# A Novel Approach to the Estimation of Causal Effects of Multiple Event Point Interventions in Continuous Time

**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 $\tau$ 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 will discuss a new sequential regression type estimator 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 machine learning methods from survival analysis and point processes, 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. In such cases, one may be interested in estimating the effects of initiating or adhering to treatment over time on a medical outcome such as the time to an event of interest.

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. MSMs typically require that data be recorded on a discrete time scale, capturing all relevant information available to the clinician at each treatment decision point and for the outcome.

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 results 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 advantage of using iterative regressions is that the resulting estimator will be less sensitive to/near practical positivity violations.

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 then allows us to apply simple regression techniques for the nuisance parameters. Moreover, our estimator addresses challenges associated with the high dimensionality of the target parameter by employing inverse probability of censoring weighting (IPCW). The distinction between event-based and time-based updating is illustrated in Figure 1 and Figure 2. To the best of our knowledge, no general estimation procedure has yet been proposed for the components involved in the efficient influence function.

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)).

Continuous-time methods for causal inference in event history analysis have also been explored by Røysland (2011) and Lok (2008). Røysland (2011) developed identification criteria using a formal 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.

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.

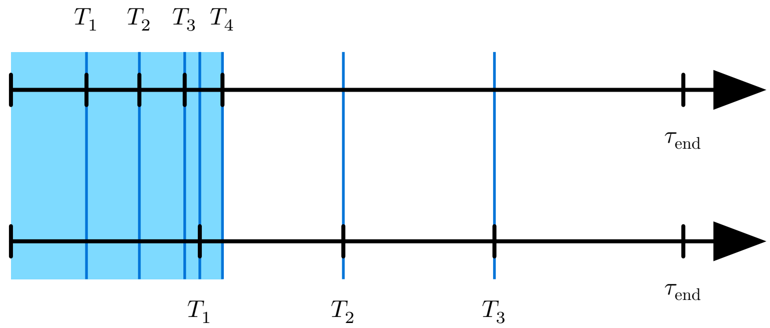


Figure 1: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let $t_1 < \ldots < t_m$ be all the event times in the sample. Let $P^{G^*}$ denote the interventional probability measure. Then, given $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{t_r}\right]$,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

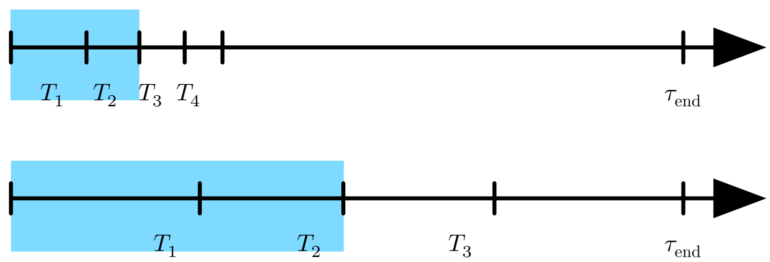


Figure 2: The figure illustrates the sequential regression approach proposed in this article. For each event number $k$ in the sample, we regress back on the history $\mathcal{F}_{T_{(k-1)}}$. Let $P^{G^*}$ denote the interventional probability measure. That is, given $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{T_{(k)}}\right]$, we regress back to $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{T_{(k-1)}}\right]$. In the figure, $k = 3$. The key difference is that we employ the stopping time $\sigma$-algebra $\mathcal{F}_{T_{(k)}}$ here instead of the filtration $\mathcal{F}_{t_r}$ which turns out to have a simpler representation.

# 2 Setting and Notation

Let $\tau_{\text{end}}$ 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 statistical experiment on which all processes and random variables are 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 $\sigma$-algebra corresponding to the baseline information. It is 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 $t \geq 0$.

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 $t \geq 0$. 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 $(N^\ell(t))_{t \geq 0}$, 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 $\Delta N^a(t) \neq 0$. Note that we allow, for practical reasons, some of the covariate values to change at the same time as the treatment values. This can occur if registrations occur only on a daily level if, for example, a patient visits the doctor, gets a blood test, and receives a treatment all on the same day. This means that we can practically assume that $\Delta N^a \Delta N^\ell \equiv 0$. For technical reasons and ease of notation, we shall assume that the number of jumps at time $\tau_{\text{end}}$ 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 the competing event $(N^d(t))_{t \geq 0}$ denote the counting processes for the event of interest and the competing event, respectively,

Thus, we have observations from a jump process $\alpha(t) = (N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t))$, 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 $\alpha$, that is $T_0 = 0$ and $T_{(k)} = \inf\{t > T_{(k-1)} \mid \alpha(t) \neq \alpha(T_{(k-1)})\} \in [0, \infty]$ 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 $\Delta N^x(T_{(k)}) = 1$. We let $T_{(k+1)} = \infty$ if $T_{(k)} = \infty$ or $\Delta_{(k-1)} \in \{y, d, c\}$. As is common in the point process literature, we define $\Delta_{(k)} = \emptyset$ if $T_{(k)} = \infty$ or $\Delta_{(k-1)} \in \{y, d, c\}$ for the empty mark.

We let $A(T_{(k)})$ $(L(T_{(k)}))$ be the treatment (covariate values) at the $k$'th event. If $T_{(k-1)} = \infty$, $\Delta_{(k-1)} \in \{y, d, c\}$, or $\Delta_{(k)} \in \{y, d, c\}$, we let $A(T_{(k)}) = \emptyset$ and $L(T_{(k)}) = \emptyset$. To the process $(\alpha(t))_{t \geq 0}$, we associate the corresponding random measure $N^\alpha$ on $(\mathbb{R}_+ \times (\{a, \ell, y, d, c\} \times \{0, 1\} \times \mathcal{L}))$ by

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

where $\delta_x$ 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 $N^\alpha$ (e.g., Theorem 2.5.10 of Last & Brandt (1995)). Thus, the random measure $N^\alpha$ carries the same information as the stochastic process $(\alpha(t))_{t \geq 0}$. This will be critical for dealing with right-censoring.

We observe $O = (T_{(K)}, \Delta_{(K)}, A(T_{(K-1)}), L(T_{(K-1)}), T_{(K-1)}, \Delta_{(K-1)}, \ldots, A(0), L(0)) \sim P \in \mathcal{M}$ where $\mathcal{M}$ is the statistical model, i.e., a set of probability measures and obtain a sample $O = (O_1, \ldots, O_n)$ 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.

| 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 $n = 3$ (3 observations). Here, the time-varying covariates only have dimension 1. 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 | $y$ | $\infty$ | Ø |
| 2 | 1 | 0 | Ø | Ø | 10 | $y$ | Ø | Ø | $\infty$ | Ø | $\infty$ | Ø |
| 3 | 3 | 1 | 4 | 1 | 2 | $\ell$ | 4 | 0 | 2.1 | $a$ | 5 | $y$ |

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

We work with the natural filtration $(\mathcal{F}_t)_{t\geq 0}$ generated by the random measure $N^\alpha$. This is needed to ensure the existence of compensators which 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 $\sigma$-algebra $\mathcal{F}_{T_{(k)}}$ – representing the information up to and including the $k$'th event – associated with stopping time $T_{(k)}$. We will interpret $\mathcal{F}_{T_{(k)}}$ as a random variable instead of a $\sigma$-algebra, whenever it is convenient to do so and also make the implicit assumption that whenever we condition on $\mathcal{F}_{T_{(k)}}$, we only consider the cases where $T_{(k)} < \infty$ and $\Delta_{(k)} \in \{a, \ell\}$.

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 $\Lambda_k^x\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right)$ be the cumulative cause-specific hazard measure (see e.g., Appendix A5.3 of Last & Brandt (1995)). Note that in many places, we will not distinguish between $\Lambda_k^x\left((0,t], \mathcal{F}_{T_{(k-1)}}\right)$ and $\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right)$. 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 $\tau$. We consider regimes which naturally act upon the treatment decisions at each visitation time but not the times at which the individuals visit the doctor. The treatment regime $g$ specifies for each event $k = 1, ..., K - 1$ with $\Delta_{(k)} = a$ (visitation time) the probability that a patient will remain treated until the next visitation time via $\pi_k^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)$ and at $k = 0$ the initial treatment probability $\pi_0^*(L(0))$.

For this, we define the likelihood ratio process as follows,

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

where $N_t = \sum_k \mathbb{1}\left\{T_{(k)} \leq t\right\}$ is random variable denoting the number of events up to time $t$. If we define the measure $P^{G^*}$ by the density,

$$\frac{\mathrm{d}P^{G^*}}{\mathrm{d}P}(\omega) = W^g(\tau_{\text{end}}, \omega), \quad \omega \in \Omega,$$

representing the interventional world in which the doctor assigns treatments according to the probability measure $\pi_k^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)$ for $k = 0, ..., K - 1$, then our target parameter is given by the mean interventional cumulative incidence function at time $\tau$,

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

where $N^y(t) = \sum_{k=1}^K \mathbb{1}\left\{T_{(k)} \leq t, \Delta_{(k)} = y\right\}$. In our application, $\pi_k^*$ 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. We that this corresponds to the intervention considered in Rytgaard et al. (2022), albeit we work with a specific *version* of the intervention.

Note that alternatively, 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 $\tau$, when the intervention enforces treatment as part of the $K - 1$ first events. We denote this target parameter by $\Psi_\tau^{g,K}$.

We now present a simple iterated representation of the data 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.

**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 $k > 0$. For $k = 0$, let $H_0 = L(0)$. Let $\bar{Q}^g_{K,\tau} : (a_k, h_k) \mapsto 0$ and recursively define for $k = K-1, ..., 1$,

$$Z^a_{k+1,\tau}(u) = \mathbb{1}\left\{T_{(k+1)} \leq u, T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\right) \bar{Q}^g_{k+1,\tau}\left(\tau, A\left(T_{(k)}\right), H_{k+1}\right)$$
$$+\mathbb{1}\left\{T_{(k+1)} \leq u, T_{(k+1)} < \tau, \Delta_{(k+1)} = a\right) \bar{Q}^g_{k+1,\tau}(\tau, 1, H_{k+1}) + \mathbb{1}\left\{T_{(k+1)} \leq u, \Delta_{(k+1)} = y\right),$$

and

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

for $u \leq \tau$. Then,

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

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

Throughout the, we will use the notation $\bar{Q}^g_{k,\tau}(u)$ to denote the value of $\bar{Q}^g_{k,\tau}\left(u, A\left(T_{(k)}\right), H_k\right)$ and $\bar{Q}^g_{k,\tau}$ to denote $\bar{Q}^g_{k,\tau}\left(\tau, A\left(T_{(k)}\right), H_k\right)$.

$\bar{Q}^g_{k,\tau}$ represents the counterfactual probability of the primary event occuring at or before time $\tau$ given the history up to and including the $k$'th event, among the people who are at risk of the event before time $\tau$ after $k$ events. Equation (3.2) then suggests that we can estimate $\bar{Q}^g_{k-1,\tau}$ via $\bar{Q}^g_{k,\tau}$: 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 $A(\bar{T}_k)$ 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.

# 4 Censoring

In this section, we allow for right-censoring. That is, we introduce a right-censoring time $C > 0$ at which we stop observing the multivariate jump process $\alpha$. Let $N^c$ be the censoring process given by $N^c(t) = \mathbb{1}\{C \leq t\}$.

We will introduce the notation necessary to discuss the algorithm for the the ICE-IPCW estimator in Section 4.1 and later discuss the assumptions necessary for consistency of the ICE-IPCW estimator in Section 4.2. In the remainder of the paper, we will assume that $C \neq T_{(k)}$ for all $k$ with probability 1.

We can 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

$$
\begin{aligned}
\bar{T}_{(k)} &= \begin{cases} T_{(k)} \text{ if } C > T_{(k)} \\ C \text{ if } C \leq T_{(k)} \text{ and } T_{(k-1)} > C \\ \infty \text{ otherwise} \end{cases} \\
\bar{\Delta}_{(k)} &= \begin{cases} \Delta_{(k)} \text{ if } C > T_{(k)} \\ \text{c if } C \leq T_{(k)} \text{ and } \bar{\Delta}_{(k-1)} \neq c \\ \emptyset \text{ otherwise} \end{cases} \\
A(\bar{T}_k) &= \begin{cases} A\left(T_{(k)}\right) \text{if } C > T_{(k)} \\ \emptyset \text{ otherwise} \end{cases} \\
L(\bar{T}_k) &= \begin{cases} L\left(T_{(k)}\right) \text{ if } C > T_{(k)} \\ \emptyset \text{ otherwise} \end{cases}
\end{aligned}
\tag{4.1}
$$

for $k = 1, ..., K$, and let $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}$, for now, heuristically be defined by

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

defining the observed history up to and including the $k$'th event. Thus $O = \left(\bar{T}_{(1)}, \bar{\Delta}_{(1)}, A(\bar{T}_1), L(\bar{T}_1), ..., \bar{T}_{(K)}, \bar{\Delta}_{(K)}, A(\bar{T}_K), L(\bar{T}_K)\right)$ is the observed data and a sample consists of $O = (O_1, ..., O_n)$ for $n$ independent and identically distributed observations with $O_i \sim P$.

Define $\tilde{\Lambda}^c_k\left(\mathrm{d}t, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)$ as the cause-specific cumulative hazard measure for censoring of the $k$'th event given the observed history $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}$ and the corresponding censoring survival functions $\tilde{S}^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) = \prod_{s \in (T_{(k-1)}, t]} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\right)$, where $\prod_{s \in (0,t]}$ is the product integral over the interval $(0, t]$ (Gill & Johansen (1990)).

## 4.1 Algorithm for ICE-IPCW Estimator

In this section, we present an algorithm for the ICE-IPCW estimator. 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.

Let $O_i = \left(\bar{T}_{(K),i}, \bar{\Delta}_{(K),i}, \bar{A}_{(K-1),i}, L_{(K-1),i}, \bar{T}_{(K-1),i}, \bar{\Delta}_{(K-1),i}, ..., \bar{A}_{0,i}, L_{0,i}\right)$ be the observed data for the $i$'th individual. We suppose that we are given an estimator of the censoring compensator $\hat{\Lambda}^c$. In particular, for $\bar{Q}^g_{k,\tau}, k = 0, ..., K-1$, we start the algorithm at $k = K-1$ by calculating $\hat{S}^c\left(\bar{T}_{(k),i} - \mid \bar{A}_{(k-1),i}, H_{k-1,i}\right) = \prod_{s \in \left(\bar{T}_{(k-1),i}, \bar{T}_{(k),i}-\right)} \left(1 - \hat{\Lambda}^c_i(s)\right)$. Given an estimator of $\bar{Q}^g_{k+1,\tau}$ denoted by $\hat{\bar{Q}}^g_{k+1,\tau}$, we then estimate the pseudo-outcome $\hat{Z}^a_{k,\tau,i}$ as follows

- If $\bar{\Delta}_{(k),i} = y$, we calculate $\hat{Z}^a_{k,\tau,i} = \frac{1}{\hat{S}^c\left(\bar{T}_{(k),i}- \mid \bar{A}_{(k-1),i}, H_{k-1,i}\right)} \mathbb{1}\left\{\bar{T}_{(k),i} \leq \tau\right\}$.
- If $\bar{\Delta}_{(k),i} = a$, evaluate $\hat{\bar{Q}}^g_{k+1,\tau}(1, H_{k,i})$ and calculate $\hat{Z}^a_{k,i} = \frac{1}{\hat{S}^c_k\left(\bar{T}_{(k),i}- \mid \bar{A}_{(k-1),i}, H_{k-1,i}\right)} \mathbb{1}\left\{\bar{T}_{(k),i} < \tau\right\} \hat{\bar{Q}}^g_{k+1,\tau}(1, \bar{H}_{k,i})$.
- If $\bar{\Delta}_{(k),i} = \ell$, evaluate $\hat{\bar{Q}}^g_{k+1,\tau}\left(\bar{A}_{(k-1),i}, H_{k,i}\right)$, and calculate $\hat{Z}^a_{k,i} = \frac{1}{\hat{S}^c_k\left(\bar{T}_{(k),i}- \mid \bar{A}_{(k-1),i}, H_{k-1,i}\right)} \mathbb{1}\left\{\bar{T}_{(k),i} < \tau\right\} \hat{\bar{Q}}^g_{k+1,\tau}\left(\bar{A}_{(k-1),i}, \bar{H}_{k,i}\right)$.

Then regress $\hat{Z}^a_{k,i}$ on $\left(\bar{A}_{(k-1),i}, H_{k-1,i}\right)$ for the observations with with $\bar{T}_{(k-1),i} < \tau$ and $\bar{\Delta}_{(k-1),i} \in \{a, \ell\}$ to obtain a prediction function $\hat{\bar{Q}}^g_{k,\tau}$. At $k = 0$, we compute $\hat{\Psi}_n = \frac{1}{n} \sum_{i=1}^n \hat{\bar{Q}}^g_{0,\tau}(1, L_i(0))$.

We mention how one may obtain an estimator of the censoring compensator, but this is a wider topic that we will not concern ourselves with here. We provide a model for the censoring that can provide estimates of the cause-specific hazard measure $\frac{1}{P\left(\bar{T}_{(k)} \geq t \mid A(\bar{T}_{k-1}), \bar{H}_{k-1}\right)} P\left(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid A(\bar{T}_{k-1}), \bar{H}_{k-1}\right)$, which is always estimable from observed data. Then, regress $\bar{E}_{(k),i} = \bar{T}_{(k),i} -$

$\bar{T}_{(k-1),i}$, known as the $k$'th *interarrival* time, with the censoring as the cause of interest on $\left(\bar{A}_{(k-1),i}, \bar{H}_{k-1,i}\right)$ among the patients who are still at risk after $k-1$ events, that is for $i$ with $\bar{\Delta}_{k-1,i} \in \{a, \ell\}$ if $k > 1$ and otherwise all $i = 1, ..., n$. This gives an estimator of the cause-specific cumulative hazard function $\hat{\Lambda}_k^c$ and provides an estimator of the compensator as follows

$$\hat{\Lambda}_i^c(t) = \sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1),i} < t \leq \bar{T}_{(k),i}\right\} \hat{\Lambda}_k^c\left(t - \bar{T}_{(k-1),i} \mid \bar{A}_{(k-1),i}, \bar{H}_{(k-1),i}\right).$$

## 4.2 Consistency of the ICE-IPCW Estimator

Now let $T^e$ further denote the (uncensored) terminal event time given by

$$T^e = \inf_{t>0}\{N^y(t) + N^d(t) = 1\}.$$

and let $\beta(t) = (\alpha(t), N^c(t))$ be the fully observable multivariate jump process in $[0, \tau_{\text{end}}]$. We assume now that we are working in the canonical setting with $\beta$ and not $\alpha$.

Then, we observe the trajectories of the process given by $t \mapsto N^\beta(t \wedge C \wedge T^e)$ and the observed filtration is given by $\mathcal{F}_t^{\tilde{\beta}} = \sigma(\beta(s \wedge C \wedge T^e) \mid s \leq t)$. The observed data is then given by Equation (4.1). Abusing notation a bit, we see that for observed histories, we have $\mathcal{F}_{T_{(k)}} = \mathcal{F}_{\bar{T}_{(k)}}^{\tilde{\beta}}$ if $\bar{\Delta}_{(k)} \neq c$.

We posit specific conditions in Theorem 2 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). A simple, sufficient condition for this to hold is e.g., that $C \perp \mathcal{F}_{T_{(K)}}$. Note that if compensator of the (observed) censoring process is absolutely continuous with respect to the Lebesgue measure, then 1. of Theorem 2 is satisfied. Our second condition in Theorem 2 is a positivity condition, ensuring that the conditional expectations are well-defined.

**Theorem 2**: Assume that the compensator $\Lambda^\alpha$ of $N^\alpha$ with respect to the filtration $\mathcal{F}_t^\beta$ is also the compensator with respect to the filtration $\mathcal{F}_t$. If

1. $\Delta\tilde{\Lambda}_k^c\left(t \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}\right) + \sum_x \Delta\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.} \Rightarrow \Delta\tilde{\Lambda}_{k+1}^c\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.} \vee$ $\sum_x \Delta\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.}.$
2. $\tilde{S}^c\left(t \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}\right) > \eta$ for all $t \in (0, \tau]$ and $k \in \{1, ..., K\}$ $P$-a.s. for some $\eta > 0$.

Let

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

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

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

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

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

# 5 Efficient inference

In this section, we derive the efficient influence function for $\Psi_\tau^g$. The overall objective is to conduct inference for this parameter. In particular, if $\hat{\Psi}_n$ is asymptotically linear at $P$ with influence function $\varphi_\tau^*(P)$,

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

then $\hat{\Psi}_n$ is regular and (locally) nonparametically efficient (Chapter 25 of van der Vaart (1998)). In this case, one can construct confidence intervals and hypothesis tests based on estimates of the influence function. Therefore, our goal is to construct an asymptotically linear estimator of $\Psi_\tau^g$ with influence function $\varphi_\tau^*(P)$.

The efficient influence function in the nonparametric setting enables the use of machine learning methods to estimate the nuisance parameters under certain regularity conditions to provide inference for the target parameter. To achieve this, we debias our initial ICE-IPCW estimator through double/debiased machine learning (Chernozhukov et al. (2018)), obtaining a one-step estimator for given estimators of the nuisance parameters appearing in the efficient influence function,

$$\hat{\Psi}_n = \hat{\Psi}_n^0 + \mathbb{P}_n \varphi_\tau^*(\cdot; \hat{\eta}),$$

where $\hat{\Psi}_n^0$ is the initial estimator, i.e., the ICE-IPCW estimator, and $\hat{\eta}$ is a collection of estimates for the nuisance parameters appearing in Equation (5.1).

Under certain regularity conditions, this estimator is asymptotically linear at $P$ with influence function $\varphi_\tau^*(\cdot; P)$; however, such conditions are not the focus of this paper. We have provided ways to estimate all the nuisance parameters appearing in the efficient influence function except $\pi_k$. These may either be estimated by a direct regression procedure or alternatively be estimated by estimating the compensators of the counting processors $N^{a1}$ 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) compensators. We only consider the first of these options. The latter choice, however, makes traditional intensity modeling methods applicable.

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 Gateaux derivative of the target parameter. Note that this does not constitute a rigorous proof that Equation (5.1) is the efficient influence function, but rather a heuristic argument.

We note the close resemblance of Equation (5.1) 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 (5.1).

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 D.1) when the compensators are absolutely continuous with respect to the Lebesgue measure and the assumption that $\Delta L(t) = 0$ whenever $\Delta N^a(t) = 1$.

Of separate interest is Theorem 4 which shows that we can adaptively select $K$ based on the observed data. For instance, we may pick $K^* = k$ such that $k$ is the largest integer such that there are at least 40 observations fulfilling that $\bar{T}_{(k-1)} < \tau$. Doing so will ensure that we will not have to estimate the terms in Equation (5.1) and the corresponding iterative regressions for which there is very little data. Therefore, our target parameter is instead

$$\Psi_\tau^{g,K^*}(P) = \mathbb{E}_P\left[\widetilde{N}^y(\tau) W^g\left(\tau \wedge \bar{T}_{(K^*)}\right)\right].$$

It holds that $\Psi_\tau^{g,K^*}(P) = \Psi_\tau^g(P)$ if $K^* = K_{\text{lim}}$. Noting that we can pool the last $K_{\text{lim}} - K^*$ events into a single terminal event, the theory discussed thus far can also be applied to the target parameter $\Psi_\tau^{g,K^*}(P)$. In practice, this means that we can essentially ignore the last $K_{\text{lim}} - K^*$ events; however, the latter will come at the cost of some finite sample size bias.

**Theorem 3** (Efficient influence function): Let for each $P \in \mathcal{M}$, $\tilde{\Lambda}_k^c\left(t \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}; P\right)$ be the corresponding cause-specific cumulative hazard function for the observed censoring for the $k$'th event. Suppose that there is a universal constant $C > 0$ such that $\tilde{\Lambda}_k^c\left(\tau_{\text{end}} \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}; P\right) \leq C$ for all $k = 1, ..., K$ and every $P \in \mathcal{M}$. The Gateaux derivative is then given by

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

where $\widetilde{M}^c(t) = \widetilde{N}^c(t) - \tilde{\Lambda}^c(t)$. Here $\widetilde{N}^c(t) = \mathbb{1}\{C \leq t, T^e > t\} = \sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k)} \leq t, \bar{\Delta}_{(k)} = c\right\}$ is the censoring counting process, $\tilde{\Lambda}^c(t) = \sum_{k=1}^K \mathbb{1}\left\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\right\} \tilde{\Lambda}_k^c\left(t \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}\right)$ is the cumulative censoring hazard process given in Section 4, and $S\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = \prod_{s \in (0,t]}\left(1 - \sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$.

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

**Theorem 4** (Adaptive selection of K): Let $K_{nc} = \max_{v: \sum_{i=1}^n \mathbb{1}\{N_{\tau i}(\tau) \geq v\} > c}$ denote the maximum number of events with at least $c$ people at risk. Suppose that we have the decomposition of the estimator $\hat{\Psi}_n$, such that

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

Suppose that there is a number $K_{\text{lim}} \in \mathbb{N}$, such that $P(N_\tau = K_{\text{lim}}) > 0$, but $P(N_\tau > K_{\text{lim}}) = 0$. 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 F.1). □

# 6 Simulation study

We consider 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. In the censored setting, we do not consider the estimation of the censoring martingale due to the computational difficulties and leave this as a future research topic. The second overall objective is to compare with existing methods in causal inference, such as discrete time methods (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 $\tau_{\text{end}} = 900 \text{ days}$ or until death. During follow-up, patients may experience (at most) one stroke, stop treatment (irreversibly), and die, that is $N^x(t) \leq 1$ for $x = a, \ell, y$. The primary outcome is the *risk of death within* $\tau = 720$ *days*. We simulate baseline data as $\text{age} \sim \text{Unif}(40, 90)$, $L(0) = 0$, and $A(0) = 1$. To simulate the time-varying data, we generate data according to the following intensities

$$
\begin{aligned}
\Lambda^\alpha(\mathrm{d}(t, x, a, \ell)) = \mathbb{1}\{t \leq T^e \wedge \tau_{\text{end}}\}\big(&\delta_y(x)\lambda^y \exp\big(\beta^y_{\text{age}}\,\text{age} + \beta^y_A A(t-) + \beta^y_L L(t-)\big) \\
&+ \mathbb{1}\big\{N^\ell(t-) = 0\big\}\delta_\ell(x)\delta_1(\ell)\lambda^\ell \exp\big(\beta^\ell_{\text{age}}\,\text{age} + \beta^\ell_A A(t-)\big) \\
&+ \mathbb{1}\{N^a(t-) = 0\}\delta_a(x)\big(\big(1 - \mathbb{1}\big\{N^\ell(t-) = 0\big\}\big)\gamma_0 \exp\big(\gamma_{\text{age}}\,\text{age}\big) + \mathbb{1}\big\{N^\ell(t-) = 0\big\}h_z(t; 360)\big) \\
&\quad \times \big(\delta_1(a)\pi(t \mid \mathcal{F}_{t-}) + \delta_0(a)(1 - \pi(t \mid \mathcal{F}_{t-}))\big)\big),
\end{aligned}
\tag{6.1}
$$

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 $\pi(t \mid \mathcal{F}_{t-}) = \text{expit}\,(\alpha_0 + \alpha_{\text{age}}\,\text{age} + \alpha_L L(t-))$ is the treatment assignment probability. Our intervention is $\pi^*(t \mid \mathcal{F}_{t-}) = 1$ which corresponds to sustained treatment throughout the follow-up period. Note that Equation (6.1) states that the intensities for $N^\ell$ 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, 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 $\alpha_L$ 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. 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 $\pi_k\Big(t, \mathcal{F}_{T_{(k-1)}}\Big)$ 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. The censoring variable is simply generated as $C \sim \text{Exp}(\lambda_c)$ and we vary the degree of censoring $\lambda_c \in \{0.0002, 0.0005, 0.0008\}$. 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. A 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** | $\alpha_0$ | $\alpha_{\text{age}}$ | $\alpha_{\text{L}}$ | $\beta^{y}_{\text{age}}$ | $\beta^{\ell}_{\text{age}}$ | $\beta^{y}_{A}$ | $\beta^{\ell}_{A}$ | $\beta^{y}_{\text{L}}$ | $\lambda^{y}$ | $\lambda^{\ell}$ | $\gamma_{\text{age}}$ | $\gamma_0$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Values (varying effects)** | 0.3 | 0.02 | $\mathbf{-0.2}$, $\underline{0}$, 0.2 | 0.025 | 0.015 | $\mathbf{-0.3}$, 0, 0.3 | $\mathbf{-0.2}$, 0, 0.2 | $-0.5$, $\underline{0}$, **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 study. Each value is varied, holding the others fixed. The values with bold font correspond to the values used when fixed. The cases with no effect of time-varying confounders are marked with an underline.

## 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. The standard errors obtained from the debiased procedure also appear slightly more biased than the standard errors obtained from the LTMLE procedure, but this difference may be negligible. Note that the choice of nuisance parameter model for the iterative regressions is misspecified, so we may encounter bias in the standard errors but do not see substantial bias in the estimates as the method is doubly robust.

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 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.

| $\beta_A^y$ | 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 confounding.

| $\beta_A^y$ | $\alpha_L$ | 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 confounding.

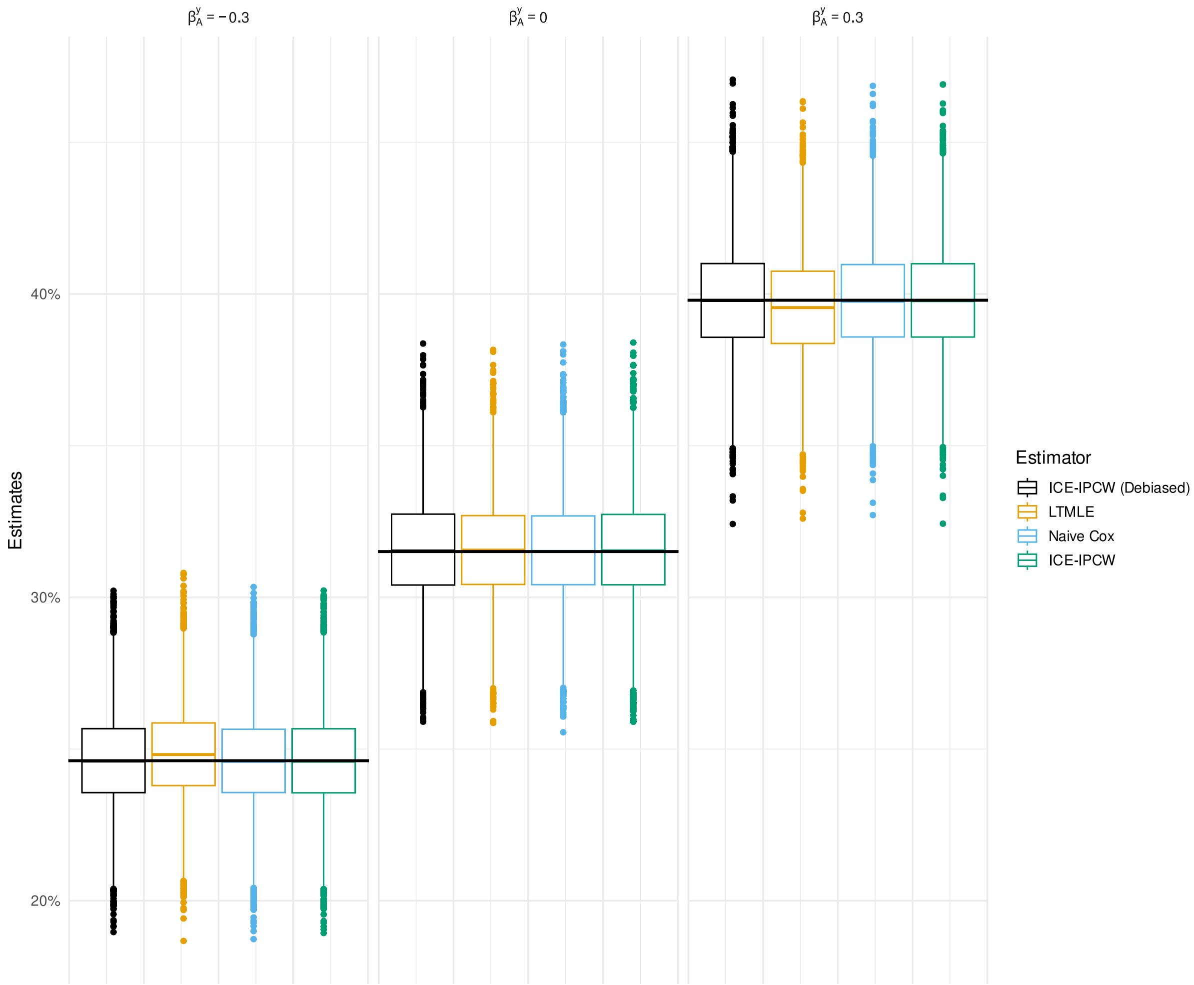


Figure 3: Boxplots of the results for the case without time confounding. The lines indicates the true value of the parameter.

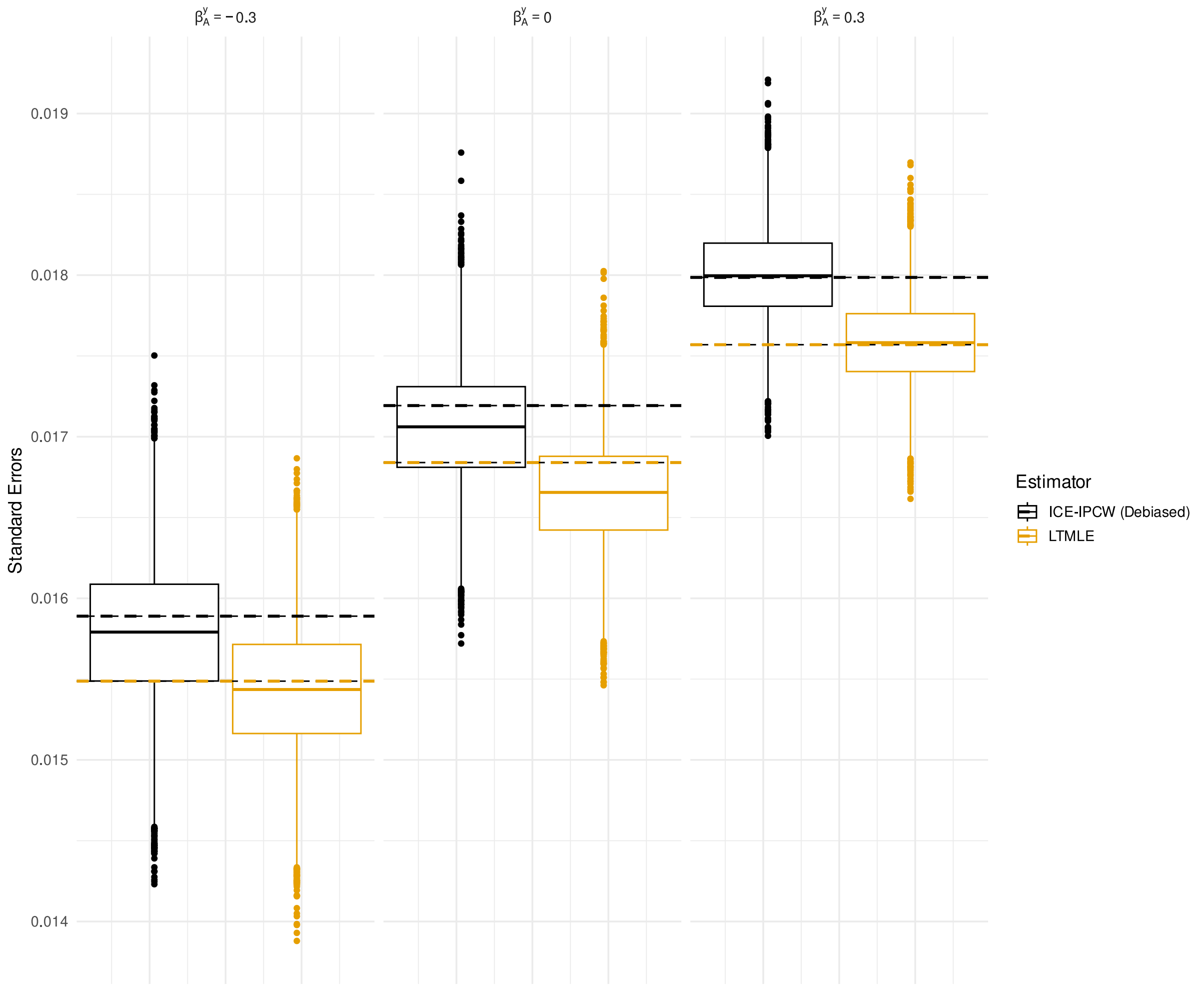


Figure 4: Boxplots of the standard errors for the case without time confounding. The red line indicates the empirical standard error of the estimates for each estimator.

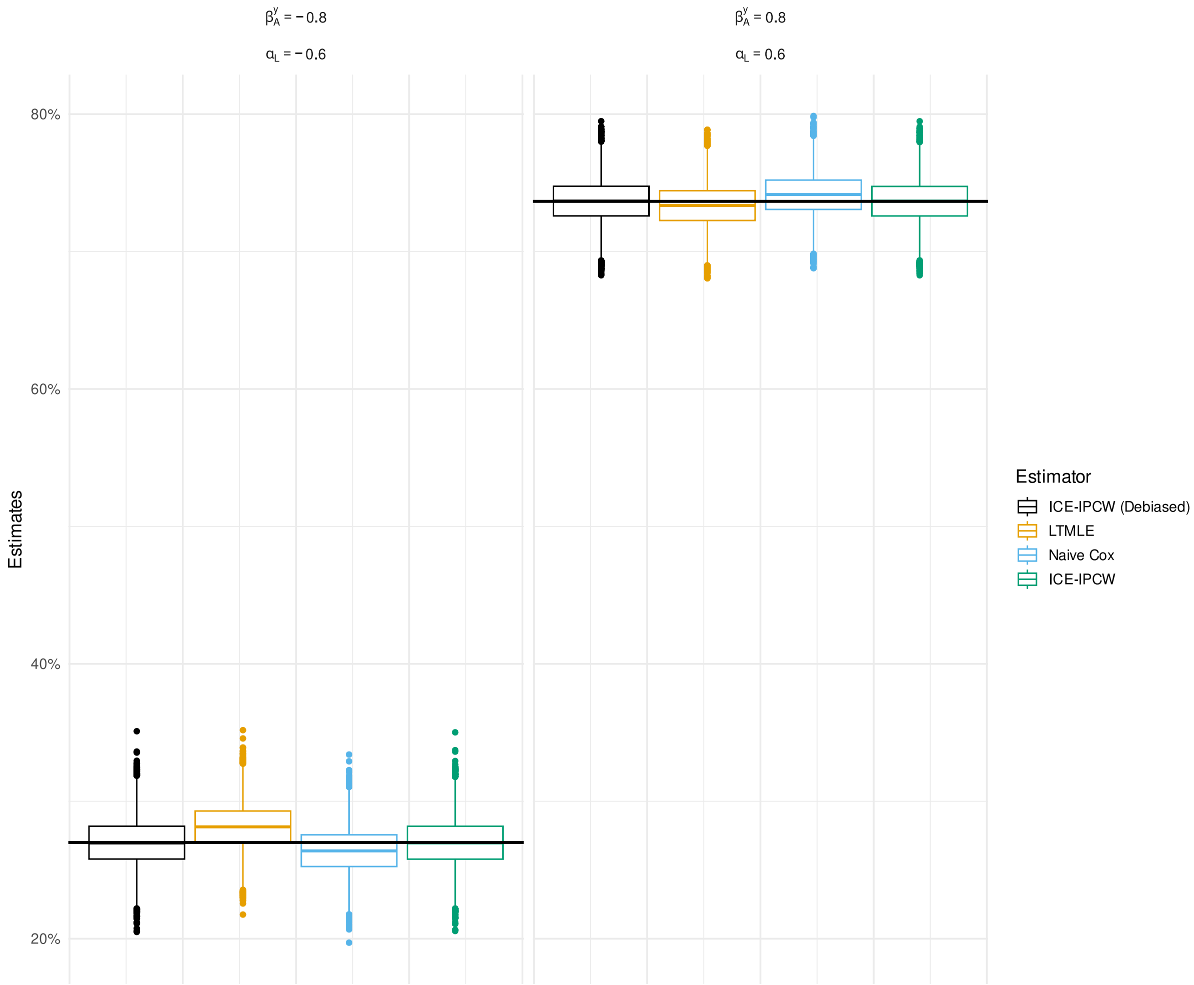


Figure 5: Boxplots of the results for the case with strong time confounding. The lines indicates the true value of the parameter.

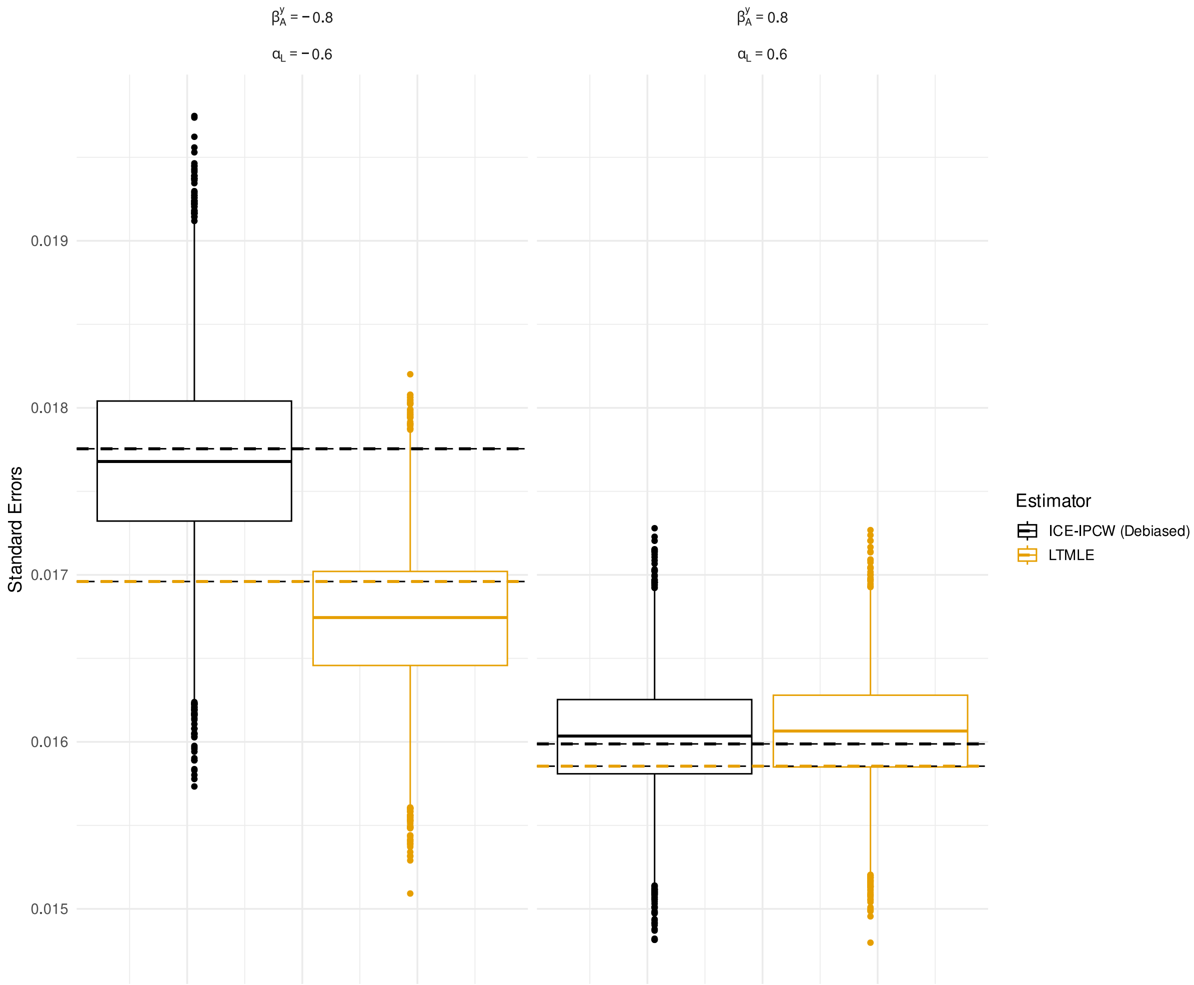


Figure 6: Boxplots of the standard errors for the case with strong time confounding. The red line indicates the empirical standard error of the estimates for each estimator.

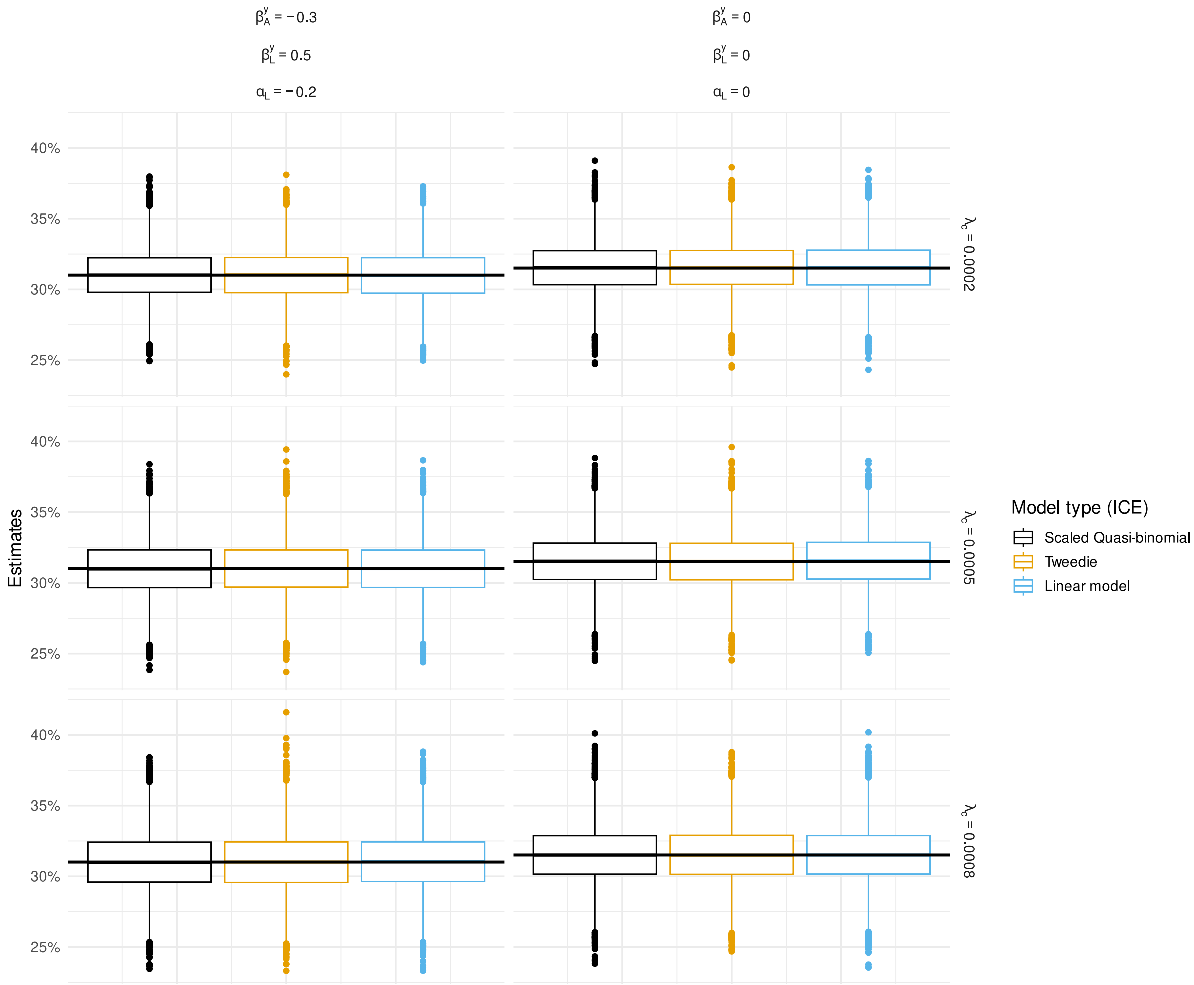


Figure 7: Boxplots of the results for the case with censoring. Different degrees of censoring are considered as well different model types for the pseudo-outcomes. Only the debiased ICE-IPCW estimator is shown.

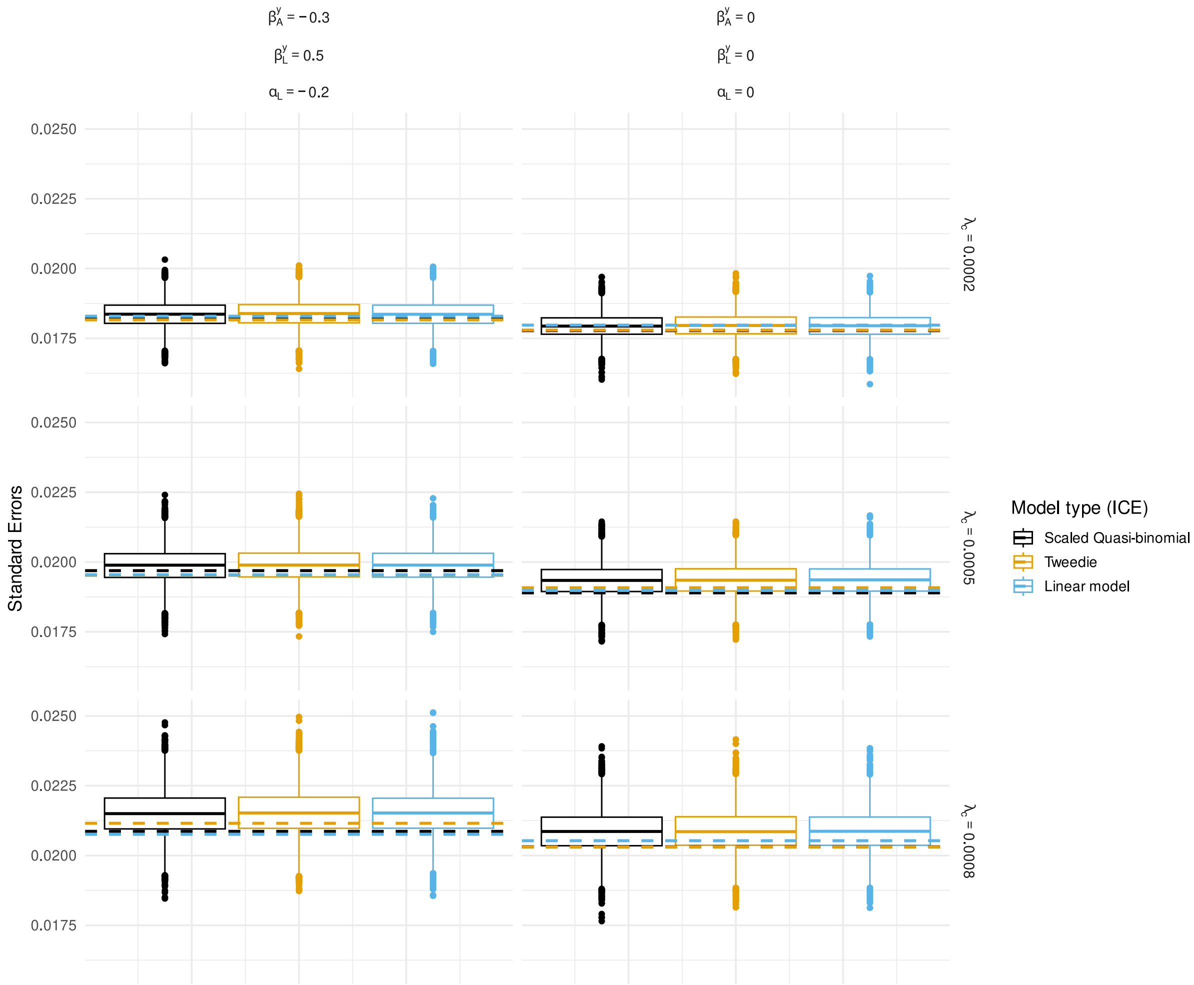


Figure 8: Boxplots of the standard errors for the case with censoring. The red line indicates the empirical standard error of the estimates for each estimator.

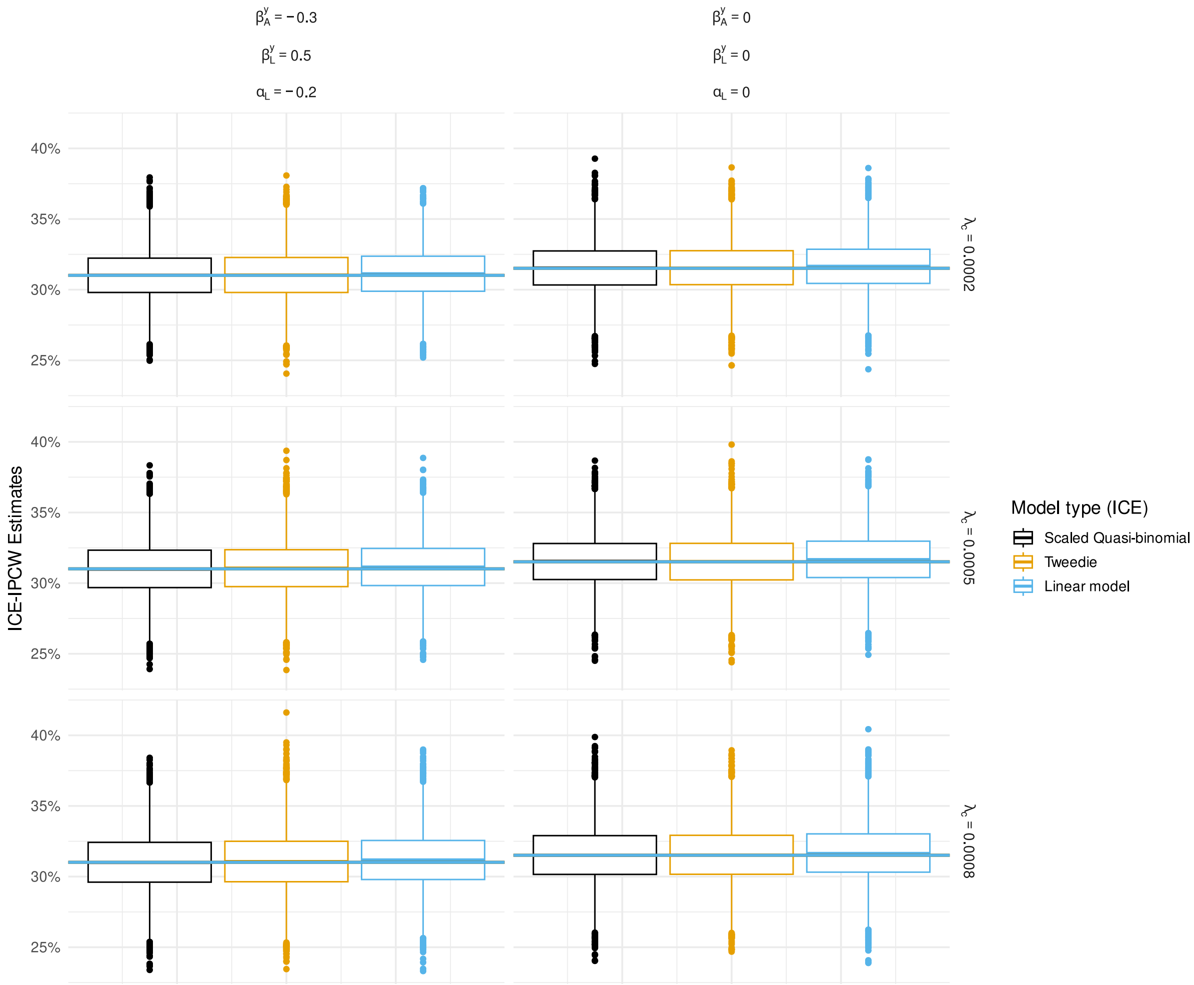


Figure 9: Boxplots of the results for the case with censoring. Different degrees of censoring are considered as well different model types for the pseudo-outcomes. Here, the (not debiased) ICE-IPCW estimator is shown.

# 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 DPP4 inhibitors or SGLT2 inhibitors between 2012 and 2022.

At baseline (time zero), patients were required to have redeemed such a prescription and to have an HbA1c 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.
- $N^\ell$ 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 and sex).

- 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 DPP4 inhibitors prior to time $t$.

Our target parameter is the risk of death, and we enforce treatment as part of the 20 first events.

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 a Cox proportional hazards model using only baseline covariates. As in the simulation study, we omitted the censoring martingale term, yielding conservative confidence intervals.

Detailed figures are 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.

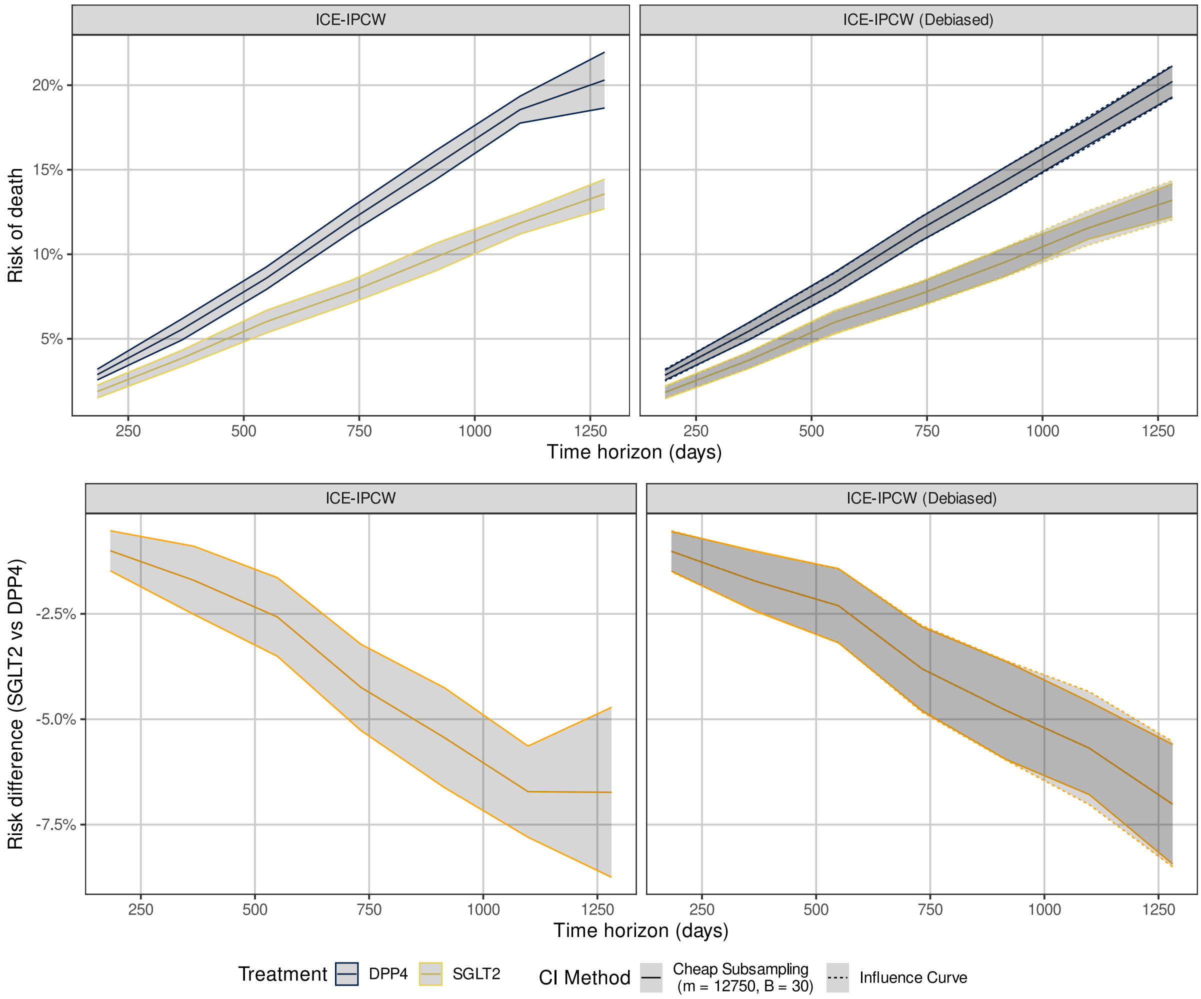


Figure 10: The causal risk of death (upper plot) and risk difference (lower plot) under treatment with SGLT2 inhibitors compared to DPP4 inhibitors shown as a function of time since initiation of treatment and 95% confidence intervals based on the efficient influence function and Cheap Subsampling confidence 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. Two main issues arise from this: One is that we have not proposed flexible intensity estimation for both the censoring intensity and the propensity scores. The literature on this is fairly limited (especially so in the presence of right-censoring), and are mostly based on neural networks (Liguori et al. (2023) for an overview). Other choices include flexible parametric models/highly adaptive LASSO using piecewise constant intensity models where the likelihood is based on Poisson regression (e.g., Laird & Olivier (1981)).

We could have also opted to use a TMLE (van der Laan & Rubin (2006)) instead. Here, we can use an iterative TMLE procedure for the $\bar{Q}^g_{k,\tau}$'s where we undersmooth the estimation of the censoring compensator to not deal with the problem of estimating the censoring martingale term. This yields conservative but valid inference when the censoring distribution is flexibly estimated.

Another potential issue with the estimation of the nuisance parameters is that the history is high dimensional and that the covariates in the history are highly correlated, since many covariates may not change between events. 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 the high dimensionality of the history. Another alternative method for inference within the TMLE framework is to use temporal difference learning to avoid iterative estimation of $\bar{Q}^g_{k,\tau}$ altogether (Shirakawa et al. (2024)) by appropriately extending it to the continuous-time setting; however, we can no longer apply just *any* machine learning method.

# 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.

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*.

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

Laird, N., & Olivier, D. (1981). Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques. *Journal of the American Statistical Association*, *76*(374), 231–240. http://www.jstor.org/stable/2287816

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

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., & Laan, M. J. van der. (2022). Continuous-Time Targeted Minimum Loss-Based Estimation of Intervention-Specific Mean Outcomes. *The Annals of Statistics*, *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., Li, Y., Wu, Y., Qiu, S., Li, Y., Zhao, M., Iso, H., & van der Laan, M. (2024, April). *Longitudinal Targeted Minimum Loss-based Estimation with Temporal-Difference Heterogeneous Transformer* (Issue arXiv:2404.04399). arXiv.

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., & 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.

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. (2025). Type 2 Diabetes, sodium-glucose cotransporter-2 inhibitors and cardiovascular outcomes: Real world evidence versus a randomised clinical trial. *Cardiovascular Diabetology*, *?*, ?.

# APPENDIX A

## A.1 Proof of Theorem 1

Let $W_{k,j} = \frac{W^g\left(T_{(j)}\right)}{W^g\left(T_{(k)}\right)}$ for $k < j$ (defining $\frac{0}{0} = 0$). We show that

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

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

*Base case*: The case $k = K-1$ is trivial $\left(W_{K-1,K} \mathbb{1}\left\{T_{(K)} \leq \tau, \Delta_{(K)} = y\right\} = \mathbb{1}\left\{T_{(K)} \leq \tau, \Delta_{(K)} = y\right\}\right)$.

*Inductive step*: Assume that the claim holds for $k+1$. We find

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

Repeatedly throughout, we use the law of iterated expectations. In a, we use that

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

for all $j = 1, \ldots, k-1$ and $k = 1, \ldots, K$. In b, we use the induction hypothesis.

The desired statement about $\bar{Q}^g_{k,\tau}$ Equation (3.2) now follows from the fact that

$$
\begin{aligned}
&\mathbb{E}_P\Big[W_{k-1,k}\bar{Q}^g_{k,\tau}\big(A\big(T_{(k)}\big),H_k\big)\mid T_{(k)},\Delta_{(k)}=a,\mathcal{F}_{T_{(k-1)}}\Big]\\
&=\mathbb{E}_P\left[\frac{\mathbb{1}\big\{A\big(T_{(k)}\big)=1\big\}}{\pi_k\Big(T_{(k)},L\big(T_{(k)}\big),\mathcal{F}_{T_{(k-1)}}\Big)}\bar{Q}^g_{k,\tau}(1,H_k)\mid T_{(k)},\Delta_{(k)}=a,\mathcal{F}_{T_{(k-1)}}\right]\\
&=\mathbb{E}_P\left[\frac{\mathbb{E}_P\Big[\mathbb{1}\big\{A\big(T_{(k)}\big)=1\big\}\mid T_{(k)},\Delta_{(k)}=a,\mathcal{F}_{T_{(k-1)}}\Big]}{\pi_k\Big(T_{(k)},L\big(T_{(k)}\big),\mathcal{F}_{T_{(k-1)}}\Big)}\bar{Q}^g_{k,\tau}(1,H_k)\mid T_{(k)},\Delta_{(k)}=a,\mathcal{F}_{T_{(k-1)}}\right]\\
&=\mathbb{E}_P\Big[\bar{Q}^g_{k,\tau}(1,H_k)\mid T_{(k)},\Delta_{(k)}=a,\mathcal{F}_{T_{(k-1)}}\Big]\\
&=\bar{Q}^g_{k,\tau}(1,H_k)
\end{aligned}
\tag{A.1}
$$

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

# APPENDIX B

## B.1 Finite dimensional distributions and compensators

Let $\big(\widetilde{X}(t)\big)_{t\geq 0}$ be a $d$-dimensional càdlàg jump process, where each component $i$ is two-dimensional such that $\widetilde{X}_i(t)=(N_i(t),X_i(t))$ and $N_i(t)$ is the counting process for the measurements of the $i$'th component $X_i(t)$ such that $\Delta X_i(t)\neq 0$ only if $\Delta N_i(t)\neq 0$ and $X(t)\in\mathcal{X}$ for some Euclidean space $\mathcal{X}\subseteq\mathbb{R}^m$. Assume that the counting processes $N_i$ with probability 1 have no simultaneous jumps and that the number of event times is bounded by a finite constant $K<\infty$. Furthermore, let $\mathcal{F}_t=\sigma\big(\widetilde{X}(s)\mid s\leq t\big)\vee\sigma(W)\in\mathcal{W}\subseteq\mathbb{R}^w$ be the natural filtration. Let $T_k$ be the $k$'th jump time of $t\mapsto\widetilde{X}(t)$ and let a random measure on $\mathbb{R}_+\times\mathcal{X}$ be given by

$$N(d(t,x))=\sum_{k:T_{(k)}<\infty}\delta_{\left(T_{(k)},X\left(T_{(k)}\right)\right)}(d(t,x)).$$

Let $\mathcal{F}_{T_{(k)}}$ be the stopping time $\sigma$-algebra associated with the $k$'th event time of the process $\widetilde{X}$. Furthermore, let $\Delta_{(k)}=j$ if $\Delta N_j\big(T_{(k)}\big)\neq 0$ and let $\mathbb{F}_k=\mathcal{W}\times(\mathbb{R}_+\times\{1,...,d\}\times\mathcal{X})^k$.

**Theorem 5** (Finite-dimensional distributions): Under the stated conditions of this section:

(i). We have $\mathcal{F}_{T_{(k)}}=\sigma\big(T_{(k)},\Delta_{(k)},X\big(T_{(k)}\big)\big)\vee\mathcal{F}_{T_{(k-1)}}$ and $\mathcal{F}_0=\sigma(W)$. Furthermore, $\mathcal{F}^{\bar{N}}_t=\sigma(\bar{N}((0,s],\cdot)\mid s\leq t)\vee\sigma(W)=\mathcal{F}_t$, where

$$\bar{N}(\mathrm{d}(t,m,x))=\sum_{k:T_{(k)}<\infty}\delta_{\left(T_{(k)},\Delta_{(k)},X\left(T_{(k)}\right)\right)}(d(t,m,x)).$$

We refer to $\bar{N}$ as the *associated* random measure.

(ii). There exist stochastic kernels $\Lambda_{k,i}$ from $\mathbb{F}_{k-1}$ to $\mathbb{R}$ and $\zeta_{k,i}$ from $\mathbb{R}_+\times\mathbb{F}_{k-1}$ to $\mathbb{R}_+$ such that the compensator for $N$ is given by,

$$\Lambda(\mathrm{d}(t,m,x))=\sum_{k:T_{(k)}<\infty}\mathbb{1}\big\{T_{(k-1)}<t\leq T_{(k)}\big\}\sum_{i=1}^d\delta_i(dm)\zeta_{k,i}\Big(dx,t,\mathcal{F}_{T_{(k-1)}}\Big)\Lambda_{k,i}\Big(dt,\mathcal{F}_{T_{(k-1)}}\Big)\prod_{l\neq j}\delta_{\left(X_l\left(T_{(k-1)}\right)\right)}(dx_l)$$

Here $\Lambda_{k,i}$ is the cause-specific hazard measure for $k$'th event of the $i$'th type, and $\zeta_{k,i}$ is the conditional distribution of $X_i\big(T_{(k)}\big)$ given $\mathcal{F}_{T_{(k-1)}}$, $T_{(k)}$ and $\Delta_{(k)}=i$.

*Proof*: To prove (i), we first note that since the number of events are bounded, we have the *minimality* condition of Theorem 2.5.10 of Last & Brandt (1995), so the filtration $\mathcal{F}^N_t=\sigma(N((0,s],\cdot)\mid s\leq t)\vee\sigma(W)=\mathcal{F}_t$ where

$$N(d(t,\tilde{x})) = \sum_{k:T_{(k)}<\infty} \delta_{(T_{(k)}, \widetilde{X}(T_{(k)}))}(d(t,\tilde{x}))$$

Thus $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \widetilde{X}\left(T_{(k)}\right)\right) \vee \mathcal{F}_{T_{(k-1)}}$ and $\mathcal{F}_0 = \sigma(W)$ in view of Equation (2.2.44) of Last & Brandt (1995). To get (i), simply note that since the counting proceses do not jump at the same time, there is a one-to-one corresponding between $\Delta_{(k)}$ and $N^i\left(T_{(k)}\right)$ for $i = 1, ..., d$, implying that $\bar{N}$ generates the same filtration as $N$, i.e., $\mathcal{F}_t^N = \mathcal{F}_t^{\bar{N}}$ for all $t \geq 0$.

To prove (ii), simply use Theorem 4.1.11 of Last & Brandt (1995) which states that

$$\Lambda(d(t,m,x)) = \sum_{k:T_{(k)}<\infty} \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \frac{P\left(\left(T_{(k)}, \Delta_{(k)}, X\left(T_{(k)}\right)\right) \in d(t,m,x) \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)}$$

is a $P$-$\mathcal{F}_t$ martingale. Then, we find by the "no simultaneous jumps" condition,

$$P\left(X\left(T_{(k)}\right) \in dx \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t, \Delta_{(k)} = j\right) = P\left(X_j\left(T_{(k)}\right) \in dx_j \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t, \Delta_{(k)} = j\right) \prod_{l \neq j} \delta_{(X_l(T_{(k-1)}))}(dx_l)$$

We then have,

$$\begin{aligned}
&\frac{P\left(\left(T_{(k)}, \Delta_{(k)}, X\left(T_{(k)}\right)\right) \in d(t,m,x) \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)} \\
&= \sum_{j=1}^d \delta_j(dm) P\left(X\left(T_{(k)}\right) \in dx \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t, \Delta_{(k)} = j\right) \frac{P\left(T_{(k)} \in dt, \Delta_{(k)} = j \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\right)}.
\end{aligned}$$

Letting

$$\zeta_{k,j}(dx, t, f_{k-1}) := P\left(X_j\left(T_{(k)}\right) \in dx \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}, T_{(k)} = t, \Delta_{(k)} = j\right)$$

$$\Lambda_{k,j}(dt, f_{k-1}) := \frac{P\left(T_{(k)} \in dt, \Delta_{(k)} = j \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\right)}$$

completes the proof of (ii).

□

# APPENDIX C

## C.1 Proof of Theorem 2

Note that the theorem can now be used to show the consistency of the ICE-IPCW estimator. We proceed by backwards induction. We apply the lemmas stated below (Lemma 1) and (Lemma 2). So, we simply assume that the theorem holds for $k+1$ and show that it also holds for $k$ as follows:

$$\begin{aligned}
\mathbb{E}_P\left[\bar{Z}^a_{k+1,\tau}(u) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right] &= \mathbb{1}\left\{\bar{\Delta}_{(1)} \in \{a,\ell\}, ..., \bar{\Delta}_{(k)} \in \{a,\ell\}\right\} \mathbb{E}_P\left[\bar{Z}^a_{k+1,\tau}(u) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right] + 0 \\
&\overset{(*)}{=} \mathbb{1}\left\{\bar{\Delta}_{(1)} \in \{a,\ell\}, ..., \bar{\Delta}_{(k)} \in \{a,\ell\}\right\} \\
&\quad \times \Bigg\{ \int_{\bar{T}_{(k)}}^u S\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) \mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}(u, 1, \bar{H}_{k+1}) \mid T_{(k)} = s, \Delta_{(k)} = a, \mathcal{F}_{T_{(k)}} = \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right] \Lambda^a_{k+1}\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad + \int_{\bar{T}_{(k)}}^u S\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) \mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}(u) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k)}} = \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right] \Lambda^\ell_{k+1}\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad + \int_{\bar{T}_{(k)}}^u S\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) \Lambda^y_{k+1}\left(\mathrm{d}s, \mathcal{F}_{T_{(k)}}\right) \Bigg\} \\
&= \mathbb{1}\left\{\bar{\Delta}_{(1)} \in \{a,\ell\}, ..., \bar{\Delta}_{(k)} \in \{a,\ell\}\right\} \bar{Q}^g_{k,\tau}\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right).
\end{aligned}$$

In $(*)$, we apply Lemma 1 and Lemma 2.

**Lemma 1**: Assume that the compensator $\Lambda^\alpha$ of $N^\alpha$ with respect to the filtration $\mathcal{F}_t^\beta$ is also the compensator with respect to the filtration $\mathcal{F}_t$. Then, we have

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

where

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

and

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

and $\mu_k^\ell\left(t, \cdot, \mathcal{F}_{T_{(k-1)}}\right)$ is the probability measure for the covariate value given $\Delta_{(k)} = \ell, T_{(k)} = t$, and $\mathcal{F}_{T_{(k-1)}}$ and $\mu_k^a\left(t, \cdot, \mathcal{F}_{T_{(k-1)}}\right)$ is the probability measure for the covariate value given $\Delta_{(k)} = a, T_{(k)} = t$, and $\mathcal{F}_{T_{(k-1)}}$.

*Proof*: Under the local independence condition, a version of the compensator of the random measure $N^\alpha(\mathrm{d}(t,m,a,l))$ with respect to the filtration $\mathcal{F}_t^\beta$, can be given by Theorem 4.2.2 (ii) of Last & Brandt (1995),

$$
\Lambda^\alpha(\mathrm{d}(t,m,a,l)) = K'((L(0),A(0)), N^\alpha, t, \mathrm{d}(m,a,l)) V'((A(0),L(0)), N^\alpha, \mathrm{d}t)
$$

for some kernel $K'$ from $\{0,1\} \times \mathcal{L} \times N_{\boldsymbol{X}} \times \mathbb{R}_+$ to $\boldsymbol{X}$ and some predictable kernel $V'$ from $\{0,1\} \times \mathcal{L} \times N_{\boldsymbol{X}} \times \mathbb{R}_+$ to $\mathbb{R}_+ \times \boldsymbol{X}$, because this *canonical* compensator is uniquely determined. Here $N_{\boldsymbol{X}}$ denotes the canonical point process space with mark space $\boldsymbol{X}$ (Last & Brandt (1995)). In this case, $\boldsymbol{X} = \mathcal{A} \times \mathcal{L} \times \{a,\ell,d,y,c\}$, where $\mathcal{A} = \{0,1\}$ and $\mathcal{L} \subseteq \mathbb{R}^d$.

Similarly, we find a compensator of the process $N^c(t)$ with respect to the filtration $\mathcal{F}_t^\beta$ given by

$$
\Lambda^c(\mathrm{d}(t,m,a,l)) = \delta_{(c,\varnothing,\varnothing)}(\mathrm{d}(m,a,l)) V''\left((A(0),L(0)), N^\beta, \mathrm{d}t\right)
$$

for some kernel $V''$ from $\{0,1\} \times \mathcal{L} \times N_{\boldsymbol{X}} \times \mathbb{R}_+$ to $\mathbb{R}_+ \times \boldsymbol{X}$. We now find the *canonical* compensator of $N^\beta$, given by

$$
\begin{aligned}
\rho\left((l_0,a_0), \varphi^\beta, d(t,m,a,l)\right) = \mathbb{1}\{m \in \{a,\ell,d,y\}\} K'((l_0,a_0), \varphi^\alpha, t, \mathrm{d}(m,a,l)) V'((a_0,l_0), \varphi^\alpha, \mathrm{d}t) \\
+ V'\left((a_0,l_0), \varphi^\beta, \mathrm{d}t\right)\delta_{(c,\varnothing,\varnothing)}(\mathrm{d}(m,a,l)).
\end{aligned}
$$

Then $\rho((L(0),A(0)), N^\beta, d(t,m,a,l))$ is a compensator, so it is by definition the canonical compensator. By construction,

$$
V''\left((A(0),L(0)), \left(N^\beta\right)^{T_{(k-1)}}, \mathrm{d}t\right) = \tilde{\Lambda}_k^c\left(\mathrm{d}t \mid \mathcal{F}^\beta_{T_{(k-1)}}\right)
$$

where superscript denotes the stopped process at the $(k-1)$'th event time. and similarly, we see find

$$K'\Big((l_0,a_0),(N^\beta)^{T_{(k-1)}},t,\mathrm{d}(m,a,l)\Big)V'\Big((a_0,l_0),(N^\beta)^{T_{(k-1)}},\mathrm{d}t\Big)=\sum_{x=a,\ell,d,y}\psi_{k,x}\Big(t,d(a,l),\mathcal{F}_{T_{(k-1)}}\Big)\Lambda_k^x\Big(\mathrm{d}t\mid\mathcal{F}_{T_{(k-1)}}\Big)$$

(see for instance (ii) of Theorem 5).

Let $\left(T^*_{(K)},\Delta^*_{(K)},...,A(0),L(0)\right)$ be random variables generating process $N^\beta$. With $S:=T^e\wedge C\wedge T_{(k)}$, we have $T_{S,0}=T^e\wedge C\wedge T_{(k)}=\bar{T}_{(k)}$ and $T_{S,1}=\bar{T}_{(k+1)}$. Using Theorem 4.3.8 of Last & Brandt (1995), it therefore holds that

$$\begin{aligned}
&P\Big((T^*_{S,1},\Delta^*_{S,1},A(T^*_{S,1}),L(T^*_{S,1}))\in d(t,m,a,l)\mid\mathcal{F}^{\tilde\beta}_{\bar{T}_{(k)}}\Big)\\
&\overset{(*)}{=}P\Big((T^*_{S,1},\Delta^*_{S,1},A(T^*_{S,1}),L(T^*_{S,1}))\in d(t,m,a,l)\mid\mathcal{F}^{\beta}_{T_{S,0}}\Big)\\
&=\mathbb{1}\{T_{S,0}<t\}\\
&\quad\times\prod_{s\in(T_{S,0},t)}\Big(1-\rho\Big((L(0),A(0)),(N^\beta)^{T_{S,0}},ds,\{a,y,l,d,y\}\times\{0,1\}\times\mathcal{L}\Big)\Big)\rho\Big((L(0),A(0)),(N^\beta)^{T_{S,0}},d(t,m,a,l)\Big)\\
&=\mathbb{1}\Big\{\bar{T}_{(k-1)}<t\Big\}\\
&\quad\times\prod_{s\in(T_{S,0},t)}\left(1-\rho\left((L(0),A(0)),\left(N^{\tilde\beta}\right)^{\bar{T}_{(k)}},ds,\{a,y,l,d,y\}\times\{0,1\}\times\mathcal{L}\right)\right)\rho\left(A(0),L(0),\left(N^{\tilde\beta}\right)^{\bar{T}_{(k)}},d(t,m,a,l)\right).
\end{aligned}$$

In $(*)$, we use that $\mathcal{F}^\beta_{T_{S,0}}\overset{(**)}{=}\sigma\Big((A(0),L(0)),(N^\beta)^{T_{S,0}}\Big)=\sigma\left((A(0),L(0)),\left(N^{\tilde\beta}\right)^{\bar{T}_{(k)}}\right)=\mathcal{F}^{\tilde\beta}_{\bar{T}_{(k)}}$ where $(**)$ follows from Theorem 2.1.14 of Last & Brandt (1995). Further note that $T^*_k=\bar{T}_{(k)}$ whenever $T_{(k-1)}<C$. By definition, $T_{S,1}=T^*_{k+1}$ if $N^\beta\big(T_{(k)}\wedge C\wedge T^e\big)=k$. If $\bar\Delta_{(1)}\notin\{c,y,d\},...,\bar\Delta_{(k)}\notin\{c,y,d\}$, then $N^\beta\big(T_{(k)}\wedge C\wedge T^e\big)=k$ and furthermore $T^*_{k+1}=\bar{T}_{(k+1)}$, so

$$\begin{aligned}
&\mathbb{1}\Big\{\bar\Delta_{(1)}\notin\{c,y,d\},...,\bar\Delta_{(k)}\notin\{c,y,d\}\Big\}P\Big((\bar{T}_{k+1},\bar\Delta_{k+1},A(\bar{T}_{k+1}),L(\bar{T}_{k+1}))\in d(t,m,a,l)\mid\mathcal{F}^{\tilde\beta}_{\bar{T}_{(k)}}\Big)\\
&=\mathbb{1}\Big\{\bar\Delta_{(1)}\notin\{c,y,d\},...,\bar\Delta_{(k)}\notin\{c,y,d\}\Big\}P\Big((T^*_{S,1},\Delta^*_{S,1},A(T^*_{S,1}),L(T^*_{S,1}))\in d(t,m,a,l)\mid\mathcal{F}^{\tilde\beta}_{\bar{T}_{(k)}}\Big)\\
&=\mathbb{1}\Big\{\bar\Delta_{(1)}\notin\{c,y,d\},...,\bar\Delta_{(k)}\notin\{c,y,d\}\Big\}\mathbb{1}\Big\{\bar{T}_{(k-1)}<t\Big\}\\
&\qquad\prod_{s\in(T_{S,0},t)}\left(1-\rho\left((L(0),A(0)),\left(N^{\tilde\beta}\right)^{\bar{T}_{(k)}},ds,\{a,y,l,d,y\}\times\{0,1\}\times\mathcal{L}\right)\right)\rho\left(A(0),L(0),\left(N^{\tilde\beta}\right)^{\bar{T}_{(k)}},d(t,m,a,l)\right),
\end{aligned}$$

and we are done. From this, we get Equation (C.1). □

**Lemma 2**: Assume independent censoring as in Theorem 2. Then the left limit of the survival function factorizes on $(0,\tau]$, i.e.,

$$\mathbb{1}\Big\{\bar\Delta_{(k-1)}\neq c\Big\}\tilde{S}\Big(t-\mid\mathcal{F}_{T_{(k-1)}}\Big)=\mathbb{1}\Big\{\bar\Delta_{(k-1)}\neq c\Big\}\prod_{s\in(0,t)}\left(1-\sum_{x=a,\ell,y,d}\Lambda_k^x\Big(\mathrm{d}s,\mathcal{F}_{T_{(k-1)}}\Big)\right)\prod_{s\in(0,t]}\Big(1-\tilde\Lambda_k^c\Big(\mathrm{d}t\mid\mathcal{F}^{\tilde\beta}_{\bar{T}_{(k-1)}}\Big)\Big)$$

if for all $t\in(0,\tau)$,

$$\begin{aligned}
&\Delta\tilde\Lambda_k^c\Big(t\mid\mathcal{F}^{\tilde\beta}_{\bar{T}_{(k-1)}}\Big)+\sum_x\Delta\Lambda_k^x\Big(t,\mathcal{F}_{T_{(k-1)}}\Big)=1\quad P-\text{a.s.}\\
&\Rightarrow\Delta\tilde\Lambda_{k+1}^c\Big(t\mid\mathcal{F}_{T_{(k-1)}}\Big)=1\quad P-\text{a.s.}\vee\sum_x\Delta\Lambda_k^x\Big(t,\mathcal{F}_{T_{(k-1)}}\Big)=1\quad P-\text{a.s.}
\end{aligned}$$

*Proof*: Consider the quadratic covariation process $\langle\cdot,\cdot\rangle$, which by the no simultaneous jump condition of the observed censoring process and the observed other processes (Andersen et al. (1993), p. 69) by and (2.4.2) of Andersen et al. (1993) fulfill

$$0=\langle M^c(\cdot\wedge T^e),\sum_x M^x(\cdot\wedge C)\rangle_t=\int_0^t\Delta\tilde\Lambda_c\sum_{x=a,\ell,y,d}d\Lambda_x$$

where $\tilde{\Lambda}_c$ and $\Lambda_x$ are the compensators of the censoring process and the rest of the counting processes, respectively. The latter is by the independent censoring assumption the one given as

$$\Lambda_x(\mathrm{d}t) = \sum_{k=1}^{K} \mathbb{1}\left\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\right\} \Lambda_k^x\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right).$$

Furthermore, $\tilde{\Lambda}_c(\mathrm{d}t) = \sum_{k=1}^{K} \mathbb{1}\left\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\right\} \Delta\tilde{\Lambda}_c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$. Using these, we have

$$\begin{aligned}
0 = \sum_{k=1}^{K} \mathbb{1}\left\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\right\} &\Bigg(\int_{\left(T_{(k-1)} \wedge C, t\right]} \Delta\tilde{\Lambda}_k^c\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\left(\sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right) \\
&+ \sum_{j=1}^{k-1} \int_{\left(T_{(j-1)} \wedge C, T_{(j)} \wedge C\right]} \Delta\tilde{\Lambda}_c\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(j-1)}}\right)\left(\sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right)\Bigg),
\end{aligned}$$

so that $\mathbb{1}\left\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\right\} \int_{\left(T_{(k-1)}, t\right]} \Delta\tilde{\Lambda}_k^c\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\left(\sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right) = 0$ since each term is non-negative. Taking the (conditional) expectations on both sides given $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}$, we have

$$\mathbb{1}\left\{T_{(k-1)} \wedge C < t\right\} \tilde{S}\left(t- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \sum_{\bar{T}_{(k)} < s \leq t} \Delta\tilde{\Lambda}_k^c\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\left(\sum_{x=a,\ell,y,d} \Delta\Lambda_k^x\left(s, \mathcal{F}_{T_{(k-1)}}\right)\right) = 0, \tag{C.2}$$

where we also use that $\Delta\tilde{\Lambda}_k^c\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right) \neq 0$ for only a countable number of $s$'s. This already means that the continuous part of the Lebesgue-Steltjes integral is zero, and thus the integral is evaluated to the sum in Equation (C.2). It follows that for every $t$ with $\tilde{S}\left(t- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\right) > 0$ and $\bar{T}_{k-1} < t$, we have

$$\sum_{\bar{T}_k < s \leq t} \Delta\tilde{\Lambda}_k^c\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\left(\sum_{x=a,\ell,y,d} \Delta\Lambda_k^x\left(s, \mathcal{F}_{T_{(k-1)}}\right)\right) = 0.$$

This entails that $\Delta\tilde{\Lambda}_{k+1}^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$ and $\sum_x \Delta\Lambda_{k+1}^x\left(t, \mathcal{F}_{T_{(k+1-1)}}\right)$ cannot be both non-zero at the same time. To keep notation brief, let $\gamma(v) = \Delta\tilde{\Lambda}_{k+1}^c\left(v \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$ and $\zeta(v) = \sum_x \Delta\Lambda_{k+1}^x\left(v, \mathcal{F}_{T_{(k+1-1)}}\right)$ and $s = \bar{T}_{k-1}$.

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

$$\begin{aligned}
&\mathbb{1}\left\{\prod_{v \in (s,t)} (1 - \mathrm{d}(\zeta + \gamma)(v)) > 0\right\} \prod_{v \in (s,t)} (1 - \mathrm{d}(\zeta + \gamma)(v)) \\
&= \mathbb{1}\left\{\prod_{v \in (s,t)} (1 - \mathrm{d}(\zeta + \gamma)(v)) > 0\right\} \prod_{v \in (s,t)} (1 - \mathrm{d}\zeta(v)) \prod_{v \in (s,t]} (1 - \mathrm{d}\gamma(v))
\end{aligned}$$

since

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

under the assumption $\prod_{v \in (s,t)} (1 - \mathrm{d}(\zeta + \gamma)(v)) > 0$. So we just need to make sure that $\prod_{v \in (s,t)} (1 - \mathrm{d}(\zeta + \gamma)(v)) = 0$ if and only if $\prod_{v \in (s,t)} (1 - \mathrm{d}\zeta(v)) = 0$ or $\prod_{v \in (s,t)} (1 - \mathrm{d}\gamma(v)) = 0$. Splitting the product integral into the continuous and discrete parts as before, we have

$$\prod_{v\in(s,t)}(1-d(\zeta+\gamma)(v))=0 \Leftrightarrow \exists u\in(s,t) \text{ s.t. } \Delta\gamma(u)+\Delta\zeta(u)=1$$

$$\prod_{v\in(s,t)}(1-d\gamma(v))\prod_{v\in(s,t)}(1-\zeta(v))=0 \Leftrightarrow \exists u\in(s,t) \text{ s.t. } \Delta\gamma(u)=1 \vee \exists u\in(s,t) \text{ s.t. } \Delta\zeta(u)=1$$

from which the result follows. □

# APPENDIX D

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

Let us define in the censored setting

$$W^g(t)=\prod_{k=1}^{\tilde{N}_t}\frac{\mathbb{1}\{A(\bar{T}_k)=1\}}{\pi_k\left(\bar{T}_{(k)},\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))}\prod_{k=1}^{N_t}\frac{\mathbb{1}\left\{\Delta_{(k)}\neq c\right\}}{\prod_{u\in\left(\bar{T}_{(k-1)},\bar{T}_{(k)}\right)}\left(1-\tilde{\Lambda}^c_k\left(\mathrm{d}u,\mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\right)}$$

in alignment with Equation (3.1). We verify that our efficient influence function is the same as Rytgaard et al. (2022) in the case with absolutely continuous compensators. For simplicity, we also assume that $L\left(T_{(k)}\right)=L\left(T_{(k-1)}\right)$ $\left(L(\bar{T}_k)=L(\bar{T}_{k-1})\right)$ whenever $\Delta_{(k)}=a$ $(\bar{\Delta}_{(k)}=a)$. The efficient influence function of Rytgaard et al. (2022) is given in Theorem 1 of Rytgaard et al. (2022) in our notation by

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

Here, we use $\sim$ to denote that the martingales and counting processes are the observed ones. We note, for instance, for $x=\ell$ that

$$
\begin{aligned}
&\mathbb{E}_{PG^*}\left[N_y(\tau) \mid \Delta N^x(t)=1, \mathcal{F}_{t-}\right] \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \mathbb{E}_{PG^*}\left[N_y(\tau) \mid T_{(k)}=t, \Delta_{(k)}=x, \mathcal{F}_{T_{(k-1)}}\right] \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \lim_{\varepsilon \to 0} \mathbb{E}_{PG^*}\left[N_y(\tau) \mid T_{(k)} \in (t, t+\varepsilon), \Delta_{(k)}=x, \mathcal{F}_{T_{(k-1)}}\right] \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \lim_{\varepsilon \to 0} \frac{\mathbb{E}_{PG^*}\left[N_y(\tau)\mathbb{1}\left\{T_{(k)} \in (t, t+\varepsilon), \Delta_{(k)}=x\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]}{\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{T_{(k)} \in (t, t+\varepsilon), \Delta_{(k)}=x\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]} \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \\
&\qquad \lim_{\varepsilon \to 0} \frac{\mathbb{E}_{PG^*}\left[\mathbb{E}_{PG^*}\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)}, \Delta_{(k)}=x, \mathcal{F}_{T_{(k-1)}}\right]\mathbb{1}\left\{T_{(k)} < \tau\right\}\mathbb{1}\left\{T_{(k)} \in (t, t+\varepsilon), \Delta_{(k)}=x\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]}{\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{T_{(k)} \in (t, t+\varepsilon), \Delta_{(k)}=x\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]} \qquad \text{(D.2)} \\
&\overset{*}{=} \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \frac{\mathbb{E}_P\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)}=t, \Delta_{(k)}=x, \mathcal{F}_{T_{(k-1)}}\right] S\left(t \mid \mathcal{F}_{T_{(k-1)}}\right)\lambda^x_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)}{S\left(t \mid \mathcal{F}_{T_{(k-1)}}\right)\lambda^x_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)} \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \mathbb{E}_P\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)}=t, \Delta_{(k)}=x, \mathcal{F}_{T_{(k-1)}}\right]
\end{aligned}
$$

where we, in $(*)$, apply dominated convergence. Similarly, we may find that

$$
\begin{aligned}
\mathbb{E}_{PG^*}\left[N_y(\tau) \mid \Delta N^y(t)=1, \mathcal{F}_{t-}\right] &= 1, \\
\mathbb{E}_{PG^*}\left[N_y(\tau) \mid \Delta N^d(t)=1, \mathcal{F}_{t-}\right] &= 0, \qquad \text{(D.3)} \\
\mathbb{E}_{PG^*}\left[N_y(\tau) \mid \Delta N^a(t)=1, \mathcal{F}_{t-}\right] &= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \bar{Q}^g_{k,\tau}(\tau, 1, H_k)
\end{aligned}
$$

For the first term in Equation (D.1), we find that

$$
\begin{aligned}
&\mathbb{E}_{PG^*}\left[N_y(\tau) \mid L(t), N^\ell(t), \mathcal{F}_{t-}\right] - \mathbb{E}_{PG^*}\left[N_y(\tau) \mid N^\ell(t), \mathcal{F}_{t-}\right] \\
&= \mathbb{E}_{PG^*}\left[N_y(\tau) \mid L(t), \Delta N^\ell(t)=0, \mathcal{F}_{t-}\right] - \mathbb{E}_{PG^*}\left[N_y(\tau) \mid \Delta N^\ell(t)=0, \mathcal{F}_{t-}\right] \\
&\quad + \mathbb{E}_{PG^*}\left[N_y(\tau) \mid L(t), \Delta N^\ell(t)=1, \mathcal{F}_{t-}\right] - \mathbb{E}_{PG^*}\left[N_y(\tau) \mid \Delta N^\ell(t)=1, \mathcal{F}_{t-}\right] \\
&= 0 \qquad \text{(D.4)} \\
&\quad + \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\}\left(\mathbb{E}_{PG^*}\left[N_y(\tau) \mid L\left(T_{(k)}\right), T_{(k)}=t, \Delta_{(k)}=\ell, \mathcal{F}_{T_{(k-1)}}\right] - \mathbb{E}_{PG^*}\left[N_y(\tau) \mid T_{(k)}=t, \Delta_{(k)}=\ell, \mathcal{F}_{T_{(k-1)}}\right]\right) \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\}\mathbb{1}\left(T_{(k)} < \tau, \Delta_{(k)}=\ell, k<K\right)\left(\bar{Q}^g_{k,\tau}(\tau) - \mathbb{E}_P\left[\bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)}=t, \Delta_{(k)}=\ell, \mathcal{F}_{T_{(k-1)}}\right]\right)
\end{aligned}
$$

Next, note that

$$
\begin{aligned}
&\mathbb{E}_{PG^*}\left[N_y(\tau) \mid \Delta N^x(t)=0, \mathcal{F}_{t-}\right] \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \mathbb{E}_{PG^*}\left[N_y(\tau) \mid \left(T_{(k)} > t \vee T_{(k)}=t, \Delta_{(k)} \neq x\right), \mathcal{F}_{T_{(k-1)}}\right] \\
&\overset{**}{=} \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \mathbb{E}_{PG^*}\left[N_y(\tau) \mid T_{(k)} > t, \mathcal{F}_{T_{(k-1)}}\right] \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \frac{\mathbb{E}_{PG^*}\left[N_y(\tau)\mathbb{1}\left\{T_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]}{\mathbb{E}_{PG^*}\left[\mathbb{1}\left\{T_{(k)} > t\right\} \mid \mathcal{F}_{T_{(k-1)}}\right]} \qquad \text{(D.5)} \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \frac{\bar{Q}^g_{k-1,\tau}\left(\tau, \mathcal{F}_{T_{(k-1)}}\right) - \bar{Q}^g_{k-1,\tau}\left(t, \mathcal{F}_{T_{(k-1)}}\right)}{S\left(t \mid \mathcal{F}_{T_{(k-1)}}\right)},
\end{aligned}
$$

where in $(**)$, we use that the event $\left(T_{(k)}=t, \Delta_{(k)} \neq x\right)$ has probability zero. Let $B_{k-1}(u) = \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(u)\right)\frac{1}{S\left(u \mid \mathcal{F}^{\tilde{\beta}}_{T_{(k-1)}}\right)}$. Combining Equation (D.2), Equation (D.3), Equation (D.4), and Equation (D.5) with Equation (D.1), we find that the efficient influence function can also be written as:

$$
\begin{aligned}
\varphi_\tau^*(P) &= \frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))} \sum_{k=1}^K \prod_{j=1}^{k-1} \left( \frac{\mathbb{1}\{A(\bar{T}_j)=1\}}{\pi_j\left(\bar{T}_{(j)}, L(\bar{T}_j), \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(j-1)}}\right)} \right)^{\mathbb{1}\{\bar{\Delta}_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} \tilde{S}^c\left(\bar{T}_{(j)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(j-1)}}\right)} \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \in \{\ell, a\}, \bar{T}_{(k-1)} < \tau\right\} \\
&\Bigg[ \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \left( \bar{Q}^g_{k,\tau}\left(\tau, 1, u, a, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) - B_{k-1}(u) \right) \widetilde{M}^a(du) \\
&+ \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \left( \mathbb{E}_P\left[ \bar{Q}^g_{k,\tau}(\tau) \mid T_{(k)} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] - B_{k-1}(u) \right) \widetilde{M}^\ell(du) \\
&+ \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} (1 - B_{k-1}(u)) \widetilde{M}^y(du) + \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} (0 - B_{k-1}(u)) \widetilde{M}^d(du) \\
&+ \frac{1}{\tilde{S}^c\left(\bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \mathbb{1}\left(\bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = \ell, k < K\right) \left( \bar{Q}^g_{k,\tau}(\tau) - \mathbb{E}_P\left[ \bar{Q}^g_{k,\tau}(\tau) \mid \bar{T}_{(k)}, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \Bigg] \\
&+ \bar{Q}^g_{0,\tau}(\tau, 1, L(0)) - \Psi^g_\tau(P)
\end{aligned} \tag{D.6}
$$

$$
\begin{aligned}
&= \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}^{\tilde{\beta}}_{\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}^{\tilde{\beta}}_{\bar{T}_{(j-1)}}\right)} \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \in \{\ell, a\}, \bar{T}_{(k-1)} < \tau\right\} \Bigg[ \\
&- \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \bar{Q}^g_{k,\tau}\left(1, L(\bar{T}_{k-1}), u, a, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \lambda^a_k\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}u \\
&- \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \mathbb{E}_P\left[ \bar{Q}^g_{k,\tau}(\tau) \mid \bar{T}_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} \right] \lambda^\ell_k\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}u \\
&- \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} (1) \lambda^y_k\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}u - \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} (0) \lambda^d_k\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}u \\
&- \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} B_{k-1}(u) M^\bullet(\mathrm{d}u) + \bar{Z}^a_{k,\tau}(\tau) + \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} B_{k-1}(u) \widetilde{M}^c(du) \Bigg] \\
&+ \bar{Q}^g_{0,\tau}(\tau, 1, L(0)) - \Psi^g_\tau(P),
\end{aligned}
$$

where $M^\bullet(t) = \sum_{x=a,\ell,d,y,c} \widetilde{M}^x(t)$. Similarly define $\lambda^\bullet_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) = \sum_{x=a,\ell,d,y,c} \lambda^x_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$. Now note that

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

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

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

where the last line holds by the Duhamel equation (2.6.5) of Andersen et al. (1993). The second integral of Equation (D.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 \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \lambda^{\bullet}_k\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}u \\
&= \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \Bigg[ \int_0^u S\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \bar{Q}^g_{k,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \lambda^a_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}s \\
&\quad + \int_0^u S\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\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^\ell_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}s \\
&\quad + \int_0^u S\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \lambda^y_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}s \Bigg] \times \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \lambda^{\bullet}_k\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}u \\
&= \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 \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \lambda^{\bullet}_k\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}u \\
&\quad \times \Bigg[ S\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \bar{Q}^g_{k,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \lambda^a_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \\
&\quad + S\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\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^\ell_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \\
&\quad + S\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \lambda^y_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \Bigg] \mathrm{d}s \\
&\overset{*}{=} \frac{1}{\tilde{S}^c\left(\tau \wedge \bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(\tau \wedge \bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \bar{Q}^g_{k-1,\tau}\left(\tau \wedge \bar{T}_{(k)}\right) \\
&\quad - \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \Bigg[ \bar{Q}^g_{k,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \lambda^a_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\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(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \\
&\quad + \lambda^y_k\left(s, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \Bigg] \mathrm{d}s
\end{aligned} \tag{D.9}$$

In $(*)$, we use that

$$\begin{aligned}
&\int_s^{\tau \wedge \bar{T}_{(k)}} \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \lambda^{\bullet}_k\left(u, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \mathrm{d}u \\
&= \frac{1}{\tilde{S}^c\left(\tau \wedge \bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(\tau \wedge \bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} - \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)},
\end{aligned}$$

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

$$
\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(u)\right) \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} M^\bullet(\mathrm{d}u)
$$

$$
= \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}\left(\bar{T}_{(k)}\right)\right) \frac{1}{\tilde{S}^c\left(\bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(\bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \mathbb{1}\left\{\bar{T}_{(k)} \le \tau\right\}
$$

$$
-\bar{Q}^g_{k-1,\tau}(\tau)\left(\frac{1}{\tilde{S}^c\left(\bar{T}_{(k)}\wedge\tau \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(\bar{T}_{(k)}\wedge\tau \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} - 1\right)
$$

$$
\frac{1}{\tilde{S}^c\left(\tau\wedge\bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(\tau\wedge\bar{T}_{(k)} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \bar{Q}^g_{k-1,\tau}\left(\tau\wedge\bar{T}_{(k)}\right)
$$

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

This now shows that Equation (D.6) is equal to Equation (5.1).

# APPENDIX E

## E.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 $\bar{Z}^a_{k,\tau}(u; P)$. We compute the efficient influence function by calculating the derivative (Gateaux derivative) of $\Psi^g_\tau(P_\varepsilon)$ with $P_\varepsilon = P + \varepsilon(\delta_O - P)$ at $\varepsilon = 0$, where $\delta$ denotes the Dirac measure at the point $O$. First note that

$$
\left.\frac{\partial}{\partial\varepsilon}\right|_{\varepsilon=0} \Psi_\tau(P_\varepsilon) = \bar{Q}^g_{0,\tau}(\tau) - \Psi^g_\tau(P) + \int \left.\frac{\partial}{\partial\varepsilon}\right|_{\varepsilon=0} \bar{Q}^g_{0,\tau}(\tau, 1, l; P_\varepsilon) P_L(\mathrm{d}l). \tag{E.1}
$$

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\Big(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)}{P_\varepsilon\Big(\bar{T}_{(k)} \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} \\
&= \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \frac{P_\varepsilon\Big(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)}{P_\varepsilon\Big(\bar{T}_{(k)} \geq t, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} \\
&= \frac{\delta_{\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(f_{k-1})}{P\Big(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} \frac{\mathbb{1}\Big\{\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c\Big\} - P\Big(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)}{P\Big(\bar{T}_{(k)} \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} \\
&\quad - \frac{\delta_{\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(f_{k-1})}{P\Big(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} \Big(\mathbb{1}\Big\{\bar{T}_{(k)} \geq t\Big\} - P\Big(\bar{T}_{(k)} \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)\Big) \frac{P\Big(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)}{\Big(P\Big(\bar{T}_{(k)} \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)\Big)^2} \\
&= \frac{\delta_{\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(f_{k-1})}{P\Big(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} \frac{\mathbb{1}\Big\{\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c\Big\}}{P\Big(\bar{T}_{(k)} \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} - \mathbb{1}\Big\{\bar{T}_{(k)} \geq t\Big\} \frac{P\Big(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)}{\Big(P\Big(\bar{T}_{(k)} \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)\Big)^2} \\
&= \frac{\delta_{\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(f_{k-1})}{P\Big(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} \frac{1}{P\Big(\bar{T}_{(k)} \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)} \Big(\mathbb{1}\Big\{\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c\Big\} - \mathbb{1}\Big\{\bar{T}_{(k)} \geq t\Big\} \tilde{\Lambda}^c_k\Big(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\Big)\Big)
\end{aligned}
$$

so that

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \prod_{u \in (s,t)} \Big(1 - \tilde{\Lambda}^c_{k,\varepsilon}(\mathrm{d}t \mid f_{k-1})\Big) \\
&= \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]} \Big(1 - \tilde{\Lambda}^c_{k,\varepsilon}(\mathrm{d}t \mid f_{k-1})\Big) \\
&\overset{(*)}{=} -\frac{1}{1 - \Delta\tilde{\Lambda}^c_{k,\varepsilon}(t \mid f_{k-1})} \prod_{u \in (s,t]} \Big(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\Big) \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(\mathrm{d}u \mid f_{k-1}) \\
&\quad + \prod_{u \in (s,t]} \Big(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\Big) \frac{1}{\Big(1 - \Delta\tilde{\Lambda}^c_k(t \mid f_{k-1})\Big)^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]} \Big(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\Big) \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(\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]} \Big(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\Big) \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(\mathrm{d}u \mid f_{k-1}) \\
&\quad + \prod_{u \in (s,t]} \Big(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\Big) \frac{1}{\Big(1 - \Delta\tilde{\Lambda}^c_k(t \mid f_{k-1})\Big)^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)} \Big(1 - \tilde{\Lambda}^c_k(\mathrm{d}t \mid f_{k-1})\Big) \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{E.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 $\mu$ 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 $\int_{\{t\}} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k(u \mid f_{k-1})} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \tilde{\Lambda}^c_k(\mathrm{d}u \mid f_{k-1}) = \frac{1}{1 - \Delta\tilde{\Lambda}^c_k(t \mid f_{k-1})} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \tilde{\Lambda}^c_{k,\varepsilon}(t \mid f_{k-1})$. Furthermore, for $P_\varepsilon = P + \varepsilon\big(\delta_{(X,Y)} - P\big)$, a simple calculation yields the well-known result

$$\frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \mathbb{E}_{P_\varepsilon}[Y \mid X=x] = \frac{\delta_X(x)}{P(X=x)}(Y-\mathbb{E}_P[Y \mid X=x]). \tag{E.3}$$

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

$$\begin{aligned}
&\frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \bar{Q}^g_{k-1,\tau}(\tau, f_{k-1}; P_\varepsilon) \\
&= \frac{\delta_{\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(f_{k-1})}{P\left(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \Bigg( \bar{Z}^a_{k,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(\tau) + \\
&\quad + \underbrace{\int_{\bar{T}_{(k-1)}}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{(\bar{T}_{(k-1)}, t_k)} \frac{1}{1-\Delta\tilde{\Lambda}^c_k(s \mid f_{k-1})} \frac{1}{\tilde{S}\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(ds) - \tilde{\Lambda}^c(ds)\right) P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}(d(t_k, d_k, l_k, a_k) \mid f_{k-1})}_{\mathrm{A}} \Bigg) \\
&\quad + \underbrace{\int_{\bar{T}_{(k-1)}}^{\tau} \frac{\mathbb{1}\{t_k < \tau, d_k \in \{a, \ell\}\}}{\tilde{S}^c(t_k- \mid f_{k-1})} \left( \frac{\mathbb{1}\{a_k = 1\}}{\pi_k\left(t_k, L(\bar{T}_k), \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \right)^{\mathbb{1}\{d_k = a\}} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \bar{Q}^g_{k,\tau}(P_\varepsilon \mid a_k, l_k, t_k, d_k, f_{k-1}) P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}(d(t_k, d_k, l_k, a_k) \mid f_{k-1})}_{\mathrm{B}}
\end{aligned}$$

for $k = 1, \dots, K+1$, where in the notation with $\bar{Z}^a_{k,\tau}$, we have made the dependence on $T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right), \mathcal{F}_{T_{(k-1)}}$ explicit. To get B, we use a similar derivation to the one given in Equation Equation (A.1). Now note that for simplifying A, we can write

$$\begin{aligned}
&\int_{\bar{T}_{(k-1)}}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{(\bar{T}_{(k-1)}, t_k)} \frac{1}{1-\Delta\tilde{\Lambda}^c_k(s \mid f_{k-1})} \frac{1}{\tilde{S}\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(ds) - \tilde{\Lambda}^c(ds)\right) \\
&\qquad P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}\left(d(t_k, d_k, l_k, a_k) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&= \int_{\bar{T}_{(k-1)}}^{\tau} \int_s^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}\left(d(t_k, d_k, l_k, a_k) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&\qquad \frac{1}{1-\Delta\tilde{\Lambda}^c_k(s \mid f_{k-1})} \frac{1}{\tilde{S}\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(ds) - \tilde{\Lambda}^c(ds)\right) \\
&= \int_{\bar{T}_{(k-1)}}^{\tau} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \\
&\qquad \frac{1}{1-\Delta\tilde{\Lambda}^c_k(s \mid f_{k-1})} \frac{1}{\tilde{S}\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(ds) - \tilde{\Lambda}^c(ds)\right) \\
&= \int_{\bar{T}_{(k-1)}}^{\tau} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \\
&\qquad \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(ds) - \tilde{\Lambda}^c(ds)\right)
\end{aligned}$$

by an exchange of integrals. Here, we apply the result of Theorem 2 to get the last equation. Hence, we have

$$
\begin{aligned}
&\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \bar{Q}^g_{k-1,\tau}(\tau, f_{k-1}; P_\varepsilon) \\
&= \frac{\delta_{\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}(f_{k-1})}}{P\left(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \Bigg( \bar{Z}^a_{k,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(\tau) + \\
&\quad + \int_{\bar{T}_{(k-1)}}^{\tau} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right) \\
&\qquad \frac{1}{\tilde{S}^c\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(\mathrm{d}s) - \tilde{\Lambda}^c(\mathrm{d}s)\right) \Bigg) \\
&\quad + \int_{\bar{T}_{(k-1)}}^{\tau} \frac{\mathbb{1}\{t_k < \tau, d_k \in \{a, \ell\}\}}{\tilde{S}^c(t_k - \mid f_{k-1})} \left( \frac{\mathbb{1}\{a_k = 1\}}{\pi_k\left(t_k, L(\bar{T}_k), \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \right)^{\mathbb{1}\{d_k = a\}} \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \bar{Q}^g_{k,\tau}(P_\varepsilon \mid a_k, l_k, t_k, d_k, f_{k-1}) \\
&\qquad P_{(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, L(\bar{T}_k), A(\bar{T}_k))}(\mathrm{d}(t_k, d_k, l_k, a_k) \mid f_{k-1})
\end{aligned} \tag{E.4}
$$

Note that for $k = K+1$, the last term vanishes. Therefore, we can combine the results from Equation (E.4) and Equation (E.1) iteratively to obtain the result.

# APPENDIX F

## F.1 Proof of Theorem 4

We find the following decomposition,

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

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

Assume that we have shown that $P(K_{nc} \neq K_{\lim}) \to 0$ as $n \to \infty$. We have that

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

and

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

as $n \to \infty$. A similar conclusion holds for 2. We now show that $P(K_{nc} \neq K_{\lim}) \to 0$. First define that $K_n = \max_i N_{\tau i}$. Then, we can certainly write that

$$
K_{nc} - K_{\lim} = K_{nc} - K_n + K_n - K_{\lim}, \tag{F.1}
$$

By independence and definition of $K_n$, we have

$$
P(K_n \neq K_{\lim}) = P(K_n < K_{\lim}) = P(N_\tau < K_{\lim})^n \to 0
$$

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 (F.1). We have,

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

as $n \to \infty$. Here, we use that

$$
\sum_{i=1}^n \mathbb{1}\{N_{\tau i} \geq v\}
$$

diverges almost surely to $\infty$. Too see this, note that $\sum_{i=1}^n \mathbb{1}\{N_{\tau i} \geq v\}$ is almost surely monotone in $n$, and

$$\sum_{i=1}^n P(N_{\tau i} \geq v) = nP(N_\tau \geq v) \to \infty$$

From this and Kolmogorov's three series theorem, we conclude that

$$\sum_{i=1}^n \mathbb{1}\{N_{\tau i} \geq v\} \to \infty$$

almost surely as $n \to \infty$ and that

$$\sum_{i=1}^n \mathbb{1}\{N_{\tau i} \geq v\} \leq c$$

has probability tending to zero as $n \to \infty$ as desired.

# APPENDIX G

## G.1 Additional simulation results

### G.1.1 Tables

**Uncensored case**

**Varying effects (A on Y, L on Y, A on L, L on A)**

| $\beta_A^y$ | Estimator | Coverage | MSE | Bias | sd($\widehat{\Psi}_n$) | Mean($\widehat{\mathrm{SE}}$) |
|---|---|---|---|---|---|---|
| -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 time confounding (vary $\beta_A^y$).

| $\beta^y_L$ | **Estimator** | **Coverage** | **MSE** | **Bias** | **sd($\widehat{\Psi}_n$)** | **Mean($\widehat{\mathrm{SE}}$)** |
|---|---|---|---|---|---|---|
| -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 7: Results for the case with time confounding (vary $\beta^y_L$)

| $\beta^L_A$ | **Estimator** | **Coverage** | **MSE** | **Bias** | **sd($\widehat{\Psi}_n$)** | **Mean($\widehat{\mathrm{SE}}$)** |
|---|---|---|---|---|---|---|
| -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 8: Results for the case with time confounding (vary $\beta^L_A$)

| $\alpha_L$ | 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 9: Results for the case with time confounding (vary $\beta_A^L$)

**Sample size**

| $n$ | 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 ($n \in \{100, 200, 500, 1000\}$)

**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 $(\beta_A^y, \beta_L^y, \alpha_L, \lambda_c) =$ (-0.3, 0.5, -0.2, 0.0002).

| 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).

| 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).

| 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).

| 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 $(\beta^y_A, \beta^y_L, \alpha_L, \lambda_c) = (0, 0, 0, 0.0005)$.

| ICE model | Estimator | Cov. | MSE | Bias | sd($\widehat{\Psi}_n$) | Mean($\widehat{\text{SE}}$) |
|---|---|---|---|---|---|---|
| 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 $(\beta^y_A, \beta^y_L, \alpha_L, \lambda_c) = (0, 0, 0, 0.0008)$.

#### G.1.2 Boxplots

**Uncensored case**

**Varying effects (A on Y, L on Y, A on L, L on A)**

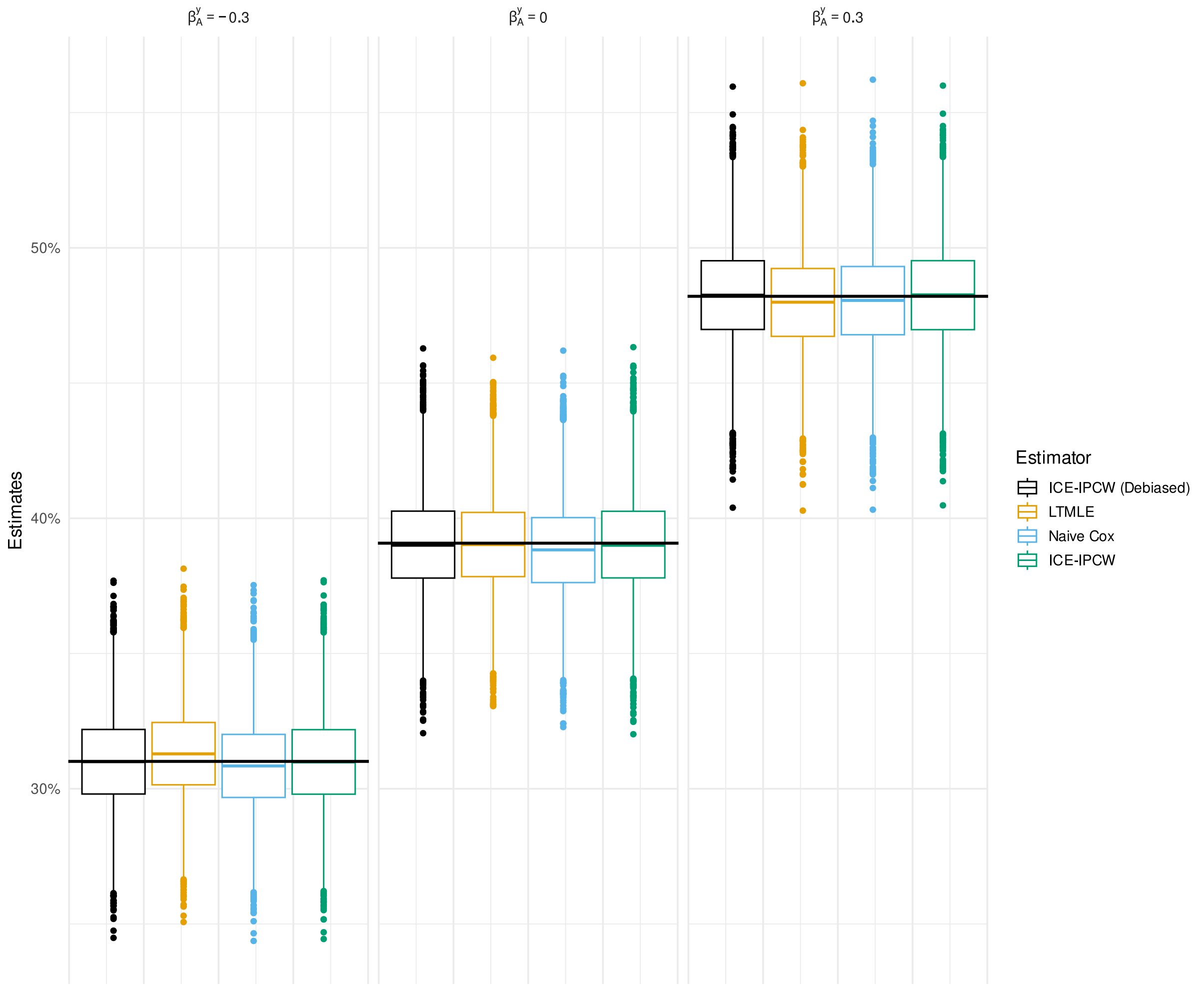


Figure 11: Boxplots of the results for the case with varying effect of $A$ on $Y$. The lines indicates the true value of the parameter.

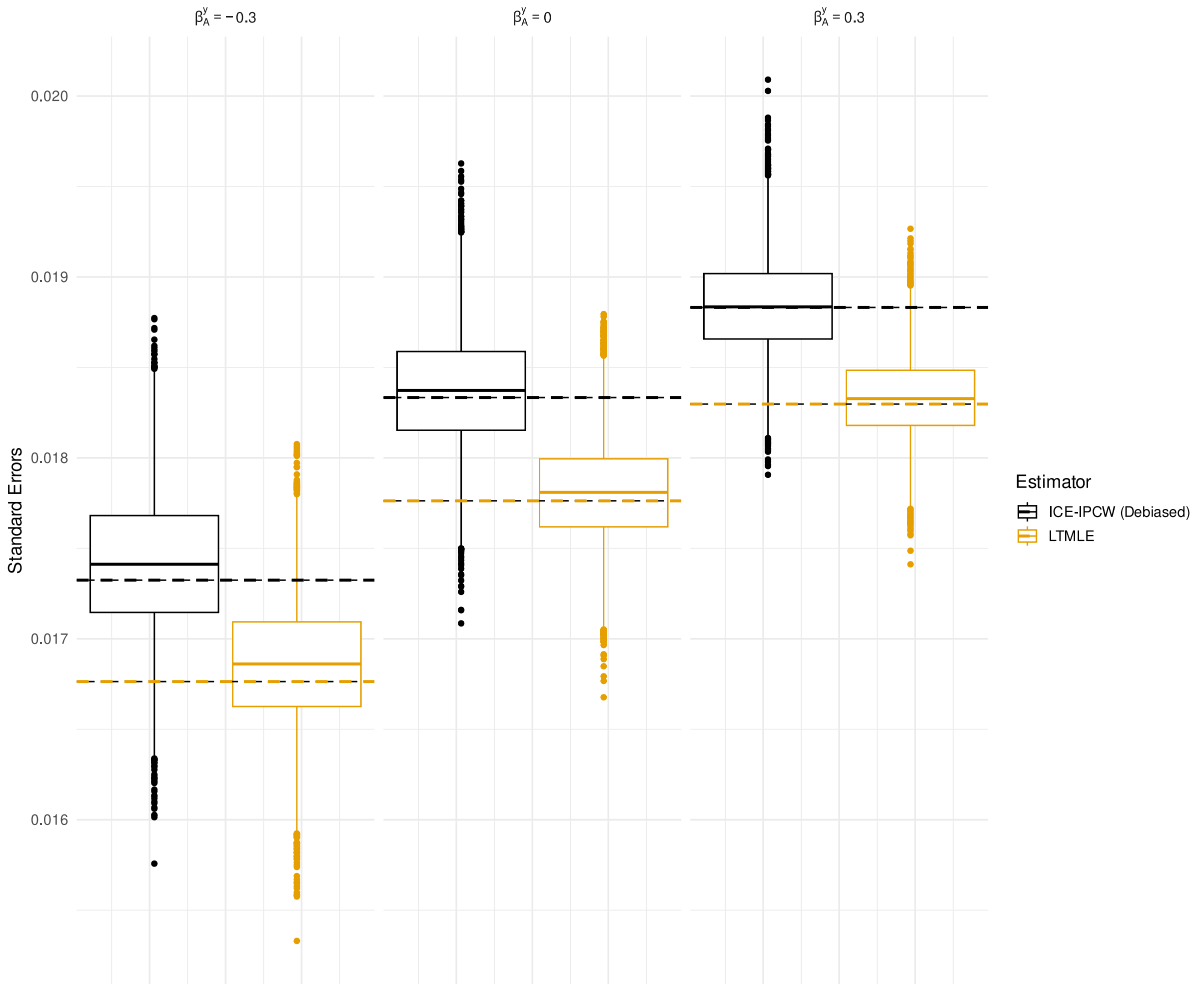


Figure 12: Boxplots of the standard errors for the case with varying effect of $A$ on $Y$. The red line indicates the empirical standard error of the estimates for each estimator.

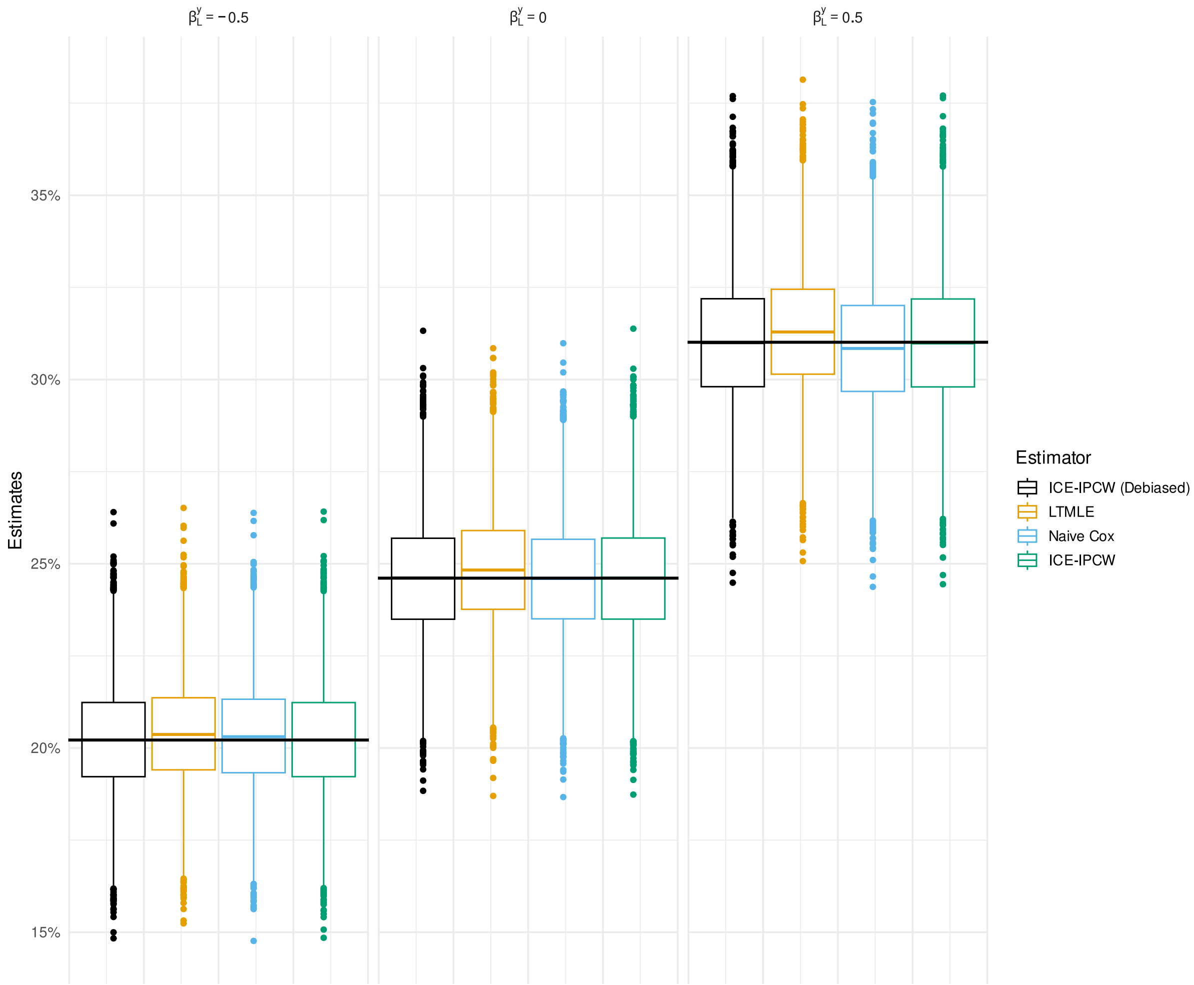


Figure 13: Boxplots of the results for the case with varying effect of $L$ on $Y$. The lines indicates the true value of the parameter.

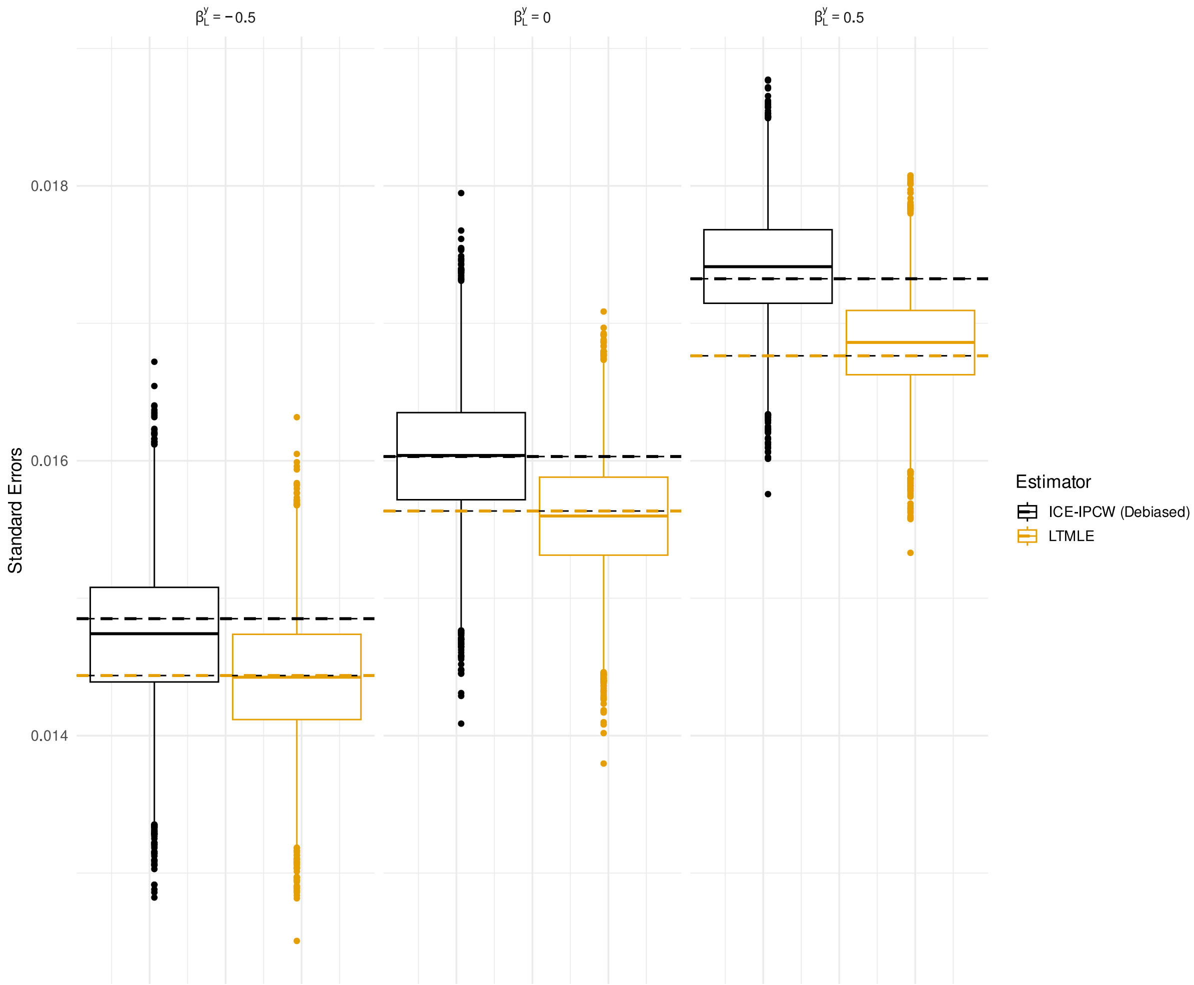


Figure 14: Boxplots of the standard errors for the case with varying effect of $L$ on $Y$. The red line indicates the empirical standard error of the estimates for each estimator.

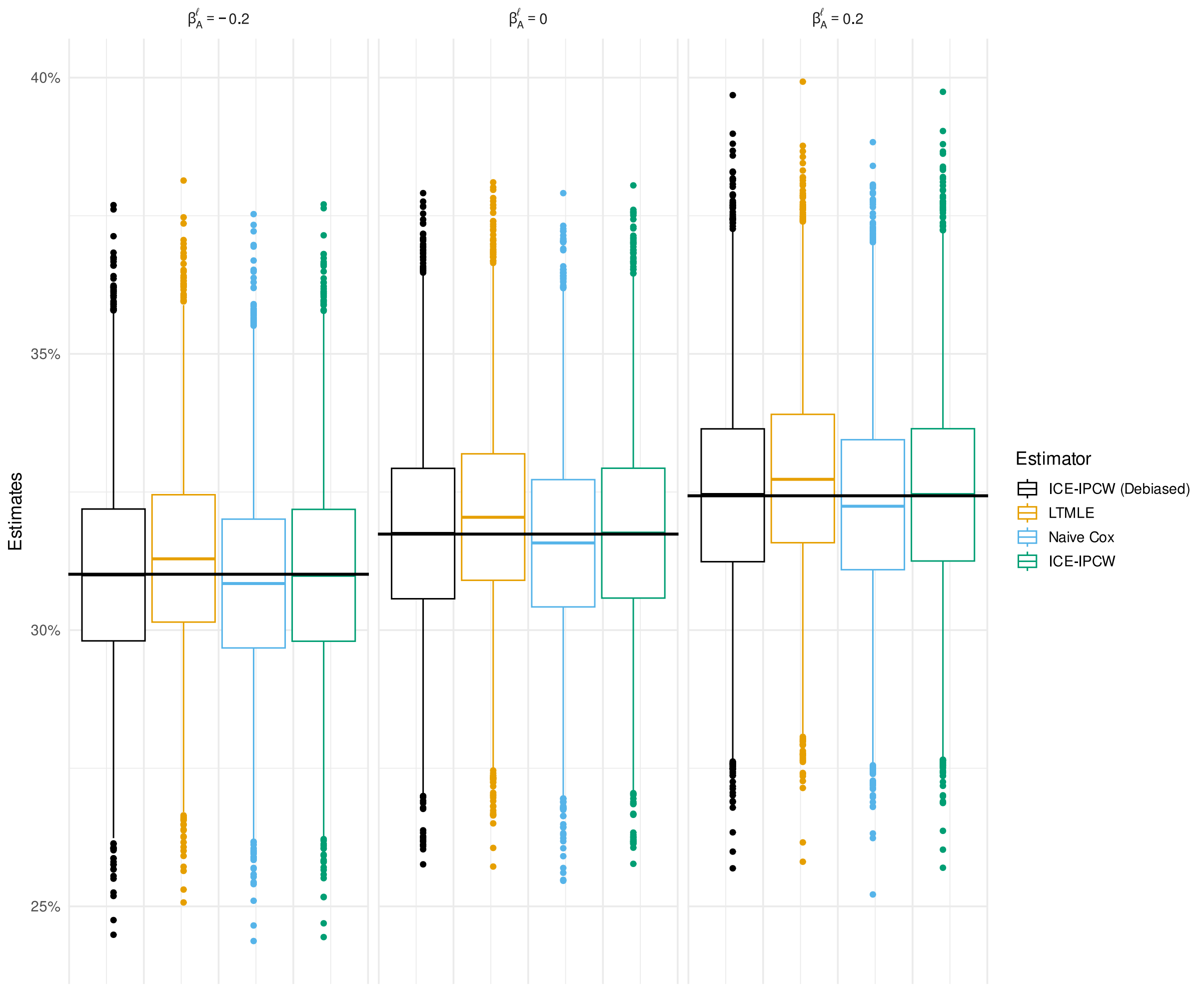


Figure 15: Boxplots of the results for the case with varying effect of $A$ on $L$. The lines indicates the true value of the parameter.

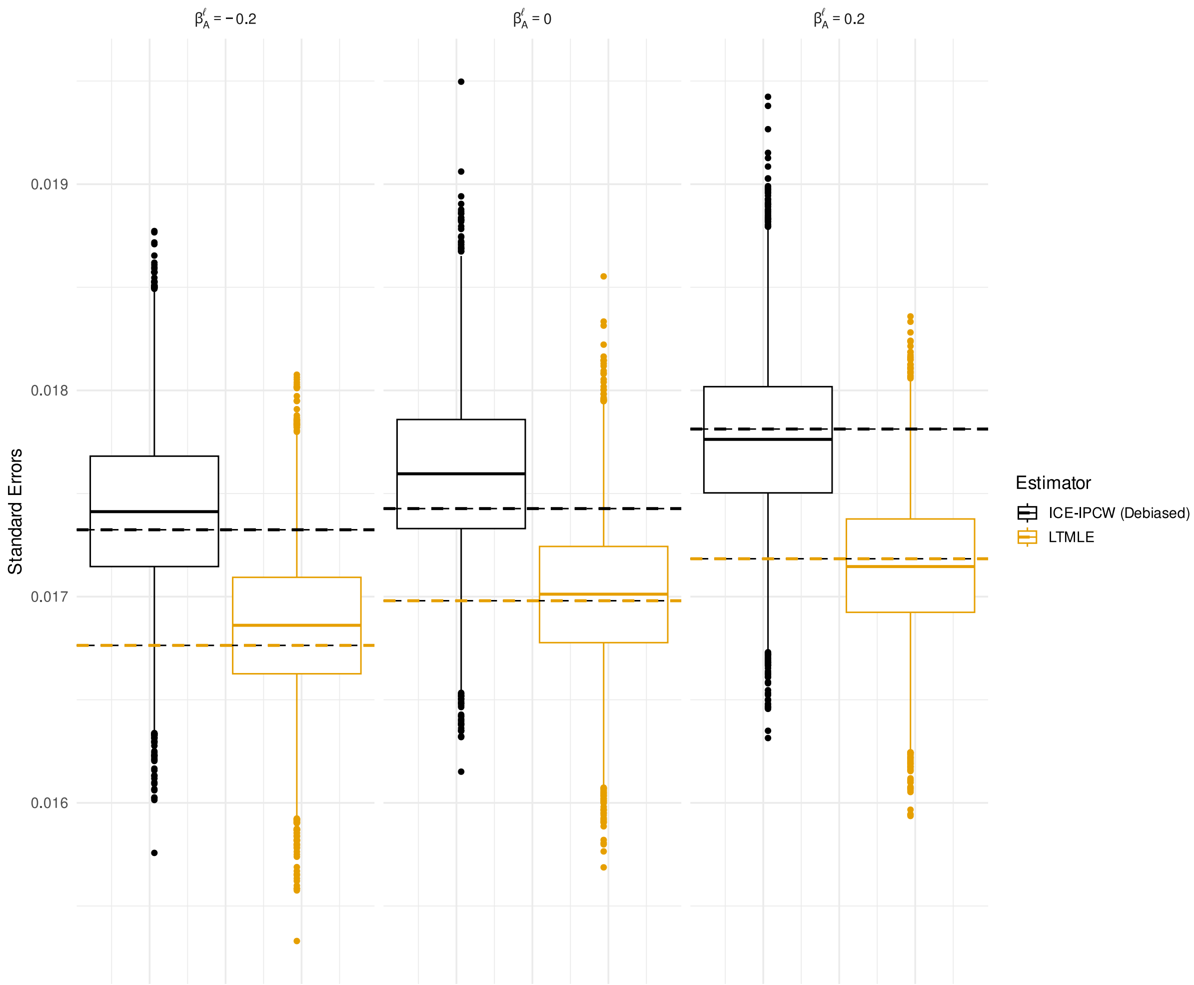


Figure 16: Boxplots of the standard errors for the case with varying effect of $A$ on $L$. The red line indicates the empirical standard error of the estimates for each estimator.

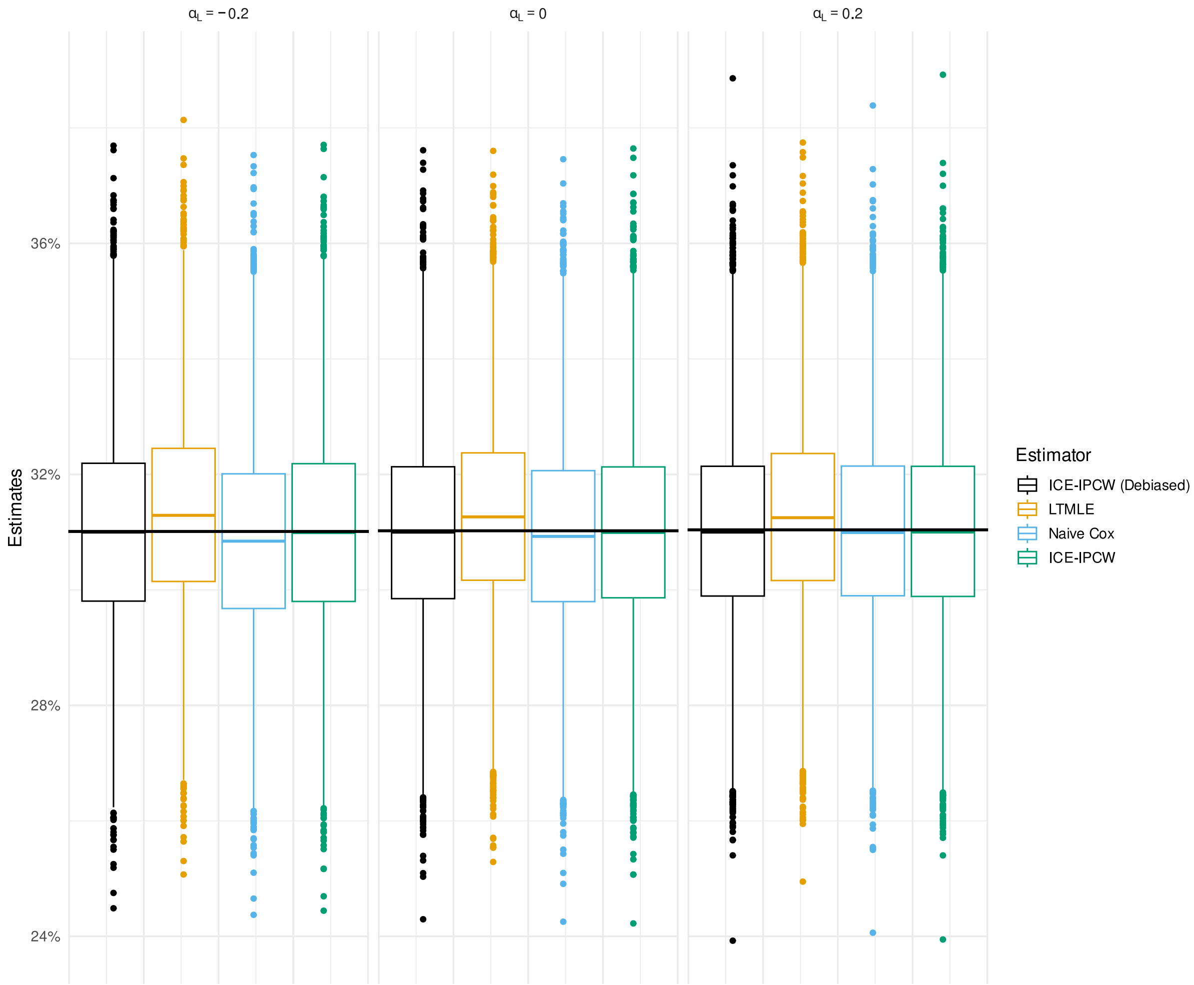


Figure 17: Boxplots of the results for the case with varying effect of $L$ on $A$. The lines indicates the true value of the parameter.

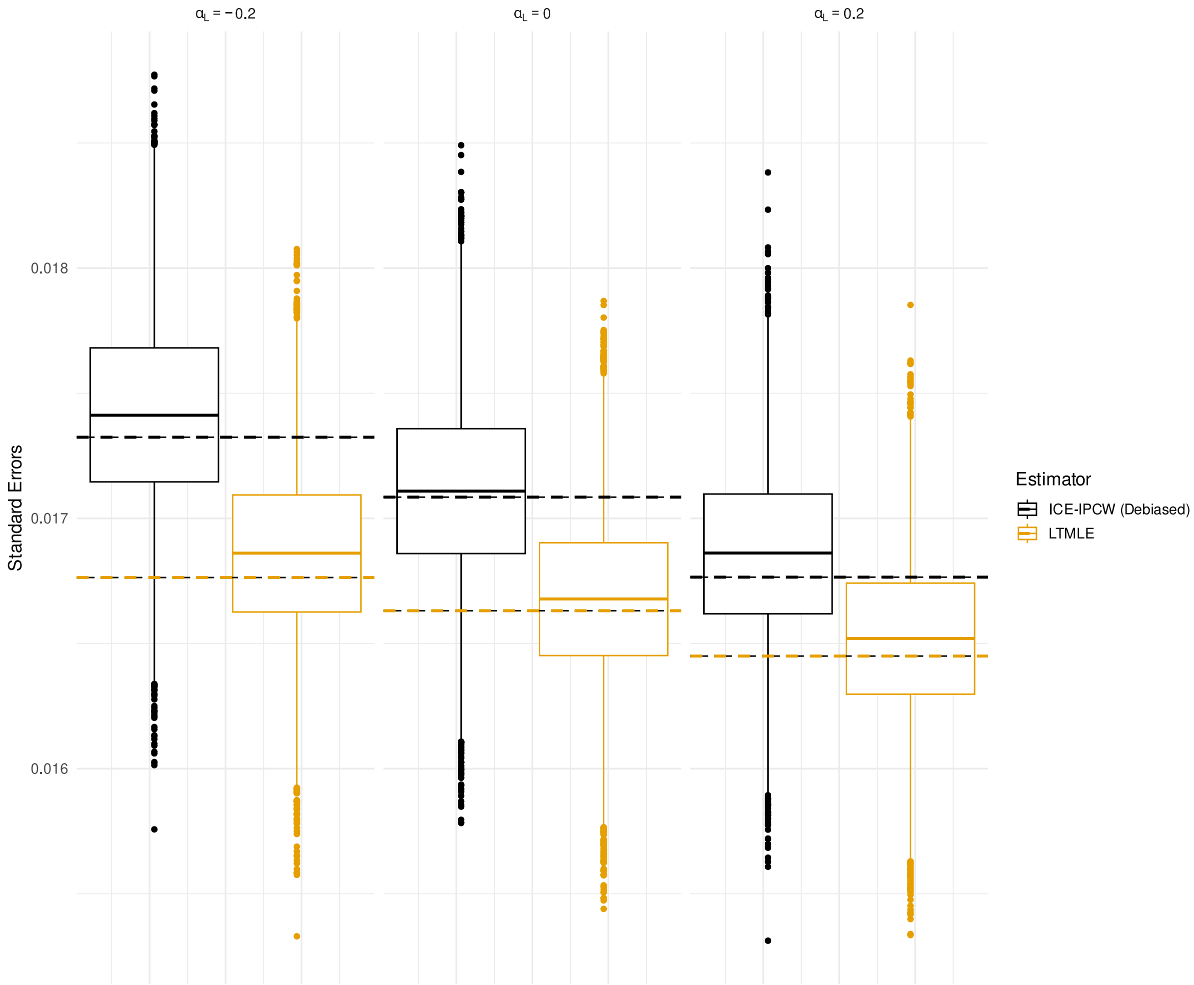


Figure 18: Boxplots of the standard errors for the case with varying effect of $L$ on $A$. The red line indicates the empirical standard error of the estimates for each estimator.

**Sample size**

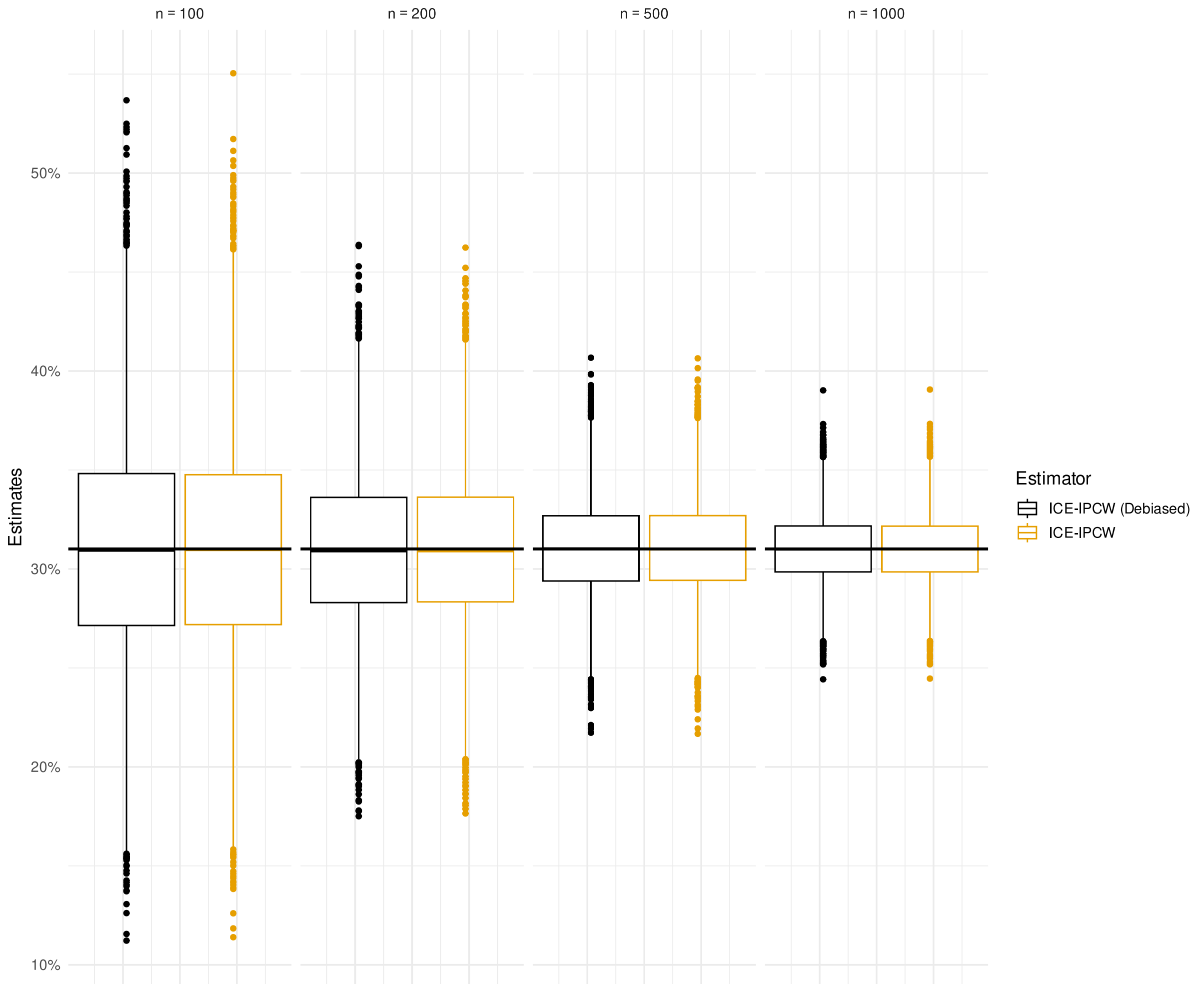


Figure 19: Boxplots of the results for the case with varying sample size. The lines indicates the true value of the parameter.

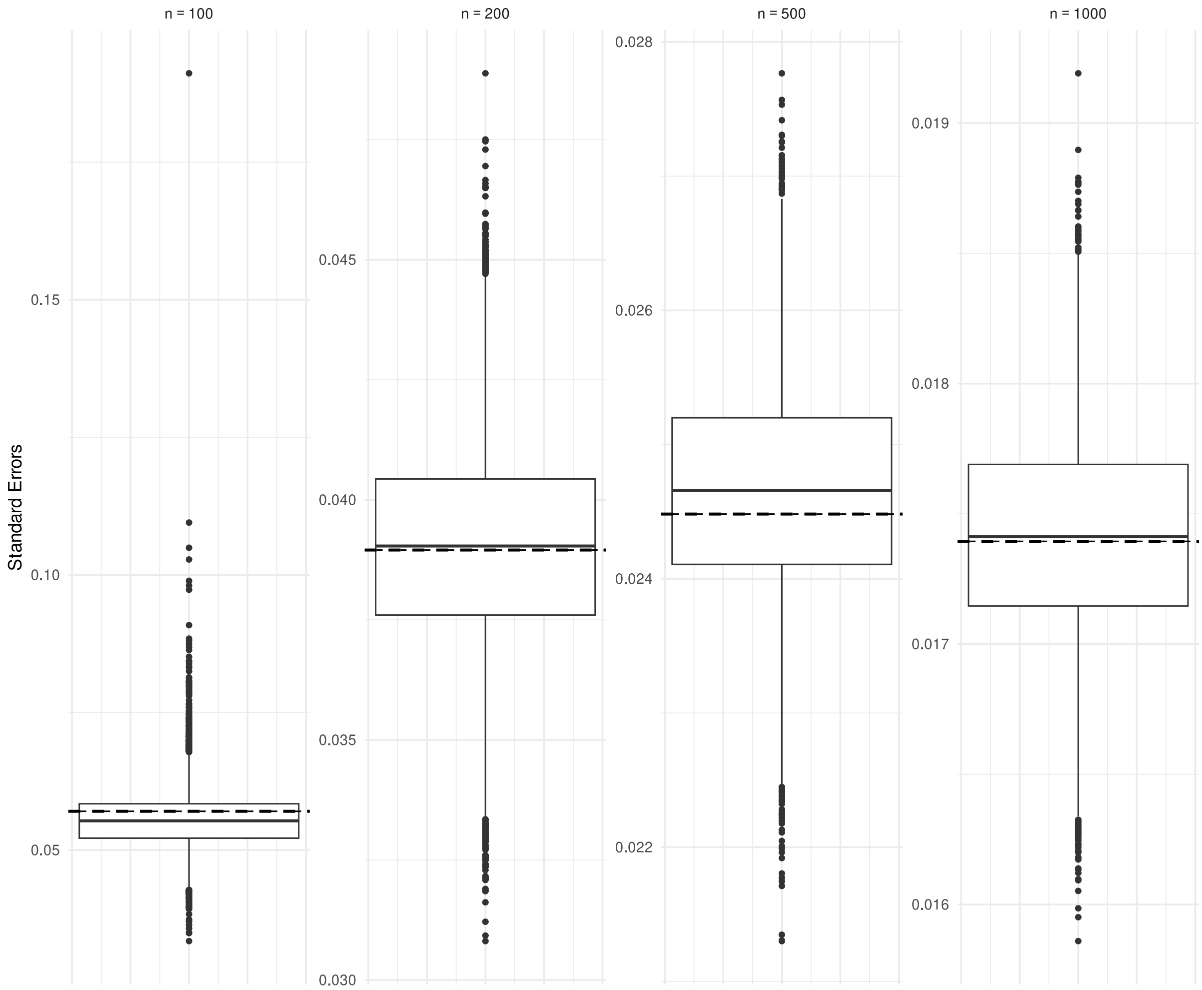


Figure 20: Boxplots of the standard errors for the case with varying sample size. The black line indicates the empirical standard error of the estimates for each estimator.