Sequential Regressions for Efficient Continuous-Time Causal Inference with Treatment-Assigned-At-Visit Interventions

Johan S. Ohlendorff
johan.ohlendorff@sund.ku.dk
University of Copenhagen

Anders Munch
a.munch@sund.ku.dk
University of Copenhagen

Thomas A. Gerds
tag@biostat.ku.dk
University of Copenhagen

Abstract

In medical research, causal effects of treatments that may change over time on an outcome can be defined in the context of an emulated target trial. We are concerned with estimands that are defined as contrasts of the absolute risk that an outcome event occurs before a given time horizon τ under prespecified treatment regimens. Most of the existing estimators based on observational data require a projection onto a discretized time scale (Rose & van der Laan (2011)). We consider a recently developed continuous-time approach to causal inference in this setting (Rytgaard et al. (2022)), which theoretically allows preservation of the precise event timing on a subject level. Working on a continuous-time scale may improve the predictive accuracy and reduce the loss of information. However, continuous-time extensions of the standard estimators comes at the cost of increased computational burden. We provide a new estimator based on sequential regressions for the continuous-time framework which estimates the nuisance parameter models by backtracking through the number of events. This estimator significantly reduces the computational complexity and allows for efficient, single-step targeting using standard machine learning methods from survival analysis, enabling robust continuous-time causal effect estimation.

Keywords continuous-time causal inference · electronic health records · survival analysis · iterative conditional expectations estimator

1 Introduction

Randomized controlled trials (RCTs) are widely regarded as the gold standard for estimating the causal effects of treatments on clinical outcomes. However, RCTs are often expensive, time-consuming, and in many cases infeasible or unethical to conduct. As a result, researchers frequently turn to observational data as an alternative. Even in RCTs, challenges such as treatment noncompliance and time-varying confounding — due to factors like side effects or disease progression — can complicate causal inference.

Marginal structural models (MSMs), introduced by Robins (1986), are a widely used approach for estimating causal effects from observational data, particularly in the presence of time-varying confounding and treatment. Then, once the target parameter has been specified, Longitudinal Targeted Maximum Likelihood Estimation (LTMLE) (van der Laan & Gruber (2012)) can estimate these causal effects. These discrete-time MSMs require that data be recorded on a discrete time scale.

However, many real-world datasets — such as health registries — are collected in continuous time, with patient characteristics updated at irregular, subject-specific times. These datasets often include detailed, timestamped information on events and biomarkers, such as drug purchases, hospital visits, and laboratory results. Analyzing data in its native continuous-time form avoids the need for discretization which can result in bias (Adams et al. (2020); Ferreira Guerra et al. (2020); Kant & Krijthe (2025); Ryalen et al. (2019); Sofrygin et al. (2019); Sun & Crawford (2023)).

In this paper, we consider a longitudinal continuous-time framework similar to that of Rytgaard et al. (2022) and Røysland (2011). Like Rytgaard et al. (2022), we define the parameter of interest nonparametrically and focus on estimation and inference through the efficient influence function, yielding nonparametrically locally efficient estimators via a one-step procedure (Bickel et al. (1993); Tsiatis (2006); van der Vaart (1998)).

To this end, we propose an inverse probability of censoring iterative conditional expectation (ICE-IPCW) estimator, which, like the iterative regression of Rytgaard et al. (2022), iteratively updates nuisance parameters by regressing back through the history. Both methods extend the original discrete-time iterative regression method introduced by Bang & Robins (2005).

A key innovation in our method is that these updates are performed by indexing backwards through the number of events rather than through calendar time. This allows us to apply standard regression techniques to estimate nuisance parameters. Moreover, our estimator accounts for the temporal complexity of the target parameter through inverse probability of censoring weighting (IPCW). To draw an analogy, in multi-state settings, state occupation probabilities are expressed as multiple—often numerous—integrals over the transition intensities between states. In contrast, in non-Markov settings, even if these transition intensities can be estimated, the corresponding integrals are typically not numerically tractable to compute without resorting to Monte Carlo methods. The distinction between event-based and time-based updating is illustrated in Figure 1 and Figure 2.

For electronic health records (EHRs), the number of registrations for each patient can be enormous. However, for finely discretized time grids in discrete time, it has been demonstrated that inverse probability of treatment weighting (IPW) estimators become increasingly biased and inefficient as the number of time points increases whereas iterative regression methods appear to be less sensitive to this issue (Adams et al. (2020)). Yet, many existing methods for estimating causal effects in continuous time apply inverse probability of treatment weighting (IPW) to estimate the target parameter (e.g., Røysland (2011); Røysland et al. (2024)).

Earlier work on continuous-time methods for causal inference in event history analysis are Røysland (2011) and Lok (2008). Røysland (2011) developed identification criteria using a martingale framework based on local independence graphs, enabling causal effect estimation in continuous time via a change of measure. We shall likewise employ a change of measure to define the target parameter.

Most of the existing literature on continuous-time causal inference have been concerned with theoretical identification formulas. In contrast, our work focuses on deriving a feasible, efficient, and debiased estimator. In this sense, our work is related to Shirakawa (2024); however, they restrict attention to a specific type of nuisance parameter estimator and use a discrete-time approximation of the efficient influence function. Our method is agnostic to the type of nuisance parameter estimator used and relies on the exact continuous-time efficient influence function.

In Section 2, we introduce the setting and notation used throughout the paper. In Section 3, we present the estimand of interest and provide the iterative representation of the target parameter. In Section 4, we introduce right-censoring, discuss the implications for inference, and present the algorithm for estimation. In Section 5, we use the Gateaux derivative to find the efficient influence function and present the debiased ICE-IPCW estimator. In Section 6 we present the results of a simulation study and in Section 7 we apply the method to electronic health records data from the Danish registers, emulating a diabetes trial.

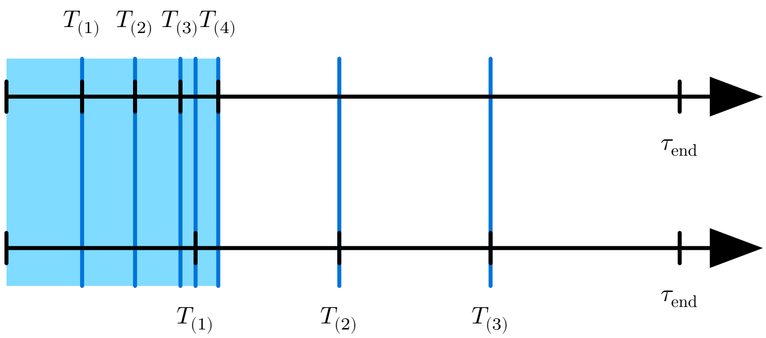

Figure 1: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let $t_1 < ... < t_m$ denotes the event times in the sample. Let P^{G^*} denote the interventional probability measure. Then, given $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{t_r}\right]$, we regress back to $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{t_{r-1}}\right]$.

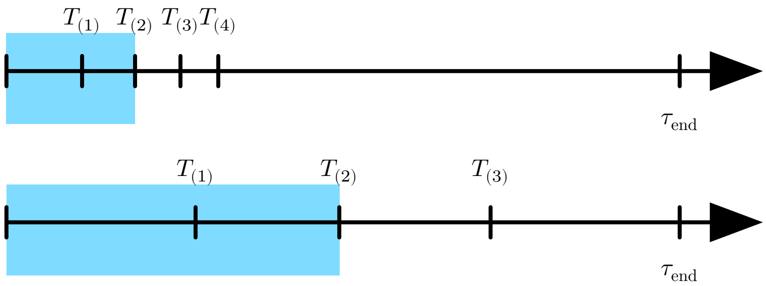

Figure 2: The figure illustrates the sequential regression approach proposed in this article. For each event number k in the sample, we regress back on the history $\mathcal{F}_{T_{(k-1)}}$. Let P^{G^*} denote the interventional probability measure. That is, given $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{T_{(k)}}\right]$, we regress back to $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{T_{(k-1)}}\right]$.

2 Setting and Notation

Let τ_{end} be the end of the observation period. We will focus on the estimation of the interventional absolute risk in the presence of time-varying confounding at a specified time horizon $\tau < \tau_{\text{end}}$. We let $(\Omega, \mathcal{F}, P)$ be a probability space on which all processes and random variables will be defined.

At baseline, we record the values of the treatment $A(0)$ and the time-varying covariates $L(0)$ and let $\mathcal{F}_0 = \sigma(A(0), L(0))$ be the σ-algebra corresponding to the baseline information. We assume that there are a finite number of treatment options. It is then not a loss of generality to assume that we have two treatment options over time so that $A(t) \in \{0, 1\}$ (e.g., placebo and active treatment), where $A(t)$ denotes the treatment at time $t \geq 0$ as contrasts are often of interest.

The time-varying confounders $L(t)$ at time $t > 0$ are assumed to take values in a finite subset $\mathcal{L} \subset \mathbb{R}^m$, so that $L(t) \in \mathcal{L}$ for all $t \geq 0$ with probability 1. We assume that the stochastic processes $(L(t))_{t\geq 0}$ and $(A(t))_{t\geq 0}$ are càdlàg (right-continuous with left limits), jump processes. Furthermore, we require that the times at which the treatment and covariate values may change are dictated entirely by the counting processes $(N^a(t))_{t\geq 0}$ and $\left(N^\ell(t)\right)_{t\geq 0}$, respectively in the sense that $\Delta A(t) \neq 0$ only if $\Delta N^a(t) \neq 0$ and $\Delta L(t) \neq 0$ only if $\Delta N^\ell(t) \neq 0$ or $\Delta N^a(t) \neq 0$. The process N^a determines when the doctor may decide on treatment, but we allow for the time-varying covariates to change at the same time as the treatment values. N^ℓ determines when only the covariate values may change, but not the treatment values. This means that we can practically assume that $\Delta N^a \Delta N^\ell \equiv 0$, i.e., that there are no common jumps between N^a and N^ℓ. For technical reasons and ease of notation, we shall assume that the number of jumps at time τ_{end} for the processes L and A satisfies $N^a(\tau_{\text{end}}) + N^\ell(\tau_{\text{end}}) \leq K - 1$ P-a.s. for some $K \geq 1$. Let further $(N^y(t))_{t\geq 0}$ and $\left(N^d(t)\right)_{t\geq 0}$ denote the counting processes for the event of interest and the competing event, respectively. It is reasonable to also assume that $\Delta N^y \Delta N^x \equiv 0$ for $x \in \{a, \ell, d\}$ and $\Delta N^d \Delta N^x \equiv 0$ for $x \in \{a, \ell, y\}$; therefore (N^y, N^d, N^a, N^ℓ) is a multivariate counting process in the sense of Andersen et al. (1993).

Thus, we have observations from a jump process $\alpha(t) = \left(N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t)\right)$, and the natural filtration $(\mathcal{F}_t)_{t \geq 0}$ is given by $\mathcal{F}_t = \sigma(\alpha(s) \mid s \leq t) \vee \mathcal{F}_0$. Let $T_{(k)}$ be the k'th ordered jump time of α, that is $T_0 = 0$ and $T_{(k)} = \inf\left\{t > T_{(k-1)} \mid \alpha(t) \neq \alpha\left(T_{(k-1)}\right)\right\} \in [0, \infty]$ be the time of the k'th event and let $\Delta_{(k)} \in \{c, y, d, a, \ell\}$ be the status of the k'th event, i.e., $\Delta_{(k)} = x$ if $\Delta N^x\left(T_{(k)}\right) = 1$. We let $T_{(k+1)} = \infty$ if $T_{(k)} = \infty$ or $\Delta_{(k)} \in \{y, d, c\}$. As is common in the point process literature, we define $\Delta_{(k)} = A(\infty) = L(\infty) = \emptyset$ if $T_{(k)} = \infty$ for the empty mark.

The observed process can without loss of information be encoded as $O = \left(T_{(K)}, \Delta_{(K)}, A\left(T_{(K-1)}\right), L\left(T_{(K-1)}\right), T_{(K-1)}, \Delta_{(K-1)}, ..., A(0), L(0)\right) \sim P \in \mathcal{M}$ where $\mathcal{M}$ is the statistical model, i.e., a set of probability measures and obtain a sample $O = (O_1, ..., O_n)$ of size n. As a concrete example, we refer to Table 1, which provides the long format of a hypothetical longitudinal dataset with time-varying covariates and treatment registered at irregular time points, and its conversion to wide format in Table 2.

To the process $(\alpha(t))_{t \geq 0}$, we associate the corresponding random measure N on $(\mathbb{R}_+ \times (\{a, \ell, y, d, c\} \times \{0, 1\} \times \mathcal{L}))$ by

$$N(\mathrm{d}(t, x, a, \ell)) = \sum_{k : T_{(k)} < \infty} \delta_{\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right)\right)}(\mathrm{d}(t, x, a, \ell)), \tag{2.1}$$

where δ_x denotes the Dirac measure on $(\mathbb{R}_+ \times (\{a, \ell, y, d, c\} \times \{0, 1\} \times \mathcal{L}))$. It follows that the filtration $(\mathcal{F}_t)_{t \geq 0}$ is the natural filtration of the random measure N (e.g., Theorem 2.5.10 of Last & Brandt (1995)) under a no explosion assumption, i.e., that the number of jumps in $[0, \tau_{\text{end}}]$ is finite P-a.s. Thus, the random measure N carries the same information as the stochastic process $(\alpha(t))_{t \geq 0}$. This point is important for dealing with right-censoring.

We work with the natural filtration $(\mathcal{F}_t)_{t \geq 0}$ generated by the random measure N within the so-called canonical setting for technical reasons (Last & Brandt (1995), Section 2.2). This is needed to ensure that the compensators can be explicitly written via by the regular conditional distributions of the jump times and marks but also to ensure that $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right)\right) \vee \mathcal{F}_{T_{(k-1)}}$ and $\mathcal{F}_0 = \sigma(A(0), L(0))$, where $\mathcal{F}_{T_{(k)}}$ stopping time σ-algebra $\mathcal{F}_{T_{(k)}}$ – representing the information up to and including the k'th event – associated with stopping time $T_{(k)}$.

id	time	event	L	A
1	0	baseline	2	1
1	0.5	visitation time; stay on treatment	2	1
1	8	primary event	Ø	Ø
2	0	baseline	1	0
2	10	primary event	Ø	Ø
3	0	baseline	3	1
3	2	side effect (L)	4	1
3	2.1	visitation time; discontinue treatment	4	0
3	5	primary event	Ø	Ø

Table 1: An example of a longitudinal dataset from electronic health records or a clinical trial with $\tau_{\text{end}} = 15$ with $K = 2$ for $n = 3$. Events are registered at irregular/subject-specific time points and are presented in a long format.

id	$L(0)$	$A(0)$	$L(T_{(1)})$	$A(T_{(1)})$	$T_{(1)}$	$\Delta_{(1)}$	$L(T_{(2)})$	$A(T_{(2)})$	$T_{(2)}$	$\Delta_{(2)}$	$T_{(3)}$	$\Delta_{(3)}$
1	2	1	2	1	0.5	a	Ø	Ø	8	y	∞	Ø
2	1	0	Ø	Ø	10	y	Ø	Ø	∞	Ø	∞	Ø
3	3	1	4	1	2	ℓ	4	0	2.1	a	5	y

Table 2: The same example as in Table 1, but presented in a wide data format.

Let $\pi_k\left(t, L\left(T_{(k)}\right), \mathcal{F}_{T_{(k-1)}}\right)$ be the probability of being treated at the k'th event given $\Delta_{(k)} = a$, $T_{(k)} = t$, $L\left(T_{(k)}\right)$, and $\mathcal{F}_{T_{(k-1)}}$. Let also $\Lambda_k^x\left(dt, \mathcal{F}_{T_{(k-1)}}\right)$ be the cumulative cause-specific hazard measure, that is the measure given by $\mathbb{1}\{t > 0\}\left(P\left(T_{(k)} \in dt, \Delta_{(k)} = x \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\left(P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)\right)^{-1}$. At baseline, we let $\pi_0(L(0))$ be the probability of being treated given $L(0)$ and $\mu_0(\cdot)$ be the probability measure for the covariate value.

3 Estimand of interest and iterative representation

We are interested in the causal effect of a treatment regime g on the cumulative incidence function of the event of interest y at time τ. We consider regimes which naturally affects the treatment decisions at each visitation time but not the times at which the patients visit the doctor. The treatment regime g specifies for each event $k = 1, \dots, K-1$ with $\Delta_{(k)} = a$ (visitation time) the probability that a patient receives treatment at the current visitation time via $\pi_k^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)$ and at $k = 0$ the initial treatment probability $\pi_0^*(L(0))$.

Define the likelihood ratio process

$$
\begin{aligned}
W^g(t) = &\prod_{k=1}^{N_t} \left(\frac{\pi_k^*\left(T_{(k)}, L\left(T_{(k)}\right), \mathcal{F}_{T_{(k-1)}}\right)^{A\left(T_{(k)}\right)} \left(1 - \pi_k^*\left(T_{(k)}, L\left(T_{(k)}\right), \mathcal{F}_{T_{(k-1)}}\right)\right)^{1-A\left(T_{(k)}\right)}}{\pi_k\left(T_{(k)}, L\left(T_{(k)}\right), \mathcal{F}_{T_{(k-1)}}\right)^{A\left(T_{(k)}\right)} \left(1 - \pi_k\left(T_{(k)}, L\left(T_{(k)}\right), \mathcal{F}_{T_{(k-1)}}\right)\right)^{1-A\left(T_{(k)}\right)}} \right)^{\mathbb{1}\left\{\Delta_{(k)}=a\right\}} \\
&\times \frac{\pi_0^*(L(0))^{A(0)}(1-\pi_0^*(L(0)))^{1-A(0)}}{\pi_0(L(0))^{A(0)}(1-\pi_0(L(0)))^{1-A(0)}},
\end{aligned}
\tag{3.1}
$$

where $N_t = \sum_k \mathbb{1}\left\{T_{(k)} \leq t\right\}$ is random variable denoting the number of events up to time t. We define the measure P^{G^*} by $\frac{\mathrm{d}P^{G^*}}{\mathrm{d}P} = W^g(\tau_{\text{end}})$, which represents the interventional world in which the doctor assigns treatments according to the probability measure π_k^* for $k = 0, ..., K-1$. Our target parameter is the mean interventional cumulative incidence function at time τ under P^{G^*}, which can by identified through the formula

$$
\Psi_\tau^g(P) = \mathbb{E}_{P^{G^*}}[N^y(\tau)] = \mathbb{E}_P[N^y(\tau)W^g(\tau)],
$$

where $N^y(t) = \sum_{k=1}^K \mathbb{1}\left\{T_{(k)} \leq t, \Delta_{(k)} = y\right\}$.

In our application, π_k^* may be chosen arbitrarily, so that, in principle, *stochastic*, *dynamic*, and *static* treatment regimes can be considered. However, for simplicity of presentation, we use the static observation plan $\pi_k^* = 1$ for all $k = 0, ..., K-1$, and the methods we present can easily be extended to more complex treatment regimes and contrasts. This corresponds to the interventions considered in Rytgaard et al. (2022).

If we do not want to limit ourselves to K events, we can interpret the target parameter $\Psi_\tau^g(P)$ as the counterfactual cumulative incidence function of the event of interest y at time τ, when the intervention enforces treatment as part of the $K-1$ first events. We denote this target parameter by $\Psi_\tau^{g,K}(P)$ and return to this interpretation later in Section 5.

We now present a simple iterated representation of the target parameter $\Psi_\tau^g(P)$ in the case with no censoring. We discuss more thoroughly the implications for inference of this representation, the algorithm for estimation and examples in Section 4 where we also deal with right-censoring. Note that the quantities given in Theorem 1 are also defined for $u < \tau$ as we need those definitions later for the efficient influence function.

Theorem 1: Let $H_k = \left(L\left(T_{(k)}\right), T_{(k)}, \Delta_{(k)}, A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right), T_{(k-1)}, \Delta_{(k-1)}, ..., A(0), L(0)\right)$ be the history up to and including the k'th event, but excluding the k'th treatment values for $k > 0$. Let $H_0 = L(0)$. Let $\bar{Q}_{K,\tau}^g : (u, a_k, h_k) \mapsto 0$ and recursively define for $k = K-1, ..., 1$,

$$
\begin{aligned}
Z_{k+1,\tau}^a(u) = &\mathbb{1}\left\{T_{(k+1)} \leq u, T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\right\} \bar{Q}_{k+1,\tau}^g\left(\tau, A\left(T_{(k)}\right), H_{k+1}\right) \\
&+ \mathbb{1}\left\{T_{(k+1)} \leq u, T_{(k+1)} < \tau, \Delta_{(k+1)} = a\right\} \bar{Q}_{k+1,\tau}^g(\tau, 1, H_{k+1}) \\
&+ \mathbb{1}\left\{T_{(k+1)} \leq u, \Delta_{(k+1)} = y\right\},
\end{aligned}
$$

and

$$
\bar{Q}_{k,\tau}^g : (u, a_k, h_k) \mapsto \mathbb{E}_P\left[Z_{k+1,\tau}^a(u) \mid A\left(T_{(k)}\right) = a_k, H_k = h_k\right], \tag{3.2}
$$

for $u \leq \tau$. Then,

$$
\Psi_\tau^g(P) = \mathbb{E}_P\left[\bar{Q}_{0,\tau}^g(\tau, 1, L(0))\right]. \tag{3.3}
$$

Proof: The proof is given in the Appendix (Section A.1). □

Throughout the article, we will use the notation $\bar{Q}_{k,\tau}^g(u)$ to denote the value of $\bar{Q}_{k,\tau}^g\left(u, A\left(T_{(k)}\right), H_k\right)$ and $\bar{Q}_{k,\tau}^g$ to denote $\bar{Q}_{k,\tau}^g\left(\tau, A\left(T_{(k)}\right), H_k\right)$. Note that $\bar{Q}_{k,\tau}^g$ represents the counterfactual probability of the primary event occuring at or before time τ given the history up to and including the k'th event, among the people who are at risk of the event before time τ after k events. Equation (3.2) then suggests that we can estimate $\bar{Q}_{k-1,\tau}^g$ via $\bar{Q}_{k,\tau}^g$: For each individual in the sample, we calculate the integrand in Equation (3.2) depending on their value of $T_{(k)}$ and $\Delta_{(k)}$, and apply the intervention by setting $A\left(T_{(k)}\right)$ to 1 if $\Delta_{(k)} =$

a. Then, we regress these values directly on $\left(A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right), T_{(k-1)}, \Delta_{(k-1)}, ..., A(0), L(0)\right)$ to obtain an estimator of $\bar{Q}^g_{k-1,\tau}$.

Note that here, we only set the current value of $A\left(T_{(k)}\right)$ to 1, instead of replacing all prior values with 1. The latter is certainly closer to the original iterative conditional expectation estimator (Bang & Robins (2005)), and mathematically equivalent, but computationally more demanding. Sofrygin et al. (2019) demonstrated this in the discrete-time setting.

4 Censoring

In this section, we consider the situation where the observed data is subject to right-censoring. Let $C > 0$ denote the right-censoring time at which we stop observing the multivariate jump process α. Let N^c be the censoring process given by $N^c(t) = \mathbb{1}\{C \leq t\}$. In Section 3, we proposed an estimation strategy based on fitting a sequence of iterated regressions backward in time for each event. When the data is subject to right-censoring, standard regression techniques cannot immediately be applied. To handle this, we use inverse probability of censoring weights. Our algorithm is presented in Algorithm 1 and we later present the assumptions necessary for validity of the ICE-IPCW estimator in Section 4.1. To do so, we first need additional notation.

In the remainder of the paper, we will assume that $\Delta N^c \Delta N^x \equiv 0$ for $x \in \{y, d, a, \ell\}$. We now let $\left(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, A(\bar{T}_k), L(\bar{T}_k)\right)$ for $k = 1, ..., K$ be the observed data given by

$$\bar{T}_{(k)} = \begin{cases} T_{(k)} \text{ if } C > T_{(k)} \\ C \text{ if } C \leq T_{(k)} \text{ and } T_{(k-1)} > C \\ \infty \text{ otherwise} \end{cases}$$

$$\bar{\Delta}_{(k)} = \begin{cases} \Delta_{(k)} \text{ if } C > T_{(k)} \\ \text{c if } C \leq T_{(k)} \text{ and } T_{(k-1)} > C \\ \emptyset \text{ otherwise} \end{cases}$$

$$A(\bar{T}_k) = \begin{cases} A\left(T_{(k)}\right) \text{if } C > T_{(k)} \\ A(C) \text{ if } \bar{\Delta}_{(k)} = c \\ \emptyset \text{ otherwise} \end{cases}$$

$$L(\bar{T}_k) = \begin{cases} L\left(T_{(k)}\right) \text{ if } C > T_{(k)} \\ L(C) \text{ if } \bar{\Delta}_{(k)} = c \\ \emptyset \text{ otherwise} \end{cases}$$

for $k = 1, ..., K$, and $\mathcal{F}_{\bar{T}_{(k)}}$ is given by

$$\mathcal{F}_{\bar{T}_{(k)}} = \sigma\left(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, A(\bar{T}_k), L(\bar{T}_k), ..., \bar{T}_{(1)}, \bar{\Delta}_{(1)}, A(\bar{T}_1), L(\bar{T}_1), A(0), L(0)\right),$$

defining the observed history up to and including the k'th event. Thus $\tilde{O} = \left(\bar{T}_{(1)}, \bar{\Delta}_{(1)}, A(\bar{T}_1), L(\bar{T}_1), ..., \bar{T}_{(K)}, \bar{\Delta}_{(K)}, A(\bar{T}_K), L(\bar{T}_K)\right)$ is the observed data and a sample consists of $\tilde{O} = \left(\tilde{O}_1, ..., \tilde{O}_n\right)$ for n independent and identically distributed observations with $\tilde{O}_i \sim P$.

Define $\tilde{\Lambda}^c_k\left(\mathrm{d}t, \mathcal{F}_{\bar{T}_{(k-1)}}\right)$ as the cause-specific hazard measure for censoring of the k'th event given the observed history $\mathcal{F}_{\bar{T}_{(k-1)}}$, that is $\mathbb{1}\{t > 0\}\left(P\left(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)\right)\left(P\left(\bar{T}_{(k)} \geq t \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)\right)^{-1}$ and the corresponding conditional censoring survival functions $\tilde{S}^c\left(t \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) = \prod_{s \in \left(T_{(k-1)}, t\right]}\left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}s, \mathcal{F}_{\bar{T}_{(k-1)}}\right)\right)$, where $\prod_{s \in (0,t]}$ is the product integral over the interval $(0, t]$ (Gill & Johansen (1990)). Also let $\widetilde{M}^c(t) = \widetilde{N}^c(t) - \tilde{\Lambda}^c(t)$ where $\widetilde{N}^c(t) = \mathbb{1}\{C \leq t, T^e > t\} = \sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k)} \leq t, \bar{\Delta}_{(k)} = c\right\}$ is the censoring counting process, $\tilde{\Lambda}^c(t) = \sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\}\tilde{\Lambda}^c_k\left(t, \mathcal{F}_{\bar{T}_{(k-1)}}\right)$ is the censoring compensator and $S\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = \prod_{s \in (0,t]}\left(1 - \sum_{x=a,\ell,y,d} \Lambda^x_k\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right)$ is the overall (uncensored) survival function for the k'th event given $\mathcal{F}_{T_{(k-1)}}$.

With these definitions, we can now present the ICE-IPCW procedure in Algorithm 1. Note that ideally the model for iterative regressions should be chosen flexibly, since even with full knowledge of the data-generating mechanism, the true functional form of the regression model cannot typically be derived in closed form. We also recommend that the model should also be chosen such that the predictions are $[0,1]$ -valued.

Algorithm 1 (ICE-IPCW procedure):

Input: Observed data $\tilde{O}_i$, $i = 1, ..., n$, estimator of the censoring compensator $\hat{\Lambda}^c$, time horizon $\tau < \tau_{\text{end}}$, and K.

Output: ICE-IPCW estimator $\hat{\Psi}_n$ of $\Psi_\tau^g(P)$.

1. Initialize $(\hat{Q}_K(\tau, a_k, h_k) := 0)$.
2. For each event $k = K-1, ..., 1$:

   a. For each observation, compute $\hat{S}^c\big(\bar{T}_{(k)} - \mid A(\bar{T}_{k-1}), H_{k-1}\big) = \prod_{s \in (\bar{T}_{(k-1)}, \bar{T}_{(k)})} \big(1 - \hat{\Lambda}^c(s)\big)$.

   b. For each observation with $\bar{T}_{(k-1)} < \tau$ and $\bar{\Delta}_{(k-1)} \in \{a, \ell\}$ compute the pseudo-outcome $\hat{Z}^a_{k,\tau}$ as follows:

   - If $\bar{\Delta}_{(k)} = y$, calculate $\hat{Z}^a_{k,\tau} = \frac{1}{\hat{S}^c(\bar{T}_{(k)}- \mid A(\bar{T}_{k-1}), H_{k-1})} \mathbb{1}\big\{\bar{T}_{(k)} \le \tau\big\}$.
   - Else if $\bar{\Delta}_{(k)} = a$, evaluate $\hat{Q}_{k+1}(\tau, 1, H_k)$ and calculate
   $$\hat{Z}^a_{k,\tau} = \frac{1}{\hat{S}^c\big(\bar{T}_{(k)} - \mid A(\bar{T}_{k-1}), H_{k-1}\big)} \mathbb{1}\big\{\bar{T}_{(k)} < \tau\big\} \hat{Q}_{k+1}(\tau, 1, H_k).$$
   - Else if $\bar{\Delta}_{(k)} = \ell$, evaluate $\hat{Q}_{k+1}(\tau, A(\bar{T}_{k-1}), H_k)$, and calculate
   $$\hat{Z}^a_{k,\tau} = \frac{1}{\hat{S}^c\big(\bar{T}_{(k)} - \mid A(\bar{T}_{k-1}), H_{k-1}\big)} \mathbb{1}\big\{\bar{T}_{(k)} < \tau\big\} \hat{Q}_{k+1}(\tau, A(\bar{T}_{k-1}), H_k).$$

   c. Regress $\hat{Z}^a_{k,\tau}$ on $(A(\bar{T}_{k-1}), \bar{H}_{k-1})$ for the observations with $\bar{T}_{(k-1)} < \tau$ and $\bar{\Delta}_{(k-1)} \in \{a, \ell\}$ to obtain a prediction function $\hat{Q}_k$.
3. Compute $\hat{\Psi}_n = \frac{1}{n} \sum_{i=1}^n \hat{Q}_0(\tau, 1, L_i(0))$.

4.1 Validity of the ICE-IPCW Estimator

Let T^e denote the (uncensored) terminal event time, i.e., $T^e = \inf_{t>0}\{N^y(t) + N^d(t) = 1\}$ and let $\beta = (\alpha, N^c)$ be the full unobserved multivariate jump process in $[0, \tau_{\text{end}}]$. Its natural filtration is denoted $\mathcal{F}^{\text{full}}_t = \sigma(\beta(s) \mid s \le t) \vee \mathcal{F}_0$. Thus, we observe the trajectories of the process given by $[0, \tau_{\text{end}}] \ni t \mapsto \beta(t \wedge C \wedge T^e)$ and the observed filtration is given by $\bar{\mathcal{F}}_t = \sigma(\beta(s \wedge C \wedge T^e) \mid s \le t) \vee \mathcal{F}_0$. Let $N^{\text{full}}(\mathrm{d}(t, m, a, l)) = \mathbb{1}\{m \ne c\} N(\mathrm{d}(t, m, a, l)) + \delta_{(c, A(C), L(C))}(\mathrm{d}(m, a, l)) N^c(\mathrm{d}t)$ be the *full* random measure, where N^c is the counting process for censoring events.

The *canonical compensator* ρ for a random measure N^* is a stochastic kernel from $\mathcal{Y} \times N_{\boldsymbol{X}} \times \mathbb{R}_+$ to $\mathbb{R}_+ \times \boldsymbol{X}$ satisfying that $\rho(Y, N^*, \mathrm{d}(t, x))$ is a P-$\mathcal{F}^*_t$ compensator of N^*, where $\mathcal{F}^*_t = \sigma(Y, N^*_t)$. Here $N_{\boldsymbol{X}}$ denotes the canonical point process space with mark space $\boldsymbol{X}$, and measurable element $Y \in \mathcal{Y}$ at $t = 0$. For $N^* = N^{\text{full}}$, we take $\boldsymbol{X} = \{0,1\} \times \mathcal{L} \times \{a, \ell, d, y, c\}$ and $\mathcal{Y} = \{0,1\} \times \mathcal{L}$.

We posit an independent censoring condition that enables the use of regression techniques similar to those that may be found the literature based on independent censoring (Andersen et al. (1993); Definition III.2.1) or local independence conditions (Røysland et al. (2024); Definition 4). Note that the second is a slight strengthening of the fact that the martingales $\widetilde{M}^c$ and M^x are orthogonal for $x \in \{a, \ell, d, y\}$. For example if the compensator of the (observed) censoring process is absolutely continuous with respect to the Lebesgue measure, then the second of Theorem 2 is satisfied. Our third condition in Theorem 2 is a positivity condition, ensuring that the conditional expectations are well-defined.

Theorem 2: If

1. Let ρ denote the canonical compensator of N^{full} and ρ' denote the canonical compensator of N. Suppose that $\varphi = (t_1, d_1, a_1, l_1, ..., t_K, d_K)$ and define the censoring time accordingly as $c(\varphi)$. We have $\mathbb{1}\{m \neq c\}\rho((a_0, l_0), \varphi, \mathrm{d}(t, m, a, l)) = \rho'((a_0, l_0), \varphi', \mathrm{d}(t, m, a, l))$, where $\varphi' = (t'_1, d'_1, a'_1, l'_1, ..., t'_K, d'_K)$, $(t'_k, d'_k, a'_k, l'_k) = (t_k, d_k, a_k, l_k)$ if $t_k \leq c(\varphi)$; $(t'_k, d'_k, a'_k, l'_k) = (c(\varphi), d'_{k-1}, a'_{k-1}, l'_{k-1})$ if $t_{k-1} < c(\varphi) < t_k$; $(t'_k, d'_k, a'_k, l'_k) = (t_{k-1}, d_{k-1}, a_{k-1}, l_{k-1})$ otherwise.
2. $\Delta\tilde{\Lambda}^c_k\left(\cdot, \mathcal{F}_{\bar{T}_{(k-1)}}\right)\Delta\Lambda^x_k\left(\cdot, \mathcal{F}_{T_{(k-1)}}\right) \equiv 0$ for $x \in \{a, \ell, y, d\}$ and $k \in \{1, ..., K\}$.
3. $\tilde{S}^c\left(t \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) > \eta$ for all $t \in (0, \tau]$ and $k \in \{1, ..., K\}$ P-a.s. for some $\eta > 0$.

Let

$$\begin{aligned}\bar{Z}^a_{k,\tau}(u) = \frac{1}{\tilde{S}^c\left(\bar{T}_{(k)} - \mid A(\bar{T}_{k-1}), \bar{H}_{k-1}\right)}\Big(&\mathbb{1}\left\{\bar{T}_{(k)} \leq u, \bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = a\right\}\bar{Q}^g_{k,\tau}(1, \bar{H}_k) \\ &+\mathbb{1}\left\{\bar{T}_{(k)} \leq u, \bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = \ell\right\}\bar{Q}^g_{k,\tau}(A(\bar{T}_k), \bar{H}_k) \\ &+\mathbb{1}\left\{\bar{T}_{(k)} \leq u, \bar{\Delta}_{(k)} = y\right\}\Big)\end{aligned} \tag{4.1}$$

Then with $h_k = (a_k, l_k, t_k, d_k, ..., a_0, l_0)$,

$$\mathbb{1}\{d_1 \in \{a, \ell\}, ..., d_k \in \{a, \ell\}\}\bar{Q}^g_{k,\tau}(u, a_k, h_k) = \mathbb{E}_P\left[\bar{Z}^a_{k+1,\tau}(u) \mid A(\bar{T}_k) = a_k, \bar{H}_k = h_k\right], \tag{4.2}$$

where $\bar{H}_k = \left(L(\bar{T}_k), A(\bar{T}_{k-1}), \bar{T}_{(k-1)}, \bar{\Delta}_{(k-1)}, ..., A(0), L(0)\right)$ is the history up to and including the k'th event. Hence, $\Psi^g_\tau(P)$ is identifiable from the observed data.

Proof: Proof is given in the Appendix (Section B.1). □

5 Inference

In this section, we derive the efficient influence function for Ψ^g_τ which we denote by φ^*_τ. The effient influence function is a P-indexed function which is square integrable and has mean zero under P. The efficient influence function φ^*_τ can be interpreted as the functional derivative of Ψ^g_τ (Hines et al. (2022); Kennedy (2024)). The efficient influence function can be used to construct efficient estimators and construct asymptotically valid tests and confidence interval. In addition, influence function-based estimators can use of machine learning methods to estimate the nuisance parameters under certain regularity. To achieve this, we debias our initial ICE-IPCW estimator (Algorithm 2 and Algorithm 3) using double/debiased machine learning (Chernozhukov et al. (2018)). This leads to the one-step estimator,

$$\hat{\Psi}_n = \hat{\Psi}^0_n + \mathbb{P}_n\varphi^*_\tau\left(\cdot; \hat{P}\right),$$

where $\hat{\Psi}^0_n$ is the initial estimator ICE-IPCW estimator, and $\hat{P}$ is a collection of estimates for the nuisance parameters appearing in Equation (5.3). Under certain regularity conditions, this estimator is asymptotically linear at P with influence function $\varphi^*_\tau(\cdot; P)$. We derive the efficient influence function using the iterative representation given in Equation (4.2), working under the conclusions of Theorem 2, by finding the pathwise derivative of the target parameter.

We also provide an algorithm for the one-step estimator in the uncensored situation or to obtain conservative inference when censoring is present (Algorithm 2). Alternatively, we may use the algorithm in Algorithm 3 of Section F.1 to obtain valid non-conservative inference although this estimator is not further here. We leave non-conservative inference and doubly robust estimation in right-censored settings as a future research topic.

We note the close resemblance of Equation (5.3) to the well-known efficient influence function for the discrete time case which was established in Bang & Robins (2005), with the most notable difference being the presence of the martingale term $\widetilde{M}^c(\mathrm{d}u)$ in Equation (5.3).

A key feature of our approach is that the efficient influence function is expressed in terms of the martingale for the censoring process. This representation is computationally simpler, as it avoids the need to estimate multiple martingale terms. For a detailed comparison, we refer the reader to the appendix, where we show that our efficient influence function simplifies to the same as the one derived by Rytgaard et al. (2022) (Section C.1).

Theorem 3 (Efficient influence function): Suppose that there is a universal constant $C^* > 0$ such that $\tilde{\Lambda}_k^c\left(\tau_{\text{end}}, \mathcal{F}_{\bar{T}_{(k-1)}}; P\right) \le C^*$ for all $k = 1, ..., K$ and every $P \in \mathcal{M}$.

Define

$$\begin{aligned}\varphi_\tau^{*,d}(P) = \frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))} \sum_{k=1}^K \prod_{j=1}^{k-1} & \left(\frac{\mathbb{1}\{A(\bar{T}_j)=1\}}{\pi_j\left(\bar{T}_{(j)}, L(\bar{T}_j), \mathcal{F}_{\bar{T}_{(j-1)}}\right)}\right)^{\mathbb{1}\{\bar{\Delta}_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} \tilde{S}^c\left(\bar{T}_{(j)} - \mid \mathcal{F}_{\bar{T}_{(j-1)}}\right)} \\ & \times \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \in \{\ell, a\}, \bar{T}_{(k-1)} < \tau\right\}\left(\bar{Z}_{k,\tau}^a(\tau) - \bar{Q}_{k-1,\tau}^g(\tau)\right) \\ & + \bar{Q}_{0,\tau}^g(\tau, 1, L(0)) - \Psi_\tau^g(P)\end{aligned} \tag{5.1}$$

and

$$\begin{aligned}\varphi_\tau^{*,\widetilde{M}^c}(P) = \frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))} \sum_{k=1}^K \prod_{j=1}^{k-1} & \left(\frac{\mathbb{1}\{A(\bar{T}_j)=1\}}{\pi_j\left(\bar{T}_{(j)}, L(\bar{T}_j), \mathcal{F}_{\bar{T}_{(j-1)}}\right)}\right)^{\mathbb{1}\{\bar{\Delta}_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} \tilde{S}^c\left(\bar{T}_{(j)} - \mid \mathcal{F}_{\bar{T}_{(j-1)}}\right)} \\ & \int_{\left(\bar{T}_{(k-1)}, \tau\right)} \mathbb{1}\left\{s \le \bar{T}_{(k)}\right\}\left(\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(u)\right) \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) S\left(u - \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} \widetilde{M}^c(\mathrm{d}u)\end{aligned} \tag{5.2}$$

The efficient influence function at $P \in \mathcal{M}$ is given by

$$\varphi_\tau^*(P) = \varphi_\tau^{*,\widetilde{M}^c}(P) + \varphi_\tau^{*,d}(P). \tag{5.3}$$

Proof: The proof is given in the Appendix (Section D.1). □

Algorithm 2 (Debiased ICE-IPCW estimator (conservative)): Input: Observed data $\tilde{O}_i$, $i = 1, ..., n$, time horizon $\tau < \tau_{\text{end}}$, and K. Estimators of the propensity score $\hat{\pi}_0$, $\hat{\pi}_k$ for $k = 1, ..., K-1$ and the censoring compensator $\hat{\Lambda}^c$.

Output: One-step estimator $\hat{\Psi}_n$ of $\Psi_\tau^g(P)$; estimate of influence function $\varphi_\tau^*\left(\tilde{O}; \hat{P}\right)$.

1. Compute the ICE-IPCW estimator $\hat{\Psi}_n^0$ and obtain estimators of $\bar{Q}_{k,\tau}^g$ for $k = 0, ..., K-1$

using Algorithm 1.

2. Let $\hat{Q}_k$ and $\hat{S}^c\left(\cdot \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)$ be the estimates obtained in Algorithm 1. Compute Equation (5.1)

using these estimates and $\hat{\pi}_k$ to obtain estimates of $\varphi^{*,d}\left(\tilde{O}_i; \hat{P}\right)$ for $i = 1, ..., n$.

3. Compute the one-step estimator

$$\hat{\Psi}_n = \hat{\Psi}_n^0 + \frac{1}{n} \sum_{i=1}^n \varphi_\tau^{*,d}\left(\tilde{O}_i; \hat{P}\right).$$

4. Return $\hat{\Psi}_n$ and $\varphi_\tau^{*,d}\left(\tilde{O}_i; \hat{P}\right)$ for $i = 1, ..., n$.

We may not know the maximum number of events that can occur a priori. Furthermore, the maximal number of events may be enormous and it may be difficult to estimate $\bar{Q}_{k,\tau}^g$ for large k due to the limited number of observations with many events. We shall show that we can data-adaptively choose this number such that we get valid inference for our target estimand.

Let $\Psi_\tau^{g,k^*}(P) = \mathbb{E}_P\left[N^y(\tau) W^g\left(\tau \wedge T_{(k^*)}\right)\right]$ for $k^* \in \mathbb{N}$ denote the target parameter when we only intervene up to the k^*'th event, that is among the k^* first events, if the patient visits the doctor, the doctor enforces treatment, but not after k^*'th event. The theory developed thus far describes this parameter with $k^* = K-1$. Let $N_{\tau-} = \sum_k \mathbb{1}\left\{T_{(k)} < \tau, \Delta_{(k)} \in \{a, \ell\}\right\}$ and $\widetilde{N}_{\tau-} = \sum_k \mathbb{1}\left\{T_{(k)} < \tau, \Delta_{(k)} \in \{a, \ell\}, T_{(k)} < C\right\}$ be the number of non-terminal events before time τ in the uncensored and censored data, respectively. Also let $K_{nc} = \max_{v: \sum_{i=1}^n \mathbb{1}\{\widetilde{N}_{\tau-,i} \geq v\} > c}$ denote the maximum number of events with at least c people at risk, where $\widetilde{N}_{\tau-,i}$ denotes $\widetilde{N}_{\tau-}$ for subject i. Suppose that we have the decomposition of the estimator $\hat{\Psi}_n$ of the data-adaptive target parameter $\Psi_\tau^{g,K_{nc}}(P)$, such that

$$\hat{\Psi}_n - \Psi_\tau^{g,K_{nc}}(P) = (\mathbb{P}_n - P)\varphi_\tau^{*,K_{nc}}(\cdot; P) + o_P\left(n^{-\frac{1}{2}}\right), \tag{5.4}$$

where $\varphi_\tau^{*,k^*}(\cdot; P)$ is the efficient influence function for the target parameter $\Psi_\tau^{g,k^*}(P)$. Then as Theorem 4 shows, we actually obtain inference for the original parameter of interest if the total number of events is bounded. Note that here we use the term data-adaptive target parameter in the sense of Hubbard et al. (2016).

Theorem 4 (Adaptive selection of K): Suppose that Equation (5.4) holds and that there is a (non-random) number $K_{\text{lim}} \in \mathbb{N}$, such that $P(N_{\tau-} = K_{\text{lim}}) > 0$, but $P(N_{\tau-} > K_{\text{lim}}) = 0$ and that $P\left(\widetilde{N}_{\tau-} = K_{\text{lim}}\right) > 0$. Then, the estimator $\hat{\Psi}_n$ satisfies

$$\hat{\Psi}_n - \Psi_\tau^{g,K_{\text{lim}}}(P) = (\mathbb{P}_n - P)\varphi_\tau^{*,K_{\text{lim}}}(\cdot; P) + o_P\left(n^{-\frac{1}{2}}\right).$$

Proof: The proof is given in the Appendix (Section E.1). □

6 Simulation study

We conduct a simulation study to evaluate the performance of our ICE-IPCW estimator and its debiased version. Overall, the purpose of the simulation study is to establish that our estimating procedure provides valid inference with varying degrees of confounding.

Additionally, the objective is to compare with existing methods in causal inference, such as discrete time methods (van der Laan & Gruber (2012)) and a naive Cox model which treats deviation as censoring, not addressing time-varying confounding. In the censored setting, we also address the choice of nuisance parameter models for the iterative regressions.

Simulation scenario: We simulate a cohort of patients who initiate treatment at time $t = 0$, denoted by $A(0) = 1$ and who are initially stroke-free, $L(0) = 0$. All individuals are followed for up to $\tau_{\text{end}} = 900$ days or until death. During follow-up, patients may experience (at most) one stroke, stop treatment (irreversibly), die, or drop out that is $\widetilde{N}^x(t) \leq 1$ for $x = a, \ell, y, c$. The primary outcome is the *risk of death within* $\tau =$ 720 *days*. We simulate baseline data as age $\sim \text{Unif}(40, 90)$, $L(0) = 0$, and $A(0) = 1$. To simulate the time-varying data, we generate data according to the following compensator

$$\begin{aligned}\tilde{\Lambda}(\mathrm{d}(t, x, a, \ell)) = {} & \delta_{(y, A(t-), L(t-))}(\mathrm{d}(x, a, \ell))\lambda^y(t)\,\mathrm{d}t \\ & + \delta_{(\ell, A(t-), 1)}(\mathrm{d}(x, a, \ell))\lambda^\ell(t)\,\mathrm{d}t \\ & + \delta_{(y, L(t-))}(\mathrm{d}(x, \ell))\pi_t(\mathrm{d}a)\lambda^a(t)\,\mathrm{d}t \\ & + \delta_{(c, A(t-), L(t-))}(\mathrm{d}(x, a, \ell))\lambda_c(t)\,\mathrm{d}t\end{aligned} \tag{6.1}$$

where

$$\begin{aligned}\lambda^y(t) &= \mathbb{1}\{t \leq T^e \wedge C \wedge \tau_{\text{end}}\}\lambda^y \exp\left(\beta^y_{\text{age}}\text{ age} + \beta^y_A A(t-) + \beta^y_L L(t-)\right) \\ \lambda^\ell(t) &= \mathbb{1}\{t \leq T^e \wedge C \wedge \tau_{\text{end}}\}\mathbb{1}\left\{\widetilde{N}^\ell(t-) = 0\right\}\lambda^\ell \exp\left(\beta^\ell_{\text{age}}\text{ age} + \beta^\ell_A A(t-)\right) \\ \lambda^a(t) &= \mathbb{1}\{t \leq T^e \wedge C \wedge \tau_{\text{end}}\}\mathbb{1}\left\{\widetilde{N}^a(t-) = 0\right\} \\ & \quad \times \left(\left(1 - \mathbb{1}\left\{\widetilde{N}^\ell(t-) = 0\right\}\right)\gamma_0 \exp\left(\gamma_{\text{age}}\text{ age}\right) + \mathbb{1}\left\{\widetilde{N}^\ell(t-) = 0\right\}h_z(t; 360; \varepsilon)\right) \\ \pi_t(\mathrm{d}a) &= \mathbb{1}\{t \leq T^e \wedge C \wedge \tau_{\text{end}}\}\mathbb{1}\left\{\widetilde{N}^a(t-) = 0\right\}\left(\delta_1(\mathrm{d}a)\pi(t \mid \bar{\mathcal{F}}_{t-}) + \delta_0(\mathrm{d}a)(1 - \pi(t \mid \bar{\mathcal{F}}_{t-}))\right) \\ \lambda_c(t) &= \mathbb{1}\{t \leq T^e \wedge C \wedge \tau_{\text{end}}\}\lambda_c\end{aligned}$$

where $h_z(t; 360; 5; \varepsilon)$ is the hazard function for a Normal distribution with mean 360 and standard deviation 5, truncated from some small value $\varepsilon > 0$ and $\pi(t \mid \bar{\mathcal{F}}_{t-}) = \text{expit}\left(\alpha_0 + \alpha_{\text{age}}\ \text{age} + \alpha_L L(t-)\right)$ is the treatment assignment probability. Our intervention is $\pi^*(t \mid \bar{\mathcal{F}}_{t-}) = 1$ which corresponds to sustained treatment throughout the follow-up period and, in the censored setting, $\lambda_c^* = 0$.

Note that Equation (6.1) states that the intensities for N^ℓ and N^y correspond to multiplicative intensity models. The case $x = a$ requires a bit more explanation: The visitation intensity depends on whether the patient has had a stroke or not. If the patient has not had a stroke, the model specifies that the patient can be expected to visit the doctor within 360 days (i.e., the patient is scheduled). If the patient has had a stroke, the visitation intensity is multiplicative, depending on age, and reflects the fact that a patient, who has had a stroke, is expected to visit the doctor within the near future.

In the uncensored setting ($\lambda_c = 0$), we vary the treatment effect on the outcome corresponding to $\beta^y_A > 0$, $\beta^y_A = 0$, and $\beta^y_A < 0$ and the effect of stroke on the outcome $\beta^y_L > 0$, $\beta^y_L = 0$, and $\beta^y_L < 0$. We also vary the effect of a stroke on the treatment propensity α_L and the effect of treatment on stroke $\beta^\ell_A > 0$, $\beta^\ell_A = 0$, and $\beta^\ell_A < 0$. Furthermore, when applying LTMLE, we discretize time into 8 intervals (see e.g., Section G.2). We consider both the debiased ICE estimator and the ICE estimator without debiasing. For modeling of the nuisance parameters, we select a logistic regression model for the treatment propensity $\pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ and a generalized linear model (GLM) with the option `family = quasibinomial()` for the conditional counterfactual probabilities $\bar{Q}^g_{k,\tau}$. For the LTMLE procedure, we use an undersmoothed LASSO (Tibshirani (1996)) estimator. Additionally, we vary sample size in the uncensored setting ($n \in \{100, 2000, 500, 1000\}$); otherwise $n = 1000$.

For the censored setting, we consider a simulation involving *completely* independent censoring, where we vary the degree of censoring $\lambda_c \in \{0.0002, 0.0005, 0.0008\}$ in Equation (6.1). We consider only two parameter settings for the censoring martingale as outlined in Table 3. Three types of models are considered for the estimation of the counterfactual probabilities $\bar{Q}^g_{k,\tau}$:

1. A linear model, which is a simple linear regression of the pseudo-outcomes $\hat{Z}^a_{k,\tau}$ on the treatment and history variables.
2. An ad-hoc scaled quasibinomial GLM, which is a generalized linear model with the `quasibinomial` as a family argument, where the outcomes are scaled down to $[0, 1]$ by dividing with the largest value of $\hat{Z}^a_{k,\tau}$ in the sample. Afterwards, the predictions are scaled back up to the original scale by multiplying with the largest value of $\hat{Z}^a_{k,\tau}$ in the sample.
3. A tweedie GLM, which is a generalized linear model with the `tweedie` family, as the pseudo-outcomes $\hat{Z}^a_{k,\tau}$ may appear marginally as a mixture of a continuous random variable and a point mass at 0.

Parameters	$\boldsymbol{\alpha_0}$	$\boldsymbol{\alpha_{\text{age}}}$	$\boldsymbol{\alpha_{\text{L}}}$	$\boldsymbol{\beta^y_{\text{age}}}$	$\boldsymbol{\beta^\ell_{\text{age}}}$	$\boldsymbol{\beta^y_A}$	$\boldsymbol{\beta^\ell_A}$	$\boldsymbol{\beta^y_{\text{L}}}$	$\boldsymbol{\lambda^y}$	$\boldsymbol{\lambda^\ell}$	$\boldsymbol{\gamma_{\text{age}}}$	$\boldsymbol{\gamma_0}$
Values (varying effects)	0.3	0.02	**−0.2**, <u>0</u>, 0.2	0.025	0.015	**−0.3**, 0, 0.3	**−0.2**, 0, 0.2	−0.5, <u>0</u>, **0.5**	0.0001	0.001	0	0.005
Values (strong confounding)	0.3	0.02	−0.6, 0.6	0.025	0.015	−0.8, 0.8	−0.2	1	0.0001	0.001	0	0.005
Values (censoring)	0.3	0.02	−0.6, 0.6	0.025	0.015	−0.8, 0.8	−0.2	1	0.0001	0.001	0	0.005

Table 3: Simulation parameters for the simulation studies. Each value is varied while holding the others fixed. Bold values indicate fixed reference values, and underlined values denote the scenarios without time-varying confounding.

6.1 Results

We present the results of the simulation study in Table 4 and Table 5 in the strong and no time confounding cases, respectively. In the tables, we report the mean squared error (MSE), mean bias, standard deviation of the estimates, and the mean of the estimated standard error, as well as coverage of 95% confidence intervals. We also present boxplots of the results, showing bias (Figure 3, Figure 5, Figure 7, and Figure 9), as well as standard errors (Figure 4, Figure 6, and Figure 7), depending on the parameters. Additional results, such as those involving sample size, can be found in the appendix (Section G.1).

Across all scenarios considered in the uncensored setting (Table 4 and Table 5 and Figure 3, Figure 4, Figure 5, and Figure 6), it appears that the debiased ICE-IPCW estimator has good performance with respect to bias, coverage, and standard errors. The debiased ICE-IPCW estimator is unbiased even in settings with substantial time-varying confounding and consistently matches or outperforms both the naive Cox method and the LTMLE estimator.

Interestingly, when strong time-varying confounding is present, LTMLE estimates are biased, but the mean squared errors are about the same as for the debiased ICE-IPCW estimator, likely owing to the fact that LTMLE has generally smaller standard errors. This reflects a bias–variance trade-off between continuous-time and discrete-time approaches.

In the presence of right-censoring (Figure 7, Figure 8, and Figure 9), we see that the debiased ICE-IPCW estimator remains unbiased across all simulation scenarios and all choices of nuisance parameter models. Moreover, standard errors are (slightly) conservative with the trend that standard errors become more biased as the degree of censoring increases as is to be expected.

When looking at the selection of nuisance parameter models for the pseudo-outcomes, we find that the linear model provides the most biased estimates for the non-debiased ICE-IPCW estimator (Figure 9), though the differences are not substantial. In Figure 7, we see that for the debiased ICE-IPCW estimator, there is no substantial difference between the linear, scaled quasibinomial, and tweedie models. Also note that the Tweedie model produces slightly larger standard errors for the debiased ICE-IPCW estimator than the linear or scaled quasibinomial models. However, the differences are otherwise minor.

β^y_A	**Estimator**	**Coverage**	**MSE**	**Bias**	**sd($\widehat{\Psi}_n$)**	**Mean($\widehat{\text{SE}}$)**
-0.3	ICE-IPCW (deb.)	0.947	0.000252	0.0000732	0.0159	0.0158
	LTMLE	0.946	0.000245	0.00217	0.0155	0.0154
	Naive Cox		0.000244	-0.0000436	0.0156	
	ICE-IPCW		0.000252	0.0000597	0.0159	
0	ICE-IPCW (deb.)	0.946	0.000296	0.00055	0.0172	0.0171
	LTMLE	0.947	0.000284	0.000553	0.0168	0.0167
	Naive Cox		0.000287	0.000419	0.0169	
	ICE-IPCW		0.000295	0.000555	0.0172	
0.3	ICE-IPCW (deb.)	0.949	0.000323	0.0000806	0.018	0.018
	LTMLE	0.946	0.000314	-0.00221	0.0176	0.0176
	Naive Cox		0.000314	-0.000105	0.0177	
	ICE-IPCW		0.000323	0.0000917	0.018	

Table 4: Results for the case without time-varying confounding. The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error and the estimator applied are provided.

β_A^y	α_L	**Estimator**	**Coverage**	**MSE**	**Bias**	**sd($\widehat{\Psi}_n$)**	**Mean($\widehat{\text{SE}}$)**
-0.8	-0.6	ICE-IPCW (deb.)	0.947	0.000315	-0.000168	0.0178	0.0177
		LTMLE	0.898	0.00042	0.0115	0.017	0.0167
		Naive Cox		0.000323	-0.00599	0.017	
		ICE-IPCW		0.000314	-0.000202	0.0177	
0.8	0.6	ICE-IPCW (deb.)	0.95	0.000256	0.00014	0.016	0.016
		LTMLE	0.951	0.000261	-0.00315	0.0159	0.0161
		Naive Cox		0.000273	0.00476	0.0158	
		ICE-IPCW		0.000255	0.000124	0.016	

Table 5: Results for the case with strong time-varying confounding. The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error and the estimator applied are provided.

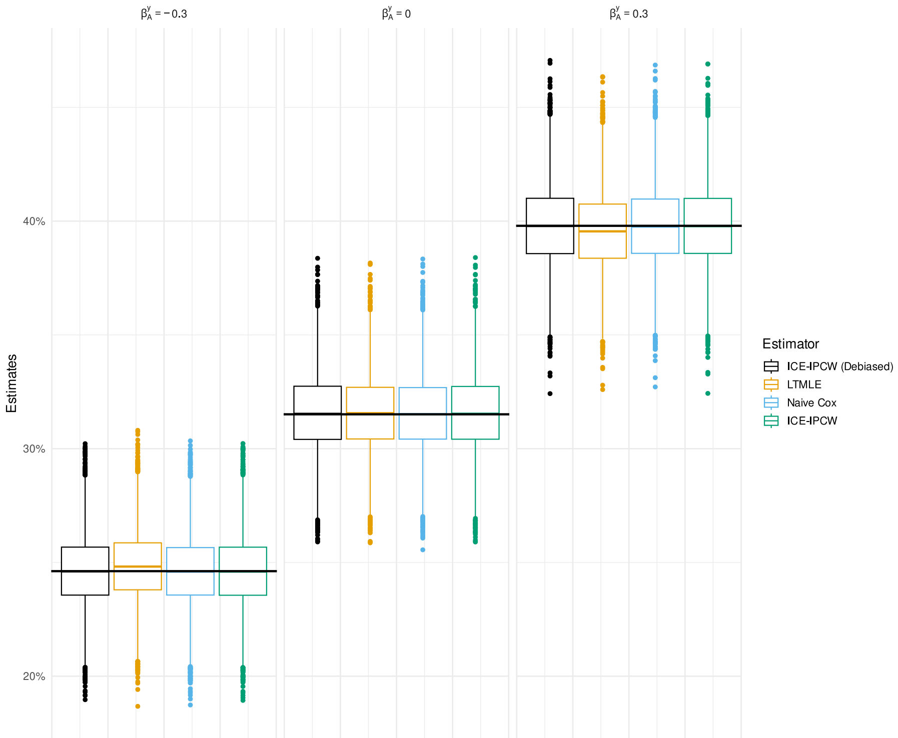

Figure 3: Boxplots of the estimates for the case without time-varying confounding for each estimator in each parameter setting for the cases without time confounding. The lines indicates the true value of the target parameter $\Psi_\tau^g(P)$.

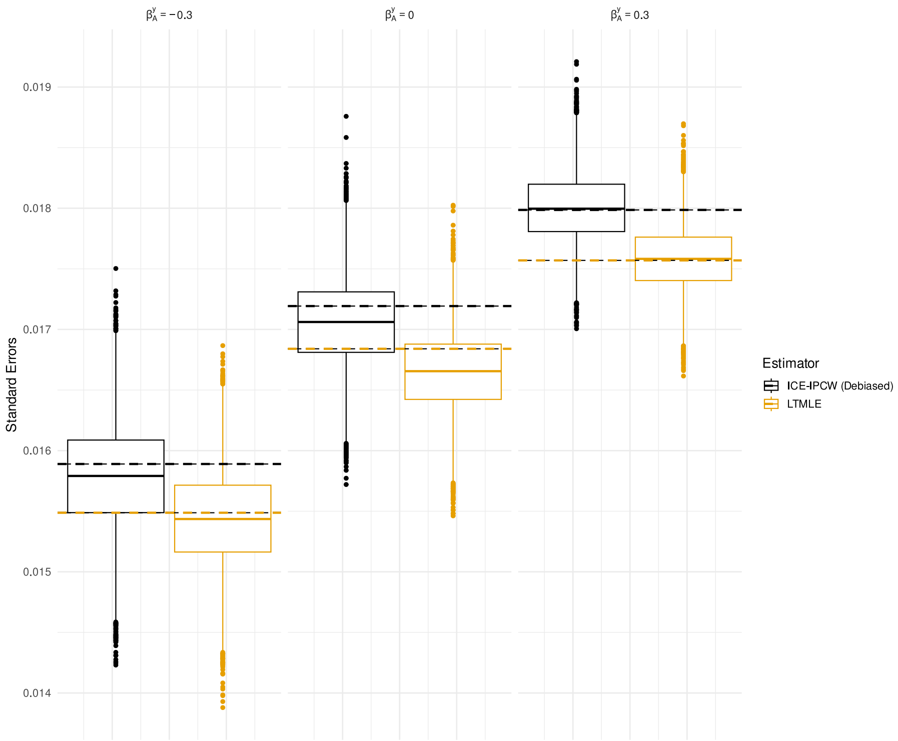

Figure 4: Boxplots of the standard errors for the case without time-varying confounding for each estimator (LTMLE and debiased ICE-IPCW) in each parameter setting for the cases without time-varying confounding. The lines indicates the empirical standard error of the estimates for each estimator.

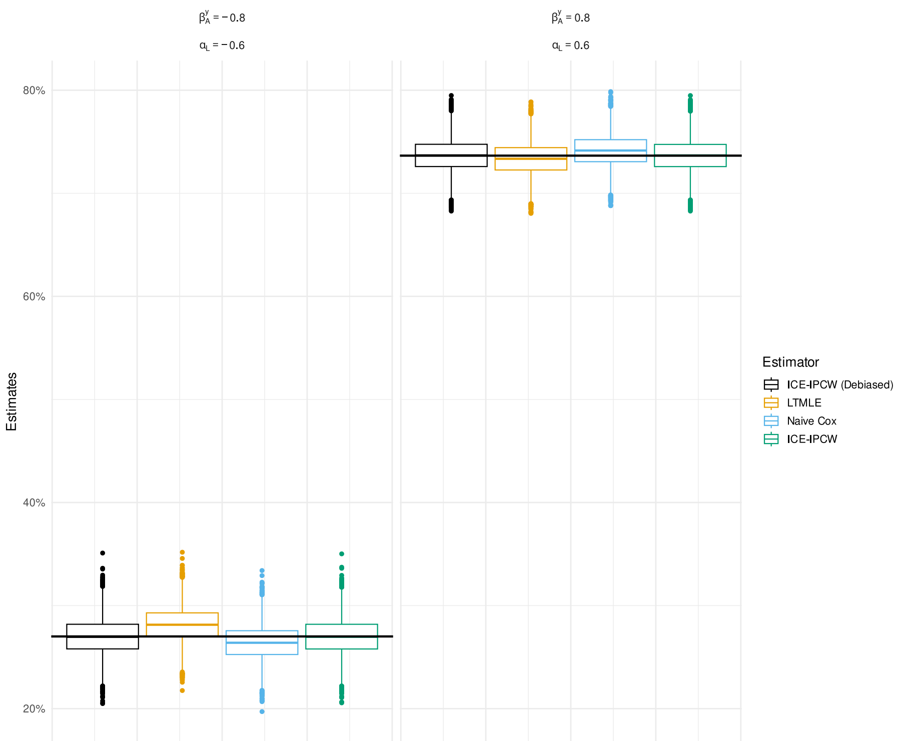

Figure 5: Boxplots of the estimates for the case with strong time-varying confounding for each estimator in each parameter setting for the cases with strong time-varying confounding. The lines indicates the true value of the target parameter $\Psi^g_\tau(P)$.

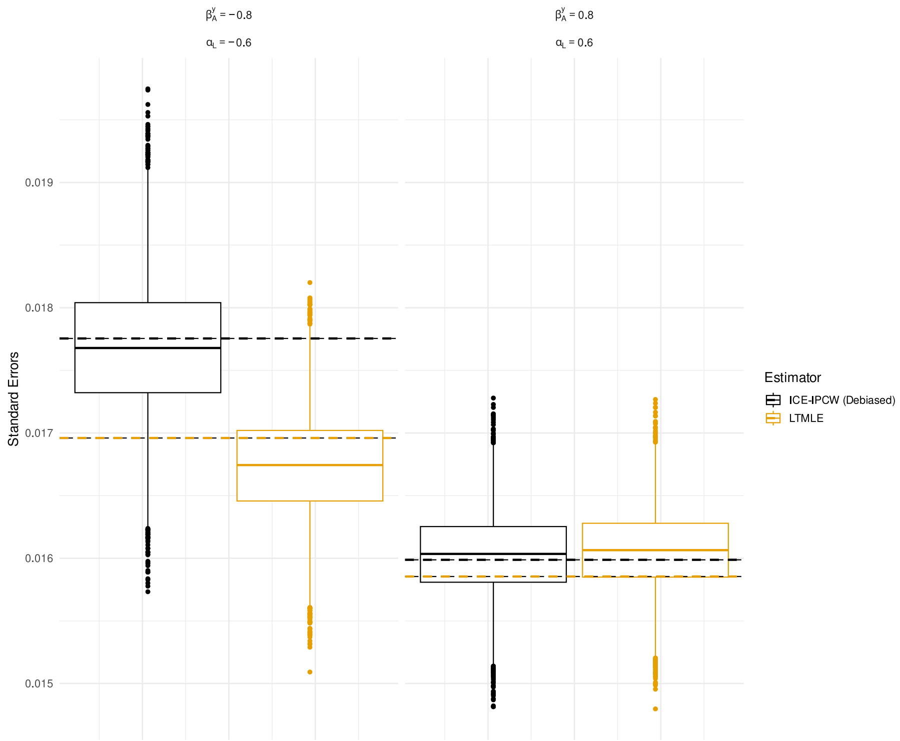

Figure 6: Boxplots of the standard errors for the case with strong time-varying confounding for each estimator (LTMLE and debiased ICE-IPCW) in each parameter setting for the cases with strong time-varying confounding. The lines indicates the empirical standard error of the estimates for each estimator.

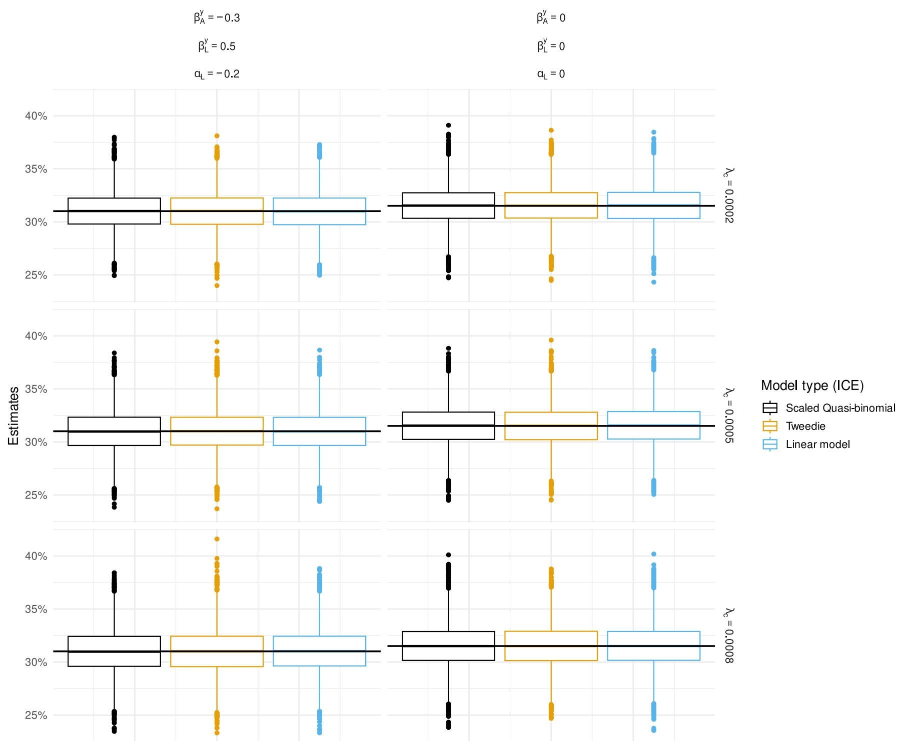

Figure 7: Boxplots of the estimates for the case with censoring for different pseudo-outcome models (linear, scaled quasibinomial, and tweedie) with varying degrees of censoring for the debiased ICE-IPCW estimator. The lines indicates the true value of the target parameter $\Psi^g_\tau(P)$.

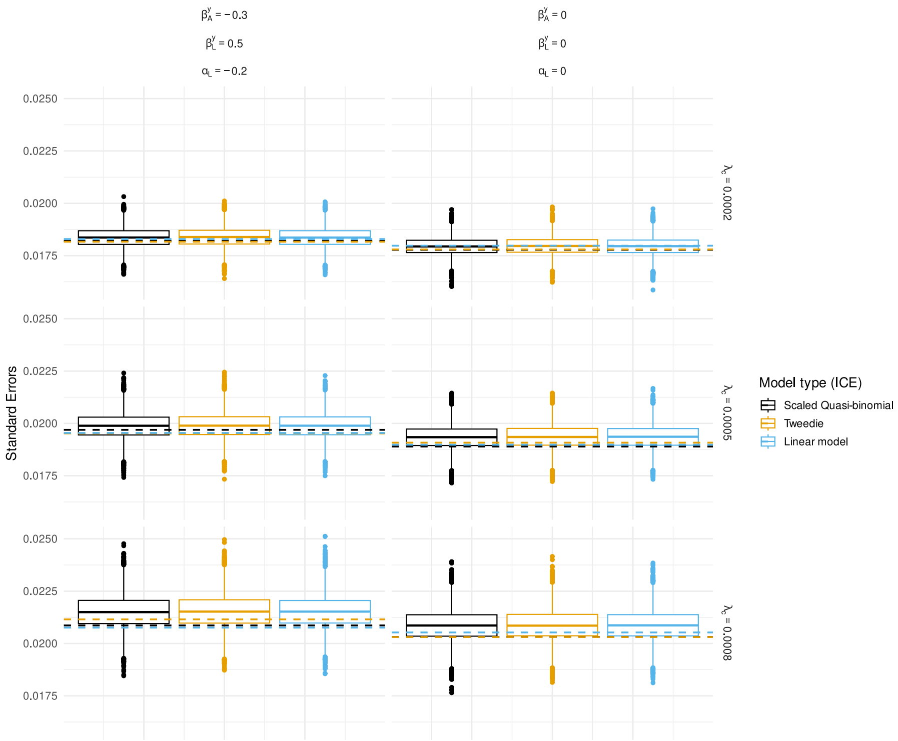

Figure 8: Boxplots of the standard errors for the case with censoring for different pseudo-outcome models (linear, scaled quasibinomial, and tweedie) with varying degrees of censoring for the debiased ICE-IPCW estimator. The lines indicates the empirical standard error of the estimates.

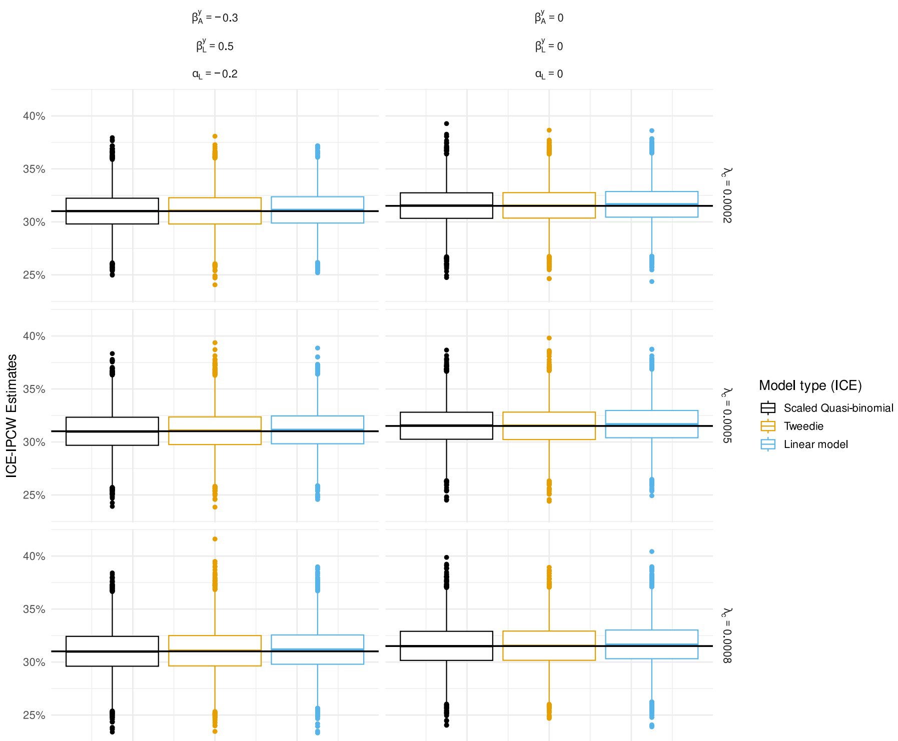

Figure 9: Boxplots of the estimates for the case with censoring for different pseudo-outcome models (linear, scaled quasibinomial, and tweedie) with varying degrees of censoring for the (not debiased) ICE-IPCW estimator. The lines indicates the true value of the target parameter $\Psi^g_\tau(P)$.

7 Application to Danish Registry Data

To illustrate the applicability of our methods, we applied them to Danish registry data emulating a target trial in diabetes research. The dataset consisted of $n = 15,937$ patients from the Danish registers who redeemed a prescription for either DPP-4 (Dipeptidyl peptidase-4) inhibitors or SGLT2 (sodium/glucose cotransporter 2) inhibitors between 2012 and 2022. The emulated target trial specifies two regimes for each patient. One is to start treatment with DPP-4 inhibitors and do not add or switch to SGLT2 inhibitors during follow-up. The other with SGLT2 inhibitors is defined analogously. We are interested on the effect of enforcing treatment and the outcome of interest is all-cause mortality within 1260 days (approximately 3.5 years). For computational reasons, we enforce treatment for the first 20 events/registrations.

At baseline (time zero), patients were required to have redeemed such a prescription and to have an HbA1c (hemoglobin A1c) measurement recorded prior to their first prescription redemption. Additionally, certain exclusion criteria were applied (Yazdanfard et al. (2025)). Within our framework, we defined:

- N^y be the counting process for the event of death.
- N^c the counting process for the event of censoring (e.g., end of study period or emigration).
- N^a the counting process counting drug purchases.
- N^ℓ the counting process for the measurement dates at which the HbA1c was measured.
- $L(t)$ denote the (latest) HbA1c measurement at time t and with the baseline HbA1c measurement at time zero (age, sex, education level, income and duration of diabetes at baseline).
- For each treatment regime (say SGLT2), we defined $A(0) = 1$ if the patient redeemed a prescription for SGLT2 inhibitors first. For $t > 0$, we defined $A(t) = 1$ if the patient has not purchased DPP-4 inhibitors prior to time t.

For the nuisance parameter estimation, we used a logistic regression model for the treatment propensity the `scaled_quasibinomial` option for the conditional counterfactual probabilities $\bar{Q}^g_{k,\tau}$. Censoring was modeled with a Cox proportional hazards model using only baseline covariates. As in the simulation study, we omitted the censoring martingale term, yielding conservative confidence intervals.

A figure of the results is provided in Figure 10. For comparison, we also applied the Cheap Subsampling confidence interval (Ohlendorff et al. (2025)) to see how robust the confidence intervals provided by our procedure are. The method was considered since bootstrapping the data is computationally expensive. With 30 bootstrap repetitions and subsample size $m = 12,750$ (approximately 80% of the data), we found that the Cheap Subsampling confidence intervals appear very similar to the ones provided by the influence function across all time horizons. For the ICE-IPCW estimator (without debiasing), we see estimates and confidence intervals are very similar to the ones of the debiased ICE-IPCW estimator.

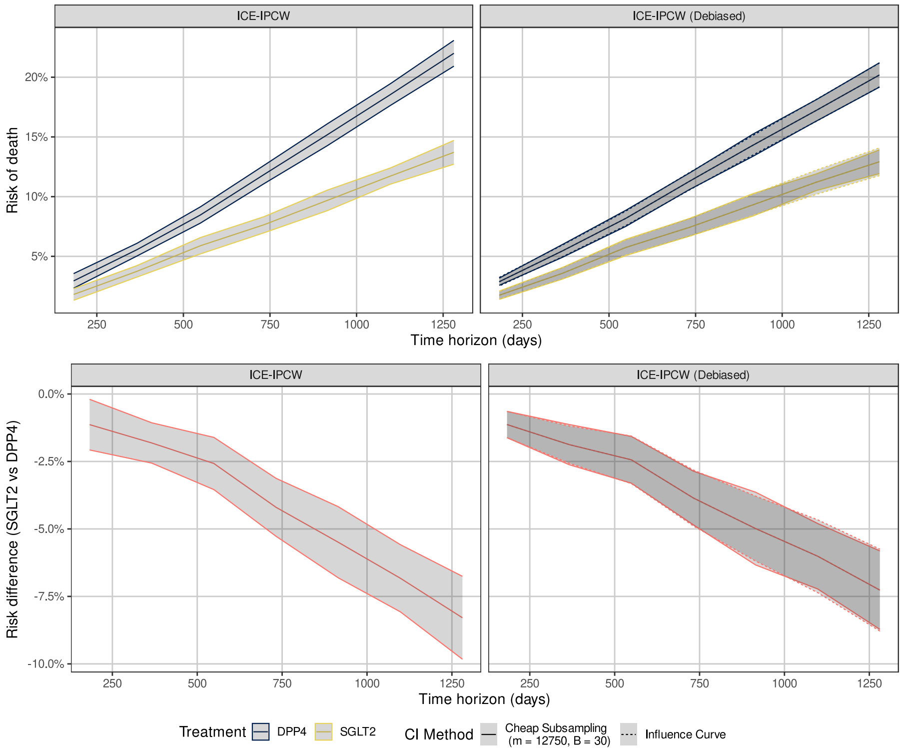

Figure 10: The causal risk of death (upper plot) and risk difference (lower plot) under sustained treatment with SGLT2 inhibitors compared to DPP4 inhibitors shown as a function of time since initiation of treatment and 95% confidence intervals based on the efficient influence function and Cheap Subsampling confidence intervals ($B = 30, m = 12,750$).

8 Discussion

In this article, we have presented a new method for causal inference in continuous-time settings with competing events and censoring. We have shown that the ICE-IPCW estimator is consistent for the target parameter, and provided inference for the target parameter using the efficient influence function. However, we have not addressed the issue of model misspecification, which is likely to occur in practice as we have not proposed flexible intensity estimators for both the censoring intensity and the propensity scores.

There are a few available options for flexible intensity estimation. For instance, neural networks (see Liguori et al. (2023) for an overview), forest based methods (Weiss & Page (2013)) and gradient boosting (Lee et al. (2021)). Other choices include flexible parametric models/highly adaptive LASSO using piecewise constant intensity models where the likelihood is based on Poisson regression (e.g., Rytgaard et al. (2022)). Another possibility is to learn the intensities sequentially tracing back through the number of events, and using standard survival methods at each event point. We consider this possibility in future work.

We also applied a simple logistic regression model for the propensity scores. However, an alternative in the absolutely continuous case is to estimate the intensities of the counting processors $N^{a,1}$ and N^a, counting the number of times that the doctor has prescribed treatment and the number of times the patient has visited the doctor, respectively, and calculating the ratio of the two (estimated) intensities.

Additionally, future work should thoroughly investigate how one should model the iterative regressions of the pseudo-outcomes since they include people observed at different times and include event times as a regression covariate. We stress, however, that any regression method may be used for this.

To obtain valid inference, we could have also opted to estimate the inverse probability weights weighted with the outcome and regressed that on $\mathcal{F}_{T_{(k)}}$. However, a key advantage of using iterative regressions is that the resulting estimator will be less sensitive to/near practical positivity violations compared to inverse probability of treatment weighting (IPW) estimators.

We could have also opted to use the TMLE framework (van der Laan & Rubin (2006)) in lieu of a one-step estimator. Here, we can use an iterative TMLE procedure for $\bar{Q}^g_{k,\tau}$ where we undersmooth the estimation of the censoring compensator to avoid estimating the censoring martingale term. This will then yield conservative but valid inference when the censoring distribution is flexibly estimated.

Another potential issue with the estimation of the nuisance parameters is the high dimensionality of the history and the variables in the history are highly correlated. This may yield issues with regression-based methods. If we adopt a TMLE approach, we may be able to use collaborative TMLE (van der Laan & Gruber (2010)) to deal with these issues.

Bibliography

Adams, R., Saria, S., & Rosenblum, M. (2020). The impact of time series length and discretization on longitudinal causal estimation methods. *Arxiv Preprint Arxiv:2011.15099*.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer US. https://doi.org/10.1007/978-1-4612-4348-9

Bang, H., & Robins, J. M. (2005). Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, *61*(4), 962–973. https://doi.org/10.1111/j.1541-0420.2005.00377.x

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., & Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models* (Vol. 4). Johns Hopkins University Press Baltimore.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Ferreira Guerra, S., Schnitzer, M. E., Forget, A., & Blais, L. (2020). Impact of discretization of the timeline for longitudinal causal inference methods. *Statistics in Medicine*, *39*(27), 4069–4085. https://doi.org/https://doi.org/10.1002/sim.8710

Gill, R. D., & Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics*, 1501–1555.

Hines, O., Dukes, O., Diaz-Ordaz, K., & Vansteelandt, S. (2022). Demystifying Statistical Learning Based on Efficient Influence Functions. *The American Statistician*, *76*(3), 292–304. https://doi.org/10.1080/00031305.2021.2021984

Hubbard, A. E., Kherad-Pajouh, S., & van der Laan, M. J. (2016). Statistical inference for data adaptive target parameters. *The International Journal of Biostatistics*, *12*(1), 3–19.

Kant, W. M., & Krijthe, J. H. (2025). Irregular measurement times in estimating time-varying treatment effects: Categorizing biases and comparing adjustment methods. *Arxiv Preprint Arxiv:2501.11449*.

Kennedy, E. H. (2024). Semiparametric Doubly Robust Targeted Double Machine Learning: A Review. In *Handbook of Statistical Methods for Precision Medicine: Handbook of Statistical Methods for Precision Medicine*. Chapman, Hall/CRC.

Last, G., & Brandt, A. (1995). *Marked Point Processes on the Real Line: The Dynamical Approach*. Springer. https://link.springer.com/book/9780387945477

Lee, D. K. K., Chen, N., & Ishwaran, H. (2021). Boosted nonparametric hazards with time-dependent covariates. *The Annals of Statistics*, *49*(4), 2101–2128. https://doi.org/10.1214/20-AOS2028

Liguori, A., Caroprese, L., Minici, M., Veloso, B., Spinnato, F., Nanni, M., Manco, G., & Gama, J. (2023, July). *Modeling Events and Interactions through Temporal Processes – A Survey* (Issue arXiv:2303.06067). arXiv. https://doi.org/10.48550/arXiv.2303.06067

Lok, J. J. (2008). Statistical modeling of causal effects in continuous time. *The Annals of Statistics*, *36*(3), 1464–1507. https://doi.org/10.1214/009053607000000820

Ohlendorff, J. S., Munch, A., Sørensen, K. K., & Gerds, T. A. (2025,). *Cheap Subsampling bootstrap confidence intervals for fast and robust inference.* https://arxiv.org/abs/2501.10289

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, *7*(9), 1393–1512. https://doi.org/https://doi.org/10.1016/0270-0255(86)90088-6

Rose, S., & van der Laan, M. J. (2011). Introduction to TMLE. In *Targeted Learning: Causal Inference for Observational and Experimental Data* (pp. 67–82). Springer New York. https://doi.org/10.1007/978-1-4419-9782-1_4

Ryalen, P. C., Stensrud, M. J., & Røysland, K. (2019). The additive hazard estimator is consistent for continuous-time marginal structural models. *Lifetime Data Analysis*, *25*, 611–638.

Rytgaard, H. C., Gerds, T. A., & van der Laan, M. J. (2022). Continuous-Time Targeted Minimum Loss-Based Estimation of Intervention-Specific Mean Outcomes. *The Annals of Stat Istics*, *50*(5), 2469–2491. https://doi.org/10.1214/21-AOS2114

Røysland, K. (2011). *A martingale approach to continuous-time marginal structural models.*

Røysland, K., C. Ryalen, P., Nygård, M., & Didelez, V. (2024). Graphical criteria for the identification of marginal causal effects in continuous-time survival and event-history analyses. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, qkae56. https://doi.org/10.1093/jrsssb/qkae056

Shirakawa, T. (2024). *Deep LTMLE+: Scalable Causal Survival Analysis with Continuous Time, Natural Visits, and Competing Risks.*

Sofrygin, O., Zhu, Z., Schmittdiel, J. A., Adams, A. S., Grant, R. W., van der Laan, M. J., & Neugebauer, R. (2019). Targeted learning with daily EHR data. *Statistics in Medicine*, *38*(16), 3073–3090.

Sun, J., & Crawford, F. W. (2023). The role of discretization scales in causal inference with continuous-time treatment. *Arxiv Preprint Arxiv:2306.08840.*

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. http://www.jstor.org/stable/2346178

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

van der Laan, M. J., & Gruber, S. (2010). Collaborative double robust targeted maximum likelihood estimation. *The International Journal of Biostatistics*, *6*(1).

van der Laan, M. J., & Gruber, S. (2012). *The International Journal of Biostatistics*, *8*(1). https://doi.org/doi:10.1515/1557-4679.1370

van der Laan, M. J., & Rubin, D. (2006). Targeted Maximum Likelihood Learning. *The International Journal of Biostatistics*, *2*(1). https://doi.org/10.2202/1557-4679.1043

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

Weiss, J. C., & Page, D. (2013). Forest-Based Point Process for Event Prediction from Electronic Health Records. In H. Blockeel, K. Kersting, S. Nijssen, & F. Železný (Eds.), *Machine Learning and Knowledge Discovery in Databases: Machine Learning and Knowledge Discovery in Databases.* https://doi.org/10.1007/978-3-642-40994-3_35

Yazdanfard, P. D. W., Sørensen, K. K., Zareini, B., Pedersen-Bjergaard, U., Ohlendorff, J. S., Munch, A., Andersen, M. P., Hasselbalch, R. B., Imberg, H., Tasseleus, V., Lind, M., Valabhji, J., Choudhary, P., Khunti, K., Schmid, S., Lanzinger, S., Mader, J., Gerds, T. A., Torp-Pedersen, C., & The REDDIE consortium. (2025). Type 2 diabetes, sodium-glucose cotransporter-2 inhibitors and cardiovascular outcomes: real world evidence versus a randomised clinical trial. *Cardiovascular Diabetology*, *24*(1), 371. https://doi.org/10.1186/s12933-025-02924-0

APPENDIX A

A.1 Proof of Theorem 1

Let $W_{k,j} = \prod_{v=k+1}^{j} \left(\frac{\mathbb{1}\left\{A\left(T_{(v)}\right)=1\right\}}{\pi_v\left(T_{(v)}, L\left(T_{(v)}\right), \mathcal{F}_{T_{(v-1)}}\right)} \right)^{\mathbb{1}\left\{\Delta_{(v)}=a\right\}}$ denote the treatment weights for $k < j$ (taking $\pi_0\left(T_0, L(0), \mathcal{F}_{T_{(0-1)}}\right) := \pi_0(L(0))$). We show that

$$\bar{Q}^g_{k,\tau}\left(\tau, A\left(T_{(k)}\right), H_k\right) = \mathbb{E}_P\left[\sum_{j=k+1}^{K} W_{k,j} \mathbb{1}\left\{T_{(j)} \leq \tau, \Delta_{(j)} = y\right\} \mid A\left(T_{(k)}\right), H_k\right] \tag{A.1}$$

for $k = 0, ..., K-1$ by backwards induction:

Base case: We consider the case $k = K-1$. First note that

$$Z^a_{K,\tau}(u) = \mathbb{1}\left\{T_{(K)} \leq u, \Delta_{(K)} = y\right\},$$

and

$$\begin{aligned} W_{K-1,K} \mathbb{1}\left\{T_{(K)} \leq \tau, \Delta_{(K)} = y\right\} &= \left(\frac{\mathbb{1}\left\{A\left(T_{(K)}\right) = 1\right\}}{\pi_K\left(T_{(K)}, L\left(T_{(K)}\right), \mathcal{F}_{T_{(K-1)}}\right)} \right)^{\mathbb{1}\left\{\Delta_{(K)}=a\right\}} \mathbb{1}\left\{T_{(K)} \leq \tau, \Delta_{(K)} = y\right\} \\ &= \mathbb{1}\left\{T_{(K)} \leq \tau, \Delta_{(K)} = y\right\} \end{aligned}$$

so we have

$$\begin{aligned} &\mathbb{E}_P\left[W_{K-1,K} \mathbb{1}\left\{T_{(K)} \leq \tau, \Delta_{(j)} = y\right\} \mid A\left(T_{(K-1)}\right), H_{K-1}\right] \\ &= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(K)} \leq \tau, \Delta_{(j)} = y\right\} \mid A\left(T_{(K-1)}\right), H_{K-1}\right] \\ &= \mathbb{E}_P\left[Z^a_{K,\tau}(\tau) \mid A\left(T_{(K-1)}\right), H_{K-1}\right], \\ &= \bar{Q}^g_{K-1,\tau}\left(\tau, A\left(T_{(K-1)}\right), H_{K-1}\right). \end{aligned}$$

Inductive step: Assume that the claim holds for $k+1$. Now, we first note that

$$
\begin{aligned}
&\mathbb{1}\Big\{T_{(k+1)} < \tau, \Delta_{(k+1)} \in \{a, \ell\}\Big\} \\
&\quad \times \mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\Big\{T_{(j)} \leq \tau, \Delta_{(j)} = y\Big\} \mid A\Big(T_{(k+1)}\Big), H_{k+1}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, A\Big(T_{(k)}\Big), H_k\right] \\
&= \mathbb{1}\Big\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\Big\} \\
&\qquad \times \mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\Big\{T_{(j)} \leq \tau, \Delta_{(j)} = y\Big\} \mid A\Big(T_{(k+1)}\Big), H_{k+1}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, A\Big(T_{(k)}\Big), H_k\right] \\
&\quad + \mathbb{1}\Big\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\Big\} \\
&\qquad \times \mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\Big\{T_{(j)} \leq \tau, \Delta_{(j)} = y\Big\} \mid A\Big(T_{(k+1)}\Big), H_{k+1}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, A\Big(T_{(k)}\Big), H_k\right] \\
&= \mathbb{1}\Big\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\Big\} \\
&\qquad \times \mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\Big\{T_{(j)} \leq \tau, \Delta_{(j)} = y\Big\} \mid A\Big(T_{(k+1)}\Big), H_{k+1}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, A\Big(T_{(k)}\Big), H_k\right] \\
&\quad + \mathbb{1}\Big\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\Big\} \\
&\qquad \times \mathbb{E}_P\left[\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\Big\{T_{(j)} \leq \tau, \Delta_{(j)} = y\Big\} \mid A\Big(T_{(k+1)}\Big), H_{k+1}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, A\Big(T_{(k)}\Big), H_k\right] \\
&\overset{(a)}{=} \mathbb{1}\Big\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\Big\} \\
&\qquad \times \mathbb{E}_P\Big[W_{k,k+1}\bar{Q}^g_{k+1,\tau}\Big(\tau, A\Big(T_{(k+1)}\Big), H_{k+1}\Big) \mid T_{(k+1)}, \Delta_{(k+1)}, A\Big(T_{(k)}\Big), H_k\Big] \\
&\quad + \mathbb{1}\Big\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\Big\} \\
&\qquad \times \mathbb{E}_P\Big[\bar{Q}^g_{k+1,\tau}\Big(\tau, A\Big(T_{(k)}\Big), H_{k+1}\Big) \mid T_{(k+1)}, \Delta_{(k+1)}, A\Big(T_{(k)}\Big), H_k\Big]
\end{aligned}
\tag{A.2}
$$

In (a), we use the induction hypothesis. Using Equation (A.2) then gives,

$$
\begin{aligned}
&\mathbb{E}_P\left[\sum_{j=k+1}^{K} W_{k,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid A\left(T_{(k)}\right), H_k\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} W_{k,k+1} \mid A\left(T_{(k)}\right), H_k\right] \\
&\quad + \mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid A\left(T_{(k)}\right), H_k\right] \\
&\overset{(b)}{=} \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} W_{k,k+1} \mid A\left(T_{(k)}\right), H_k\right] \\
&\quad + \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} \in \{a, \ell\}\right\} \sum_{j=k+2}^{K} W_{k,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid A\left(T_{(k)}\right), H_k\right] \\
&= \mathbb{E}_P\Bigg(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} W_{k,k+1} \\
&\quad + \mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} \in \{a, \ell\}\right\} \\
&\quad \times \mathbb{E}_P\Bigg[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid A\left(T_{(k+1)}\right), H_{k+1}\right] \\
&\qquad \mid T_{(k+1)}, \Delta_{(k+1)}, A\left(T_{(k)}\right), H_k\Bigg] \mid A\left(T_{(k)}\right), H_k\Bigg)
\end{aligned}
\tag{A.3}
$$

$$
\begin{aligned}
&\overset{\text{Equation (A.2)}}{=} \mathbb{E}_P\Big(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} \\
&\quad + \mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\right\} \\
&\qquad \times \mathbb{E}_P\left[W_{k,k+1}\bar{Q}^g_{k+1,\tau}\left(\tau, A\left(T_{(k+1)}\right), H_{k+1}\right) \mid T_{(k+1)}, \Delta_{(k+1)}, A\left(T_{(k)}\right), H_k\right] \\
&\quad + \mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\right\} \\
&\qquad \times \mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}\left(\tau, A\left(T_{(k)}\right), H_{k+1}\right) \mid T_{(k+1)}, \Delta_{(k+1)}, A\left(T_{(k)}\right), H_k\right] \mid A\left(T_{(k)}\right), H_k\Big) \\
&= \mathbb{E}_P\Big(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} \\
&\quad + \mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\right\} \\
&\qquad \times \mathbb{E}_P\left[W_{k,k+1}\bar{Q}^g_{k+1,\tau}\left(\tau, A\left(T_{(k+1)}\right), H_{k+1}\right) \mid L\left(T_{(k+1)}\right), T_{(k+1)}, \Delta_{(k+1)}, A\left(T_{(k)}\right), H_k\right] \\
&\quad + \mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\right\} \\
&\qquad \times \bar{Q}^g_{k+1,\tau}\left(\tau, A\left(T_{(k)}\right), H_{k+1}\right) \mid A\left(T_{(k)}\right), H_k\Big).
\end{aligned}
$$

Throughout, we use the law of iterated expectations. In (b), we use that

$$\left(T_{(k)} \le \tau, \Delta_{(k)} = y\right) \subseteq \left(T_{(j)} < \tau, \Delta_{(j)} \in \{a, \ell\}\right)$$

for all $j = 1, ..., k-1$ and $k = 1, ..., K$ as otherwise the corresponding term would be zero.

The desired statement (Equation (A.1)) now follows from the fact that

$$
\begin{aligned}
&\mathbb{E}_P\left[W_{k,k+1}\bar{Q}^g_{k+1,\tau}\left(\tau, A\left(T_{(k+1)}\right), H_{k+1}\right) \mid L\left(T_{(k+1)}\right), T_{(k+1)}, \Delta_{(k+1)} = a, A\left(T_{(k)}\right), H_k\right] \\
&= \mathbb{E}_P\left[\frac{\mathbb{1}\left\{A\left(T_{(k+1)}\right) = 1\right\}}{\pi_{k+1}\left(T_{(k+1)}, L\left(T_{(k+1)}\right), A\left(T_{(k)}\right), H_k\right)}\bar{Q}^g_{k+1,\tau}(\tau, 1, H_{k+1}) \mid L\left(T_{(k+1)}\right), T_{(k+1)}, \Delta_{(k+1)} = a, A\left(T_{(k)}\right), H_k\right] \\
&= \mathbb{E}_P\Bigg[\frac{\mathbb{E}_P\left[\mathbb{1}\left\{A\left(T_{(k+1)}\right) = 1\right\} \mid T_{(k+1)}, L\left(T_{(k+1)}\right), \Delta_{(k+1)} = a, A\left(T_{(k)}\right), H_k\right]}{\pi_{k+1}\left(T_{(k+1)}, L\left(T_{(k+1)}\right), A\left(T_{(k)}\right), H_k\right)} \\
&\qquad \times \bar{Q}^g_{k+1,\tau}(\tau, 1, H_{k+1}) \mid L\left(T_{(k+1)}\right), T_{(k+1)}, \Delta_{(k+1)} = a, A\left(T_{(k)}\right), H_k\Bigg] \\
&= \mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}(\tau, 1, H_{k+1}) \mid L\left(T_{(k+1)}\right), T_{(k+1)}, \Delta_{(k+1)} = a, A\left(T_{(k)}\right), H_k\right] \\
&= \bar{Q}^g_{k+1,\tau}(\tau, 1, H_{k+1}),
\end{aligned}
\tag{A.4}
$$

and Equation (A.3).

A similar calculation to Equation (A.4) shows that $\Psi_\tau^g(P) = \mathbb{E}_P[\bar{Q}_{0,\tau}^g(\tau, 1, L(0))]$ and so Equation (3.3) follows.

APPENDIX B

B.1 Proof of Theorem 2

We show that if $\bar{Q}_{k+1,\tau}^g$ is identified, then $\bar{Q}_{k,\tau}^g$ is identifiable via Equation (4.1),

$$
\begin{aligned}
\mathbb{E}_P\left[\bar{Z}_{k+1,\tau}^a(u) \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right] &= \mathbb{1}\left\{\bar{\Delta}_{(1)} \in \{a,\ell\}, \dots, \bar{\Delta}_{(k)} \in \{a,\ell\}\right\} \mathbb{E}_P\left[\bar{Z}_{k+1,\tau}^a(u) \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right] + 0 \\
&\overset{\text{Lemma 1}}{=} \mathbb{1}\left\{\bar{\Delta}_{(1)} \in \{a,\ell\}, \dots, \bar{\Delta}_{(k)} \in \{a,\ell\}\right\} \\
&\quad \times \Bigg\{ \int_{\bar{T}_{(k)}}^u \frac{\tilde{S}\left(s- \mid \mathcal{F}_{T_{(k)}}\right)}{\tilde{S}^c\left(s- \mid \mathcal{F}_{T_{(k)}}\right)} \mathbb{E}_P\left[\bar{Q}_{k+1,\tau}^g(u, 1, \bar{H}_{k+1}) \mid T_{(k)} = s, \Delta_{(k)} = a, \mathcal{F}_{T_{(k)}}\right] \Lambda_{k+1}^a\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad + \int_{\bar{T}_{(k)}}^u \frac{\tilde{S}\left(s- \mid \mathcal{F}_{T_{(k)}}\right)}{\tilde{S}^c\left(s- \mid \mathcal{F}_{T_{(k)}}\right)} \mathbb{E}_P\left[\bar{Q}_{k+1,\tau}^g(u) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k)}}\right] \Lambda_{k+1}^\ell\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad + \int_{\bar{T}_{(k)}}^u \frac{\tilde{S}\left(s- \mid \mathcal{F}_{T_{(k)}}\right)}{\tilde{S}^c\left(s- \mid \mathcal{F}_{T_{(k)}}\right)} \Lambda_{k+1}^y\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \Bigg\} \\
&\overset{\text{Lemma 2}}{=} \mathbb{1}\left\{\bar{\Delta}_{(1)} \in \{a,\ell\}, \dots, \bar{\Delta}_{(k)} \in \{a,\ell\}\right\} \\
&\quad \times \Bigg\{ \int_{\bar{T}_{(k)}}^u S\left(s- \mid \mathcal{F}_{T_{(k)}}\right) \mathbb{E}_P\left[\bar{Q}_{k+1,\tau}^g(u, 1, \bar{H}_{k+1}) \mid T_{(k)} = s, \Delta_{(k)} = a, \mathcal{F}_{T_{(k)}}\right] \Lambda_{k+1}^a\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad + \int_{\bar{T}_{(k)}}^u S\left(s- \mid \mathcal{F}_{T_{(k)}}\right) \mathbb{E}_P\left[\bar{Q}_{k+1,\tau}^g(u) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k)}}\right] \Lambda_{k+1}^\ell\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad + \int_{\bar{T}_{(k)}}^u S\left(s- \mid \mathcal{F}_{T_{(k)}}\right) \Lambda_{k+1}^y\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \Bigg\} \\
&\overset{(*)}{=} \mathbb{1}\left\{\bar{\Delta}_{(1)} \in \{a,\ell\}, \dots, \bar{\Delta}_{(k)} \in \{a,\ell\}\right\} \bar{Q}_{k,\tau}^g\left(u, \mathcal{F}_{T_{(k)}}\right).
\end{aligned}
$$

To obtain $(*)$, we use the definition of $\bar{Q}_{k,\tau}^g$ in Equation (3.3), writing out the conditional expectation with respect to the densities/cause-specific cumulative hazards. This completes the proof.

Lemma 1: Assume condition 1. of Theorem 2. $\mu_k^\ell\left(t, \cdot, \mathcal{F}_{T_{(k-1)}}\right)$ be the probability measure for $L\left(T_{(k)}\right)$ given $\Delta_{(k)} = \ell, T_{(k)} = t$, and $\mathcal{F}_{T_{(k-1)}}$, and $\mu_k^a\left(t, \cdot, \mathcal{F}_{T_{(k-1)}}\right)$ be the probability measure for $L\left(T_{(k)}\right)$ given $\Delta_{(k)} = a, T_{(k)} = t$, and $\mathcal{F}_{T_{(k-1)}}$. Then, we have

$$\begin{aligned}
&\mathbb{1}\left\{\bar{\Delta}_{(k-1)} \in \{a, \ell\}\right\} P\left(\left(\bar{T}_{(k)}, \bar{\Delta}_k, A\left(\bar{T}_{(k)}\right), L\left(\bar{T}_{(k)}\right)\right) \in \mathrm{d}(t, m, a, l) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \mathbb{1}\left\{\bar{T}_{(k-1)} < t, \bar{\Delta}_{(k-1)} \in \{a, \ell\}\right\} \Bigg(\tilde{S}\left(t- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \sum_{x=a,\ell,d,y} \delta_x(\mathrm{d}m) \psi_{k,x}\left(\mathcal{F}_{T_{(k-1)}}, t, \mathrm{d}(a,l)\right) \Lambda_k^x\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right) \\
&\qquad + \delta_{(c, A(C), L(C))}(\mathrm{d}(m, a, l)) \tilde{\Lambda}_k^c\left(\mathrm{d}t, \mathcal{F}_{\bar{T}_{(k-1)}}\right)\Bigg),
\end{aligned} \tag{B.1}$$

where

$$\begin{aligned}
\psi_{k,x}\left(\mathcal{F}_{T_{(k-1)}}, t, \mathrm{d}(a,l)\right) &= \mathbb{1}\{x = a\}\Big(\delta_1(\mathrm{d}a)\pi_k\left(t, l, \mathcal{F}_{T_{(k-1)}}\right) \\
&\qquad + \delta_0(\mathrm{d}a)\left(1 - \pi_k\left(t, l, \mathcal{F}_{T_{(k-1)}}\right)\right)\mu_k^a\left(\mathrm{d}l, t, \mathcal{F}_{T_{(k-1)}}\right)\Big) \\
&\quad + \mathbb{1}\{x = \ell\}\mu_k^\ell\left(\mathrm{d}l, t, \mathcal{F}_{T_{(k-1)}}\right)\delta_{A\left(T_{(k-1)}\right)}(\mathrm{d}a) \\
&\quad + \mathbb{1}\{x \in \{y, d\}\}\delta_{A(T^e)}(\mathrm{d}a)\delta_{L(T^e)}(\mathrm{d}l),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{1}\left\{\bar{\Delta}_{(k-1)} \neq c, \bar{T}_{(k-1)} < t\right\}\tilde{S}\left(t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \neq c, \bar{T}_{(k-1)} < t\right\} \prod_{s \in \left(T_{(k-1)}, t\right]} \left(1 - \left(\sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right) + \tilde{\Lambda}_k^c\left(\mathrm{d}s, \mathcal{F}_{\bar{T}_{(k-1)}}\right)\right)\right).
\end{aligned}$$

Proof: Let $\Phi = \left(T_{(n)}, \Delta_{(n)}, A\left(T_{(n)}\right), L\left(T_{(n)}\right)\right)_n$ denote the marked point process associated with N. A version of the compensator of the random measure N with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ is by Theorem 4.1.11 (ii) of Last & Brandt (1995) is

$$\begin{aligned}
\Lambda(\mathrm{d}(t, m, a, l)) &= \sum_k \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \frac{P\left(\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right)\right) \in \mathrm{d}(t, m, a, l) \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)} \\
&= \sum_k \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} P\left(\left(L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right) \in \mathrm{d}(l, a) \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t, \Delta_{(k)} = m\right) \\
&\quad \times \frac{P\left(\left(T_{(k)}, \Delta_{(k)}\right) \in \mathrm{d}(t, m)\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)} \\
&= \sum_k \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \sum_{x=a,\ell,y,d} P\left(\left(L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right) \in \mathrm{d}(l, a) \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t, \Delta_{(k)} = x\right) \\
&\quad \times \delta_x(\mathrm{d}m)\Lambda_k^x\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right) \\
&= \sum_k \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \sum_{x=a,\ell,y,d} \psi_{k,x}\left(\mathcal{F}_{T_{(k-1)}}, t, \mathrm{d}(a,l)\right)\delta_x(\mathrm{d}m)\Lambda_k^x\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right) \\
&\overset{*}{=} \sum_k \mathbb{1}\{v_{k-1}(\Phi) < t \leq v_{k-1}(\Phi)\} \sum_{x=a,\ell,y,d} \psi_{k,x}(h_{k-1}(\Phi), t, \mathrm{d}(a,l))\delta_x(\mathrm{d}m)\Lambda_k^x(\mathrm{d}t, h_{k-1}(\Phi)),
\end{aligned} \tag{B.2}$$

where we use explicit conditioning and the definition of the cause-specific hazard measures. In $(*)$, we use that $T_{(k)} = v_k(\Phi)$ for some measurable function v_k and that $\mathcal{F}_{T_{(k)}} = h_k(\Phi)$ for some measurable function h_k.

We let $\widetilde{N}(\mathrm{d}(t, m, a, l))$ denote the observed random measure, i.e., the random measure corresponding to the observed data and $\tilde{\Phi}$ its corresponding marked point process. Let $T^*_{(k)}, \Delta^*_{(k)}, A\left(T^*_{(k)}\right), L\left(T^*_{(k)}\right)$ denote the event times and marks of the random measure N^{full}.

To complete the proof, we first need some additional definitions. Let Φ^* be a marked point process with baseline element Y and let U be a given P-$\sigma(Y, \Phi^*_t)$ stopping time. For $n \in \mathbb{N}$, we let

$$(T^*_{U,n}, X^*_{U,n}) = \begin{cases} (T^*_{n+k}, X^*_{n+k}) \text{ if } \Phi^*(U) = k < \infty \\ (\infty, \emptyset) \text{ otherwise} \end{cases}.$$

Note that $T^*_{U,n}$ denotes the n'th event time of Φ^* after U. Also recall that the restriction of Φ^* to S denoted by Φ^*_S is given by

$$(T(S)^*_k, X(S)^*_k) = \begin{cases} (T^*_n, X^*_n) \text{ if } T^*_n \leq S \\ (\infty, \emptyset) \text{ otherwise} \end{cases}$$

Note that $T^*_{(k+1)} = \bar{T}_{(k+1)}$ whenever $T_{(k)} < C$. Let $T_{S,1}$ denote the first event time of N^{full} after $T_{S,0} := S$, where S is a stopping time with respect to the filtration $\mathcal{F}^{\text{full}}$. With $S := T^e \wedge C \wedge T_{(k)}$, we have $T_{S,0} := S = T^e \wedge C \wedge T_{(k)}$.

Importantly, we have that

$$\mathbb{1}\left\{\bar{\Delta}_{(1)} \notin \{c, y, d\}, ..., \bar{\Delta}_{(k)} \notin \{c, y, d\}\right\} \Phi^{\text{full}}_{T_{S,0}} = \mathbb{1}\left\{\bar{\Delta}_{(1)} \notin \{c, y, d\}, ..., \bar{\Delta}_{(k)} \notin \{c, y, d\}\right\} \tilde{\Phi}_{\bar{T}_{(k)}} \quad P- \text{.a.s} \tag{B.3}$$

Using Theorem 4.3.8 of Last & Brandt (1995), it therefore holds that

$$\begin{aligned}
&\mathbb{1}\left\{\bar{\Delta}_{(1)} \notin \{c, y, d\}, ..., \bar{\Delta}_{(k)} \notin \{c, y, d\}\right\} \\
&\quad \times P\left(\left(\bar{T}_{k+1}, \bar{\Delta}_{k+1}, A(\bar{T}_{k+1}), L(\bar{T}_{k+1})\right) \in \mathrm{d}(t, m, a, l) \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right) \\
&= \mathbb{1}\left\{\bar{\Delta}_{(1)} \notin \{c, y, d\}, ..., \bar{\Delta}_{(k)} \notin \{c, y, d\}\right\} \\
&\quad \times P\left(\left(T^*_{S,1}, \Delta^*_{S,1}, A(T^*_{S,1}), L(T^*_{S,1})\right) \in \mathrm{d}(t, m, a, l) \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right) \\
&\overset{(*)}{=} \mathbb{1}\left\{\bar{\Delta}_{(1)} \notin \{c, y, d\}, ..., \bar{\Delta}_{(k)} \notin \{c, y, d\}\right\} \\
&\quad \times P\left(\left(T^*_{S,1}, \Delta^*_{S,1}, A(T^*_{S,1}), L(T^*_{S,1})\right) \in \mathrm{d}(t, m, a, l) \mid \mathcal{F}^{\text{full}}_{T_{S,0}}\right) \\
&\overset{\text{Thm. 4.3.8}}{=} \mathbb{1}\left\{\bar{\Delta}_{(1)} \notin \{c, y, d\}, ..., \bar{\Delta}_{(k)} \notin \{c, y, d\}\right\} \mathbb{1}\{T_{S,0} < t\} \\
&\quad \times \prod_{s \in (T_{S,0}, t)} \left(1 - \rho\left((L(0), A(0)), \Phi^{\text{full}}_{T_{S,0}}, \mathrm{d}s, \{a, y, l, d, y\} \times \{0, 1\} \times \mathcal{L}\right)\right) \\
&\quad \times \rho\left((L(0), A(0)), \Phi^{\text{full}}_{T_{S,0}}, \mathrm{d}(t, m, a, l)\right) \\
&= \mathbb{1}\left\{\bar{\Delta}_{(1)} \notin \{c, y, d\}, ..., \bar{\Delta}_{(k)} \notin \{c, y, d\}\right\} \mathbb{1}\left\{\bar{T}_{(k)} < t\right\} \\
&\quad \times \prod_{s \in (\bar{T}_{(k)}, t)} \left(1 - \rho\left((L(0), A(0)), \tilde{\Phi}_{\bar{T}_{(k)}}, \mathrm{d}s, \{a, y, l, d, y\} \times \{0, 1\} \times \mathcal{L}\right)\right) \\
&\quad \times \rho\left(A(0), L(0), \tilde{\Phi}_{\bar{T}_{(k)}}, \mathrm{d}(t, m, a, l)\right).
\end{aligned} \tag{B.4}$$

In $(*)$, we use that $\mathcal{F}^{\text{full}}_{T_{S,0}} = \sigma\left((A(0), L(0)), \Phi^{\text{full}}_{T_{S,0}}\right)$, but also $\mathcal{F}_{\bar{T}_{(k)}} = \sigma\left((A(0), L(0)), \tilde{\Phi}_{\bar{T}_{(k)}}\right)$ by Theorem 2.1.14 of Last & Brandt (1995). Hence, the conditional expectations are almost surely the same by Equation (B.3). From Equation (B.4), we get Equation (B.1) by noting that $\psi_{k,x}(t, \{a, y, l, d, y\} \times \{0, 1\} \times \mathcal{L}) = 1$. □

Lemma 2: Assume condition 1. and 2. of Theorem 2. Then the left limit of the survival function factorizes on $(0, \tau]$, i.e.,

$$\begin{aligned}
&\mathbb{1}\left\{\bar{\Delta}_{(k-1)} \neq c, \bar{T}_{(k-1)} < t\right\} \tilde{S}\left(t - \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \neq c, \bar{T}_{(k-1)} < t\right\} \prod_{s \in (0,t)} \left(1 - \sum_{x = a, \ell, y, d} \Lambda^x_k\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right) \prod_{s \in (0,t)} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)
\end{aligned}$$

for $x = a, \ell, y, d$.

Proof: To keep notation for the proof brief, let $\gamma(v) = \Delta\tilde{\Lambda}_k^c\left(v \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)$ and $\zeta(v) = \sum_x \Delta\Lambda_k^x\left(v, \mathcal{F}_{T_{(k-1)}}\right)$ and $s = \bar{T}_{k-1}$.

Recall that $\tilde{S}\left(t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) = \prod_{v\in(s,t]}\left(1 - \left(\sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}v, \mathcal{F}_{T_{(k-1)}}\right) + \tilde{\Lambda}_k^c\left(\mathrm{d}v, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)\right)$.

$$
\begin{aligned}
\prod_{v\in(s,t)} (1 - \mathrm{d}(\zeta+\gamma)(v)) &= \exp(-\beta^c)\exp(-\gamma^c) \prod_{\substack{v\in(s,t) \\ \gamma(v)\neq\gamma(v-)\vee\zeta(v)\neq\zeta(v-)}} (1 - \Delta(\zeta+\gamma)) \\
&\overset{(*)}{=} \exp(-\beta^c)\exp(-\gamma^c) \prod_{\substack{v\in(s,t) \\ \gamma(v)\neq\gamma(v-)}} (1-\Delta\gamma) \prod_{\substack{v\in(s,t) \\ \zeta(v)\neq\zeta(v-)}} (1-\Delta\zeta) \\
&= \prod_{v\in(s,t)} (1-\mathrm{d}\zeta(v)) \prod_{v\in(s,t)} (1-\mathrm{d}\gamma(v)),
\end{aligned}
$$

where we apply condition 2. of Theorem 2 in $(*)$. □

APPENDIX C

C.1 Comparison with the EIF in Rytgaard et al. (2022)

Let us define in the censored setting

$$
W^g(t) = \prod_{k=1}^{\tilde{N}_t} \frac{\mathbb{1}\{A(\bar{T}_k) = 1\}}{\pi_k\left(\bar{T}_{(k)}, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)} \prod_{k=1}^{\tilde{N}_t} \frac{\mathbb{1}\{\Delta_{(k)} \neq c\}}{\prod_{u\in\left(\bar{T}_{(k-1)},\bar{T}_{(k)}\right)}\left(1 - \tilde{\Lambda}_k^c\left(\mathrm{d}u, \mathcal{F}_{\bar{T}_{(k-1)}}\right)\right)},
$$

in alignment with Equation (3.1) with perfect compliance at time zero. We verify that our efficient influence function is the same as Rytgaard et al. (2022), but first we need to make sure that the setting is the same as theirs. Rytgaard et al. (2022) assumes orthogonal martingales and that $L\left(T_{(k)}\right) = L\left(T_{(k-1)}\right)$ whenever $\Delta_{(k)} = a$ (in our manuscript, we allowed $L\left(T_{(k)}\right)$ to change). The efficient influence function of Rytgaard et al. (2022) is given in Theorem 1 of Rytgaard et al. (2022) in our notation by

$$
\begin{aligned}
\varphi_\tau^*(P) &= \mathbb{E}_{PG^*}\left[\tilde{N}^y(\tau) \mid \mathcal{F}_0\right] - \Psi_\tau(P) \\
&+ \int_0^\tau W^g(t-)\left(\mathbb{E}_{PG^*}\left[\tilde{N}^y(\tau) \mid \tilde{L}(t), \tilde{N}^\ell(t), \bar{\mathcal{F}}_{t-}\right] - \mathbb{E}_{PG^*}\left[\tilde{N}_y(\tau) \mid \tilde{N}^\ell(t), \bar{\mathcal{F}}_{t-}\right]\right)\tilde{N}^\ell(\mathrm{d}t) \\
&+ \int_0^\tau W^g(t-)\left(\mathbb{E}_{PG^*}\left[\tilde{N}^y(\tau) \mid \Delta\tilde{N}^\ell(t) = 1, \bar{\mathcal{F}}_{t-}\right] - \mathbb{E}_{PG^*}\left[\tilde{N}_y(\tau) \mid \Delta\tilde{N}^\ell(t) = 0, \bar{\mathcal{F}}_{t-}\right]\right)\tilde{M}^\ell(\mathrm{d}t) \\
&+ \int_0^\tau W^g(t-)\left(\mathbb{E}_{PG^*}\left[\tilde{N}^y(\tau) \mid \Delta\tilde{N}^a(t) = 1, \bar{\mathcal{F}}_{t-}\right] - \mathbb{E}_{PG^*}\left[\tilde{N}_y(\tau) \mid \Delta\tilde{N}^a(t) = 0, \bar{\mathcal{F}}_{t-}\right]\right)\tilde{M}^a(\mathrm{d}t) \\
&+ \int_0^\tau W^g(t-)\left(1 - \mathbb{E}_{PG^*}\left[\tilde{N}^y(\tau) \mid \Delta\tilde{N}^y(t) = 0, \bar{\mathcal{F}}_{t-}\right]\right)\tilde{M}^y(\mathrm{d}t) \\
&+ \int_0^\tau W^g(t-)\left(0 - \mathbb{E}_{PG^*}\left[\tilde{N}^y(\tau) \mid \Delta\tilde{N}^d(t) = 0, \bar{\mathcal{F}}_{t-}\right]\right)\tilde{M}^d(\mathrm{d}t).
\end{aligned}
\tag{C.1}
$$

In Equation (C.1), we need to define the sample paths of the process $\mathbb{E}_{PG^*}\left[\tilde{N}_y(\tau) \mid \Delta\tilde{N}^x(t) = 1, \bar{\mathcal{F}}_{t-}\right]$. This is a non-trivial matter in continuous time since $\Delta\tilde{N}^x(t) = 1$ is a null-set in continuous time. We verify that we can do this in discrete time and provide a simple presentation of it. On the other hand,Corollary 2.2.3 of Last & Brandt (1995) shows that $\mathbb{E}_{PG^*}\left[\tilde{N}_y(\tau) \mid \Delta\tilde{N}^x(t) = 0, \bar{\mathcal{F}}_{t-}\right]$ can be defined as a P-$\bar{\mathcal{F}}_t$-predictable process, and the term involving the counting process integral obviously only needs to be defined at the event times. We now provide these details.

Here, $\tilde{M}^x$ denotes the observed martingales with respect to the observed filtration. We find $\mathbb{E}_{PG^*}\left[\tilde{N}_y(\tau) \mid \Delta\tilde{N}^x(t) = 1, \bar{\mathcal{F}}_{t-}\right]$ in a situation where it can be relatively easily found. Assuming piecewise-constant compensators we have that for $x = \ell$,

$$
\begin{aligned}
&\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \Delta\widetilde{N}^x(t) = 1, \bar{\mathcal{F}}_{t-}\right] \\
&= \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\Delta\widetilde{N}^x(t) = 1\right\} \mid \bar{\mathcal{F}}_{t-}\right]}{\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\Delta\widetilde{N}^x(t) = 1\right\} \mid \bar{\mathcal{F}}_{t-}\right]} \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{t \le \bar{T}_{(k)}\right\}\mathbb{1}\left\{\Delta\widetilde{N}^x(t) = 1\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{t \le \bar{T}_{(k)}\right\}\mathbb{1}\left\{\Delta\widetilde{N}^x(t) = 1\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]} \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{t = \bar{T}_{(k)}, \Delta_{(k)} = x\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{t = \bar{T}_{(k)}, \Delta_{(k)} = x\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]} \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \frac{\mathbb{E}_{PG^*}\left[\mathbb{E}_{P}\left[\bar{Q}^g_{k,\tau}(\tau) \mid \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} = x, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]\mathbb{1}\left\{t = \bar{T}_{(k)}, \Delta_{(k)} = x\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{t = \bar{T}_{(k)}, \Delta_{(k)} = x\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]} \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \mathbb{E}_{PG^*}\left[\bar{Q}^g_{k,\tau}(\tau) \mid \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} = x, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right].
\end{aligned}
\tag{C.2}
$$

by Corollary 2.2.3 of Last & Brandt (1995). Because of the assumption of piecewise-constant compensators, this process need only be defined whenever the denominator of Equation (C.2) is non-zero. Similarly, we may find that

$$
\begin{aligned}
\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \Delta\widetilde{N}^y(t) = 1, \bar{\mathcal{F}}_{t-}\right] &= 1, \\
\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \Delta\widetilde{N}^d(t) = 1, \bar{\mathcal{F}}_{t-}\right] &= 0, \\
\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \Delta\widetilde{N}^a(t) = 1, \bar{\mathcal{F}}_{t-}\right] &= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \bar{Q}^g_{k,\tau}(\tau, 1, H_k).
\end{aligned}
\tag{C.3}
$$

For the first term in Equation (C.1), at $t = \bar{T}_{(k-1)}$,

$$
\begin{aligned}
&\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \tilde{L}(t), \widetilde{N}^\ell(t), \bar{\mathcal{F}}_{t-}\right] - \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \widetilde{N}^\ell(t), \bar{\mathcal{F}}_{t-}\right] \\
&= \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \tilde{L}(t), \Delta\widetilde{N}^\ell(t) = 0, \bar{\mathcal{F}}_{t-}\right] - \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \Delta\widetilde{N}^\ell(t) = 0, \bar{\mathcal{F}}_{t-}\right] \\
&\quad + \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \tilde{L}(t), \Delta\widetilde{N}^\ell(t) = 1, \bar{\mathcal{F}}_{t-}\right] - \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \Delta\widetilde{N}^\ell(t) = 1, \bar{\mathcal{F}}_{t-}\right] \\
&= 0 + \left[\widetilde{N}_y(\tau) \mid \tilde{L}\left(\bar{T}_{(k)}\right), \bar{T}_{(k)} = t, \Delta_{(k)} = \ell, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] - \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \bar{T}_{(k)} = t, \Delta_{(k)} = \ell, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]
\end{aligned}
\tag{C.4}
$$

Next, note that under the integral, by Corollary 2.2.3 of Last & Brandt (1995)

$$
\begin{aligned}
&\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \Delta\widetilde{N}^x(t) = 0, \bar{\mathcal{F}}_{t-}\right] \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \left(\bar{T}_{(k)} > t \vee \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} \neq x\right), \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} = t, \bar{\Delta}_{(k)} \neq x\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\left(\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} = t, \bar{\Delta}_{(k)} \neq x\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]\right)} \\
&\overset{**}{=} \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \bar{T}_{(k)} > t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]} \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{\bar{T}_{(k-1)} < t \le \bar{T}_{(k)}\right\} \frac{\bar{Q}^g_{k-1,\tau}\left(\tau, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) - \bar{Q}^g_{k-1,\tau}\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{S\left(t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)},
\end{aligned}
\tag{C.5}
$$

where in $(**)$, we use that

$$
\begin{aligned}
&\int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad \times \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} = t, \bar{\Delta}_{(k)} \neq x\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\left(\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right] + \mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} = t, \bar{\Delta}_{(k)} \neq x\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]\right)} \widetilde{M}^x(\mathrm{d}t) \\
&= \int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad \times \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \sum_{z\neq x} \mathbb{E}_{PG^*}\left[\widetilde{N}^y(\tau) \mid \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} = z, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\left(\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right] + \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)} \widetilde{M}^x(\mathrm{d}t) \\
&= \int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau) \mid \bar{T}_{(k)} > t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] \widetilde{M}^x(\mathrm{d}t).
\end{aligned}
$$

To see the last equality, note that by orthogonality of the martingales (Equation (2.4.2) and (2.4.3) of Andersen et al. (1993)), we have that

$$
0 = \sum_k \int_0^s \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \Lambda_k^x\left(\mathrm{d}t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)
$$

Conclude that

$$
\begin{aligned}
&\int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad \times \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \sum_{z\neq x} \mathbb{E}_{PG^*}\left[\widetilde{N}^y(\tau) \mid \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} = z, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\left(\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right] + \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)} \\
&\quad \times \Lambda_k^x\left(\mathrm{d}t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\left(\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]\right)} \Lambda_k^x\left(\mathrm{d}t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad\times \frac{\mathbb{E}_{P^{G^*}}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \sum_{z\neq x} \mathbb{E}_{P^{G^*}}\left[\widetilde{N}^y(\tau) \mid \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} = z, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\left(\mathbb{E}_{P^{G^*}}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right] + \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)} \\
&\quad\times \widetilde{N}^x\left(\mathrm{d}t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad\times \frac{\mathbb{E}_{P^{G^*}}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\left(\mathbb{E}_{P^{G^*}}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]\right)}\mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\}\mathbb{1}\left\{\sum_{z\neq x}\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) = 0\right\}\widetilde{N}^x(\mathrm{d}t) \\
&+ \int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad\times \frac{\mathbb{E}_{P^{G^*}}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \sum_{z\neq x} \mathbb{E}_{P^{G^*}}\left[\widetilde{N}^y(\tau) \mid \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} = z, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\left(\mathbb{E}_{P^{G^*}}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right] + \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)} \\
&\quad\times \mathbb{1}\left\{\sum_{z\neq x}\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) > 0\right\}\widetilde{N}^x(\mathrm{d}t) \\
&\overset{***}{=} \int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad\times \frac{\mathbb{E}_{P^{G^*}}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\left(\mathbb{E}_{P^{G^*}}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]\right)}\mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\}\mathbb{1}\left\{\sum_{z\neq x}\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) = 0\right\}\widetilde{N}^x(\mathrm{d}t) \\
&\overset{****}{=} \int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad\times \frac{\mathbb{E}_{P^{G^*}}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]}{\left(\mathbb{E}_{P^{G^*}}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]\right)}\mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\}\widetilde{N}^x(\mathrm{d}t)
\end{aligned}
$$

Here, we use in $(***)$ that the second term is seen to be almost surely zero since have by the definition of a compensator

$$
\begin{aligned}
&\mathbb{E}_P\Bigg[\int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad\times \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \sum_{z\neq x}\mathbb{E}_{PG^*}\left[\widetilde{N}^y(\tau) \mid \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} = z, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\left(\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right] + \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)} \\
&\quad\times \mathbb{1}\left\{\sum_{z\neq x}\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) > 0\right\}\widetilde{N}^x\left(\mathrm{d}t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Bigg] \\
&= \mathbb{E}_P\Bigg[\int_0^\tau W^g(t-)\sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \\
&\quad\times \frac{\mathbb{E}_{PG^*}\left[\widetilde{N}_y(\tau)\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] + \sum_{z\neq x}\mathbb{E}_{PG^*}\left[\widetilde{N}^y(\tau) \mid \bar{T}_{(k)} = t, \bar{\Delta}_{(k)} = z, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\left(\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{\bar{T}_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right] + \Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)} \\
&\quad\times \mathbb{1}\left\{\sum_{z\neq x}\Delta\Lambda_k^z\left(t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) > 0\right\}\Lambda_k^x\left(\mathrm{d}t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Bigg] \\
&= 0,
\end{aligned}
$$

where we use the assumption of orthogonal martingales. Next, in $(****)$ we use the same line of argument to show that the difference of the last line and previous line is almost surely zero. Let $B_{k-1}(u) = \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(u)\right)\frac{1}{S\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}$. Combining Equation (C.2), Equation (C.3), Equation (C.4), and Equation (C.5) with Equation (C.1), we find that the efficient influence function can also be written as:

$$
\begin{aligned}
\varphi^*_\tau(P) &= \sum_{k=1}^K \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{A(\bar T_j)=1\}}{\pi_j\left(\bar T_{(j)}, L(\bar T_j), \bar{\mathcal{F}}_{\bar T_{(j-1)}}\right)} \right)^{\mathbb{1}\left\{\bar\Delta_{(j)}=a\right\}} \frac{1}{\prod_{j=1}^{k-1} \tilde S^c\left(\bar T_{(j)}- \mid \bar{\mathcal{F}}_{\bar T_{(j-1)}}\right)} \mathbb{1}\left\{\bar\Delta_{(k-1)} \in \{\ell, a\}, \bar T_{(k-1)} < \tau\right\} \\
&\quad \Bigg[\int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \left(\bar Q^g_{k,\tau}\left(L(\bar T_k), 1, a, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) - B_{k-1}(u)\right) \widetilde M^a(du) \\
&\quad + \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \left(\mathbb{E}_P\left[\bar Q^g_{k,\tau}(\tau) \mid T_{(k)} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right] - B_{k-1}(u)\right) \widetilde M^\ell(du) \\
&\quad + \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} (1 - B_{k-1}(u)) \widetilde M^y(du) + \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} (0 - B_{k-1}(u)) \widetilde M^d(du) \\
&\quad + \frac{1}{\tilde S^c\left(\bar T_{(k)}- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \mathbb{1}\left(\bar T_{(k)} < \tau, \bar\Delta_{(k)} = \ell, k < K\right) \left(\bar Q^g_{k,\tau}(\tau) - \mathbb{E}_P\left[\bar Q^g_{k,\tau}(\tau) \mid \bar T_{(k)}, \Delta_{(k)} = \ell, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right]\right) \Bigg] \\
&\quad + \bar Q^g_{0,\tau}(\tau, 1, L(0)) - \Psi^g_\tau(P) \\
&= \sum_{k=1}^K \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{A(\bar T_j)=1\}}{\pi_j\left(\bar T_{(j)}, L(\bar T_j), \bar{\mathcal{F}}_{\bar T_{(j-1)}}\right)} \right)^{\mathbb{1}\left\{\bar\Delta_{(j)}=a\right\}} \frac{1}{\prod_{j=1}^{k-1} \tilde S^c\left(\bar T_{(j)}- \mid \bar{\mathcal{F}}_{\bar T_{(j-1)}}\right)} \mathbb{1}\left\{\bar\Delta_{(k-1)} \in \{\ell, a\}, \bar T_{(k-1)} < \tau\right\} \Bigg[\\
&\quad - \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \bar Q^g_{k,\tau}\left(1, L(\bar T_{k-1}), u, a, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) \Lambda^a_k\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) \\
&\quad - \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \mathbb{E}_P\left[\bar Q^g_{k,\tau}(\tau) \mid \bar T_{(k)} = s, \Delta_{(k)} = \ell, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right] \Lambda^\ell_k\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) \\
&\quad - \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} (1) \Lambda^y_k\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) - \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} (0) \Lambda^d_k\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) \\
&\quad - \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} B_{k-1}(u) M^\bullet(\mathrm{d}u) + \bar Z^a_{k,\tau}(\tau) \\
&\quad + \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \left(\bar Q^g_{k-1,\tau}(\tau) - \bar Q^g_{k-1,\tau}(u)\right) \frac{1}{S\left(u- \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \widetilde M^c(du) \Bigg] \\
&\quad + \bar Q^g_{0,\tau}(\tau, 1, L(0)) - \Psi^g_\tau(P),
\end{aligned}
\tag{C.6}
$$

where $M^\bullet(t) = \sum_{x=a,\ell,d,y,c} \widetilde M^x(t)$ and $\Lambda^\bullet_k\left(s, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) = \sum_{x=a,\ell,d,y,c} \Lambda^x_k\left(s, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)$. In the second equality, we apply orthogonality of martingales again to conclude that we can replace the u's with $u-$ in all the relevant places. Now note that

$$
\begin{aligned}
&\int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \left(\bar Q^g_{k-1,\tau}(\tau) - \bar Q^g_{k-1,\tau}(u)\right) \frac{1}{\tilde S^c\left(u \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) S\left(u \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} M^\bullet(\mathrm{d}u) \\
&= \left(\bar Q^g_{k-1,\tau}(\tau) - \bar Q^g_{k-1,\tau}\left(\bar T_{(k)}\right)\right) \frac{1}{\tilde S^c\left(\bar T_{(k)} \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) S\left(\bar T_{(k)} \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \mathbb{1}\left\{\bar T_{(k)} \le \tau\right\} \\
&\quad - \bar Q^g_{k-1,\tau}(\tau) \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{1}{\tilde S^c\left(u \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) S\left(u \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \Lambda^\bullet_k\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) \\
&\quad + \int_{\bar T_{(k-1)}}^{\tau \wedge \bar T_{(k)}} \frac{\bar Q^g_{k-1,\tau}(u)}{\tilde S^c\left(u \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right) S\left(u \mid \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right)} \Lambda^\bullet_k\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar T_{(k-1)}}\right).
\end{aligned}
\tag{C.7}
$$

Let us calculate the first integral of Equation (C.7). We have,

$$
\begin{aligned}
&\bar{Q}^g_{k-1,\tau}(\tau)\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\Lambda_k^\bullet\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \bar{Q}^g_{k-1,\tau}(\tau)\left(\frac{1}{\tilde{S}^c\left(\bar{T}_{(k)}\wedge\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(\bar{T}_{(k)}\wedge\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)} - 1\right),
\end{aligned} \tag{C.8}
$$

where the last line holds by the Duhamel equation (2.6.5) of Andersen et al. (1993). The second integral of Equation (C.7) is equal to

$$
\begin{aligned}
&\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \frac{\bar{Q}^g_{k-1,\tau}(u)}{\tilde{S}^c\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\Lambda_k^\bullet\left(u, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \Bigg[\int_{\bar{T}_{(k-1)}}^{u} S\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\bar{Q}^g_{k,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Lambda_k^a\left(\mathrm{d}s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&+ \int_{\bar{T}_{(k-1)}}^{u} S\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\mathbb{E}_P\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\Lambda_k^\ell\left(\mathrm{d}s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&+ \int_{\bar{T}_{(k-1)}}^{u} S\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Lambda_k^y\left(s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\mathrm{d}s\Bigg] \times \frac{1}{\tilde{S}^c\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\Lambda_k^\bullet\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \int_{[s,\tau\wedge\bar{T}_{(k)}]} \frac{1}{\tilde{S}^c\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\Lambda_k^\bullet\left(\mathrm{d}u, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&\times \Big[S\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\bar{Q}^g_{k,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Lambda_k^a\left(\mathrm{d}s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&\quad + S\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\mathbb{E}_P\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\Lambda_k^\ell\left(\mathrm{d}s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&\quad + S\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Lambda_k^y\left(\mathrm{d}s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Big] \\
&\overset{(*)}{=} \frac{1}{\tilde{S}^c\left(\tau\wedge\bar{T}_{(k)} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(\tau\wedge\bar{T}_{(k)} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\bar{Q}^g_{k-1,\tau}\left(\tau\wedge\bar{T}_{(k)}\right) \\
&- \int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\Big[\bar{Q}^g_{k,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Lambda_k^a\left(\mathrm{d}s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&\quad + \mathbb{E}_P\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\Lambda_k^\ell\left(\mathrm{d}s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) + \Lambda_k^y\left(\mathrm{d}s, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\Big].
\end{aligned} \tag{C.9}
$$

In $(*)$, we use that

$$
\begin{aligned}
&\int_{[s,\tau\wedge\bar{T}_{(k)}]} \frac{1}{\tilde{S}^c\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\Lambda_k^\bullet\left(u, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \\
&= \frac{1}{\tilde{S}^c\left(\tau\wedge\bar{T}_{(k)} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(\tau\wedge\bar{T}_{(k)} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)} - \frac{1}{\tilde{S}^c\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(s- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)},
\end{aligned}
$$

which, again, follows by the Duhamel equation. Thus, we find by Equation (C.7) Equation (C.8), Equation (C.9)

$$
\begin{aligned}
&\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(u)\right) \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} M^\bullet(\mathrm{d}u) \\
&= \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}\left(\bar{T}_{(k)}\right)\right) \frac{1}{\tilde{S}^c\left(\bar{T}_{(k)} \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) S\left(\bar{T}_{(k)} \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} \mathbb{1}\left\{\bar{T}_{(k)} \le \tau\right\} \\
&- \bar{Q}^g_{k-1,\tau}(\tau) \left(\frac{1}{\tilde{S}^c\left(\bar{T}_{(k)} \wedge \tau \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) S\left(\bar{T}_{(k)} \wedge \tau \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} - 1 \right) \\
&\quad \times \frac{1}{\tilde{S}^c\left(\tau \wedge \bar{T}_{(k)} \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) S\left(\tau \wedge \bar{T}_{(k)} \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} \bar{Q}^g_{k-1,\tau}\left(\tau \wedge \bar{T}_{(k)}\right) \\
&- \int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} \Big[\bar{Q}^g_{k,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \mathcal{F}_{\bar{T}_{(k-1)}}\right) \Lambda^a_k\left(\mathrm{d}s, \mathcal{F}_{\bar{T}_{(k-1)}}\right) \\
&\quad + \mathbb{E}_P\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right] \Lambda^\ell_k\left(\mathrm{d}s, \mathcal{F}_{\bar{T}_{(k-1)}}\right) + \Lambda^y_k\left(\mathrm{d}s, \mathcal{F}_{\bar{T}_{(k-1)}}\right)\Big] \\
&= - \int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} \Big[\bar{Q}^g_{k,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \mathcal{F}_{\bar{T}_{(k-1)}}\right) \Lambda^a_k\left(\mathrm{d}s, \mathcal{F}_{\bar{T}_{(k-1)}}\right) \\
&\quad + \mathbb{E}_P\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right] \Lambda^\ell_k\left(\mathrm{d}s, \mathcal{F}_{\bar{T}_{(k-1)}}\right) + \Lambda^y_k\left(\mathrm{d}s, \mathcal{F}_{\bar{T}_{(k-1)}}\right)\Big] + \bar{Q}^g_{k-1,\tau}(\tau).
\end{aligned}
$$

This now shows that Equation (C.6) is equal to Equation (5.3).

APPENDIX D

D.1 Proof of Theorem 3

We let $\bar{Q}^g_{k,\tau}(u, a_k, h_k; P)$ denote the right-hand side of Equation (4.2), with P being explicit in the notation and likewise define the notation with $\bar{Z}^a_{k,\tau}(u; P)$. Let $\{P_\varepsilon \mid |\varepsilon| \le \delta\}$ denote a bounded, Hellinger differentiable path with $P_\varepsilon = P$ at $\varepsilon = 0$ and score function $\dot{\ell}$ at $\varepsilon = 0$. We compute the efficient influence function by calculating $\frac{\partial}{\partial \varepsilon}\big|_{\varepsilon=0} \Psi_\tau(P_\varepsilon)$. Assuming boundedness, first note that

$$
\begin{aligned}
\frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \Psi_\tau(P_\varepsilon) &= \int \dot{\ell}(l) \bar{Q}^g_{0,\tau}(\tau, 1, l) P_L(\mathrm{d}l) + \int \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \bar{Q}^g_{0,\tau}(\tau, 1, l; P_\varepsilon) P_L(\mathrm{d}l) \\
&= \int \dot{\ell}(l) \left(\bar{Q}^g_{0,\tau}(\tau, 1, l) - \Psi_\tau(P)\right) P_L(\mathrm{d}l) + \int \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \bar{Q}^g_{0,\tau}(\tau, 1, l; P_\varepsilon) P_L(\mathrm{d}l).
\end{aligned}
\tag{D.1}
$$

Then note that

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \Lambda^c_{k,\varepsilon}(\mathrm{d}t \mid f_{k-1}) \\
&= \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \frac{P_\varepsilon\left(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}{P_\varepsilon\left(\bar{T}_{(k)} \geq t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \\
&= \frac{\frac{\partial}{\partial \varepsilon}\big|_{\varepsilon=0} P_\varepsilon\left(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}{P\left(\bar{T}_{(k)} \geq t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \\
&\qquad - \frac{\frac{\partial}{\partial \varepsilon}\big|_{\varepsilon=0} P_\varepsilon\left(\bar{T}_{(k)} \geq t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) P\left(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}{\left(P\left(\bar{T}_{(k)} \geq t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right)^2} \\
&= \frac{1}{P\left(\bar{T}_{(k)} \geq t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \\
&\quad \times \underbrace{\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} P_\varepsilon\left(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) - \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} P_\varepsilon\left(\bar{T}_{(k)} \geq t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \Lambda^c_k(\mathrm{d}t \mid f_{k-1})}_{:= h_P(\mathrm{d}t)}
\end{aligned}
$$

We also have

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} P_\varepsilon\left(\bar{T}_{(k)} \in \cdot, \bar{\Delta}_{(k)} = c \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&= \int_{\cdot \times \{a,\ell,d,y,c\}} \mathbb{1}\{\delta = c\} \dot{\ell}(s, \delta \mid f_{k-1}) P\left(\bar{T}_{(k)} \in \mathrm{d}s, \bar{\Delta}_{(k)} \in \mathrm{d}\delta \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} P_\varepsilon\left(\bar{T}_{(k)} \geq t \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&= \int \mathbb{1}\{t \geq s\} \dot{\ell}(s, \delta \mid f_{k-1}) P\left(\bar{T}_{(k)} \in \mathrm{d}s, \bar{\Delta}_{(k)} \in \mathrm{d}\delta \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)
\end{aligned}
$$

so that

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \prod_{u \in (s,t)} \left(1 - \tilde{\Lambda}^c_{k,\varepsilon}(\mathrm{d}t \mid f_{k-1})\right) \\
&= \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \frac{1}{1 - \Delta\tilde{\Lambda}^c_{k,\varepsilon}(t \mid f_{k-1})} \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_{k,\varepsilon}(\mathrm{d}t \mid f_{k-1})\right) \\
&\overset{(*)}{=} -\frac{1}{1 - \Delta\tilde{\Lambda}^c_k(t \mid f_{k-1})} \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\right) \int_{(s,t]} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k(u \mid f_{k-1})} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \tilde{\Lambda}^c_{k,\varepsilon}(\mathrm{d}u \mid f_{k-1}) \\
&\quad + \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\right) \frac{1}{\left(1 - \Delta\tilde{\Lambda}^c_k(t \mid f_{k-1})\right)^2} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \Delta\tilde{\Lambda}^c_{k,\varepsilon}(t \mid f_{k-1}) \\
&= -\frac{1}{1 - \Delta\tilde{\Lambda}^c_k(t \mid f_{k-1})} \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\right) \int_{(s,t)} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k(u \mid f_{k-1})} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \tilde{\Lambda}^c_{k,\varepsilon}(\mathrm{d}u \mid f_{k-1}) \\
&\quad - \frac{1}{1 - \Delta\tilde{\Lambda}^c_k(t \mid f_{k-1})} \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\right) \int_{\{t\}} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k(u \mid f_{k-1})} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \tilde{\Lambda}^c_{k,\varepsilon}(\mathrm{d}u \mid f_{k-1}) \\
&\quad + \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\right) \frac{1}{\left(1 - \Delta\tilde{\Lambda}^c_k(t \mid f_{k-1})\right)^2} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \Delta\tilde{\Lambda}^c_{k,\varepsilon}(t \mid f_{k-1}) \\
&\overset{(**)}{=} -\prod_{u \in (s,t)} \left(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\right) \int_{(s,t)} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k(u \mid f_{k-1})} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \tilde{\Lambda}^c_{k,\varepsilon}(\mathrm{d}u \mid f_{k-1}).
\end{aligned} \tag{D.2}
$$

In $(*)$, we use the product rule of differentiation and a result for product integration (Theorem 8 of Gill & Johansen (1990)), which states that the (Hadamard) derivative of the product integral $\mu \mapsto \prod_{u \in (s,t]}(1 + \mu(u))$ in the direction h is given by (for μ of uniformly bounded variation)

$$\int_{(s,t]} \prod_{v\in(s,u)} (1+\mu(\mathrm{d}v)) \prod_{v\in(u,t]} (1+\mu(\mathrm{d}v)) h(\mathrm{d}u) = \prod_{v\in(s,t]} (1+\mu(\mathrm{d}v)) \int_{(s,t]} \frac{1}{1+\Delta\mu(u)} h(\mathrm{d}u).$$

In $(**)$, we use that $\int_{\{t\}} \frac{1}{1-\Delta\tilde{\Lambda}_k^c(u \mid f_{k-1})} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \tilde{\Lambda}_k^c(\mathrm{d}u \mid f_{k-1}) = \frac{1}{1-\Delta\tilde{\Lambda}_k^c(t \mid f_{k-1})} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \Delta\tilde{\Lambda}_{k,\varepsilon}^c(t \mid f_{k-1})$. Furthermore, a simple calculation yields the well-known result

$$\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \mathbb{E}_{P_\varepsilon}[Y \mid X=x] = \int y\dot{\ell}(y \mid x) P(\mathrm{d}y \mid x) = \int (y-\mathbb{E}_P[Y \mid X=x])\dot{\ell}(y \mid x) P(\mathrm{d}y \mid x) \tag{D.3}$$

Using Equation (4.2) with Equation (D.3) and Equation (D.2), we have

$$\begin{aligned}
&\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \bar{Q}_{k-1,\tau}^g(\tau, f_{k-1}; P_\varepsilon) \\
&= \int \mathbb{1}\{t_{k-1} < t_k \le \tau\} \dot{\ell}(t_k, d_k, l_k, a_k \mid f_{k-1}) \\
&\qquad \left(\bar{Z}_{k,\tau}^a(t_k, d_k, l_k, a_k, f_{k-1}) - \bar{Q}_{k-1,\tau}^g(\tau, f_{k-1})\right) P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}(\mathrm{d}(t_k, d_k, l_k, a_k) \mid f_{k-1}) \\
&\quad + A + B,
\end{aligned} \tag{D.4}$$

for $k = 1, \dots, K$, where in the notation with $\bar{Z}_{k,\tau}^a$, we have made the dependencies explicit, and

$$\begin{aligned}
A &:= \int \mathbb{1}\{t_{k-1} < t_k \le \tau\} \bar{Z}_{k,\tau}^a(t_k, d_k, l_k, a_k, f_{k-1}) \int_{(t_{k-1}, t_k)} \frac{1}{1-\Delta\tilde{\Lambda}_k^c(s \mid f_{k-1})} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \Lambda_{k,\varepsilon}^c(\mathrm{d}s \mid f_{k-1}) \\
&\quad P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}(\mathrm{d}(t_k, d_k, l_k, a_k) \mid f_{k-1}) \\
B &:= \int \mathbb{1}\{t_{k-1} < t_k \le \tau\} \frac{\mathbb{1}\{t_k < \tau, d_k \in \{a, \ell\}\}}{\tilde{S}^c(t_k - \mid f_{k-1})} \left(\frac{\mathbb{1}\{a_k = 1\}}{\pi_k(t_k, l_k, f_{k-1})}\right)^{\mathbb{1}\{d_k = a\}} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \bar{Q}_{k,\tau}^g\left(P_\varepsilon \mid a_k, l_k, t_k, d_k, f_{k-1}\right) \\
&\quad P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}(\mathrm{d}(t_k, d_k, l_k, a_k) \mid f_{k-1}),
\end{aligned}$$

To get B, we use the same derivation as the one given in Equation (A.4). Now note that for simplifying A, we can write

$$\begin{aligned}
&\int \mathbb{1}\{t_{k-1} < t_k \le \tau\} \bar{Z}_{k,\tau}^a(t_k, d_k, l_k, a_k, f_{k-1}) \int_{(t_{k-1}, t_k)} \frac{1}{1-\Delta\tilde{\Lambda}_k^c(s \mid f_{k-1})} \frac{1}{\tilde{S}\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} h_P(\mathrm{d}s) \\
&\qquad P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}\left(\mathrm{d}(t_k, d_k, l_k, a_k) \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&= \int_{(t_{k-1}, \tau)} \int \mathbb{1}\{s < t_k \le \tau\} \bar{Z}_{k,\tau}^a(t_k, d_k, l_k, a_k, f_{k-1}) P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}\left(\mathrm{d}(t_k, d_k, l_k, a_k) \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&\qquad \frac{1}{1-\Delta\tilde{\Lambda}_k^c(s \mid f_{k-1})} \frac{1}{\tilde{S}\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} h_P(\mathrm{d}s) \\
&= \int_{(t_{k-1}, \tau)} \left(\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(s)\right) \\
&\qquad \frac{1}{1-\Delta\tilde{\Lambda}_k^c(s \mid f_{k-1})} \frac{1}{\tilde{S}\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} h_P(\mathrm{d}s) \\
&= \int_{(t_{k-1}, \tau)} \left(\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(s)\right) \\
&\qquad \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} h_P(\mathrm{d}s),
\end{aligned}$$

by an exchange of integrals. Then,

$$
\begin{aligned}
&\int \mathbb{1}\{t_{k-1} < s < \tau\}\left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} h_P(\mathrm{d}s) \\
&= \int \mathbb{1}\{t_{k-1} < s < \tau\}\left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \\
&\quad \times \mathbb{1}\{\delta = c\} \dot{\ell}(s, \delta \mid f_{k-1}) P\left(\bar{T}_{(k)} \in \mathrm{d}s, \bar{\Delta}_{(k)} \in \mathrm{d}\delta\right) \\
&\quad - \int \mathbb{1}\{t_{k-1} < s < \tau\}\left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \\
&\quad \times \int \mathbb{1}\{u \geq s\} \dot{\ell}(u, \delta' \mid f_{k-1}) P\left(\bar{T}_{(k)} \in \mathrm{d}u, \bar{\Delta}_{(k)} \in \mathrm{d}\delta' \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \Lambda^c_k(\mathrm{d}s \mid f_{k-1}) \\
&= \int \mathbb{1}\{t_{k-1} < s < \tau\}\left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \\
&\quad \times \mathbb{1}\{\delta = c\} \dot{\ell}(s, \delta \mid f_{k-1}) P\left(\bar{T}_{(k)} \in \mathrm{d}s, \bar{\Delta}_{(k)} \in \mathrm{d}\delta\right) \\
&\quad - \int \dot{\ell}(u, \delta' \mid f_{k-1}) \int \mathbb{1}\{t_{k-1} < s < \tau\}\left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \\
&\quad \times \mathbb{1}\{u \geq s\} \Lambda^c_k(\mathrm{d}s \mid f_{k-1}) P\left(\bar{T}_{(k)} \in \mathrm{d}u, \bar{\Delta}_{(k)} \in \mathrm{d}\delta' \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&= \int \dot{\ell}(u, \delta' \mid f_{k-1}) \int \mathbb{1}\{t_{k-1} < s < \tau\}\left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \\
&\quad \times \left(\widetilde{N}^c(\mathrm{d}s) - \mathbb{1}\{u \geq s\} \Lambda^c_k(\mathrm{d}s \mid f_{k-1})\right) P\left(\bar{T}_{(k)} \in \mathrm{d}u, \bar{\Delta}_{(k)} \in \mathrm{d}\delta' \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right)
\end{aligned} \tag{D.5}
$$

Note that due to orthogonality, every place with a conditional score can be replaced with the full score. Therefore, we can combine the results from iteratively from Equation (D.4), Equation (D.5), and Equation (D.1) to obtain the result, i.e. Equation (5.3).

APPENDIX E

E.1 Proof of Theorem 4

We find the following decomposition,

$$
\begin{aligned}
\hat{\Psi}_n - \Psi^{g,K_{\lim}}_\tau(P) &= \hat{\Psi}_n - \Psi^{g,K_{nc}}_\tau(P) + \Psi^{g,K_{nc}}_\tau(P) - \Psi^{g,K_{\lim}}_\tau(P) \\
&= (\mathbb{P}_n - P)\varphi^{*,K_{nc}}_\tau(\cdot; P) + \Psi^{g,K_{nc}}_\tau(P) - \Psi^{g,K_{\lim}}_\tau(P) + o_P\left(n^{-\frac{1}{2}}\right) \\
&= (\mathbb{P}_n - P)\varphi^{*,K_{\lim}}_\tau(\cdot; P) + (\mathbb{P}_n - P)\left(\varphi^{*,K_{nc}}_\tau(\cdot; P) - \varphi^{*,K_{\lim}}_\tau(\cdot; P)\right) + \Psi^{g,K_{nc}}_\tau(P) - \Psi^{g,K_{\lim}}_\tau(P) + o_P\left(n^{-\frac{1}{2}}\right).
\end{aligned} \tag{E.1}
$$

Equation (E.1) shows that we will have shown the result if

1. $\Psi^{g,K_{nc}}_\tau(P) - \Psi^{g,K_{\lim}}_\tau(P) = o_P\left(n^{-\frac{1}{2}}\right)$.
2. $(\mathbb{P}_n - P)\left(\varphi^{*,K_{nc}}_\tau(\cdot; P) - \varphi^{*,K_{\lim}}_\tau(\cdot; P)\right) = o_P\left(n^{-\frac{1}{2}}\right)$.

To do so, we now show that $P(K_{nc} \neq K_{\lim}) \to 0$. First define $K_n = \max_i \widetilde{N}_{\tau-,i}$. Then, we can certainly write that $K_{nc} - K_{\lim} = K_{nc} - K_n + K_n - K_{\lim}$ so that $P(K_{nc} = K_{\lim}) \geq P(K_{nc} = K_n, K_n = K_{\lim})$ whence

$$
P(K_{nc} \neq K_{\lim}) \leq P(K_{nc} \neq K_n) + P(K_n \neq K_{\lim}). \tag{E.2}
$$

By independence and definition of K_n, we have

$$
P(K_n \neq K_{\lim}) = P(K_n < K_{\lim}) + P(K_n > K_{\lim}) \overset{(a)}{=} P(K_n < K_{\lim}) = P\left(\widetilde{N}_{\tau-} < K_{\lim}\right)^n \overset{(b)}{\to} 0. \tag{E.3}
$$

In (a), we use that $N_{\tau-} \geq \widetilde{N}_{\tau-}$ and the fact that $P(N_{\tau-} > K_{\lim}) = 0$. In (b), we use that $P\left(\widetilde{N}_{\tau-} < K_{\lim}\right) < 1$ by the assumptions stated in the theorem. We now show that $P(K_{nc} < K_n) \to 0$ as $n \to \infty$, which will show that $P(K_{nc} \neq K_{\lim}) \to 0$ as $n \to \infty$ by Equation (E.2) and Equation (E.3). We have,

$$
P(K_{nc} \neq K_n) = P\left(\cup_{v=1}^{K_{\lim}} \left(\sum_{i=1}^n \mathbb{1}\left\{\widetilde{N}_{\tau-,i} \geq v\right\} \leq c\right)\right) \leq \sum_{v=1}^{K_{\lim}} P\left(\sum_{i=1}^n \mathbb{1}\left\{\widetilde{N}_{\tau-,i} \geq v\right\} \leq c\right) \to 0
$$

as $n \to \infty$. Here, we use that $\sum_{i=1}^n \mathbb{1}\left\{\widetilde{N}_{\tau-,i} \geq v\right\}$ diverges almost surely to ∞. Too see this, note that $\sum_{i=1}^n \mathbb{1}\left\{\widetilde{N}_{\tau-,i} \geq v\right\}$ is almost surely monotone in n, and $\sum_{i=1}^n P\left(\widetilde{N}_{\tau-,i} \geq v\right) = nP\left(\widetilde{N}_{\tau-} \geq v\right) \to \infty$. From this and Kolmogorov's three series theorem, we conclude that $\sum_{i=1}^n \mathbb{1}\left\{\widetilde{N}_{\tau-,i} \geq v\right\} \to \infty$ almost surely as $n \to \infty$ and that $\sum_{i=1}^n \mathbb{1}\left\{\widetilde{N}_{\tau-,i} \geq v\right\} \leq c$ has probability tending to zero as $n \to \infty$ as desired.

Returning to showing 1 and 2, we now have that

$$\sqrt{n}\left(\Psi_\tau^{g,K_{nc}}(P) - \Psi_\tau^{g,K_{\lim}}(P)\right) = \sqrt{n}\mathbb{1}\{K_{nc} \neq K_{\lim}\}\left(\Psi_\tau^{g,K_{nc}}(P) - \Psi_\tau^{g,K_{\lim}}(P)\right) := E_n$$

and

$$P(|E_n| > \varepsilon) \leq P(K_{nc} \neq K_{\lim}) \to 0,$$

as $n \to \infty$, so that 1 holds. A similar conclusion holds for 2, so the proof is complete.

APPENDIX F

F.1 One-step procedure (the general case)

Algorithm 3 (Debiased ICE-IPCW estimator (non-conservative)): Input: Observed data $\tilde{O}_i$, $i = 1, ..., n$, time horizon $\tau < \tau_{\text{end}}$, and K. Estimators of the propensity score $\hat{\pi}_0$, $\hat{\pi}_k$ for $k = 1, ..., K-1$ and the censoring compensator $\hat{\Lambda}^c$. Estimates of $\hat{S}\left(\cdot \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) = \prod_{s \in \left(\bar{T}_{(k-1)}, \bar{T}_{(k)}\right)}\left(1 - \sum_{x=a,\ell,d,y} \Lambda_k^x\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right)$.

Output: One-step estimator $\hat{\Psi}_n$ of $\Psi_\tau^g(P)$; estimate of influence function $\varphi_\tau^*\left(\tilde{O}; \hat{P}\right)$.

1. Compute the ICE-IPCW estimator $\hat{\Psi}_n^0$ and obtain estimators of $\bar{Q}_{k,\tau}^g$ for $k = 0, ..., K-1$.
2. Obtain estimates of $\varphi_\tau^{*,d}\left(\tilde{O}_i; \hat{P}\right)$ for $i = 1, ..., n$ via Algorithm 2.
3. For $k = K-1, ..., 1$:

   a. Estimate $\bar{Q}_{k-1,\tau}^g(u)$ for $u = t_1, ..., t_{m-1}$ where $0 < t_1 < ... < t_m = \tau$ by applying the following procedure:

   - For each observation with $\bar{T}_{(k-1)} < u$ and $\bar{\Delta}_{(k-1)} \in \{a, \ell\}$ compute the pseudo-outcome $\hat{Z}_{k,\tau}^a$ as follows:
     - If $\bar{\Delta}_{(k)} = y$, calculate $\hat{Z}_{k,\tau}^a(u) = \frac{1}{\hat{S}^c\left(\bar{T}_{(k)}- \mid A(\bar{T}_{k-1}), H_{k-1}\right)} \mathbb{1}\left\{\bar{T}_{(k)} \leq u, \bar{T}_{(k)} < \tau\right\}$.
     - Else if $\bar{\Delta}_{(k)} = a$, evaluate $\hat{Q}_{k+1}(\tau, 1, H_k)$ and calculate
     $$\hat{Z}_{k,\tau}^a(u) = \frac{1}{\hat{S}^c\left(\bar{T}_{(k)} - \mid A(\bar{T}_{k-1}), H_{k-1}\right)} \mathbb{1}\left\{\bar{T}_{(k)} \leq u, \bar{T}_{(k)} < \tau\right\} \hat{Q}_{k+1}(\tau, 1, H_k).$$
     - Else if $\bar{\Delta}_{(k)} = \ell$, evaluate $\hat{Q}_{k+1}(\tau, A(\bar{T}_{k-1}), H_k)$, and calculate
     $$\hat{Z}_{k,\tau}^a(u) = \frac{1}{\hat{S}^c\left(\bar{T}_{(k)} - \mid A(\bar{T}_{k-1}), H_{k-1}\right)} \mathbb{1}\left\{\bar{T}_{(k)} < \tau\right\} \hat{Q}_{k+1}(\tau, A(\bar{T}_{k-1}), H_k).$$
   - Regress $\hat{Z}_{k,\tau}^a(u)$ on $(A(\bar{T}_{k-1}), \bar{H}_{k-1})$ for the observations with $\bar{T}_{(k-1)} < u$ and $\bar{\Delta}_{(k-1)} \in \{a, \ell\}$ to obtain a prediction function $\hat{Q}_k(u)$.

   b. Approximate the integral in Equation (5.2) using an integral approximation along the grid $t_1, ..., t_m$, that is:

   - Obtain estimates for each observation $i = 1, ..., n$
   $$\mathbb{1}\left\{\bar{T}_{(k)} \leq \tau, \bar{\Delta}_{(k)} = c\right\}\left(\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g\left(\bar{T}_{(k)}\right)\right) \frac{1}{\hat{S}^c\left(\bar{T}_{(k)} \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)\hat{S}\left(\bar{T}_{(k)} - \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)}$$
   where we estimate $\bar{Q}_{k-1,\tau}^g\left(\bar{T}_{(k)}\right)$ as $\bar{Q}_{k-1,\tau}^g(u)$ with $u = \min_{t \in \{t_1, ..., t_m\}} |t - \bar{T}_{(k)}|$.
   - Obtain estimates of the compensator term for each observation $i = 1, ..., n$
   $$\sum_{j=1}^m \mathbb{1}\left\{\bar{T}_{(k-1)} < t_j \leq \bar{T}_{(k)} \wedge \tau\right\}\left(\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(t_j)\right) \frac{1}{\tilde{S}^c\left(t_j \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) S\left(t_j - \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)}$$
   $$\times \left(\tilde{\Lambda}_k^c\left(t_j, \mathcal{F}_{\bar{T}_{(k-1)}}\right) - \tilde{\Lambda}_k^c\left(t_{j-1}, \mathcal{F}_{\bar{T}_{(k-1)}}\right)\right)$$
   - Plug in the estimates in Equation (5.2) of the martingale term, and the propensity scores $\hat{\pi}_k$ and censoring weights $\hat{S}^c$ to obtain estimates of $\varphi_\tau^{*,\widetilde{M}^c}\left(\tilde{O}_i; \hat{P}\right)$ for $i = 1, ..., n$.
4. Compute an estimate $\varphi_\tau^*\left(\tilde{O}_i; \hat{P}\right) = \varphi_\tau^{*,\widetilde{M}^c}\left(\tilde{O}_i; \hat{P}\right) + \varphi_\tau^{*,d}\left(\tilde{O}_i; \hat{P}\right)$ for $i = 1, ..., n$.
5. Compute the one-step estimator
$$\hat{\Psi}_n = \hat{\Psi}_n^0 + \frac{1}{n} \sum_{i=1}^n \varphi_\tau^*\left(\tilde{O}_i; \hat{P}\right).$$
6. Return $\hat{\Psi}_n$ and $\varphi_\tau^*\left(\widetilde{O}_i; \hat{P}\right)$ for $i = 1, ..., n$.

APPENDIX G

G.1 Additional simulation results

G.1.1 Tables

Varying effects (A on Y, L on Y, A on L, L on A) – uncensored case

β_A^y	Estimator	Coverage	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
-0.3	ICE-IPCW (deb.)	0.951	0.0003	-0.00000649	0.0173	0.0174
	LTMLE	0.948	0.00029	0.00296	0.0168	0.0169
	Naive Cox		0.000291	-0.00158	0.017	
	ICE-IPCW		0.0003	-0.0000142	0.0173	
0	ICE-IPCW (deb.)	0.951	0.000336	-0.000497	0.0183	0.0184
	LTMLE	0.952	0.000316	-0.000303	0.0178	0.0178
	Naive Cox		0.000327	-0.00231	0.0179	
	ICE-IPCW		0.000336	-0.000491	0.0183	
0.3	ICE-IPCW (deb.)	0.95	0.000355	0.000395	0.0188	0.0188
	LTMLE	0.948	0.00034	-0.00228	0.0183	0.0183
	Naive Cox		0.000347	-0.00166	0.0185	
	ICE-IPCW		0.000354	0.000417	0.0188	

Table 6: Results for the case with varying time-varying confounding (vary β_A^y). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error and the estimator applied are provided.

β_L^y	Estimator	Coverage	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
-0.5	ICE-IPCW (deb.)	0.948	0.000221	0.000198	0.0149	0.0147
	LTMLE	0.952	0.000212	0.00185	0.0144	0.0144
	Naive Cox		0.000216	0.00119	0.0147	
	ICE-IPCW		0.00022	0.000194	0.0148	
0	ICE-IPCW (deb.)	0.948	0.000257	-0.0000468	0.016	0.016
	LTMLE	0.948	0.000249	0.00224	0.0156	0.0156
	Naive Cox		0.000248	-0.000156	0.0157	
	ICE-IPCW		0.000257	-0.000044	0.016	
0.5	ICE-IPCW (deb.)	0.951	0.0003	-0.00000649	0.0173	0.0174
	LTMLE	0.948	0.00029	0.00296	0.0168	0.0169
	Naive Cox		0.000291	-0.00158	0.017	
	ICE-IPCW		0.0003	-0.0000142	0.0173	

Table 7: Results for the case with varying time-varying confounding (vary β_L^y). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error and the estimator applied are provided.

β_A^L	Estimator	Coverage	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
-0.2	ICE-IPCW (deb.)	0.951	0.0003	-0.00000649	0.0173	0.0174
	LTMLE	0.948	0.00029	0.00296	0.0168	0.0169
	Naive Cox		0.000291	-0.00158	0.017	
	ICE-IPCW		0.0003	-0.0000142	0.0173	
0	ICE-IPCW (deb.)	0.951	0.000304	0.000162	0.0174	0.0176
	LTMLE	0.948	0.000298	0.00309	0.017	0.017
	Naive Cox		0.000297	-0.00165	0.0171	
	ICE-IPCW		0.000303	0.000153	0.0174	
0.2	ICE-IPCW (deb.)	0.949	0.000317	0.000196	0.0178	0.0178
	LTMLE	0.95	0.000305	0.00316	0.0172	0.0171
	Naive Cox		0.000305	-0.00171	0.0174	
	ICE-IPCW		0.000317	0.000198	0.0178	

Table 8: Results for the case with varying time-varying confounding (vary β_A^L). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error and the estimator applied are provided.

α_L	Estimator	Coverage	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
-0.2	ICE-IPCW (deb.)	0.951	0.0003	-0.00000649	0.0173	0.0174
	LTMLE	0.948	0.00029	0.00296	0.0168	0.0169
	Naive Cox		0.000291	-0.00158	0.017	
	ICE-IPCW		0.0003	-0.0000142	0.0173	
0	ICE-IPCW (deb.)	0.95	0.000292	-0.000207	0.0171	0.0171
	LTMLE	0.947	0.000283	0.00257	0.0166	0.0167
	Naive Cox		0.000284	-0.000892	0.0168	
	ICE-IPCW		0.000292	-0.000207	0.0171	
0.2	ICE-IPCW (deb.)	0.951	0.000281	-0.000305	0.0168	0.0169
	LTMLE	0.949	0.000276	0.00224	0.0164	0.0165
	Naive Cox		0.000275	-0.000257	0.0166	
	ICE-IPCW		0.000281	-0.000311	0.0168	

Table 9: Results for the case with varying time-varying confounding (vary α_L). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error and the estimator applied are provided.

Sample size – uncensored case

n	Estimator	Coverage	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
100	ICE-IPCW (deb.)	0.934	0.00325	0.000382	0.0571	0.0555
	ICE-IPCW		0.00316	0.000521	0.0562	
200	ICE-IPCW (deb.)	0.947	0.00152	-0.000339	0.039	0.039
	ICE-IPCW		0.00151	-0.000265	0.0388	
500	ICE-IPCW (deb.)	0.95	0.000599	0.000325	0.0245	0.0246
	ICE-IPCW		0.000598	0.000344	0.0245	
1000	ICE-IPCW (deb.)	0.952	0.000303	0.000142	0.0174	0.0174
	ICE-IPCW		0.000302	0.000136	0.0174	

Table 10: Results for varying sample size ($n \in \{100, 200, 500, 1000\}$). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error and the estimator applied are provided.

G.1.2 Censoring

ICE model	Estimator	Cov.	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
Scaled quasibinomial	ICE-IPCW (deb.)	0.952	0.000331	0.000188	0.0182	0.0184
Scaled quasibinomial	ICE-IPCW		0.00033	0.000171	0.0182	
Tweedie	ICE-IPCW (deb.)	0.953	0.00033	0.0000904	0.0182	0.0184
Tweedie	ICE-IPCW		0.000328	0.000288	0.0181	
Linear model	ICE-IPCW (deb.)	0.95	0.000335	-0.000294	0.0183	0.0184
Linear model	ICE-IPCW		0.000327	0.00113	0.018	

Table 11: The table shows the results for the censored case for $(\beta^y_A, \beta^y_L, \alpha_L, \lambda_c)$ = (-0.3, 0.5, -0.2, 0.0002). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error, the estimator (debiased ICE-IPCW and ICE-IPCW), and the pseudo-outcome model are provided

ICE model	Estimator	Cov.	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
Scaled quasibinomial	ICE-IPCW (deb.)	0.95	0.000388	-0.0000739	0.0197	0.0199
Scaled quasibinomial	ICE-IPCW		0.000387	-0.0000568	0.0197	
Tweedie	ICE-IPCW (deb.)	0.955	0.000381	0.000187	0.0195	0.0199
Tweedie	ICE-IPCW		0.000377	0.000596	0.0194	
Linear model	ICE-IPCW (deb.)	0.951	0.000382	-0.0000409	0.0195	0.0199
Linear model	ICE-IPCW		0.000375	0.00143	0.0193	

Table 12: The table shows the results for the censored case for $(\beta^y_A, \beta^y_L, \alpha_L, \lambda_c)$ = (-0.3, 0.5, -0.2, 0.0005). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error, the estimator (debiased ICE-IPCW and ICE-IPCW), and the pseudo-outcome model are provided

ICE model	Estimator	Cov.	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
Scaled quasibinomial	ICE-IPCW (deb.)	0.953	0.000435	-0.0000717	0.0209	0.0215
Scaled quasibinomial	ICE-IPCW		0.000434	0.0000358	0.0208	
Tweedie	ICE-IPCW (deb.)	0.953	0.000447	0.00000225	0.0212	0.0215
Tweedie	ICE-IPCW		0.000441	0.000636	0.021	
Linear model	ICE-IPCW (deb.)	0.956	0.000431	0.000211	0.0208	0.0215
Linear model	ICE-IPCW		0.000425	0.00176	0.0205	

Table 13: The table shows the results for the censored case for $(\beta^y_A, \beta^y_L, \alpha_L, \lambda_c)$ = (-0.3, 0.5, -0.2, 0.0008). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error, the estimator (debiased ICE-IPCW and ICE-IPCW), and the pseudo-outcome model are provided

ICE model	Estimator	Cov.	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
Scaled quasibinomial	ICE-IPCW (deb.)	0.952	0.000316	0.000438	0.0178	0.0179
Scaled quasibinomial	ICE-IPCW		0.000316	0.000438	0.0178	
Tweedie	ICE-IPCW (deb.)	0.951	0.000317	0.000434	0.0178	0.018
Tweedie	ICE-IPCW		0.000315	0.000464	0.0178	
Linear model	ICE-IPCW (deb.)	0.949	0.000323	0.00048	0.018	0.0179
Linear model	ICE-IPCW		0.000321	0.00154	0.0179	

Table 14: The table shows the results for the censored case for $(\beta^y_A, \beta^y_L, \alpha_L, \lambda_c)$ = (0, 0, 0, 0.0002). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error, the estimator (debiased ICE-IPCW and ICE-IPCW), and the pseudo-outcome model are provided

ICE model	Estimator	Cov.	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\text{SE}}$)
Scaled quasibinomial	ICE-IPCW (deb.)	0.953	0.000357	0.000352	0.0189	0.0193
Scaled quasibinomial	ICE-IPCW		0.000356	0.000385	0.0189	
Tweedie	ICE-IPCW (deb.)	0.954	0.000364	0.00027	0.0191	0.0194
Tweedie	ICE-IPCW		0.000361	0.000368	0.019	
Linear model	ICE-IPCW (deb.)	0.955	0.00036	0.000512	0.019	0.0194
Linear model	ICE-IPCW		0.000359	0.00173	0.0189	

Table 15: The table shows the results for the censored case for $(\beta^y_A, \beta^y_L, \alpha_L, \lambda_c)$ = (0, 0, 0, 0.0005). The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error, the estimator (debiased ICE-IPCW and ICE-IPCW), and the pseudo-outcome model are provided

ICE model	Estimator	Cov.	MSE	Bias	sd($\widehat{\Psi}_n$)	Mean($\widehat{\mathrm{SE}}$)
Scaled quasibinomial	ICE-IPCW (deb.)	0.955	0.000413	0.000149	0.0203	0.0209
Scaled quasibinomial	ICE-IPCW		0.000411	0.000241	0.0203	
Tweedie	ICE-IPCW (deb.)	0.958	0.000412	0.0000457	0.0203	0.0209
Tweedie	ICE-IPCW		0.000409	0.00025	0.0202	
Linear model	ICE-IPCW (deb.)	0.953	0.000421	0.00021	0.0205	0.0209
Linear model	ICE-IPCW		0.000421	0.00159	0.0205	

Table 16: The table shows the results for the censored case for $(\beta_A^y, \beta_L^y, \alpha_L, \lambda_c) = (0,\ 0,\ 0,\ 0.0008)$. The coverage, the mean squared error (MSE), average bias, standard deviation of the estimates, mean of the estimated standard error, the estimator (debiased ICE-IPCW and ICE-IPCW), and the pseudo-outcome model are provided

G.1.3 Boxplots

Varying effects (A on Y, L on Y, A on L, L on A) – uncensored case

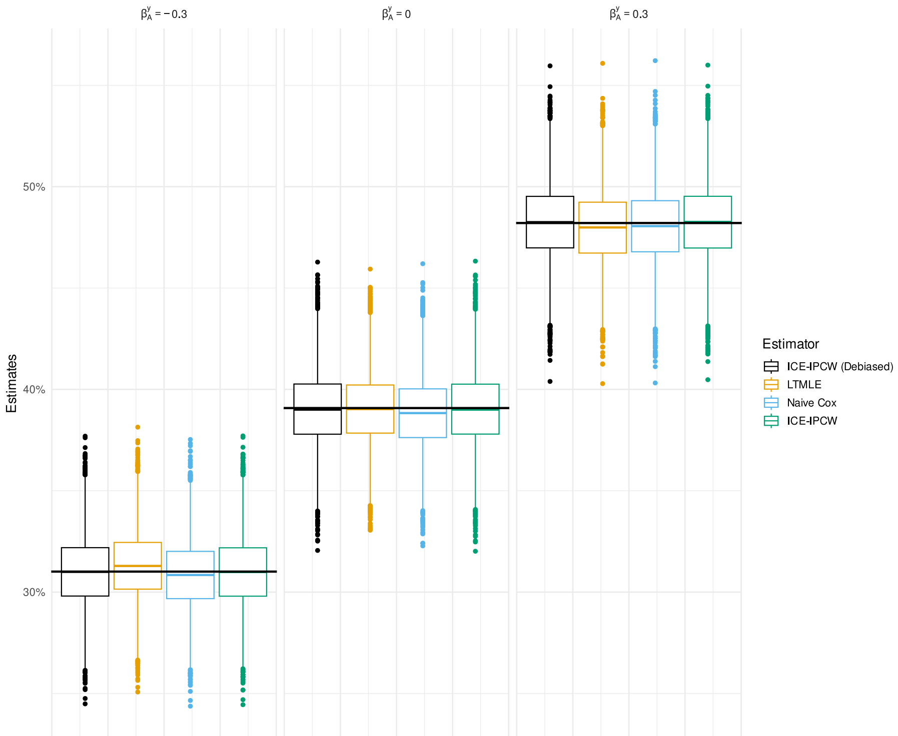

Figure 11: Boxplots of the estimates for each estimator in each parameter setting for the cases with varying effects of A on Y. The lines indicates the true value of the target parameter $\Psi_\tau^g(P)$.

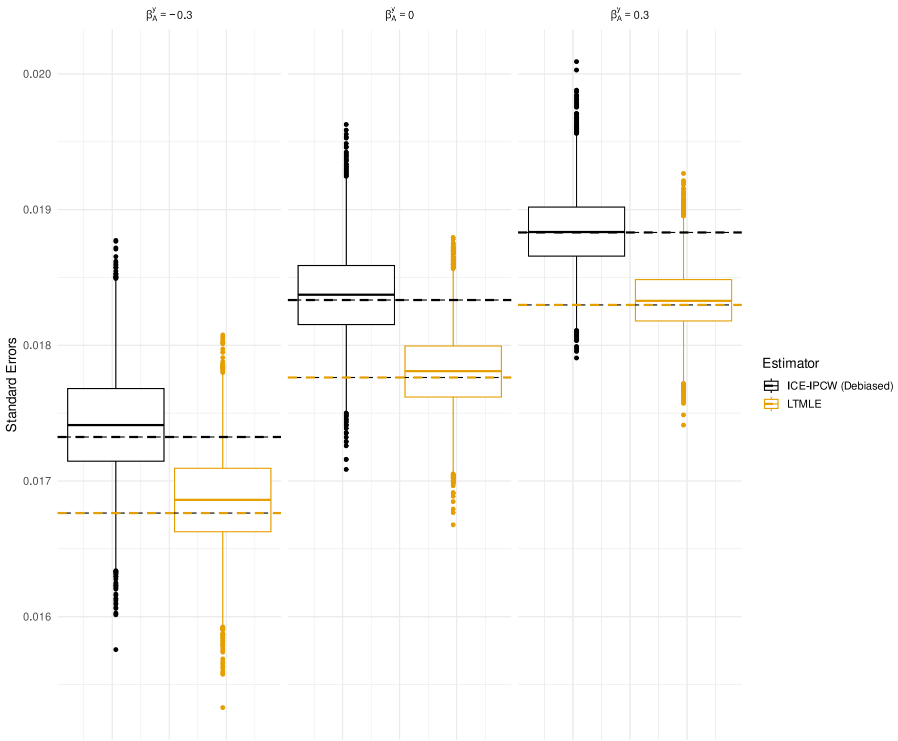

Figure 12: Boxplots of the standard errors for each estimator (LTMLE and debiased ICE-IPCW) in each parameter setting for the cases with varying effects of A on Y. The lines indicates the empirical standard error of the estimates for each estimator.

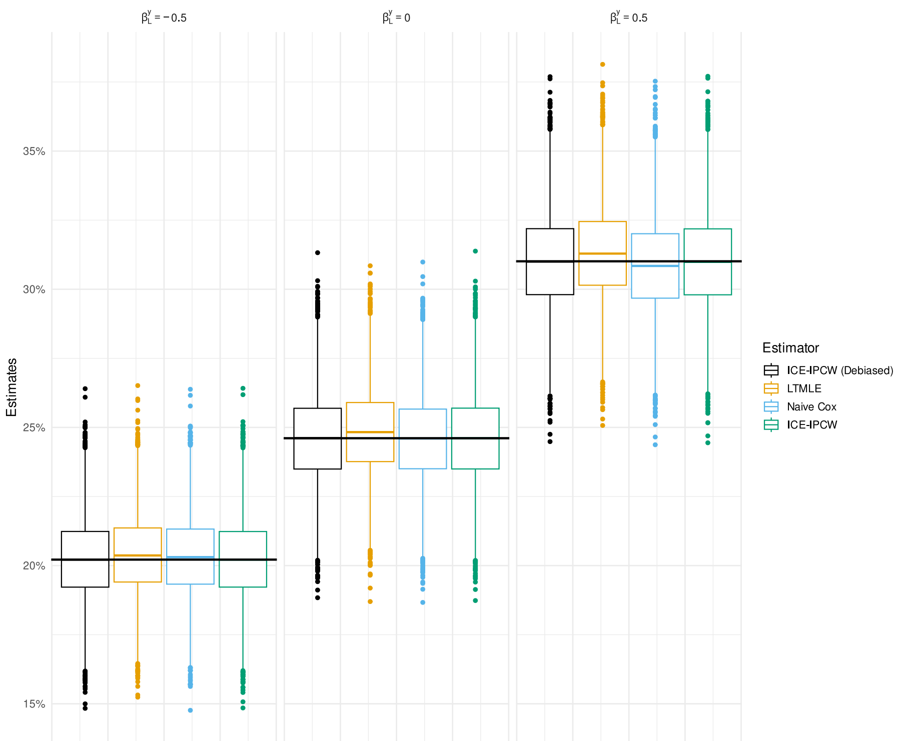

Figure 13: Boxplots of the estimates for each estimator in each parameter setting for the cases with varying effects of L on Y. The lines indicates the true value of the target parameter $\Psi^g_\tau(P)$.

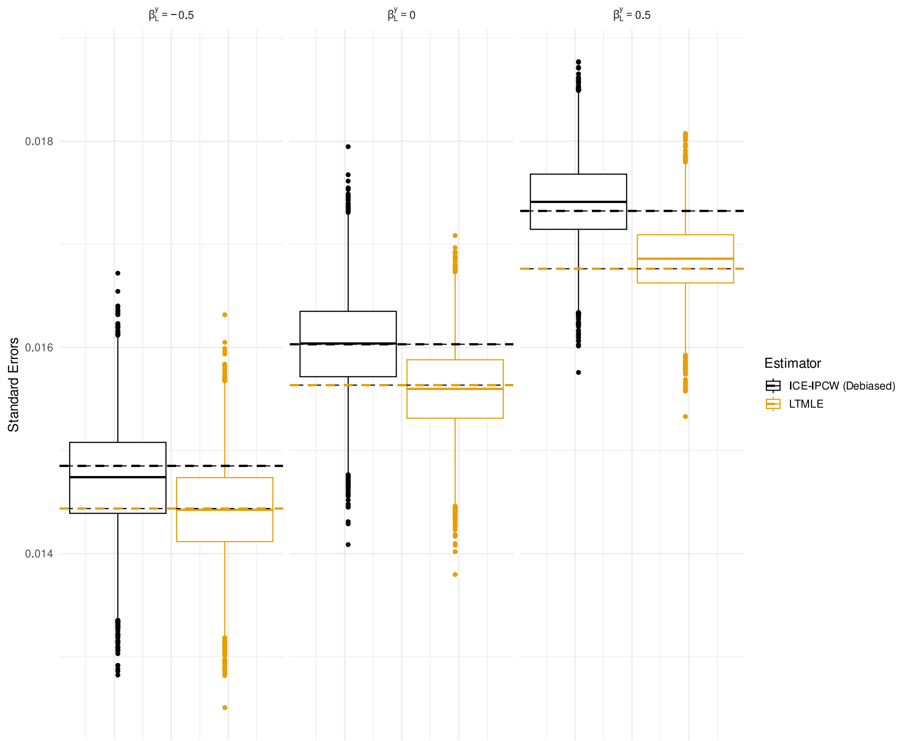

Figure 14: Boxplots of the standard errors for each estimator (LTMLE and debiased ICE-IPCW) in each parameter setting for the cases with varying effects of L on Y. The lines indicates the empirical standard error of the estimates for each estimator.

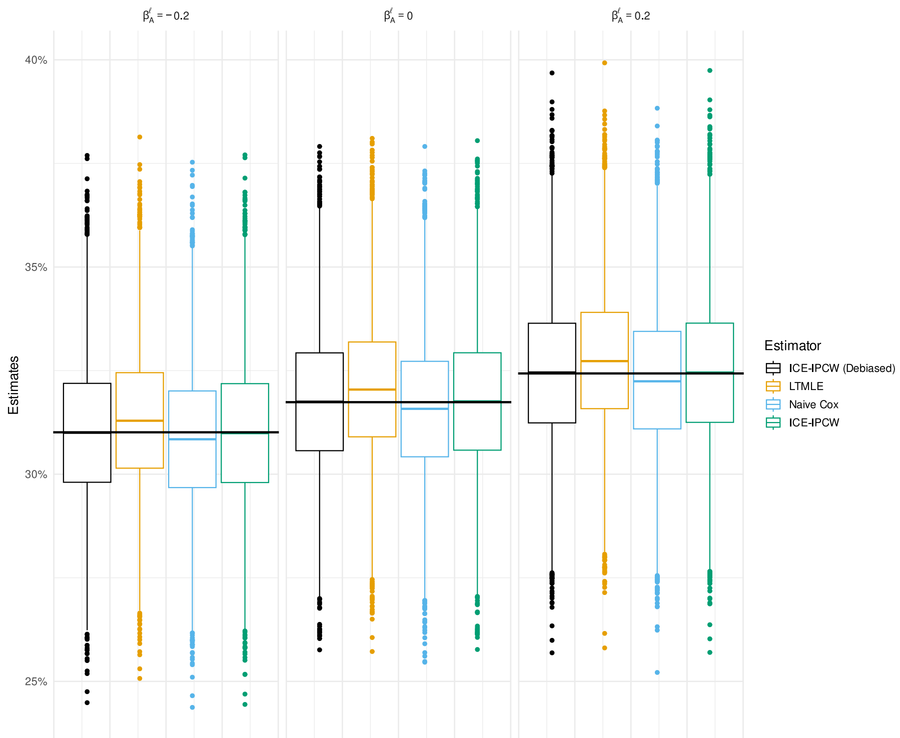

Figure 15: Boxplots of the estimates for each estimator in each parameter setting for the cases with varying effects of A on L. The lines indicates the true value of the target parameter $\Psi_\tau^g(P)$.

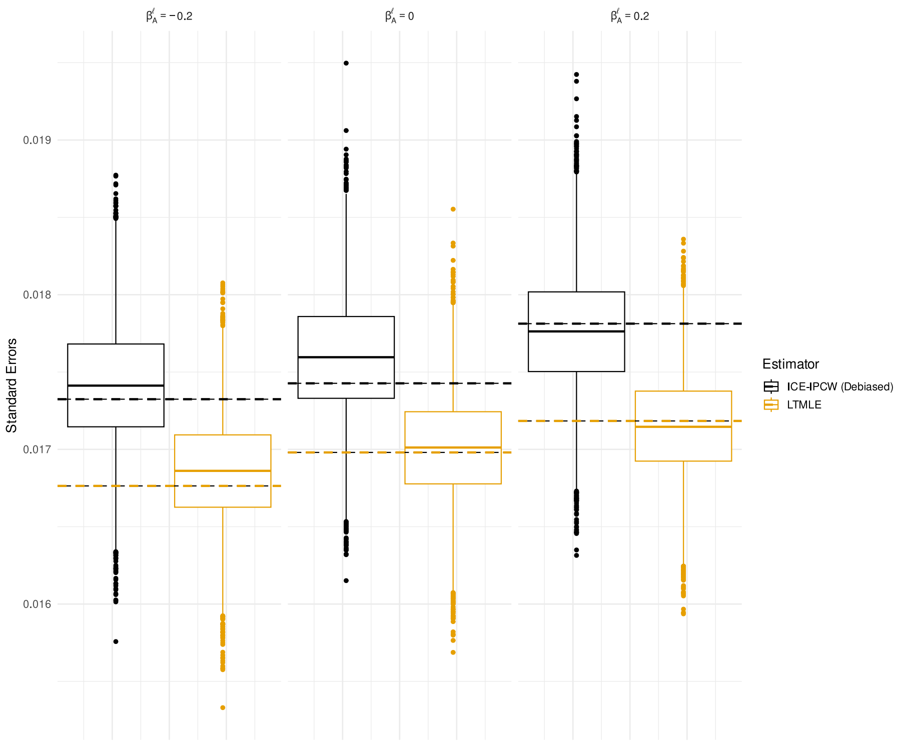

Figure 16: Boxplots of the standard errors for each estimator (LTMLE and debiased ICE-IPCW) in each parameter setting for the cases with varying effects of A on L. The lines indicates the empirical standard error of the estimates for each estimator.

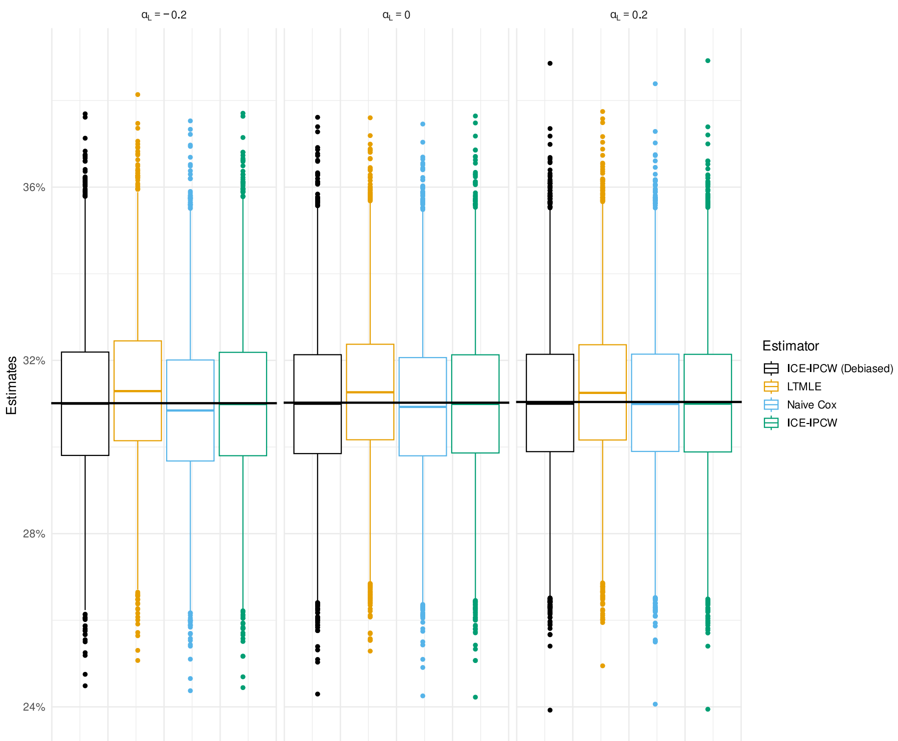

Figure 17: Boxplots of the estimates for each estimator in each parameter setting for the cases with varying effects of L on A. The lines indicates the true value of the target parameter $\Psi_\tau^g(P)$.

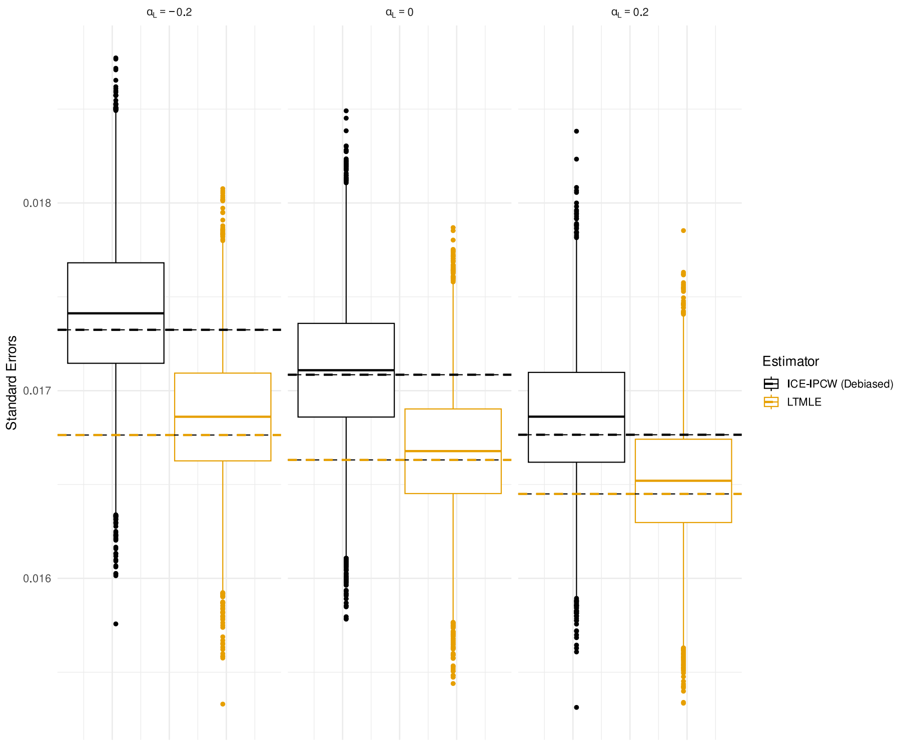

Figure 18: Boxplots of the standard errors for each estimator (LTMLE and debiased ICE-IPCW) in each parameter setting for the cases with varying effects of L on A. The lines indicates the empirical standard error of the estimates for each estimator.

Sample size – uncensored case

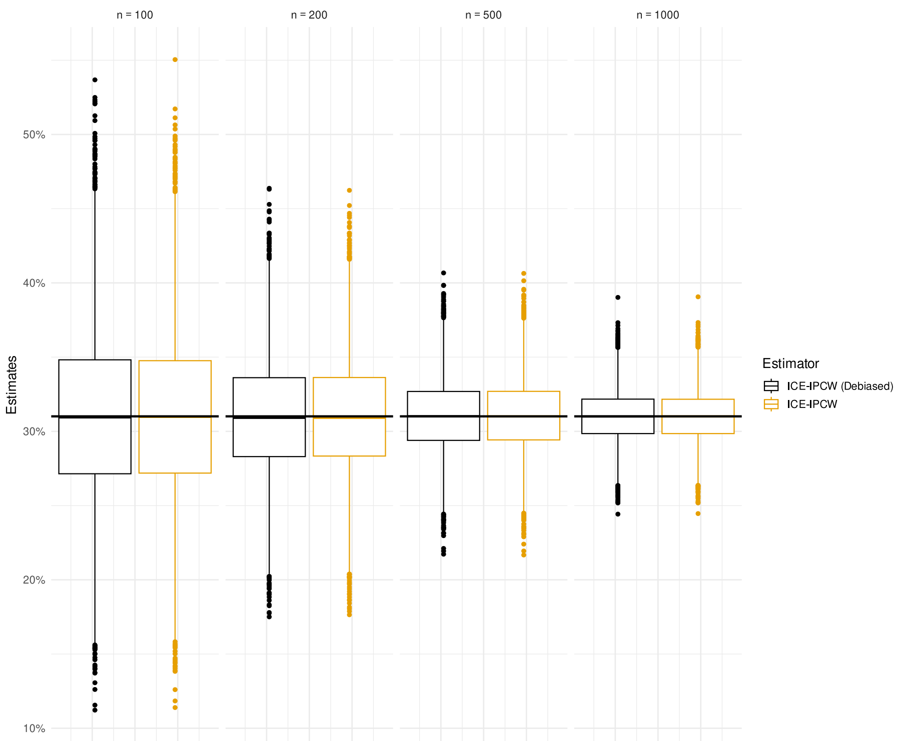

Figure 19: Boxplots of the estimates for each estimator with varying sample size ($n \in \{100, 200, 500, 1000\}$). The line indicates the true value of the target parameter $\Psi_\tau^g(P)$.

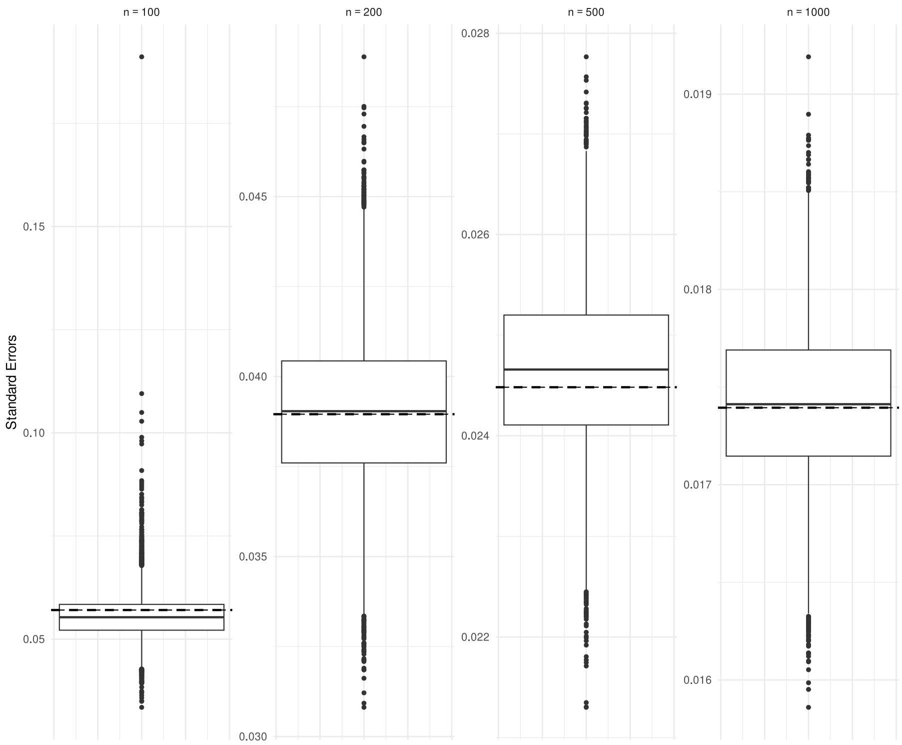

Figure 20: Boxplots of the standard errors with varying sample size ($n \in \{100, 200, 500, 1000\}$) for the debiased ICE-IPCW estimator The line indicates the empirical standard error of the estimates.

G.2 Discretizing time

We briefly illustrate how to discretize the data into discrete-time data consisting of K time points with a target parameter that is the interventional absolute risk of a specified event within time horizon τ, representing the usual longitudinal setting. To do so, suppose that we have observed the processes $(L(t))_{t\geq 0}$, $(A(t))_{t\geq 0}$, and $(N^y(t))_{t\geq 0}$ at the time points $t_k = k \times \frac{\tau}{K}$ for $k = 0, ..., K$. Then, we put

$$Y_k = N^y(t_k),$$
$$L_k = L(t_k),$$
$$A_k = A(t_k).$$

Our data set then consists of $(L_0, A_0, Y_1, L_1, A_1, ..., Y_{K-1}, L_{K-1}, A_{K-1}, Y_K)$, which may then be applied with a discrete-time longitudinal causal inference estimator such as LTMLE (van der Laan & Gruber (2012)).