. 121 implies

$$\sum_{r=1}^{R} b_{t,r} = o_{a.s.}(\sqrt{R\log R}). \tag{123}$$

Step 2: Conditional second moment of the centered nuisance fluctuation. We next control $u_{t,r} = m_{t,r} - b_{t,r}$. By construction,

$$\mathbb{E}[u_{t,r} \mid \mathcal{H}_{r-1}] = 0, \tag{124}$$

so $\{u_{t,r}, \mathcal{H}_r\}_{r\geq 1}$ is a martingale difference sequence. We now bound its conditional second moment rather than the realized value of $m_{t,r}$.

Write

$$\hat{\xi}_{t,r} := \xi_t(\mathcal{H}_r, \hat{\eta}_{t,r-1}), \qquad \xi_{t,r} := \xi_t(\mathcal{H}_r, \eta_t), \qquad W_r := \frac{A_r - \pi_r(X_r)}{\pi_r(X_r)\{1-\pi_r(X_r)\}}, \tag{125}$$

and define

$$\Delta^S_{t,r}(a) := \hat{S}_{t,r-1}(X_r, a) - S_t(X_r, a). \tag{126}$$

Then

$$m_{t,r} = \Delta^S_{t,r}(1) - \Delta^S_{t,r}(0) - W_r \left[(\hat{\xi}_{t,r} - \xi_{t,r})\hat{S}_{t,r-1}(X_r, A_r) + \xi_{t,r}\Delta^S_{t,r}(A_r)\right]. \tag{127}$$

The treatment truncation in Eq. 11 gives

$$|W_r| \leq k_r. \tag{128}$$

The product-difference identity for the survival curve gives

$$|\Delta^S_{t,r}(a)| \leq \sum_{i=0}^{t} |\hat{\lambda}^S_{i,r-1}(X_r, a) - \lambda^S_i(X_r, a)|, \tag{129}$$

and hence

$$\mathbb{E}\big[|\Delta^S_{t,r}(a)|^2 \mid \mathcal{H}_{r-1}\big] \leq (t+1)\sum_{i=0}^{t} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2^2. \tag{130}$$

Likewise, under Assumption 6.1 and Eq. 103, the mapping from the nuisance hazards to each summand of ξ_t is Lipschitz on a compact region whose denominators are bounded away from zero. Therefore,

$$\mathbb{E}\Big[|\hat{\xi}_{t,r} - \xi_{t,r}|^2 \mid \mathcal{H}_{r-1}\Big] \leq C_t \sum_{a\in\{0,1\}} \left[\sum_{i=0}^{t} \|\hat{\lambda}^S_{i,r-1}(a,\cdot) - \lambda^S_i(a,\cdot)\|_2^2 + \sum_{i=0}^{t-1} \|\hat{\lambda}^G_{i,r-1}(a,\cdot) - \lambda^G_i(a,\cdot)\|_2^2\right]. \tag{131}$$

Combining Eqs. 127–131 gives

$$\mathbb{E}[m^2_{t,r} \mid \mathcal{H}_{r-1}] \leq C_t k_r^2 \sum_{a\in\{0,1\}} \left[(e^S_{r,a})^2 + (e^G_{r,a})^2\right]. \tag{132}$$

Since centering can only reduce conditional variance,

$$\mathbb{E}[u^2_{t,r} \mid \mathcal{H}_{r-1}] \leq \mathbb{E}[m^2_{t,r} \mid \mathcal{H}_{r-1}]. \tag{133}$$

Eq. 132 is the required L_2 control: it bounds the *conditional second moment* of the nuisance perturbation by the L_2 nuisance errors. No pointwise inequality of the form $|m_{t,r}| \lesssim \|\hat{\eta}_{t,r-1} - \eta_t\|_2$ is used.

Step 3: Time-uniform martingale control. Let

$$a_r := \sqrt{r\log(e+r)}. \tag{134}$$

Using $k_r = O(r^\kappa)$, $\kappa < 1/4$, together with Eq. 114, Eqs. 132–133 imply

$$\begin{aligned}\sum_{r=1}^{\infty} \frac{\mathbb{E}[u^2_{t,r} \mid \mathcal{H}_{r-1}]}{a_r^2} &\leq C_t \sum_{r=1}^{\infty} \frac{k_r^2 \sum_{a\in\{0,1\}}\{(e^S_{r,a})^2 + (e^G_{r,a})^2\}}{r\log(e+r)} \\ &= \sum_{r=1}^{\infty} o_{a.s.}\left(\frac{r^{2\kappa-3/2}}{\sqrt{\log(e+r)}}\right) < \infty \qquad a.s., \end{aligned} \tag{135}$$

where the final series is summable because $\kappa < 1/4$. Hence, the martingale

$$\sum_{r=1}^{R} \frac{u_{t,r}}{a_r} \tag{136}$$

converges almost surely. By Kronecker's lemma,

$$\frac{1}{a_R} \sum_{r=1}^{R} u_{t,r} \longrightarrow 0 \qquad a.s., \tag{137}$$

and therefore

$$\sum_{r=1}^{R} u_{t,r} = o_{a.s.}(\sqrt{R \log R}). \tag{138}$$

Combining this with Eq. 110 proves Eq. 112. □

The preceding lemma establishes the asymptotic linearity remainder required for time-uniform inference. The remaining ingredient is consistency of the empirical variance process in Eq. 104.

Lemma H.6 (Consistency of the estimated variance process)**.** *Under the conditions of Theorem H.2,*

$$\hat{V}_{t,R} \longrightarrow V_{\mathrm{eff},t}(\pi) \qquad a.s. \tag{139}$$

Proof. Let

$$z_{t,r} := \phi_{t,r} - \tau_t, \qquad m_{t,r} := \hat{\phi}_{t,r} - \phi_{t,r}, \qquad \delta_{t,R} := \hat{\tau}_{t,R}^{\mathrm{ASE}} - \tau_t. \tag{140}$$

Then

$$\hat{\phi}_{t,r} - \hat{\tau}_{t,R}^{\mathrm{ASE}} = z_{t,r} + m_{t,r} - \delta_{t,R}. \tag{141}$$

We proceed in three steps.

Step 1: Oracle empirical second moment. As shown in Appendix F.4, $\{z_{t,r}, \mathcal{H}_r\}_{r \geq 1}$ is a martingale difference sequence. Let

$$Q_{t,r} := \mathbb{E}[z_{t,r}^2 \mid \mathcal{H}_{r-1}]. \tag{142}$$

Using the conditional variance expression from the fixed-time proof,

$$Q_{t,r} = \mathbb{E}\left[b_t(X_r)^2 + \frac{v_{t,1}(X_r)}{\pi_r(X_r)} + \frac{v_{t,0}(X_r)}{1 - \pi_r(X_r)} \;\middle|\; \mathcal{H}_{r-1} \right], \tag{143}$$

where $b_t(X) = S_t(X,1) - S_t(X,0) - \tau_t$. Since $v_{t,a}$ is uniformly bounded under Assumption 6.1, $\pi(X) \in [\varepsilon, 1-\varepsilon]$, and π_r is truncated to $[1/k_r, 1 - 1/k_r]$, the same calculation as in the conditional-variance part of Appendix F.4 yields

$$|Q_{t,r} - V_{\mathrm{eff},t}(\pi)| \leq C_t k_r \|\pi_r - \pi\|_2 = o_{a.s.}(1). \tag{144}$$

Hence,

$$\frac{1}{R} \sum_{r=1}^{R} Q_{t,r} \longrightarrow V_{\mathrm{eff},t}(\pi) \qquad a.s. \tag{145}$$

To replace the predictable second moment by the realized one, define

$$D_{t,r} := z_{t,r}^2 - Q_{t,r}. \tag{146}$$

Then $\{D_{t,r}, \mathcal{H}_r\}_{r \geq 1}$ is a martingale difference sequence. By the treatment truncation and uniform overlap, there exists $C_t < \infty$ such that

$$|z_{t,r}| \leq C_t k_r. \tag{147}$$

Therefore,

$$\mathbb{E}[D_{t,r}^2 \mid \mathcal{H}_{r-1}] \le \mathbb{E}[z_{t,r}^4 \mid \mathcal{H}_{r-1}] \le C_t k_r^4. \tag{148}$$

Because $k_r = O(r^\kappa)$ with $\kappa < 1/4$,

$$\sum_{r=1}^{\infty} \frac{\mathbb{E}[D_{t,r}^2 \mid \mathcal{H}_{r-1}]}{r^2} \le C_t \sum_{r=1}^{\infty} r^{4\kappa-2} < \infty. \tag{149}$$

The martingale strong law therefore gives

$$\frac{1}{R}\sum_{r=1}^{R} D_{t,r} \longrightarrow 0 \qquad a.s. \tag{150}$$

Together with Eq. 145, this yields

$$\frac{1}{R}\sum_{r=1}^{R} z_{t,r}^2 \longrightarrow V_{\text{eff},t}(\pi) \qquad a.s. \tag{151}$$

Step 2: Mean squared nuisance perturbation. From Eq. 132,

$$q_{t,r} := \mathbb{E}[m_{t,r}^2 \mid \mathcal{H}_{r-1}] \le C_t k_r^2 \sum_{a\in\{0,1\}} \left[(e_{r,a}^S)^2 + (e_{r,a}^G)^2\right]. \tag{152}$$

The conditions $k_r = O(r^\kappa)$ with $\kappa < 1/4$ and Eq. 114 imply

$$q_{t,r} = o_{a.s.}\Big(r^{2\kappa-1/2}\sqrt{\log r}\Big) = o_{a.s.}(1). \tag{153}$$

Hence,

$$\frac{1}{R}\sum_{r=1}^{R} q_{t,r} \longrightarrow 0 \qquad a.s. \tag{154}$$

It remains to replace $q_{t,r}$ by $m_{t,r}^2$. Define

$$H_{t,r} := m_{t,r}^2 - q_{t,r}. \tag{155}$$

Then $\{H_{t,r}, \mathcal{H}_r\}_{r\ge1}$ is a martingale difference sequence. Under the treatment truncation and the nuisance clipping, both the oracle and estimated EIF scores are bounded in absolute value by $C_t k_r$, and hence

$$|m_{t,r}| \le C_t k_r. \tag{156}$$

Therefore,

$$\begin{aligned}\mathbb{E}[H_{t,r}^2 \mid \mathcal{H}_{r-1}] &\le \mathbb{E}[m_{t,r}^4 \mid \mathcal{H}_{r-1}] \\ &\le C_t k_r^2 \mathbb{E}[m_{t,r}^2 \mid \mathcal{H}_{r-1}] \\ &\le C_t k_r^4 \sum_{a\in\{0,1\}} \left[(e_{r,a}^S)^2 + (e_{r,a}^G)^2\right]. \end{aligned} \tag{157}$$

Consequently,

$$\sum_{r=1}^{\infty} \frac{\mathbb{E}[H_{t,r}^2 \mid \mathcal{H}_{r-1}]}{r^2} < \infty \qquad a.s., \tag{158}$$

because the summand is

$$o_{a.s.}\Big(r^{4\kappa-5/2}\sqrt{\log r}\Big), \tag{159}$$

which is summable for every $\kappa < 1/4$. The martingale strong law gives

$$\frac{1}{R}\sum_{r=1}^{R} H_{t,r} \longrightarrow 0 \qquad a.s. \tag{160}$$

Combining this with Eq. 154, we obtain

$$\frac{1}{R}\sum_{r=1}^{R} m_{t,r}^2 \longrightarrow 0 \qquad a.s. \tag{161}$$

Step 3: Replacing the oracle variance by the estimated variance. Lemma H.5 implies

$$\frac{1}{R}\sum_{r=1}^{R} m_{t,r} \longrightarrow 0 \qquad a.s. \tag{162}$$

Moreover, since $Q_{t,r} \to V_{\mathrm{eff},t}(\pi)$ almost surely, the martingale strong law applied to $\{z_{t,r}\}$ gives

$$\frac{1}{R}\sum_{r=1}^{R} z_{t,r} \longrightarrow 0 \qquad a.s. \tag{163}$$

Therefore,

$$\delta_{t,R} = \frac{1}{R}\sum_{r=1}^{R}(z_{t,r} + m_{t,r}) \longrightarrow 0 \qquad a.s. \tag{164}$$

Since $\delta_{t,R}$ is the empirical mean of $z_{t,r} + m_{t,r}$,

$$\hat{V}_{t,R} = \frac{1}{R}\sum_{r=1}^{R}(z_{t,r} + m_{t,r})^2 - \delta_{t,R}^2. \tag{165}$$

Thus,

$$\left|\hat{V}_{t,R} - \frac{1}{R}\sum_{r=1}^{R} z_{t,r}^2\right| \le 2\left(\frac{1}{R}\sum_{r=1}^{R} z_{t,r}^2\right)^{1/2}\left(\frac{1}{R}\sum_{r=1}^{R} m_{t,r}^2\right)^{1/2} + \frac{1}{R}\sum_{r=1}^{R} m_{t,r}^2 + \delta_{t,R}^2. \tag{166}$$

By Eqs. 151, 161, and 164, the right-hand side converges to zero almost surely. Therefore,

$$\hat{V}_{t,R} \longrightarrow V_{\mathrm{eff},t}(\pi) \qquad a.s., \tag{167}$$

which proves the claim. □

Proof of Theorem H.2. We verify the conditions of Lemma H.4. Recall the decomposition from Eq. 43,

$$\hat{\tau}_{t,R}^{\mathrm{ASE}} - \tau_t = \frac{1}{R}\sum_{r=1}^{R} z_{t,r} + \frac{1}{R}\sum_{r=1}^{R} m_{t,r}, \tag{168}$$

where

$$z_{t,r} = \phi_t(\mathcal{H}_r; \pi_r, \eta_t) - \tau_t. \tag{169}$$

Step 1: Asymptotic linearity. By Lemma H.5,

$$\frac{1}{R}\sum_{r=1}^{R} m_{t,r} = o_{a.s.}\left(\sqrt{\frac{\log R}{R}}\right). \tag{170}$$

Therefore,

$$\hat{\tau}_{t,R}^{\mathrm{ASE}} - \tau_t = \frac{1}{R}\sum_{r=1}^{R} z_{t,r} + o_{a.s.}\left(\sqrt{\frac{\log R}{R}}\right), \tag{171}$$

which is the asymptotic linear representation required in Eq. 107.

Step 2: Oracle martingale and limiting variance. The argument in Appendix F.4 shows that

$$\mathbb{E}[z_{t,r} \mid \mathcal{H}_{r-1}] = 0, \tag{172}$$

because π_r is $\mathcal{H}_{r-1}$-measurable and

$$\mathbb{E}[\xi(\mathcal{H}_r, \eta_t) \mid A_r, X_r, \mathcal{H}_{r-1}] = 0. \tag{173}$$

Hence $\{z_{t,r}, \mathcal{H}_r\}_{r\geq 1}$ is a martingale difference sequence. Furthermore, Eq. 144 gives

$$\mathbb{E}[z_{t,r}^2 \mid \mathcal{H}_{r-1}] \longrightarrow V_{\text{eff},t}(\pi) \qquad a.s., \tag{174}$$

and therefore

$$\frac{1}{R}\sum_{r=1}^{R} \mathbb{E}[z_{t,r}^2 \mid \mathcal{H}_{r-1}] \longrightarrow V_{\text{eff},t}(\pi) \qquad a.s. \tag{175}$$

The limiting variance is finite. Indeed, for the limiting policy $\pi(X) \in [\varepsilon, 1-\varepsilon]$ and each arm a,

$$\phi_t^a(O; \pi, \eta_t) = S_t(X,a) - \frac{\mathbf{1}(A=a)}{\pi_a(X)} \sum_{j=0}^{t} \frac{S_t(X,a)}{S_j(X,a) G_{j-1}(X,a)} \left\{ \mathbf{1}(\widetilde{T}=j, \Delta=1) - \mathbf{1}(\widetilde{T}\geq j)\lambda_j^S(X,a) \right\}, \tag{176}$$

where $\pi_1(X) = \pi(X)$ and $\pi_0(X) = 1-\pi(X)$. Since $S_t(X,a)/S_j(X,a) \leq 1$, $\pi_a(X) \geq \varepsilon$, and Assumption 6.1 implies $G_{j-1}(X,a) \geq \underline{g}_t > 0$, we have

$$|\phi_t^a(O; \pi, \eta_t)| \leq 1 + \frac{t+1}{\varepsilon \underline{g}_t} < \infty. \tag{177}$$

Hence $V_{\text{eff},t}(\pi) < \infty$.

The Lindeberg condition follows from the same truncation argument used in Appendix F.4. In particular, Eq. 147 and $k_R = O(R^\kappa)$ with $\kappa < 1/4$ imply

$$\max_{1\leq r\leq R} \frac{|z_{t,r}|}{\sqrt{R}} \leq C_t \frac{k_R}{\sqrt{R}} \longrightarrow 0. \tag{178}$$

Thus, for every $\varepsilon > 0$, the indicator $\mathbb{1}\{|z_{t,r}| > \varepsilon\sqrt{R}\}$ is eventually zero uniformly over $r \leq R$.

Step 3: Consistency of the variance estimator. By Lemma H.6,

$$\hat{V}_{t,R} \longrightarrow V_{\text{eff},t}(\pi) \qquad a.s. \tag{179}$$

We have therefore verified the asymptotic linearity, martingale, predictable-variance, Lindeberg, and variance-estimation conditions of Lemma H.4. Applying Lemma H.4 yields the confidence sequence in Eq. 105. The approximate value of $\rho^\star$ follows from the same normal-mixture boundary calculation as in Waudby-Smith et al. (2024). □

I Extension to Batch Settings

The theoretical analysis in Section 6 is stated for the fully sequential setting in which the policy $\pi_r(X \mid \mathcal{H}_{r-1})$ and the nuisance estimates $\hat{\eta}_{t,r-1}$ are updated after every individual observation. In practice, however, it is common to update both the policy and the nuisance estimates only at the end of each *batch* of observations, thereby reducing computational overhead and stabilizing estimation. We show here that ASE extends naturally to this setting without any modification to the asymptotic normality result.

Batch setting: Partition the R rounds into B batches of size $m = R/B$ each. Let batch b consist of rounds $r \in \mathcal{B}_b := \{(b-1)m+1, \ldots, bm\}$. All units within batch b are assigned using the same policy $\pi_b(X) := \pi_{(b-1)m+1}(X \mid \mathcal{H}_{(b-1)m})$, computed once from the accumulated history at the end of batch $b-1$. Similarly, the nuisance estimators $\hat{\eta}_{b-1}$ are refitted once per batch using $\mathcal{H}_{(b-1)m}$ via

the sequential cross-fitting scheme of Algorithm 1. The ASE estimator is $\hat{\tau}_R^{\mathrm{ASE}} = \frac{1}{R}\sum_{r=1}^{R} \hat{\phi}_r^{\mathrm{cf}}$, as before.

The martingale difference sequence (MDS): The key structural requirement for asymptotic normality is that the oracle pseudo-outcome at horizon t, $\phi(X_{t,r}, Z_{t,r}, A_{t,r}, Y_{t,r}; \pi_{t,r}, \eta)$ forms a martingale difference sequence (MDS) with respect to the filtration $\{\mathcal{H}_{t,r}\}$. In the batch setting, π_b is computed once from $\mathcal{H}_{(b-1)m}$ before any unit in batch b is observed. Since $(b-1)m \le r-1$ for all $r \in \mathcal{B}_b$, the policy π_b is already determined by time $r-1$. Hence, the MDS property $\mathbb{E}[\phi(X_{t,r}, Z_{t,r}, A_{t,r}, Y_{t,r}; \pi_b, \eta) \mid \mathcal{H}_{r-1}] = \tau$ continues to hold exactly. The martingale CLT (Cook et al., 2024) then applies without modification to the oracle term in the near-martingale decomposition

$$\sqrt{R}\left(\hat{\tau}_{t,R}^{\mathrm{ASE}} - \tau_t\right) = \frac{1}{\sqrt{R}}\sum_{r=1}^{R} z_{t,r} + \frac{1}{\sqrt{R}}\sum_{r=1}^{R} m_{t,r}, \tag{180}$$

where $z_{t,r} = \phi(\mathcal{H}_r; \pi_b, \eta_{t,r}) - \tau$ is the oracle MDS term and $m_{t,r} = \hat{\phi}_{t,r}^{\mathrm{cf}} - \phi(\mathcal{H}_r; \pi_b, \eta_{t,r})$ is the nuisance remainder.

L_2-consistency condition in the batch setting. Theorem 6.3 requires $k_r\|\pi_r - \pi^\star\|_2 = o_p(1)$. In the batch setting, $\pi_r \equiv \pi_b$ for all $t \in \mathcal{B}_b$, so this reduces to $k_b\|\pi_b - \pi^\star\|_2 = o_p(1)$ as $b \to \infty$, where k_b is the truncation parameter for batch b. This is no harder to satisfy than in the sequential case; indeed, because each batch-level policy update uses m additional observations relative to the previous update, the plug-in estimate of $\pi^\star$ is typically more stable, making the convergence condition easier to verify in practice.

Cross-fitting within batches. Since all m units in batch b share the nuisance estimator $\hat{\eta}_{b-1}$ fitted on $\mathcal{H}_{(b-1)m}$, the standard sequential cross-fitting scheme from Algorithm 1 applies directly: the history $\mathcal{H}_{(b-1)m}$ is split into two temporal folds, nuisances are fitted on one fold and pseudo-outcomes are evaluated on the other. This ensures that the remainder term satisfies $\frac{1}{\sqrt{R}}\sum_{r=1}^{R} m_r = o_p(1)$ under the same L_2-consistency conditions as in Theorem 6.3, with no additional assumptions.

Asymptotic normality is unchanged. Combining the above, the limiting distribution of the ASE estimator under the batch protocol is identical to the fully sequential case:

$$\sqrt{R}\left(\hat{\tau}_{t,R}^{\mathrm{ASE}} - \tau_t\right) \xrightarrow{d} \mathcal{N}(0,\, V_{\mathrm{eff},t}(\pi^\star))\,, \tag{181}$$

where $V_{\mathrm{eff}}(\pi^\star)$ is the semiparametric efficiency bound from Theorem 4.1. In particular, if $\pi_b \to \pi^\star$ as $b \to \infty$, the estimator remains semiparametrically efficient. The batch size m affects only the finite-sample rate at which π_b converges to $\pi^\star$ and the computational cost of nuisance estimation, but does not alter the asymptotic variance or the form of valid confidence intervals.

J ADDITIONAL EXPERIMENT RESULTS

J.1 TWINS DATA

The dataset considers the birth weight of 11984 pairs of twins born in the USA between 1989 and 1991 with respect to mortality in the first year of life. Treatment $a = 1$ corresponds to being born the heavier twin. The dataset contains 46 confounders. For a detailed description of the dataset, see Louizos et al. (2017).

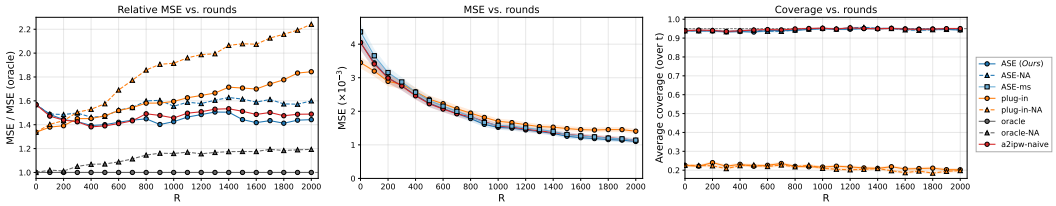


Figure 5: **Twins data:** Adaptive allocation improves estimation efficiency, accuracy, and inferential validity on the semi-synthetic Twins dataset. **Left:** relative MSE with respect to the Oracle; ASE remains closest to oracle performance among all feasible estimators. **Middle:** MSE across rounds; ASE converges fastest and ASE-MS remains consistent under nuisance misspecification. **Right:** empirical coverage of nominal 95% confidence intervals; ASE and ASE-MS maintain nominal coverage while Plugin and A2IPW-NAïve deteriorate as R grows.

(a) *How much do we benefit from adaptive allocation?* ⇒ **Effectiveness of censoring-aware A-optimal allocation**: As shown in Fig. 5 (a), adaptive design consistently delivers efficiency gains over non-adaptive alternatives. ASE achieves a relative MSE of approximately $1.5\times$ the oracle at $R = 2000$, nearly closing the gap to oracle performance despite relying on estimated nuisances. The non-adaptive baselines ASE-NA and Plugin-NA stabilize at roughly $2.3\times$ the oracle MSE, maintaining a persistent efficiency gap throughout. A2IPW-NAïve occupies an intermediate position at approximately $1.70\times$–$1.80\times$, confirming that censoring-aware A-optimal allocation, and not just adaptive assignment alone, is the key driver of efficiency gains over uniform randomization.

(b) *Does ASE improve estimation accuracy over rounds?* ⇒ **Estimation accuracy**: We plot MSE against R in Fig. 5 (b). ASE, ASE-MS, and Plugin converge toward the true τ as R grows, with ASE achieving the lowest MSE due to its variance- and censoring-aware allocation. The results further highlight the importance of *robustness*: ASE-MS tracks ASE closely despite one nuisance component being misspecified, consistent with our theoretical guarantees.

(c) *Is ASE inferentially valid?* ⇒ **Inferential validity**: We report empirical coverage of nominal 95% confidence intervals in Fig. 5 (c). Consistent with our theoretical guarantees, ASE and ASE-MS are the only estimators that maintain nominal coverage uniformly across rounds. In contrast, Plugin, Plugin-NA, and A2IPW-NAïve all degraded substantially as R grows, owing to finite-sample bias and censoring bias, respectively.

J.2 SURVIVAL CURVES

According to Eq. 4, we can write the non-centered EIF for APO via

$$\phi_a(\mathcal{O};\eta_t) = S_t(X,a) + \frac{\mathbf{1}(A=a)}{\pi_a(X)}\,\xi(\mathcal{O},\eta_t) \tag{182}$$

$$= S_t(X,a) + \frac{\mathbf{1}(A=a)}{\pi_a(X)} \sum_{i=0}^{t} \frac{\mathbf{1}(\widetilde{T}=i,\Delta=1) - \mathbf{1}(\widetilde{T}\geq i)\,\lambda_i^S(X,A)}{S_i(X,A)\,G_{i-1}(X,A)} \tag{183}$$

In the following, we show the average survival estimation of ASE over rounds.

Figure 6 plots the ASE-estimated average survival curves $\hat{S}_t(a)$ for $a \in \{0,1\}$ across $t \in \{0,\ldots,t_{\max}\}$, compared against the ground-truth oracle curves $S_t(a)$ (solid lines). Dashed lines and shaded bands report the mean and ± 1 standard error of ASE across repeated trials at increasing sample sizes. On the synthetic dataset, ASE recovers the survival curves of both arms accurately

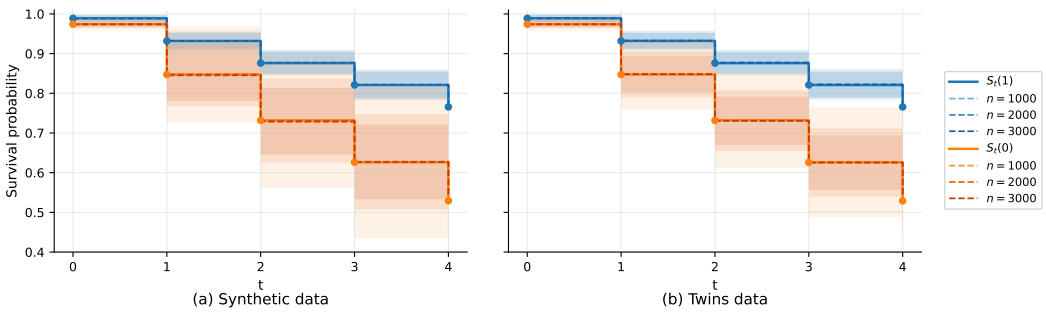


Figure 6: Estimated average survival curves $\hat{S}_t(a)$ for treatment ($a = 1$, blue) and control ($a = 0$, orange) produced by ASE at increasing sample sizes (dashed lines), compared against ground-truth oracle curves $S_t(a)$ (solid lines with dots). Shaded bands denote ± 1 standard error across repeated trials. (*a*) Synthetic data, $t \in \{0, \ldots, 4\}$, $n \in \{1000, 1500, 2000\}$. (*b*) Semi-synthetic Twins data, $t \in \{0, \ldots, 3\}$, $n \in \{50, 500, 1000\}$. Ground-truth construction is detailed in Appendix K.

even at $n = 1000$, with confidence bands tightening visibly as n grows. On the semi-synthetic Twins dataset, a modest upward bias in $\hat{S}_t(1)$ is present at $n = 50$ but resolves by $n = 1000$, consistent with the asymptotic guarantees of Theorem 6.3.

K IMPLEMENTATION DETAILS

K.1 DATA GENERATION

Synthetic data generation: We sample five baseline covariates $X = (X_1, \dots, X_5)^\top$, where $X_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, and construct the scalar index $U = \Phi\Big(5^{-1/2} \sum_{j=1}^{5} X_j\Big)$, so that $U \sim \text{Uniform}(0,1)$ while the observed covariate vector remains five-dimensional. We set $t_{\max} = 4$ and generate the event hazard as

$$\lambda_t^S(u,a) = \sigma\{\alpha_t + \eta_1(u) + a\tau_1(u)\}, \tag{184}$$

where $\sigma(z) = \{1 + \exp(-z)\}^{-1}$ and

$$\eta_f(u) = 1.45\,[0.55\sin(2\pi \times u) + 0.35\cos(4\pi \times u) + 0.25\ \text{sign}\{\sin(6\pi \times u)\}], \tag{185}$$

$$\tau_f(u) = -1.25 + 0.75\,[\text{sign}\{\sin(4\pi \times u)\} + 0.45\sin(6\pi \times u)]. \tag{186}$$

The intercept path is constructed from the reference survival trajectory $(s_0, \dots, s_4) = (0.98, 0.8725, 0.765, 0.6575, 0.55)$ via

$$\alpha_0 = \text{logit}(1 - s_0), \qquad \alpha_t = \text{logit}\left(1 - \frac{s_t}{s_{t-1}}\right), \quad t = 1, \dots, 4, \tag{187}$$

yielding $(\alpha_0, \dots, \alpha_4) = (-3.8918, -2.0939, -1.9624, -1.8110, -1.6324)$. Censoring is generated after the event at each time using the conditional censoring hazard

$$h_t^G(u,a) = \begin{cases} 0.92, & t = 0,\ a = 0, \\ 0.10, & \text{otherwise}, \end{cases} \qquad \lambda_t^G(u,a) = \{1 - \lambda_t^S(u,a)\} h_t^G(u,a), \tag{188}$$

which induces a strong early censoring imbalance in the control arm without repeatedly depleting its risk set. Ground-truth average potential survival outcomes and treatment effects are

$$S_t(u,a) = \prod_{j=0}^{t} \{1 - \lambda_j^S(u,a)\}, \qquad \mu_{a,t} = \mathbb{E}_U[S_t(U,a)], \qquad \theta_t = \mu_{1,t} - \mu_{0,t}, \tag{189}$$

where the expectation is evaluated by Gauss–Legendre quadrature over $u \in [0,1]$.

Twins semi-synthetic data generation: For the semi-synthetic experiment, we sample covariate rows with replacement from the combined Twins training and test sets, retaining `gestat10`, `wt_std`, `mager8`, `preterm`, and `meduc_std`. After standardizing each variable, let $Z = (Z_1, \dots, Z_5)^\top$ and define the scalar summary $M(Z) = 5^{-1} \sum_{j=1}^{5} Z_j$. We map this summary to an approximately uniform index through the empirical probability integral transform,

$$U_i = \widehat{F}_M\{M(Z_i)\} = \frac{\text{rank}\{M(Z_i)\} - 1/2}{N_{\text{Twins}}}, \qquad N_{\text{Twins}} = 11{,}983. \tag{190}$$

This preserves the empirical marginal distributions and dependence structure of the five Twins covariates while yielding an approximately uniform scalar index. We again set $t_{\max} = 4$ and use the same nonlinear event-hazard model, but with the higher frequency $f = 3$,

$$\lambda_t^S(u,a) = \sigma\{\alpha_t + \eta_3(u) + a\tau_3(u)\}, \tag{191}$$

where η_f, τ_f, and $\{\alpha_t\}_{t=0}^4$ are defined as above. The conditional censoring hazard is

$$h_t^G(u,a) = \begin{cases} 0.88, & t = 0,\ a = 0, \\ 0.10, & \text{otherwise}, \end{cases} \qquad \lambda_t^G(u,a) = \{1 - \lambda_t^S(u,a)\} h_t^G(u,a), \tag{192}$$

so that the control arm again experiences a strong initial censoring shock. Ground-truth quantities are computed by averaging over the full empirical Twins covariate pool,

$$\mu_{a,t} = \frac{1}{N_{\text{Twins}}} \sum_{i=1}^{N_{\text{Twins}}} S_t(U_i, a), \qquad \theta_t = \mu_{1,t} - \mu_{0,t}. \tag{193}$$

K.2 Experiment setup

For both synthetic and semi-synthetic datasets, each estimator was evaluated over 500 seeds per dataset. Simulations were run over $R = 20$ rounds with a burn-in period of $N_0 = 1000$ rounds, after which nuisance estimators were updated in mini-batches of 100 units over 20 successive rounds. For all adaptive methods, we applied the clipped optimal allocation policy from Proposition 4.3 with truncation parameter $\alpha = 0.05$. Oracle methods used ground-truth nuisance functions computed analytically from the DGP parameters, while the misspecified variant **ASE-MS** was constructed by replacing $\lambda^G(X, A)$ with a constant fit to the marginal average.

For each estimator, we report the mean squared error at round r, averaged over time horizons:

$$\mathrm{MSE}(r) := \frac{1}{t_{\max} + 1} \sum_{t=0}^{t_{\max}} \left(\hat{\tau}_t(r) - \tau_t\right)^2. \tag{194}$$

The relative MSE is computed as $\mathrm{MSE}(r)/\mathrm{MSE}_{\mathrm{Oracle}}(r)$, where $\mathrm{MSE}_{\mathrm{Oracle}}(r)$ is the MSE of the oracle estimator using ground-truth nuisance functions.

Inference coverage calculation: For each horizon t, we construct a nominal 95% confidence interval for the average survival effect $\tau_t = \mathbb{E}\{S_t(X, 1) - S_t(X, 0)\}$. The reported coverage is averaged over horizons and simulation seeds. Hence, it measures marginal coverage of each τ_t, consistent with our A-optimal trace criterion, rather than simultaneous coverage of the full survival-effect vector.

K.3 Implementation

All experiments were run on an Intel Xeon Silver 4316 CPU with 40 cores and 976GB RAM.

Model architecture and parameters: We estimate the nuisance survival and censoring hazards using LightGBM classifiers, fit separately for each time point and treatment arm, with fixed default hyperparameters.