

---

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

---

**Johan Sebastian Ohlendorff**

johan.ohlendorff@sund.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.

# 1 TODO

- ☒ Clean up figures.
- ☒ Clean up existence of compensator + integral.
- ☒ identifiability. My potential outcome approach. Add figure for potential outcome processes. Show full identification formula without reweighting
- ☒ Censoring. Independent censoring IPCW rigorously.
- ☒ Consistency of estimator. Skip not done in other papers.
- ☐ Efficient influence function. Cleanup.
- ☐ Simulation study (ML?).
- ☐ Debiased estimator
- ☐ DR properties + ML rates/criteria (rate conditions + conditions for  $\hat{\nu}^*$ )
- ☒ Cross-fitting
- ☐ Discussion. Bayesian approach + pooling/rare events.

## 2 Introduction

In medical research, the estimation of causal effects of treatments over time is often of interest. We consider a longitudinal continuous-time setting that is very similar to [Rytgaard et al. \(2022\)](#) in which patient characteristics can change at subject-specific times. This is the typical setting of registry data, which usually contains precise information about when events occur, e.g., information about drug purchase history, hospital visits, and laboratory measurements. This approach offers an advantage over discretized methods, as it eliminates the need to select a time grid mesh for discretization, which can affect both the bias and variance of the resulting estimator. A continuous-time approach would adapt to the events in the data. Furthermore, continuous-time data captures more precise information about when events occur, which may be valuable in a predictive sense. Let  $\tau_{\text{end}}$  be the end of the observation period. We will focus on the estimation of the interventional cumulative incidence function in the presence of time-varying confounding at a specified time horizon  $\tau < \tau_{\text{end}}$ .

**Assumption 1** (Bounded number of events): In the time interval  $[0, \tau_{\text{end}}]$  there are at most  $K - 1 < \infty$  many changes of treatment and covariates in total for a single individual. We let  $K - 1$  be given by the maximum number of non-terminal events for any individual in the data.

**Assumption 2** (No simultaneous jumps): The counting processes  $N^a$ ,  $N^\ell$ ,  $N^y$ ,  $N^d$ , and  $N^c$  have with probability 1 no jump times in common.

Let  $\kappa_i(\tau)$  be the number of events for individual  $i$  up to time  $\tau$ . In [Rytgaard et al. \(2022\)](#), the authors propose a continuous-time LTMLE for the estimation of causal effects in which a single step of the targeting procedure must update each of the nuisance estimators  $\sum_{i=1}^n \kappa_i(\tau)$  times. We propose an estimator where the number of nuisance parameters is reduced to  $\sim \max_i \kappa_i(\tau)$  in total, and, in principle, only one step of the targeting procedure is needed to update all nuisance parameters. We provide an iterative conditional expectation formula that, like [Rytgaard et al. \(2022\)](#), iteratively updates the nuisance parameters. The key difference is that the estimation of the nuisance parameters can be performed by going back in the number of events instead of going back in time. The different approaches are illustrated in [Figure 2](#) and [Figure 3](#) for an outcome  $Y$  of interest. Moreover, we argue that the nuisance components can be estimated with existing machine learning algorithms from the

survival analysis and point process literature. As always let  $(\Omega, \mathcal{F}, P)$  be a probability space on which all processes and random variables are defined.

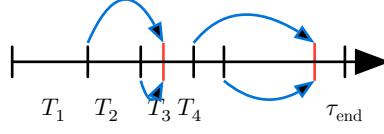


Figure 1: The “usual” approach where time is discretized. Each event time and its corresponding mark is rolled forward to the next time grid point, that is the values of the observations are updated based on the on the events occuring in the previous time interval.

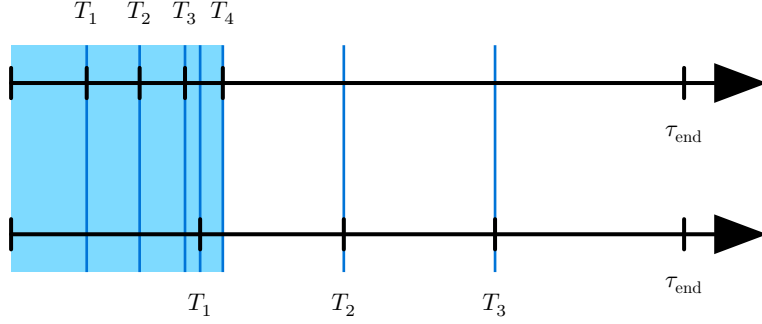


Figure 2: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let  $t_1 < \dots < t_m$  be all the event times in the sample. Then, given  $\mathbb{E}_Q[Y | \mathcal{F}_{t_r}]$ , we regress back to  $\mathbb{E}_Q[Y | \mathcal{F}_{t_{r-1}}]$  (through multiple regressions).

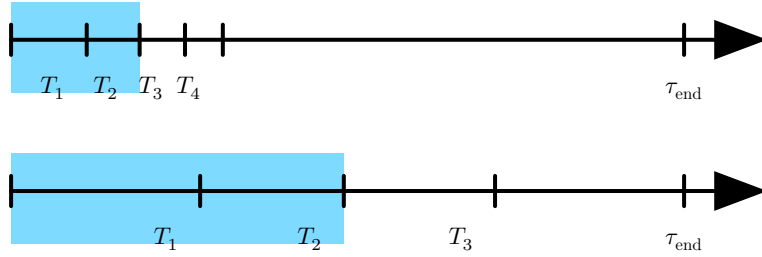


Figure 3: The figure illustrates the sequential regression approach proposed in this article. For each event  $k$  in the sample, we regress back on the history  $\mathcal{F}_{T_{(k-1)}}$ . That is, given  $\mathbb{E}_Q[Y | \mathcal{F}_{T_{(k)}}]$ , we regress back to  $\mathbb{E}_Q[Y | \mathcal{F}_{T_{(k-1)}}]$ . In the figure,  $k = 3$ .

### 3 Setting and Notation

First, we assume that at baseline, we observe the treatment  $A_0$  and the time-varying confounders at time 0,  $L_0$ . The time-varying confounders can in principle include covariates which do not change over time, but for simplicity of notation, we will include them among those that do change over time. Also, we assume that we have two treatment options,  $A(t) = 0, 1$  (e.g., placebo and active treatment). The time-varying confounders and treatment are assumed to take values in  $\mathbb{R}^m$  and  $\mathbb{R}$ , and that  $L(t) : \Omega \rightarrow \mathbb{R}^m$  and  $A(t) : \Omega \rightarrow \mathbb{R}$  are measurable for each  $t \geq 0$ , respectively. These processes are assumed to be càdlàg, i.e., right-continuous with left limits. Furthermore, the times at which the treatment and covariates may only change at the jump times of the counting processes  $N^a$  and  $N^\ell$ , respectively which makes  $L(t)$  and  $A(t)$  into jump processes (Last & Brandt (1995)). The jump times of these counting processes thus represent visitation times.

We are interested in the cumulative incidence function, so we also observe  $N^y$  and  $N^d$  corresponding to the counting processes for the primary and competing event, respectively. The outcome of interest is thus  $N_t^y$ . We initially assume no censoring, but we will later include it. We assume that the jump times differ with probability 1 (Assumption 2). Moreover, we assume that only a bounded number of events occur for each individual in the time interval  $[0, \tau_{\text{end}}]$  (Assumption 1). Thus, we have observations from a multivariate jump process  $(N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t))$ , and let  $\mathcal{F}_t = \sigma((N_s^a, A(s), N_s^\ell, L(s), N_s^y, N_s^d) \mid s \leq t)$  be the natural filtration generated by the process up to time  $t$ . We observe  $\mathcal{F}_{T_{(K)}} = (T_{(K)}, \Delta_{(K)}, T_{(K-1)}, \Delta_{(K-1)}, A(T_{(K-1)}), L(T_{(K-1)}), \dots, A_0, L_0) \sim P \in \mathcal{M}$ , where  $T_{(k)}$  may be possibly  $\infty$  and  $\mathcal{M}$  is the nonparametric set of probability measures. We assume tacitly that the last event has to be a terminal event, i.e.,  $\Delta_{(K)} = y$  or  $d$ . Let  $\pi_k(da, \mathcal{F}_{T_{(k-1)}})$  be the distribution the treatment value at the  $k$ 'th event  $\mathcal{F}_{T_{(k-1)}}$  given that the event time is  $t$  and that the  $k$ 'th event is a visitation time. Let  $\mu_k(d\ell, t, \mathcal{F}_{T_{(k-1)}})$  be the density of the covariate value at the  $k$ 'th event given  $\mathcal{F}_{T_{(k-1)}}$  given that the event time is  $t$  and that the  $k$ 'th event is a covariate event. Let  $\lambda_{k-1}^x(t, \mathcal{F}_{T_{(k-1)}})$  be the hazard of the  $k$ 'th event at time  $t$  given  $\mathcal{F}_{T_{(k-1)}}$ .

**Assumption 3** (Conditional distributions of jumps and marks): We assume that the conditional distributions  $P(T_{(k)} \in \cdot \mid \mathcal{F}_{T_{(k-1)}}) \ll m$   $P$ -a.s., and  $P(A(T_{(k)}) \in \cdot \mid T_{(k)} = t, \Delta_{(k)} = a, \mathcal{F}_{T_{(k-1)}}) \ll \nu_a$   $P$ -a.s. and  $P(L(T_{(k)}) \in \cdot \mid T_{(k)} = t, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}) \ll \nu_\ell$   $P$ -a.s., where  $m$  is the Lebesgue measure on  $\mathbb{R}_+$ ,  $\nu_a$  is a measure on  $\mathcal{A}$ , and  $\nu_\ell$  is a measure on  $\mathcal{L}$ .

## 4 Target parameter and potential outcome framework

We now take an interventionist stance to causal inference such as the one given in [Ryalen \(2024\)](#). In the interventionist school of thought, one tries to emulate a randomized controlled trial. In the continuous-time longitudinal setting, this can e.g., correspond to a trial in which there is perfect compliance. Our approach is inspired by the conditions of [Ryalen \(2024\)](#), stating the conditions without martingale theory. Moreover, the identification conditions do not require the existence of an entire potential outcome process, but just a random variable at the time horizon of interest. While our theory provides a potential outcome framework, it is unclear if they are related local independence graphs, and as such the conditions may be hard to justify. For simplicity, we presuppose that there are two treatment levels (0/1) and that we are only interested in the effect of staying on treatment ( $A(t) = 1$  for all  $t > 0$ ) and starting on treatment ( $A(0) = 1$ ).

**Definition 1** (Target parameter): Our target parameter  $\Psi_\tau^g : \mathcal{M} \rightarrow \mathbb{R}$  is the mean interventional potential outcome at time  $\tau$ ,

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

where  $\tilde{N}_t^y$  denotes the potential outcome had the treatment protocol been followed at time  $t$ .

We now define the stopping time  $T^a$  as the time of the first visitation event where the treatment plan is not followed, i.e.,

$$T^a = \inf_{t \geq 0} \{A(t) = 0\} = \inf_{k \geq 1} \{T_{(k)} \mid \Delta_{(k)} = a, A(T_{(k)}) \neq 1\} \wedge \infty \mathbb{1}\{A(0) = 1\}$$

where we use that  $\infty \cdot 0 = 0$ .

**Theorem 1:** We suppose that there exists a potential outcome  $\tilde{Y}_\tau = \tilde{N}_\tau^y$  at time  $\tau$  such that

- **Consistency:**  $\tilde{Y}_\tau \mathbb{1}\{T^a > \tau\} = Y_\tau \mathbb{1}\{T^a > \tau\}$ .
- **Exchangeability:** We have

$$\begin{aligned} \tilde{Y}_\tau \mathbb{1}\{T_{(j)} \leq \tau < T_{(j+1)}\} &\perp A(T_{(k)}) \mid \mathcal{F}_{T_{(k)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \geq k > 0, \\ \tilde{Y}_\tau \mathbb{1}\{T_{(j)} \leq \tau < T_{(j+1)}\} &\perp A_0 \mid L_0, \quad \forall j \geq 0. \end{aligned} \tag{1}$$

- **Positivity:** The measure given by  $dR = WdP$  where  $W_t^* = \prod_{k=1}^{N_t} \left( \frac{\mathbb{1}\{A(T_{(k)})=1\}}{\pi_k(A(T_{(k)}), \mathcal{F}_{T_{(k-1)}})} \right)^{\mathbb{1}\{\Delta_{(k)}=a\}} \frac{\mathbb{1}\{A_0=1\}}{\pi_0(A_0, L(0))}$  is a probability measure, where  $N_t = \sum_{k=1}^{K-1} \mathbb{1}\{T_{(k)} \leq t\}$ .

Then the estimand of interest is identifiable by

$$\Psi_\tau^g(P) = \mathbb{E}_P[\tilde{Y}_\tau] = \mathbb{E}_P[Y_\tau W_\tau].$$

*Proof:* Write  $\tilde{Y}_t = \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau$ . The theorem is shown if we can prove that  $\mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau] = \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} Y_\tau W_\tau]$  by linearity of expectation. We have that for  $k \geq 1$ ,

$$\begin{aligned}
\mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} Y_\tau W_\tau] &= \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \mathbb{1}\{T^a > \tau\} Y_\tau W_\tau] \\
&= \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \mathbb{1}\{T^a > \tau\} \tilde{Y}_\tau W_\tau] \\
&= \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau W_\tau] \\
&= \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau W_{T_{(k-1)}}] \\
&= \mathbb{E}_P \left[ \mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau \left( \frac{\mathbb{1}\{A(T_{(k-1)}) = 1\}}{\pi_k(A(T_{(k-1)}), T_{(k-1)}, \mathcal{F}_{T_{(k-2)}})} \right)^{\mathbb{1}\{\Delta_{(k-1)}=a\}} W_{T_{(k-2)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}, A(T_{(k-1)})] \right. \\
&\quad \times \left. \left( \frac{\mathbb{1}\{A(T_{(k-1)}) = 1\}}{\pi_k(A(T_{(k-1)}), T_{(k-1)}, \mathcal{F}_{T_{(k-2)}})} \right)^{\mathbb{1}\{\Delta_{(k-1)}=a\}} W_{T_{(k-2)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}] \right. \\
&\quad \times \left. \left( \frac{\mathbb{1}\{A(T_{(k-1)}) = 1\}}{\pi_k(A(T_{(k-1)}), T_{(k-1)}, \mathcal{F}_{T_{(k-2)}})} \right)^{\mathbb{1}\{\Delta_{(k-1)}=a\}} W_{T_{(k-2)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}] \right. \\
&\quad \times \left. \mathbb{E}_P \left[ \left( \frac{\mathbb{1}\{A(T_{(k-1)}) = 1\}}{\pi_k(A(T_{(k-1)}), T_{(k-1)}, \mathcal{F}_{T_{(k-2)}})} \right)^{\mathbb{1}\{\Delta_{(k-1)}=a\}} \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)} \right] W_{T_{(k-2)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}] W_{T_{(k-2)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-3)}}, \Delta_{(k-2)}, T_{(k-2)}, A(T_{(k-2)})] W_{T_{(k-2)}} \right]
\end{aligned}$$

Iteratively applying the same argument, we get that  $\mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} \tilde{Y}_\tau] = \mathbb{E}_P[\mathbb{1}\{T_{(k-1)} \leq \tau < T_{(k)}\} Y_\tau W_\tau]$  as needed.  $\square$

For the intersection property of conditional independence, a sufficient condition for [Equation 1](#) is that

$$\begin{aligned}
&\tilde{Y}_\tau \perp A(T_{(k)}) \mid T_{(j)} \leq \tau < T_{(j+1)}, \mathcal{F}_{T_{(k)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \geq k > 0, \\
&\mathbb{1}\{T_{(j)} \leq \tau < T_{(j+1)}\} \perp A(T_{(k)}) \mid \tilde{Y}_\tau, \mathcal{F}_{T_{(k)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \geq k > 0.
\end{aligned}$$

The second condition may in particular be too strong to require. Alternatively, it is possible to posit the existence of a potential outcome for each event separately and get the same conclusion. The overall exchangeability condition may be stated differently. Specifically, let  $\tilde{Y}_{\tau,k}$  be the potential outcome at event  $k$  corresponding to  $\mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = y\}$ . Then the exchangeability condition is that  $\tilde{Y}_{\tau,k} \perp A(T_{(j)}) \mid \mathcal{F}_{T_{(j-1)}}, T_{(j)}, \Delta_{(j)} = a$  for  $0 \leq j < k$  and  $k = 1, \dots, K$ . However, it has been noted (Gill & Robins, 2001) in discrete time that the existence of multiple potential outcomes can be restrictive and that the resulting exchangeability condition may be too strong. An illustration of the consistency condition is given in [Figure 4](#).

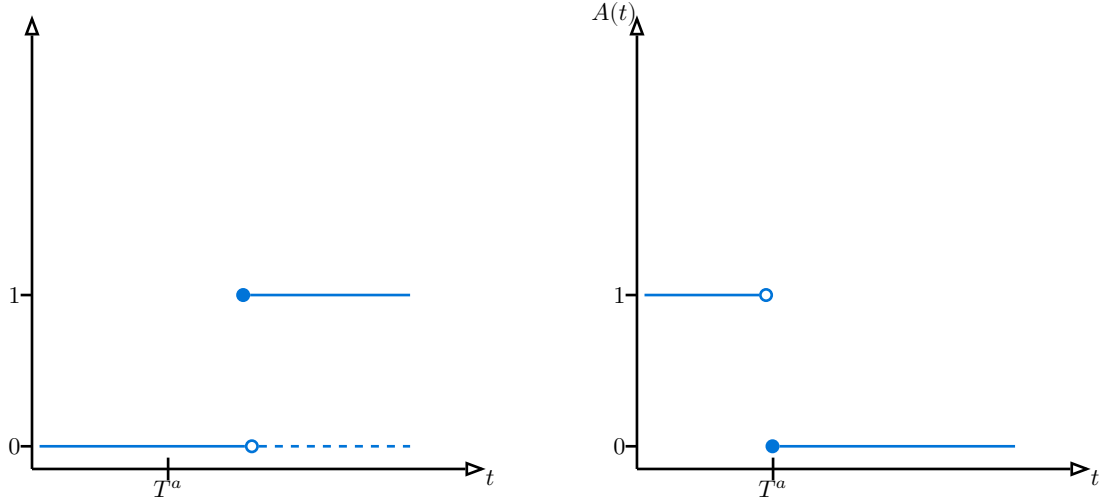


Figure 4: The figure illustrates the consistency condition for the potential outcome framework. The left panel shows the potential outcome process  $\tilde{Y}_t$  (dashed) and the observed process  $Y_t$  (solid). The right panel shows the treatment process  $A(t)$ .

We are now ready to give an iterated conditional expectations formula for the target parameter in the case with no censoring. The formula is given in Theorem 2, and we will use this to find an algorithm for estimating the target parameter and the terms appearing in the efficient influence function. First, and foremost though, we will remark that this can be used to write down the target parameter directly in terms of the event-specific cause-specific hazards and the density for the covariate process.

**Theorem 2:** Let  $W_{k,j} = \frac{W_{T(j)}}{W_{T(k)}}$  for  $k < j$  (defining  $\frac{0}{0} = 0$ ). Let  $\bar{Q}_K = \mathbb{1}\{T_{(K)} \leq \tau, \Delta_{(K)} = y\}$  and  $\bar{Q}_k = \mathbb{E}_P \left[ \sum_{j=k+1}^K W_{k,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k)}} \right]$ . Then,

$$\begin{aligned} \bar{Q}_{k-1} &= \mathbb{E}_P \left[ \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = \ell\} \bar{Q}_k \left( A(T_{(k-1)}), L(T_{(k)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}} \right) \right. \\ &\quad \left. + \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = a\} \mathbb{E}_P \left[ W_{k-1,k} \bar{Q}_k \left( A(T_{(k)}), L(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}} \right) \mid T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}} \right] \right. \\ &\quad \left. + \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = y\} \mid \mathcal{F}_{T_{(k-1)}} \right] \end{aligned}$$

and

$$\begin{aligned} \bar{Q}_{k-1} &= \mathbb{E}_P \left[ \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = \ell\} \bar{Q}_k \left( A(T_{(k-1)}), L(T_{(k)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}} \right) \right. \\ &\quad \left. + \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = a\} \bar{Q}_k \left( 1, L(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}} \right) \right. \\ &\quad \left. + \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = y\} \mid \mathcal{F}_{T_{(k-1)}} \right] \end{aligned}$$

for  $k = K, \dots, 1$ . Thus,  $\Psi_\tau^g(P) = \mathbb{E}_P \left[ \sum_{k=1}^K W_{T_{(k-1)}} \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = y\} \right] = \mathbb{E}_P [\bar{Q}_0 W_0] = \mathbb{E}_P [\mathbb{E}_P [\bar{Q}_0 \mid A_0 = 1, L_0]]$ .

*Proof:* First note that the  $\bar{Q}_k$  only need to be evaluated when the person is at risk. For the first claim, we find

$$\begin{aligned}
\bar{Q}_k &= \mathbb{E}_P \left[ \sum_{j=k+1}^K W_{k,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{E}_P \left[ \mathbb{E}_P \left[ \sum_{j=k+1}^K W_{k,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}} \right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}} \right] \mid \mathcal{F}_{T_{(k)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} W_{k,k+1} \right. \\
&\quad \left. + \mathbb{E}_P \left[ W_{k,k+1} \mathbb{E}_P \left[ \sum_{j=k+2}^K W_{k+1,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}} \right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}} \right] \mid \mathcal{F}_{T_{(k)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} W_{k,k+1} \right. \\
&\quad + \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = a\} \mathbb{E}_P \left[ W_{k,k+1} \mathbb{E}_P \left[ \sum_{j=k+2}^K W_{k+1,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}} \right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}} \right] \\
&\quad \left. + \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = \ell\} \mathbb{E}_P \left[ W_{k,k+1} \mathbb{E}_P \left[ \sum_{j=k+2}^K W_{k+1,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}} \right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}} \right] \mid \mathcal{F}_{T_{(k)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} \right. \\
&\quad + \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = a\} \mathbb{E}_P \left[ W_{k,k+1} \mathbb{E}_P \left[ \sum_{j=k+2}^K W_{k+1,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}} \right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}} \right] \\
&\quad \left. + \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = \ell\} \mathbb{E}_P \left[ \mathbb{E}_P \left[ \sum_{j=k+2}^K W_{k+1,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}} \right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}} \right] \mid \mathcal{F}_{T_{(k)}} \right] \\
&= \mathbb{E}_P \left[ \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} \right. \\
&\quad + \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = a\} \mathbb{E}_P \left[ W_{k,k+1} \mathbb{E}_P \left[ \sum_{j=k+2}^K W_{k+1,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}} \right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}} \right] \\
&\quad \left. + \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = \ell\} \mathbb{E}_P \left[ \sum_{j=k+2}^K W_{k+1,j} \mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}} \right] \mid \mathcal{F}_{T_{(k)}} \right]
\end{aligned}$$

by the law of iterated expectations and that

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

for all  $j = 1, \dots, k-1$  and  $k = 1, \dots, K$ . The second statement simply follows from the fact that

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

by the law of iterated expectations.  $\square$



**Theorem 3:** Let  $\bar{Q}_k(Q)$  be defined as  $\bar{Q}_{k,\tau}^g$  in Theorem 2. Denote by  $S_k(s - | \mathcal{F}_{T_{(k-1)}}) = \prod_{s \in (T_{(k-1)}, t)} \left(1 - \sum_{x=\ell, a, d, y} \lambda_{k-1}^x(s, \mathcal{F}_{T_{(k-1)}}) ds\right)$  the survival function of the  $k$ 'th event at time  $s$  given  $\mathcal{F}_{T_{(k-1)}}$ . Then, we have

$$\begin{aligned} p_{ka}(t | \mathcal{F}_{T_{(k-1)}}) &= \int_{\mathcal{F}_{T_{(k-1)}}}^t S_k(s - | \mathcal{F}_{T_{(k-1)}}) \bar{Q}_{k+1,\tau}^g(1, L(T_{(k-1)}), s, a, \mathcal{F}_{T_{(k-1)}}) \lambda_k^a(s, \mathcal{F}_{T_{(k)}}) ds \\ p_{k\ell}(t | \mathcal{F}_{T_{(k-1)}}) &= \int_{\mathcal{F}_{T_{(k-1)}}}^t S_k(s - | \mathcal{F}_{T_{(k-1)}}) \left( \mathbb{E}_P \left[ \bar{Q}_{k+1,\tau}^g(A(T_{(k-1)}), L(T_{(k)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \lambda_k^\ell(s, \mathcal{F}_{T_{(k)}}) ds \\ p_{ky}(t | \mathcal{F}_{T_{(k-1)}}) &= \int_{\mathcal{F}_{T_{(k-1)}}}^t S_k(s - | \mathcal{F}_{T_{(k-1)}}) \lambda_k^y(s, \mathcal{F}_{T_{(k)}}) ds \end{aligned}$$

and we can identify  $\bar{Q}_{k,\tau}^g$  via the intensities as

$$\bar{Q}_{k,\tau}^g = p_{ka}(\tau | \mathcal{F}_{T_{(k-1)}}) + p_{k\ell}(\tau | \mathcal{F}_{T_{(k-1)}}) + p_{ky}(\tau | \mathcal{F}_{T_{(k-1)}}) \quad (2)$$

*Proof:* To prove the theorem, we simply have to find the conditional density of  $(T_{(k)}, \Delta_{(k)})$  given  $\mathcal{F}_{T_{(k-1)}}$ . First note that we can write,

$$\begin{aligned} \bar{Q}_{k-1} &= \mathbb{E}_P \left[ \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = \ell\} \mathbb{E}_P \left[ \bar{Q}_k(A(T_{(k-1)}), L(T_{(k)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)}, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right. \\ &\quad + \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = a\} \bar{Q}_k(1, L(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \\ &\quad \left. + \mathbb{1}\{T_{(k)} \leq \tau, \Delta_{(k)} = y\} \mid \mathcal{F}_{T_{(k-1)}} \right] \end{aligned} \quad (3)$$

Since  $\lambda_{k-1}^x(t, \mathcal{F}_{T_{(k-1)}})$  is the cause-specific hazard of the  $k$ 'th event at time  $t$  given  $\mathcal{F}_{T_{(k-1)}}$  and that the event was of type  $x$ , it follows that the conditional density of  $(T_{(k)}, \Delta_{(k)})$  given  $\mathcal{F}_{T_{(k-1)}}$  is given by

$$p(t, d | \mathcal{F}_{T_{(k-1)}} = f_{k-1}) = \prod_{s \in (T_{(k-1)}, t)} \left( 1 - \sum_{x=\ell, a, d, y} \lambda_{k-1}^x(s, \mathcal{F}_{T_{(k-1)}}) ds \right) \lambda_{k-1}^d(t, \mathcal{F}_{T_{(k-1)}}).$$

Putting this into the expectation of Equation 3, we get the claim.  $\square$

## 5 Censoring

In this section, we introduce a right-censoring time  $C > 0$  at which we stop observing the multivariate jump process  $Z(t) = (N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t))$ . Denote by  $N^c(t)$  the counting process for the censoring process and its filtration  $\mathcal{G}_t = \sigma(N^c(s) \mid s \leq t)$ . Let  $T^e$  further denote the terminal event time,  $T^e = \inf_{t > 0} \{N^y(t) + N^d(t) = 1\}$ . Then we can view the censoring as being coarsened by the terminal event time  $T^e$ . The full data filtration is therefore given by

$$\mathcal{F}_t^{\text{full}} = \mathcal{F}_t \vee \mathcal{G}_t$$

Let  $\lambda_{k-1}^c(t, \mathcal{F}_{T_{(k-1)}})$  be the cause-specific hazard of the  $k$ 'th event at time  $t$  given the full history and that the event was a censoring event and define correspondingly  $S^c(t - | \mathcal{F}_{T_{(k-1)}}) = \prod_{s \in (T_{(k-1)}, t)} \left( 1 - \lambda_{k-1}^c(s, \mathcal{F}_{T_{(k-1)}}) ds \right)$  the censoring survival function. Unfortunately, we only ever fully observe the process  $t \mapsto (Z(t \wedge C), N^c(t \wedge T^e))$  which is adapted to the filtration  $\mathcal{F}_{t \wedge C \wedge T^e}^{\text{full}} \subseteq \mathcal{F}_t^{\text{full}}$ . From this, we get the observed data,

$$\begin{aligned}
\bar{T}_k &= C \wedge T_{(k)} \\
\bar{D}_k &= \begin{cases} \Delta_{(k)} & \text{if } C > T_{(k)} \\ c & \text{otherwise} \end{cases} \\
A(\bar{T}_k) &= \begin{cases} A(T_{(k)}) & \text{if } C > T_{(k)} \\ A(T_{(k-1)}) & \text{otherwise} \end{cases} \\
L(\bar{T}_k) &= \begin{cases} L(T_{(k)}) & \text{if } C > T_{(k)} \\ L(T_{(k-1)}) & \text{otherwise} \end{cases}
\end{aligned}$$

Denote by

$$\begin{aligned}
N^{r,a}(dt, da) &= \sum_{k=1}^K \delta_{\Delta_{(k)}}(a) \delta_{(T_{(k)}, A(T_{(k)}))}(dt, da) \\
N^{r,\ell}(dt, d\ell) &= \sum_{k=1}^K \delta_{\Delta_{(k)}}(\ell) \delta_{(T_{(k)}, L(T_{(k)}))}(dt, d\ell) \\
N^{r,y}(dt) &= \sum_{k=1}^K \delta_{\Delta_{(k)}}(y) \delta_{T_{(k)}}(dt) \\
N^{r,d}(dt) &= \sum_{k=1}^K \delta_{\Delta_{(k)}}(d) \delta_{T_{(k)}}(dt) \\
N^{r,c}(dt) &= \sum_{k=1}^K \delta_{\Delta_{(k)}}(c) \delta_{T_{(k)}}(dt)
\end{aligned}$$

the corresponding random measures of the fully observed  $Z(t)$  and  $N^c(t)$ . We provide the necessary conditions in terms of independent censoring (or local independence conditions) in the sense of [Andersen et al. \(1993\)](#). It follows from arguments given in Theorem 7 that the filtration generated by the random measures is necessarily the same as  $\mathcal{F}_t^{\text{full}}$ . We are now ready to state the main theorem which allows us to prove that the ICE IPCW estimator is valid.

**Theorem 4:** Assume that the intensity processes of  $N^{r,x}$ ,  $x = a, \ell, d, y$  with respect to the filtration  $\mathcal{F}_t$  are also the intensities with respect to the filtration  $\mathcal{F}_t^{\text{full}}$ . Additionally, assume also that the intensity process of  $N^c(t)$  with respect to the filtration  $\mathcal{G}_t$  is also the intensity process with respect to the filtration  $\mathcal{F}_t^{\text{full}}$ . Then the cause-specific hazard measure  $\tilde{\Lambda}_k^x$  for the  $k$ 'th jump of  $t \mapsto (Z(t \wedge C), N^c(t \wedge T^e))$  at time  $t$  given  $\mathcal{F}_{T_{(k-1)}}$  is given by

$$\begin{aligned}\tilde{\Lambda}_k^x(dt \mid \mathcal{F}_{T_{(k-1)}}) &= \lambda_{k-1}^x(t, \mathcal{F}_{T_{(k-1)}})dt, & x = a, \ell, d, y, c \\ \tilde{\pi}_k(a, t, \mathcal{F}_{T_{(k-1)}}) &= \pi_k(a, \mathcal{F}_{T_{(k-1)}}) \\ \tilde{\mu}_k(l, t, \mathcal{F}_{T_{(k-1)}}) &= \mu_k(l, t, \mathcal{F}_{T_{(k-1)}})\end{aligned}$$

Consequently, we have that

$$\begin{aligned}\bar{Q}_{k-1, \tau}^g = \mathbb{E}_P \left[ \frac{\mathbb{1}\{\bar{T}_{(k)} \leq \tau, \bar{\Delta}_{(k)} = \ell\}}{S_k^c(\bar{T}_{(k-1)} - \mid \mathcal{F}_{T_{(k-1)}}^{\text{obs}})} \bar{Q}_{k, \tau}^g(\bar{A}(T_{(k-1)}), \bar{L}(T_{(k)}), \bar{T}_{(k)}, \bar{\Delta}_{(k)}, \mathcal{F}_{T_{(k-1)}}^{\text{obs}}) \right. \\ \left. + \frac{\mathbb{1}\{\bar{T}_{(k)} \leq \tau, \bar{\Delta}_{(k)} = a\}}{S_k^c(\bar{T}_{(k-1)} - \mid \mathcal{F}_{T_{(k-1)}}^{\text{obs}})} \bar{Q}_{k, \tau}^g(1, \bar{L}(T_{(k-1)}), \bar{T}_{(k)}, \bar{\Delta}_{(k)}, \mathcal{F}_{T_{(k-1)}}^{\text{obs}}) \right. \\ \left. + \frac{\mathbb{1}\{\bar{T}_{(k)} \leq \tau, \bar{\Delta}_{(k)} = y\}}{S_k^c(\bar{T}_{(k-1)} - \mid \mathcal{F}_{T_{(k-1)}}^{\text{obs}})} \middle| \mathcal{F}_{T_{(k-1)}}^{\text{obs}} \right] \quad (4)\end{aligned}$$

for  $k = K - 1, \dots, 1$ . Then,

$$\Psi_\tau(Q) = \mathbb{E}_P[\bar{Q}_{0, \tau}^g(1, L_0)]. \quad (5)$$

*Proof:* The last statement (Equation 4 and Equation 5) follows from the first statement and Theorem 3. The compensator of the random measures  $N^{r,x}$  with respect to the filtration  $\mathcal{F}_t$  is given by

$$\begin{aligned}\Lambda^{r,a}(dt, da) &= \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} < t \leq T_{(k)}\} \lambda_{k-1}^a(t, \mathcal{F}_{T_{(k-1)}}) \pi_{k-1}(t, \mathcal{F}_{T_{(k-1-1)}}) dt \\ \Lambda^{r,\ell}(dt, d\ell) &= \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} < t \leq T_{(k)}\} \lambda_{k-1}^\ell(t, \mathcal{F}_{T_{(k-1)}}) \mu_{k-1}(t, d\ell, \mathcal{F}_{T_{(k-1-1)}}) dt \\ \Lambda^{r,y}(dt) &= \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} < t \leq T_{(k)}\} \lambda_{k-1}^y(t, \mathcal{F}_{T_{(k-1)}}) dt \\ \Lambda^{r,d}(dt) &= \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} < t \leq T_{(k)}\} \lambda_{k-1}^d(t, \mathcal{F}_{T_{(k-1)}}) dt\end{aligned}$$

by Theorem 7 and by assumption also the compensator for the filtration  $\mathcal{F}_t^{\text{full}}$ . By the innovation theorem Andersen et al. (1993),

$$\begin{aligned}\tilde{\Lambda}^{r,a}(dt, da) &= \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\} \lambda_{k-1}^a(t, \mathcal{F}_{T_{(k-1)}}) \pi_{k-1}(t, \mathcal{F}_{T_{(k-1-1)}}) dt \\ \tilde{\Lambda}^{r,\ell}(dt, d\ell) &= \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\} \lambda_{k-1}^\ell(t, \mathcal{F}_{T_{(k-1)}}) \mu_{k-1}(t, d\ell, \mathcal{F}_{T_{(k-1-1)}}) dt \\ \tilde{\Lambda}^{r,y}(dt) &= \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\} \lambda_{k-1}^y(t, \mathcal{F}_{T_{(k-1)}}) dt \\ \tilde{\Lambda}^{r,d}(dt) &= \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\} \lambda_{k-1}^d(t, \mathcal{F}_{T_{(k-1)}}) dt\end{aligned}$$

is the compensator of the random measures  $N^{r,x}$  with respect to the filtration  $\mathcal{F}_t^{\text{obs}}$ . This can be seen by noting that  $\mathcal{F}_t^{\text{obs}} \subseteq \mathcal{F}_t^{\text{full}}$  and that the censoring function  $C(t) = \mathbb{1}\{t \leq C\}$  is adapted to the filtration  $\mathcal{F}_t^{\text{full}}$ . On the other hand, we can apply Theorem 7 directly to the observed process to get that

$$\begin{aligned}
\tilde{\Lambda}^{r,a}(dt, da) &= \sum_{k=1}^K \mathbb{1}\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\} \tilde{\pi}_k(a, t, \mathcal{F}_{T_{(k-1)}}^{\text{obs}}) \tilde{\Lambda}_k^x(dt | \mathcal{F}_{T_{(k-1)}}^{\text{obs}}) \\
\tilde{\Lambda}^{r,\ell}(dt, d\ell) &= \sum_{k=1}^K \mathbb{1}\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\} \tilde{\mu}_k(l, t, \mathcal{F}_{T_{(k-1)}}^{\text{obs}}) \tilde{\Lambda}_k^x(dt | \mathcal{F}_{T_{(k-1)}}^{\text{obs}}) \\
\tilde{\Lambda}^{r,y}(dt) &= \sum_{k=1}^K \mathbb{1}\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\} \tilde{\Lambda}_k^x(dt | \mathcal{F}_{T_{(k-1)}}^{\text{obs}}) \\
\tilde{\Lambda}^{r,d}(dt) &= \sum_{k=1}^K \mathbb{1}\{\bar{T}_{(k-1)} < t \leq \bar{T}_{(k)}\} \tilde{\Lambda}_k^x(dt | \mathcal{F}_{T_{(k-1)}}^{\text{obs}})
\end{aligned} \tag{6}$$

Since the canonical compensator given in Equation 6 (Theorem 4.3.2 in Last & Brandt (1995)) determines uniquely the distribution of the marks and the event times, the theorem follows.  $\square$

Interestingly, Equation 2 corresponds exactly to the target parameter of Rytgaard et al. (2022) by plugging in the definitions of  $\bar{Q}_{k,\tau}^g$  and simplifying. A simple implementation of the IPCW is provided below in the simple case of a static treatment plan. The other representations of the target parameter in terms of the intensities are useful directly, but we may, as in the discrete, estimate the target parameter by Monte Carlo integration (i.e., direct simulation from the estimated intensities/densities).

## 5.1 Algorithm for IPCW Iterative Conditional Expectations Estimator

We assume that  $K$  denotes the last non-terminal event in the sample before time  $\tau$ .

- For each event point  $k = K, K-1, \dots, 1$  (starting with  $k = K$ ):
  1. Obtain  $\hat{S}^c(t | \mathcal{F}_{T_{(k-1)}})$  by fitting a cause-specific hazard model for the censoring via the interevent time  $S_{(k)} = T_{(k)} - T_{(k-1)}$ , regressing on  $\mathcal{F}_{T_{(k-1)}}$  (among the people who are still at risk after  $k-1$  events).
  2. Define the subject-specific weight:

$$\hat{\eta}_k = \frac{\mathbb{1}\{T_{(k)} \leq \tau, \Delta_k \in \{a, \ell\}, k < K\} \hat{\nu}_k(\mathcal{F}_{T_{(k)}}^{-A}, \mathbf{1})}{\hat{S}^c(T_{(k)} | \mathcal{F}_{T_{(k-1)}}^{-A}, \mathbf{1})}$$

Then calculate the subject-specific pseudo-outcome

$$\hat{R}_k = \frac{\mathbb{1}\{T_{(k)} \leq \tau, \Delta_k = y\}}{\hat{S}^c(T_{(k)} | \mathcal{F}_{T_{(k-1)}}^{-A}, \mathbf{1})} + \hat{\eta}_k$$

Regress  $\hat{R}_k$  on  $\mathcal{F}_{T_{(k-1)}}$  on the data with  $T_{(k-1)} < \tau$  and  $\Delta_k \in \{a, \ell\}$  to obtain a prediction function  $\hat{\nu}_{k-1} : \mathcal{H}_{k-1} \rightarrow \mathbb{R}_+$ .

- At baseline, we obtain the estimate  $\hat{\Psi}_n = \frac{1}{n} \sum_{i=1}^n \hat{\nu}_0(L_i(0), 1)$ .

**Note:** The  $\bar{Q}_{k,\tau}^g$  have the interpretation of the heterogenous causal effect after  $k$  events.

For now, we recommend Equation 4 for estimating  $\bar{Q}_{k,\tau}^g$  instead of direct computation Equation 2: For estimators of the hazard that are piecewise constant, we would need to compute integrals for each unique pair of history and event times occurring in the sample at each event  $k$ . On the other hand, the IPCW approach is very sensitive to the specification of the censoring distribution.

Let  $\|\cdot\|_{L^2(P)}$  denote the  $L^2(P)$ -norm, that is

$$\|f\|_{L^2(P)} = \sqrt{\mathbb{E}_P[f^2(X)]} = \sqrt{\int f^2(x) P(dx)}.$$

Then, we have the following result,

**Lemma 1:** Assume that  $\|\hat{\nu}_{k+1} - \bar{Q}_{k+1,\tau}^g\|_{L^2(P)} = o_P(1)$ ,  $\|\hat{\Lambda}_k^c - \Lambda_k^c\|_{L^2(P)} = o_P(1)$ . For the censoring, we need that  $\hat{\Lambda}_k^c$  and  $\Lambda_k^c$  are uniformly bounded, that is  $\hat{\Lambda}_k^c(t | f_{k-1}) \leq C$  and  $\Lambda_k^c(t | (f_{k-1})) \leq C$  on the interval for all  $t \in [0, \tau]$  for some constant  $C > 0$  and for  $P$ -almost all  $f_{k-1}$ . Let  $\hat{P}_k$  denote the regression estimator and assume that

$$\|\hat{P}_k[\bar{Z}_{k,\tau}^a(\hat{S}^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot] - \mathbb{E}_P[\bar{Z}_{k,\tau}^a(\hat{S}^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot]\|_{L^2(P)} = o_P(1)$$

where  $\bar{Z}_{k,\tau}^a$  is given in Equation 9. Then,

$$\|\hat{\nu}_k - \bar{Q}_{k,\tau}^g\|_{L^2(P)} = o_P(1)$$

*Proof:* By the triangle inequality,

$$\begin{aligned} \|\hat{\nu}_k - \bar{Q}_{k,\tau}^g\|_{L^2(P)} &\leq \|\hat{P}_k[\bar{Z}_{k,\tau}^a(\hat{S}^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot] - \mathbb{E}_P[\bar{Z}_{k,\tau}^a(\hat{S}^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot]\|_{L^2(P)} \\ &\quad + \|\mathbb{E}_P[\bar{Z}_{k,\tau}^a(\hat{S}^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot] - \mathbb{E}_P[\bar{Z}_{k,\tau}^a(S^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot]\|_{L^2(P)} \\ &\quad + \|\mathbb{E}_P[\bar{Z}_{k,\tau}^a(S^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot] - \mathbb{E}_P[\bar{Z}_{k,\tau}^a(S^c, \bar{Q}_{k+1,\tau}^g) | \mathcal{F}_{T_{(k)}} = \cdot]\|_{L^2(P)} \end{aligned}$$

The first and third terms are  $o_P(1)$  by assumption. That the second term is  $o_P(1)$  follows from the fact that

$$\begin{aligned} &\|\mathbb{E}_P[\bar{Z}_{k,\tau}^a(\hat{S}^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot] - \mathbb{E}_P[\bar{Z}_{k,\tau}^a(S^c, \hat{\nu}_{k+1}) | \mathcal{F}_{T_{(k)}} = \cdot]\|_{L^2(P)} \\ &\leq \|\mathbb{E}_P\left[\left(\frac{1}{S^c(T_{(k+1)} - | \mathcal{F}_{T_{(k)}})} - \frac{1}{\hat{S}^c(T_{(k+1)} - | \mathcal{F}_{T_{(k)}})}\right) \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} \in \{a, \ell\}\} \hat{\nu}_{k+1} | \mathcal{F}_{T_{(k)}} = \cdot\right]\|_{L^2(P)} \\ &\quad + \|\mathbb{E}_P\left[\left(\frac{1}{S^c(T_{(k+1)} - | \mathcal{F}_{T_{(k)}})} - \frac{1}{\hat{S}^c(T_{(k+1)} - | \mathcal{F}_{T_{(k)}})}\right) \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} | \mathcal{F}_{T_{(k)}} = \cdot\right]\|_{L^2(P)} \end{aligned} \quad (7)$$

Using that the function  $x \mapsto \exp(-x)$  is Lipschitz continuous and uniformly bounded, the second term in Equation 7 can be bounded by  $K \|\mathbb{E}_P[(\hat{\Lambda}_k^c - \Lambda_k^c) \mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} | \mathcal{F}_{T_{(k)}} = \cdot]\|_{L^2(P)}$ . The first term in Equation 7 can with a very similar argument be shown to be bounded, assuming that the estimator  $\hat{\nu}_{k+1}$  is uniformly bounded.  $\square$

## 6 The efficient influence function

We want to use machine learning estimators of the nuisance parameters, so to get inference we need to debias our estimate with the efficient influence function, e.g., double/debiased machine learning Chernozhukov et al. (2018) or targeted minimum loss estimation van der Laan & Rubin (2006). We use Equation 4 for censoring to derive the efficient influence function, because it will contain fewer martingale terms.

**Theorem 5** (Efficient influence function): The efficient influence function is given by

$$\begin{aligned} \varphi^*(P) &= \frac{\mathbb{1}\{A_0 = 1\}}{\pi_0(L(0))} \prod_{j=1}^{k-1} \left( \frac{\mathbb{1}\{A(T_{(j)}) = 1\}}{\pi_j(T_{(j)}, \mathcal{F}_{T_{(j-1)}})} \right)^{\mathbb{1}\{\Delta_{(j)} = a\}} \frac{1}{\prod_{j=1}^{k-1} S^c(T_{(j)} - | \mathcal{F}_{T_{(j-1)}})} \mathbb{1}\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\} \\ &\quad \times \left( (\bar{Z}_{k,\tau}^a - \bar{Q}_{k-1,\tau}^g) + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} (\bar{Q}_{k-1,\tau}^g(\tau | \mathcal{F}_{T_{(k-1)}}) - \bar{Q}_{k-1,\tau}^g(u | \mathcal{F}_{T_{(k-1)}})) \frac{1}{S^c(u - | \mathcal{F}_{T_{(k-1)}})} S(u | \mathcal{F}_{T_{(k-1)}}) M_k^c(du) \right) \\ &\quad + \bar{Q}_{0,\tau}^g - \Psi_\tau(P), \end{aligned} \quad (8)$$

where  $M_k^c(t) = \mathbb{1}\{T_{(k-1)} < t \leq T_{(k)}\} (N^c(t) - \Lambda^c(t | \mathcal{F}_{T_{(k-1)}}))$  is the martingale for the censoring process.

*Proof:* Define (sorry about the notation!)

$$\begin{aligned}\bar{Z}_{k,\tau}^a(s, t_k, d_k, l_k, a_k, f_{k-1}) &= \frac{I(t_k \leq s, d_k = \ell)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1})ds\right)} \bar{Q}_{k,\tau}^g(a_{k-1}, l_k, t_k, d_k, f_{k-1}) \\ &\quad + \frac{I(t_k \leq s, d_k = a)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1})ds\right)} \\ &\quad \times \int \bar{Q}_{k,\tau}^g(\tilde{a}_k, l_{k-1}, t_k, d_k, f_{k-1}) \pi_{k-1}^*\left(t_k, \tilde{a}_k, \mathcal{F}_{T_{(k-1)-1}}\right) \nu_A(d\tilde{a}_k) \\ &\quad + \frac{I(t_k \leq s, d_k = y)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1})ds\right)}, s \leq \tau\end{aligned}\tag{9}$$

and let

$$\bar{Q}_{k-1,\tau}^g(s) = \mathbb{E}_P\left[\bar{Z}_{k,s}^a\left(\tau, T_{(k)}, \Delta_{(k)}, L(T_{(k)}), A(T_{(k)}), \mathcal{F}_{T_{(k-1)}}\right) \mid \mathcal{F}_{T_{(k-1)}}\right], s \leq \tau$$

We compute the efficient influence function by taking the Gateaux derivative of the above with respect to  $P$ , by discretizing the time. We will use two well-known “results” for the efficient influence function.

$$\begin{aligned}&\frac{\partial}{\partial \varepsilon} \int_{T_{(k-1)}}^t \lambda_{\varepsilon}^x(s \mid \mathcal{F}_{T_{(k-1)}}) ds \\ &= \frac{\delta_{\mathcal{F}_{T_{(k-1)}}}(f_{k-1})}{P(\mathcal{F}_{T_{(k-1)}} = f_{k-1})} \int_{T_{(k-1)}}^t \frac{1}{\exp\left(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^s \lambda_{k-1}^x(s, \mathcal{F}_{T_{(k-1)}}) ds\right)} \left(N_k^x(ds) - \tilde{Y}_{k-1}(s) \lambda_{k-1}^x(s, \mathcal{F}_{T_{(k-1)}}) ds\right)\end{aligned}$$

and

$$\frac{\partial}{\partial \varepsilon} \mathbb{E}_{(1-\varepsilon)P + \varepsilon \delta_{(Y,X)}}[Y \mid X = x] \Big|_{\varepsilon=0} = \frac{\delta_X(x)}{P(X = x)} (Y - \mathbb{E}_P[Y \mid X = x])$$

We will recursively calculate the derivative,

$$\frac{\partial}{\partial \varepsilon} \bar{Q}_{k-1,\tau}^{a,\varepsilon}(a_{k-1}, l_{k-1}, t_{k-1}, d_{k-1}, f_{k-2}) \left( (1-\varepsilon)P + \varepsilon \delta_{\mathcal{F}_{T_{(k-1)}}} \right) \Big|_{\varepsilon=0}$$

where we have introduced the notation for the dependency on  $P$ . Then, taking the Gateaux derivative of the above yields,

$$\begin{aligned}&\frac{\partial}{\partial \varepsilon} \bar{Q}_{k-1,\tau}^{a,\varepsilon}(a_{k-1}, l_{k-1}, t_{k-1}, d_{k-1}, f_{k-2}) \left( (1-\varepsilon)P + \varepsilon \delta_{\mathcal{F}_{T_{(k-1)}}} \right) \Big|_{\varepsilon=0} \\ &= \frac{\delta_{\mathcal{F}_{T_{(k-1)}}}(f_{k-1})}{P(\mathcal{F}_{T_{(k-1)}} = f_{k-1})} \left( \bar{Z}_{k,\tau}^a - \bar{Q}_{k-1,\tau}^g(\tau, \mathcal{F}_{T_{(k-1)}}) + \right. \\ &\quad + \int_{T_{(k-1)}}^{\tau} \bar{Z}_{k,\tau}^a(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{T_{(k-1)}}^{t_k} \frac{1}{\exp\left(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^s \lambda_{k-1}^x(s, \mathcal{F}_{T_{(k-1)}}) ds\right)} \left(N_k^c(ds) - \tilde{Y}_{k-1}(s) \lambda_{k-1}^c(s, \mathcal{F}_{T_{(k-1)}}) ds\right) \\ &\quad \left. P_{(T_{(k)}, \Delta_{(k)}, L(T_{(k)}), A(T_{(k)}))}(dt_k, dd_k, dl_k, da_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}) \right) \\ &\quad + \int_{T_{(k-1)}}^{\tau} \left( \frac{I(t_k \leq \tau, d_k \in \{a, \ell\})}{\exp\left(-\int_{T_{(k-1)}}^{t_k} \lambda^c(s \mid f_{k-1})ds\right)} \cdot \left( \frac{\pi_{k-1}^*(t_k, a_k, \mathcal{F}_{T_{(k-1)-1}})}{\pi_{k-1}(t_k, \mathcal{F}_{T_{(k-1)-1}})} \right)^{I(d_k=a)} \frac{\partial}{\partial \varepsilon} \bar{Q}_{k-1,\tau}^{a,\varepsilon}(a_k, l_k, t_k, d_k, f_{k-1}) \left( (1-\varepsilon)P + \varepsilon \delta_{\mathcal{F}_{T_{(k)}}} \right) \Big|_{\varepsilon=0} \right. \\ &\quad \left. P_{(T_{(k)}, \Delta_{(k)}, L(T_{(k)}), A(T_{(k)}))}(dt_k, dd_k, dl_k, da_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}) \right)\end{aligned}$$

Now note for the second term, we can write

$$\begin{aligned}
& \int_{T_{(k-1)}}^{\tau} \bar{Z}_{k,\tau}^a(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{T_{(k-1)}}^{t_k} \frac{1}{\exp\left(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^s \lambda_{k-1}^x(s, \mathcal{F}_{T_{(k-1)}}) ds\right)} \left(N_k^c(ds) - \tilde{Y}_{k-1}(s) \lambda_{k-1}^c(s, \mathcal{F}_{T_{(k-1)}}) ds\right) \\
& P_{(T_{(k)}, \Delta_{(k)}, L(T_{(k)}), A(T_{(k)}))} \left(dt_k, dd_k, dl_k, da_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\right) \\
& = \int_{T_{(k-1)}}^{\tau} \int_s^{\tau} \bar{Z}_{k,\tau}^a(\tau, t_k, d_k, l_k, a_k, f_{k-1}) P_{(T_{(k)}, \Delta_{(k)}, L(T_{(k)}), A(T_{(k)}))} \left(dt_k, dd_k, dl_k, da_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\right) \\
& \quad \frac{1}{\exp\left(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^s \lambda_{k-1}^x(s, \mathcal{F}_{T_{(k-1)}}) ds\right)} \left(N_k^c(ds) - \tilde{Y}_{k-1}(s) \lambda_{k-1}^c(s, \mathcal{F}_{T_{(k-1)}}) ds\right) \\
& = \int_{T_{(k-1)}}^{\tau} \left(\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(s)\right) \\
& \quad \frac{1}{\exp\left(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^s \lambda_{k-1}^x(s, \mathcal{F}_{T_{(k-1)}}) ds\right)} \left(N_k^c(ds) - \tilde{Y}_{k-1}(s) \lambda_{k-1}^c(s, \mathcal{F}_{T_{(k-1)}}) ds\right)
\end{aligned}$$

by an exchange of integrals. Combining the results iteratively gives the result.  $\square$

For now, we recommend using the one step estimator and not the TMLE because the martingales are computationally intensive to estimate. This means that multiple TMLE updates may not be a good idea.

## 7 Paired ICE IPCW one-step estimator

In this section, we provide a special procedure for the purpose of one-step estimation. Though the present section is stated in the context one-step estimation, a targeted minimum loss estimator (TMLE) can be obtained by very similar considerations. Recall that the efficient influence function in [Equation 8](#) includes a censoring martingale. To estimate this martingale, we would need to have estimators  $\bar{Q}_{k,\tau}^g(t)$  at a sufficiently dense grid of time points  $t$ . Unfortunately, the event-specific cause-specific hazards  $\hat{\lambda}_k^x$  cannot readily be used to estimate  $\bar{Q}_{Q,\tau_k}^g$  due to the high dimensionality of integrals. The IPCW approach we have given in [Section 5.1](#) also would be prohibitively computationally expensive (at the very least if we use flexible machine learning estimators). Another issue altogether is that it does not explicitly give conditions on the remainder since we will integrate with respect to  $\bar{Q}_{k,\tau}^g$ . Instead, we split up the estimation the estimation into two parts for  $\bar{Q}_{k,\tau}^g$ . For each  $k$ , the procedure constructs two new estimators of  $\bar{Q}_{k,\tau}^g$ :

1.  $\hat{\nu}_{k,\tau}^*(\mathcal{F}_{T_{(k)}})$  which is obtained the same way as in [Section 5.1](#).
2. First obtain the estimates  $\tilde{\nu}_{k,\tau}$  by regressing  $\hat{R}_{k+1}$  on  $(A(T_{(k)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}})$  (i.e., we do not include the latest covariate value). Given cause-specific estimators  $\hat{\lambda}_{k+1}^x$  for  $x = a, l, d, y$ , we estimate  $\bar{Q}_{k,\tau}^g(t, \mathcal{F}_{T_{(k)}})$  by

$$\begin{aligned}
\hat{\nu}_{k,\tau}^*(t \mid \mathcal{F}_{T_{(k)}}) &= \int_0^{t-T_{(k)}} \pi_{s \in (0, u-T_{(k)})} \left(1 - \sum_{x=a,l,d,y} \hat{\lambda}_{k+1}^x(ds \mid \mathbf{1}, \mathcal{F}_{T_{(k)}}^A)\right) \left[\hat{\lambda}_{k+1}^y(du \mid \mathbf{1}, \mathcal{F}_{T_{(k)}}^A)\right. \\
&\quad + \tilde{\nu}_{k+1,\tau} \left(1, u + T_{(k)}, a, \mathbf{1}, \mathcal{F}_{T_{(k-1)}}^A\right) \hat{\lambda}_{k+1}^a(du \mid \mathbf{1}, \mathcal{F}_{T_{(k)}}^A) \\
&\quad \left. + \tilde{\nu}_{k+1,\tau} \left(1, u + T_{(k)}, \ell, \mathbf{1}, \mathcal{F}_{T_{(k-1)}}^A\right) \hat{\lambda}_{k+1}^\ell(du \mid \mathbf{1}, \mathcal{F}_{T_{(k)}}^A)\right]
\end{aligned}$$

on the interevent level.

Given also estimators of the propensity scores, we can estimate the efficient influence function as:

$$\begin{aligned} \varphi^*(\hat{P}_n^*) &= \frac{\mathbb{1}\{A_0 = 1\}}{\hat{\pi}_0(L(0))} \prod_{j=1}^{k-1} \left( \frac{\mathbb{1}\{A(T_{(j)}) = 1\}}{\hat{\pi}(T_{(j)}, A(T_{(j)}), \mathcal{F}_{T_{(j-1)}})} \right)^{\mathbb{1}\{\Delta_{(j)} = a\}} \frac{1}{\prod_{j=1}^{k-1} \hat{S}^c(T_{(j)} - | \mathcal{F}_{T_{(j-1)}})} \mathbb{1}\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\} \\ &\times \left( \bar{Z}_{k,\tau}^a(\hat{S}^c, \hat{\nu}_{k,\tau}) - \hat{\nu}_{k-1,\tau}(\mathcal{F}_{T_{(k-1)}}) + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} (\hat{\nu}_{k-1}^*(\tau | \mathcal{F}_{T_{(k-1)}}) - \hat{\nu}_{k-1,\tau}^*(u | \mathcal{F}_{T_{(k-1)}})) \frac{1}{\hat{S}^c(u - | \mathcal{F}_{T_{(k-1)}})} \hat{S}(u | \mathcal{F}_{T_{(k-1)}}) M_k^c(du) \right) \\ &+ \hat{\nu}_{0,\tau}(1, \mathcal{F}_0) - \mathbb{P}_n \hat{\nu}_{0,\tau}(1, \cdot) \end{aligned}$$

The resulting estimator is given by

$$\hat{\Psi}_n = \mathbb{P}_n \hat{\nu}_{0,\tau}(1, \cdot) + \mathbb{P}_n \varphi^*(\hat{P}_n^*)$$

We have the following rate result for  $\hat{\nu}_{k,\tau}^*$ :

**Lemma 2:** Let  $\bar{Q}_k^{-L} = \mathbb{E}_P[\bar{Q}_{k,\tau}^g | A(T_{(k)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}]$ . Assume that  $\|\tilde{\nu}_{k+1}^* - \bar{Q}_{k+1}^{-L}\|_{L^2(P)} = o_P(1)$ . If the estimators for the cause-specific hazards for the event times converge at  $n^{-\frac{1}{4}}$ -rate, that is

$$\sqrt{\int \int_{t_{k-1}}^{\tau} (\lambda^x(t | f_{k-1}) - \hat{\lambda}^x(t | f_{k-1}))^2 dt P_{\mathcal{F}_{T_{(k-1)}}}(df_{k-1})} = o_P(n^{-\frac{1}{4}})$$

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

$$\|\hat{\nu}_{k,\tau}^*(\cdot) - \bar{Q}_{k,\tau}^g(\cdot)\|_{L^2(P)} = o_P(n^{-\frac{1}{4}})$$

*Proof:* By the triangle inequality,

$$\begin{aligned} \|\hat{\nu}_{k,\tau}^*(\cdot) - \bar{Q}_{k,\tau}^g(\cdot)\|_{L^2(P)} &\leq \sqrt{\int \int_{t_k}^{\tau} (\hat{S}_{k+1}(t|f_k) \hat{\lambda}_{k+1}^y(t | f_k) - S_{k+1}(t|f_k) \lambda_{k+1}^y(t | f_k))^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)} \\ &+ \sqrt{\int \int_{t_k}^{\tau} (\hat{\lambda}_{k+1}^a(t | f_k) \hat{S}_{k+1}(t|f_k) - S_{k+1}(t|f_k) \lambda_{k+1}^a(t | f_k))^2 (\tilde{\nu}(1, t, a, f_k))^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)} \\ &+ \sqrt{\int \int_{t_k}^{\tau} (\hat{\lambda}_{k+1}^\ell(t | f_k) \hat{S}_{k+1}(t|f_k) - S_{k+1}(t|f_k) \lambda_{k+1}^\ell(t | f_k))^2 (\tilde{\nu}(a_{k-1}, t, a, f_k))^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)} \\ &+ \sqrt{\int \int_{t_k}^{\tau} (\tilde{\nu}_{k+1,\tau}^*(t, \dots) - \bar{Q}_{k+1}^{-L}(t, \dots))^2 \left( \sum_{x=a,\ell,y} S_{k+1}(t|f_k) \lambda_{k+1}^x(t | f_k) \right) dt P_{\mathcal{F}_{T_{(k-1)}}}(df_k)} \end{aligned}$$

The last term is  $o_P(n^{-\frac{1}{4}})$  by assumption. By bounding  $\tilde{\nu}$ , the first three terms are then also  $o_P(n^{-\frac{1}{4}})$ . By i.e., noting that the mapping  $(x, y) \mapsto x \exp(-(x + y))$  is Lipschitz continuous and uniformly bounded (under additional boundedness conditions on the hazards), we see that the conditions on the hazards are sufficient to show that the first three terms are  $o_P(n^{-\frac{1}{4}})$ .  $\square$

## 7.1 Remainder term

We now consider the efficient influence function, occurring in the remainder term. The following result shows that we can separate the estimation of the martingale term and the outcome term in the efficient influence function.



**Theorem 6** (Remainder term): The remainder term  $R_2 = \Psi_\tau(P) - \Psi_\tau(P_0) + \mathbb{E}_{P_0}[\varphi^*(P)]$  is given by

$$R_2 = \sum_{k=1}^K \int \mathbb{1}\{t_1 < \dots < t_k < \tau\} \prod_{j=0}^{k-2} \left( \frac{\pi_{0,j}(t_k, f_{j-1}^1)}{\pi_j(t_k, f_{j-1}^1)} \right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^{k-1} S_0^c(t_j - | f_{j-1}^1)}{\prod_{j=1}^{k-1} S^c(t_j - | f_{j-1}^1)} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} z_k(f_{j-1}^1) P_{\mathcal{F}_{T_k}}(df_k),$$

where

$$\begin{aligned} z_k(\mathcal{F}_{T(k)}) &= \left( \left( \frac{\pi_{k-1,0}(T(k), \mathcal{F}_{T(k-1)})}{\pi_{k-1}(T(k), \mathcal{F}_{T(k-1)})} \right)^{\mathbb{1}\{\Delta(k)=a\}} - 1 \right) \left( \bar{Q}_{k-1,\tau}^g(\mathcal{F}_{T(k-1)}) - \nu_{k-1,\tau}(\mathcal{F}_{T(k-1)}) \right) \\ &\quad + \left( \frac{\pi_{k-1,0}(T(k), \mathcal{F}_{T(k-1)})}{\pi_{k-1}(T(k), \mathcal{F}_{T(k-1)})} \right)^{\mathbb{1}\{\Delta(k)=a\}} \int_{T(k-1)}^\tau \left( \frac{S_0^c(u | \mathcal{F}_{T(k-1)})}{S^c(u | \mathcal{F}_{T(k-1)})} - 1 \right) \left( \nu_{k-1,\tau}^*(du | \mathcal{F}_{T(k-1)}) - \bar{Q}_{k-1,\tau}^g(du | \mathcal{F}_{T(k-1)}) \right) \\ &\quad + \left( \frac{\pi_{k-1,0}(T(k), \mathcal{F}_{T(k-1)})}{\pi_{k-1}(T(k), \mathcal{F}_{T(k-1)})} \right)^{\mathbb{1}\{\Delta(k)=a\}} \int_{T(k-1)}^\tau V_k(u, \mathcal{F}_{T(k-1)}) \nu_{k-1,\tau}^*(du | \mathcal{F}_{T(k-1)}), \end{aligned}$$

and  $V_k(u, \mathcal{F}_{T(k)}) = \int_{T(k-1)}^u \left( \frac{S_0(s | \mathcal{F}_{T(k-1)})}{S(s | \mathcal{F}_{T(k-1)})} - 1 \right) \frac{S_0^c(s - | \mathcal{F}_{T(k-1)})}{S^c(s - | \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(ds | \mathcal{F}_{T(k-1)}) - \Lambda^c(ds | \mathcal{F}_{T(k-1)}) \right)$ . Here  $f_j^1$  simply means that we insert 1 into every place where we have  $a_i, i = 1, \dots, j$  in  $f_j$ . **NOTE:** We define the empty product to be 1 and  $\pi_0(T_0, \mathcal{F}_{T(-1)}) = \pi_0(L_0)$  (and  $\pi_{0,0}$ ).

*Proof:* **NOTE:** We should write  $f_j^1$  most places instead of  $f_j$ . **Sketch:** First define

$$\begin{aligned} \varphi_k^*(P) &= \frac{\mathbb{1}\{A_0 = 1\}}{\pi_0(L_0)} \prod_{j=1}^{k-1} \left( \frac{\mathbb{1}\{A(T_{(j)}) = 1\}}{\pi_j(T_{(j)}, \mathcal{F}_{T_{(j-1)}})} \right)^{\mathbb{1}\{\Delta_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} S^c(T_{(j)} - | \mathcal{F}_{T_{(j-1)}})} \mathbb{1}\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\} \\ &\quad \times \left( \left( \bar{Z}_{k,\tau}^a - \nu_{k-1} \right) + \int_{T(k-1)}^{\tau \wedge T(k)} (\nu_{k-1}^*(\tau) - \nu_{k-1}^*(u)) \frac{1}{S^c(u - | \mathcal{F}_{T(k-1)})} S(u | \mathcal{F}_{T(k-1)}) M_k^c(du) \right) \end{aligned}$$

for  $k > 0$  and define  $\varphi_0^*(P) = \nu_0(L_0) - \Psi_\tau(P)$ , so that

$$\varphi^*(P) = \sum_{k=0}^K \varphi_k^*(P)$$

Also note that

$$\mathbb{E}_{P_0}[\varphi_0^*(P)] + \Psi_\tau(P) - \Psi_\tau(P_0) = \mathbb{E}_{P_0}[\nu_0(L_0) - \bar{Q}_{0,\tau}^g(L_0)]. \quad (10)$$

Apply the law of iterated expectation to the efficient influence function in [Equation 8](#) to get

$$\begin{aligned} \mathbb{E}_{P_0}[\varphi_k^*(P)] &= \int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left( \frac{\pi_{0,j}(t_k, f_{j-1})}{\pi_j(t_k, f_{j-1})} \right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^{k-1} S_0^c(t_j - | f_{j-1})}{\prod_{j=1}^{k-1} S^c(t_j - | f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} \\ &\quad \times (\mathbb{E}_P[h_k(\mathcal{F}_{T(k)}) | \mathcal{F}_{T(k-1)} = f_{k-1}]) P_{\mathcal{F}_{T(k-1)}}(df_{k-1}) \end{aligned}$$

where

$$h_k(\mathcal{F}_{T(k)}) = \bar{Z}_{k,\tau}^a - \nu_{k-1} + \int_{T(k-1)}^{\tau \wedge T(k)} (\nu_{k-1}^*(\tau | \mathcal{F}_{T(k-1)}) - \nu_{k-1}^*(u | \mathcal{F}_{T(k-1)})) \frac{1}{S^c(u - | \mathcal{F}_{T(k-1)})} S(u | \mathcal{F}_{T(k-1)}) M_k^c(du).$$

Now note that

$$\begin{aligned}
& \mathbb{E}_{P_0} \left[ h_k \left( \mathcal{F}_{T(k)} \right) \mid \mathcal{F}_{T(k-1)} \right] \\
&= \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S^c, \nu_k) \mid \mathcal{F}_{T(k-1)} \right] - \nu_{k-1,\tau} \left( \mathcal{F}_{T(k-1)} \right) \\
&+ \mathbb{E}_{P_0} \left[ \int_{T(k-1)}^{\tau \wedge T(k)} \left( \nu_{k-1}^* \left( \tau \mid \mathcal{F}_{T(k-1)} \right) - \nu_{k-1,\tau}^* \left( u \mid \mathcal{F}_{T(k-1)} \right) \right) \frac{1}{S^c(u - \mid \mathcal{F}_{T(k-1)}) S(u \mid \mathcal{F}_{T(k-1)})} M_k^c(du) \right] \mid \mathcal{F}_{T(k-1)} \right] \\
&= \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S_0^c, \bar{Q}_{k,\tau}^g) \mid \mathcal{F}_{T(k-1)} \right] - \nu_{k-1,\tau} \left( \mathcal{F}_{T(k-1)} \right) \\
&+ \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S^c, \nu_k) \mid \mathcal{F}_{T(k-1)} \right] - \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S^c, \bar{Q}_{k,\tau}^g) \mid \mathcal{F}_{T(k-1)} \right] \\
&+ \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S^c, \bar{Q}_{k,\tau}^g) \mid \mathcal{F}_{T(k-1)} \right] - \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S_0^c, \bar{Q}_{k,\tau}^g) \mid \mathcal{F}_{T(k-1)} \right] \\
&+ \int_{T(k-1)}^{\tau} \left( \nu_{k-1}^* \left( \tau \mid \mathcal{F}_{T(k-1)} \right) - \nu_{k-1,\tau}^* \left( u \mid \mathcal{F}_{T(k-1)} \right) \right) \frac{S_0^c(u - \mid \mathcal{F}_{T(k-1)}) S_0(u \mid \mathcal{F}_{T(k-1)})}{S^c(u - \mid \mathcal{F}_{T(k-1)}) S(u \mid \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(du \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(du \mid \mathcal{F}_{T(k-1)}) \right)
\end{aligned}$$

by a martingale argument. Noting that,

$$\begin{aligned}
& \mathbb{E}_{P_0} \left[ \int_{T(k-1)}^{\tau \wedge T(k)} \left( \nu_{k-1}^* \left( \tau \mid \mathcal{F}_{T(k-1)} \right) - \nu_{k-1,\tau}^* \left( u \mid \mathcal{F}_{T(k-1)} \right) \right) \frac{1}{S^c(u - \mid \mathcal{F}_{T(k-1)}) S(u \mid \mathcal{F}_{T(k-1)})} M_k^c(du) \mid \mathcal{F}_{T(k-1)} \right] \\
&= \int_{T(k-1)}^{\tau} \left( \nu_{k-1}^* \left( \tau \mid \mathcal{F}_{T(k-1)} \right) - \nu_{k-1,\tau}^* \left( u \mid \mathcal{F}_{T(k-1)} \right) \right) \frac{S_0^c(u - \mid \mathcal{F}_{T(k-1)}) S_0(u \mid \mathcal{F}_{T(k-1)})}{S^c(u - \mid \mathcal{F}_{T(k-1)}) S(u \mid \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(du \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(du \mid \mathcal{F}_{T(k-1)}) \right) \\
&= \int_{T(k-1)}^{\tau} \int_{T(k-1)}^u \frac{S_0^c(s - \mid \mathcal{F}_{T(k-1)}) S_0(s \mid \mathcal{F}_{T(k-1)})}{S^c(s - \mid \mathcal{F}_{T(k-1)}) S(s \mid \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(ds \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(ds \mid \mathcal{F}_{T(k-1)}) \right) \left( \nu_{k-1,\tau}^* \left( du \mid \mathcal{F}_{T(k-1)} \right) \right) \\
&= \int_{T(k-1)}^{\tau} \int_{T(k-1)}^u \left( \frac{S_0(s \mid \mathcal{F}_{T(k-1)})}{S(s \mid \mathcal{F}_{T(k-1)})} - 1 \right) \frac{S_0^c(s - \mid \mathcal{F}_{T(k-1)})}{S^c(s - \mid \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(ds \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(ds \mid \mathcal{F}_{T(k-1)}) \right) \left( \nu_{k-1,\tau}^* \left( du \mid \mathcal{F}_{T(k-1)} \right) \right) \\
&+ \int_{T(k-1)}^{\tau} \int_{T(k-1)}^u \left( \frac{S_0^c(s - \mid \mathcal{F}_{T(k-1)})}{S^c(s - \mid \mathcal{F}_{T(k-1)})} \right) \left( \Lambda_{k,0}^c(ds \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(ds \mid \mathcal{F}_{T(k-1)}) \right) \left( \nu_{k-1,\tau}^* \left( du \mid \mathcal{F}_{T(k-1)} \right) \right) \\
&= \int_{T(k-1)}^{\tau} \int_{T(k-1)}^u \left( \frac{S_0(s \mid \mathcal{F}_{T(k-1)})}{S(s \mid \mathcal{F}_{T(k-1)})} - 1 \right) \frac{S_0^c(s - \mid \mathcal{F}_{T(k-1)})}{S^c(s - \mid \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(ds \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(ds \mid \mathcal{F}_{T(k-1)}) \right) \left( \nu_{k-1,\tau}^* \left( du \mid \mathcal{F}_{T(k-1)} \right) \right) \\
&+ \int_{T(k-1)}^{\tau} \left( \frac{S_0^c(u \mid \mathcal{F}_{T(k-1)})}{S^c(u \mid \mathcal{F}_{T(k-1)})} - 1 \right) \left( \nu_{k-1,\tau}^* \left( du \mid \mathcal{F}_{T(k-1)} \right) \right)
\end{aligned}$$

where we apply the Duhamel equation in the second last equality, it follows that

$$\begin{aligned}
& \mathbb{E}_{P_0} \left[ h_k \left( \mathcal{F}_{T(k)} \right) \mid \mathcal{F}_{T(k-1)} \right] \\
&= \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S_0^c, \bar{Q}_{k,\tau}^g) \mid \mathcal{F}_{T(k-1)} \right] - \nu_{k-1,\tau} \left( \mathcal{F}_{T(k-1)} \right) \\
&+ \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S^c, \nu_k) \mid \mathcal{F}_{T(k-1)} \right] - \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S^c, \bar{Q}_{k,\tau}^g) \mid \mathcal{F}_{T(k-1)} \right] \\
&+ \int_{T(k-1)}^{\tau} \left( \frac{S_0^c(u \mid \mathcal{F}_{T(k-1)})}{S^c(u \mid \mathcal{F}_{T(k-1)})} - 1 \right) \left( \nu_{k-1,\tau}^* \left( du \mid \mathcal{F}_{T(k-1)} \right) - \bar{Q}_{k-1,\tau}^g \left( du \mid \mathcal{F}_{T(k-1)} \right) \right) \\
&+ \int_{T(k-1)}^{\tau} \int_{T(k-1)}^u \left( \frac{S_0(s \mid \mathcal{F}_{T(k-1)})}{S(s \mid \mathcal{F}_{T(k-1)})} - 1 \right) \frac{S_0^c(s - \mid \mathcal{F}_{T(k-1)})}{S^c(s - \mid \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(ds \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(ds \mid \mathcal{F}_{T(k-1)}) \right) \left( \nu_{k-1,\tau}^* \left( du \mid \mathcal{F}_{T(k-1)} \right) \right)
\end{aligned}$$

Since also

$$\begin{aligned}
& \int \mathbb{1}\{t_1 < \dots < t_k < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^k \left( \frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})} \right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^k S_0^c(t_j - | f_{j-1})}{\prod_{j=1}^k S^c(t_j - | f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
& \quad \times \left( \mathbb{E}_{P_0} \left[ \bar{Z}_{k+1,\tau}^a(S_0^c, \bar{Q}_{k+1,\tau}^g) \mid \mathcal{F}_{T(k)} = f_k \right] - \nu_{k,\tau}(f_k) \right) P_{\mathcal{F}_{T(k)}}(df_k) \\
& + \int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left( \frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})} \right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^{k-1} S_0^c(t_j - | f_{j-1})}{\prod_{j=1}^{k-1} S^c(t_j - | f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} \\
& \quad \times \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S^c, \nu_k) \mid \mathcal{F}_{T(k-1)} = f_{k-1} \right] - \mathbb{E}_{P_0} \left[ \bar{Z}_{k,\tau}^a(S^c, \bar{Q}_{k,\tau}^g) \mid \mathcal{F}_{T(k-1)} = f_{k-1} \right] P(\mathcal{F}_{T(k-1)})(df_{k-1}) \\
& = \int \mathbb{1}\{t_1 < \dots < t_k < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^k \left( \frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})} \right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^k S_0^c(t_j - | f_{j-1})}{\prod_{j=1}^k S^c(t_j - | f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
& \quad \times (\bar{Q}_{k,\tau}^g(f_k) - \nu_{k,\tau}(f_k)) P_{\mathcal{F}_{T(k)}}(df_k) \tag{11} \\
& + \int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left( \frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})} \right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^{k-1} S_0^c(t_j - | f_{j-1})}{\prod_{j=1}^{k-1} S^c(t_j - | f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} \\
& \quad \int \mathbb{1}\{t_k < \tau\} \frac{S_0^c(t_k - | f_{k-1})}{S^c(t_k - | f_{k-1})} \\
& \quad \times \sum_{d_k=a,\ell} (\nu_k(t_k, d_k, g_k(a_k, d_k, f_{k-1}), l_k, f_{k-1}) - \bar{Q}_{k,\tau}^g(t_k, d_k, g_k(a_k, d_k, f_{k-1}), l_k, f_{k-1})) P_{T(k), \Delta(k), L(T(k)) \mid \mathcal{F}_{T(k-1)}}(df_k \mid f_{k-1}) P(\mathcal{F}_{T(k-1)})(df_{k-1}) \\
& = \int \mathbb{1}\{t_1 < \dots < t_k < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left( \frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})} \right)^{\mathbb{1}\{d_k=a\}} \left( \left( \frac{\pi_{0,k}(t_k, f_{k-1})}{\pi_k(t_k, f_{k-1})} \right)^{\mathbb{1}\{d_k=a\}} - 1 \right) \frac{\prod_{j=1}^k S_0^c(t_j - | f_{j-1})}{\prod_{j=1}^k S^c(t_j - | f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
& \quad \times (\bar{Q}_{k,\tau}^g(f_k) - \nu_{k,\tau}(f_k)) P_{\mathcal{F}_{T(k)}}(df_k)
\end{aligned}$$

where we set  $g_k(a_k, d_k, f_{k-1}) = 1$  for  $k > 1$ . By combining Equation 10 and Equation 11, we are done.  $\square$

Note that by the triangle inequality

$$\begin{aligned}
& \left| \int_{T(k-1)}^\tau \int_{T(k-1)}^u \left( \frac{S_0(s \mid \mathcal{F}_{T(k-1)})}{S(s \mid \mathcal{F}_{T(k-1)})} - 1 \right) \frac{S_0^c(s - | \mathcal{F}_{T(k-1)})}{S^c(s - | \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(ds \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(ds \mid \mathcal{F}_{T(k-1)}) \right) \left( \nu_{k-1,\tau}^*(du \mid \mathcal{F}_{T(k-1)}) \right) \right| \\
& \leq \nu_{k-1,\tau}^*(\tau) \sup_{u \in (T(k-1), \tau)} \left| \int_{T(k-1)}^u \left( \frac{S_0(s \mid \mathcal{F}_{T(k-1)})}{S(s \mid \mathcal{F}_{T(k-1)})} - 1 \right) \frac{S_0^c(s - | \mathcal{F}_{T(k-1)})}{S^c(s - | \mathcal{F}_{T(k-1)})} \left( \Lambda_{k,0}^c(ds \mid \mathcal{F}_{T(k-1)}) - \Lambda^c(ds \mid \mathcal{F}_{T(k-1)}) \right) \right|
\end{aligned}$$

## 8 Real data application

How should the methods be applied to real data and what data can we use?

Should we apply the methods to trial data? In that case, the visitation times may no longer be irregular, and we may have to rederive some of the results. Another possibility is to simply ignore the fact that the visitation times are regular and apply the methods as they are stated.

We also want to compare with other methods.

- comparison with LTMLE (Laan & Gruber, 2012).
- or multi-state models

Maybe we can look at the data applications in Kjetil Røyslands papers?

## 9 Simulation study

The data generating mechanism should be based on real data given in [Section 8](#). The dag for simulating the data can be found in [Figure 6](#). Depending on the results from the data application, we should consider

- machine learning methods if misspecification of the outcome model appears to be an issue with parametric models. If this is indeed the case, we want to apply the targeted learning framework and machine learning models for the estimation of the nuisance parameters.
- performance comparison with LTMLE/other methods.

## 10 Discussion

There is one main issue with the method that we have not discussed yet: In the case of irregular data, we may have few people with many events. For example there may only be 5 people in the data with a censoring event as their 4'th event. In that case, we can hardly estimate  $\lambda_4^c(\cdot | \mathcal{F}_{T_{(3)}})$  based on the data set with observations only for the 4'th event. One immediate possibility is to only use flexible machine learning models for the effective parts of the data that have a sufficiently large sample size and to use (simple) parametric models for the parts of the data that have a small sample size. By using cross-fitting/sample-splitting for this data-adaptive procedure, we will be able to ensure that the asymptotics are still valid. Another possibility is to only consider the  $k$  first (non-terminal) events in the definition of the target parameter. In that case,  $k$  will have to be specified prior to the analysis which may be a point of contention (otherwise we would have to use a data-adaptive target parameter ([Hubbard et al. \(2016\)](#))).

Let us discuss a pooling approach to handle the issue with few events. We consider parametric maximum likelihood estimation for the cumulative cause specific censoring-hazard  $\Lambda_{\theta_k}^c$  of the  $k$ 'th event. Pooling is that we use the model  $\Lambda_{\theta_j}^c = \Lambda_{\theta^*}^c$  for all  $j \in S \subseteq \{1, \dots, K\}$  and  $\theta^* \in \Theta^*$  which is variationally independent of the parameter spaces  $\theta_k \in \Theta_k$  for  $k \notin S$ . This is directly suggested by the point process likelihood, which we can write as

$$\begin{aligned} & \prod_{i=1}^n \pi_{t \in (0, \tau_{\max}]} (d\Lambda_{\theta^*}^c(t | \mathcal{F}_{t-}^i)^{\Delta N_i^c(t)} (1 - d\Lambda_{\theta^*}^c(t | \mathcal{F}_{t-}^i))^{1 - \Delta N_i^c(t)}) \\ &= \prod_{i=1}^n \left( \prod_{k=1}^{K_i(\tau)} d\Lambda_{\theta_k}^c(T_{(k)}^i | \mathcal{F}_{T_{(k-1)}^i}) \right)_{t \in (0, \tau_{\max}]} \left( 1 - d(\mathbb{1}\{t \in (T_{(k-1)}^i, T_{(k)}^i)\}) \Lambda_{\theta_k}^c(t | \mathcal{F}_{T_{(k-1)}^i}) \right)_{t \in (0, \tau_{\max}]} \left( 1 - d(\mathbb{1}\{t \in (T_{(K_i)}^i, \tau)\}) \Lambda_{\theta_{K_i+1}}^c(t | \mathcal{F}_{T_{(K_i)}^i}) \right) \\ &= \prod_{i=1}^n \left( \prod_{k \in S, k \leq K_i(\tau)+1} \left( d\Lambda_{\theta^*}^c(T_{(k)}^i | \mathcal{F}_{T_{(k-1)}^i}) \right)^{\mathbb{1}\{k \neq K_i+1\}} \pi_{t \in (0, \tau_{\max}]} \left( 1 - d(\mathbb{1}\{t \in (T_{(k-1)}^i, T_{(k)}^i)\}) \Lambda_{\theta^*}^c(t | \mathcal{F}_{T_{(k-1)}^i}) \right) \right) \\ & \quad \times \prod_{i=1}^n \left( \prod_{k \notin S, k \leq K_i(\tau)+1} \left( d\Lambda_{\theta_k}^c(T_{(k)}^i | \mathcal{F}_{T_{(k-1)}^i}) \right)^{\mathbb{1}\{k \neq K_i+1\}} \pi_{t \in (0, \tau_{\max}]} \left( 1 - d(\mathbb{1}\{t \in (T_{(k-1)}^i, T_{(k)}^i)\}) \Lambda_{\theta_k}^c(t | \mathcal{F}_{T_{(k-1)}^i}) \right) \right) \end{aligned}$$

(Note that we take  $T_{K_i+1}^i = \tau_{\max}$ ). Thus

$$\begin{aligned} & \operatorname{argmax}_{\theta^* \in \Theta^*} \prod_{i=1}^n \pi_{t \in (0, \tau_{\max}]} (d\Lambda_{\theta^*}^c(t | \mathcal{F}_{t-}^i)^{\Delta N_i^c(t)} (1 - d\Lambda_{\theta^*}^c(t | \mathcal{F}_{t-}^i))^{1 - \Delta N_i^c(t)}) \\ &= \operatorname{argmax}_{\theta^* \in \Theta^*} \prod_{i=1}^n \left( \prod_{k \in S, k \leq K_i(\tau)+1} \left( d\Lambda_{\theta^*}^c(T_{(k)}^i | \mathcal{F}_{T_{(k-1)}^i}) \right)^{\mathbb{1}\{k \neq K_i+1\}} \pi_{t \in (0, \tau_{\max}]} \left( 1 - d(\mathbb{1}\{t \in (T_{(k-1)}^i, T_{(k)}^i)\}) \Lambda_{\theta^*}^c(t | \mathcal{F}_{T_{(k-1)}^i}) \right) \right) \\ & \quad \times \prod_{i=1}^n \left( \prod_{k \notin S, k \leq K_i(\tau)+1} \left( d\Lambda_{\theta_k}^c(T_{(k)}^i | \mathcal{F}_{T_{(k-1)}^i}) \right)^{\mathbb{1}\{k \neq K_i+1\}} \pi_{t \in (0, \tau_{\max}]} \left( 1 - d(\mathbb{1}\{t \in (T_{(k-1)}^i, T_{(k)}^i)\}) \Lambda_{\theta_k}^c(t | \mathcal{F}_{T_{(k-1)}^i}) \right) \right) \end{aligned}$$

and that

$$\begin{aligned} & \operatorname{argmax}_{\theta^* \in \Theta^*} \prod_{i=1}^n \left( \prod_{k \in S, k \leq K_i(\tau)+1} \left( d\Lambda_{\theta^*}^c(T_{(k)}^i \mid \mathcal{F}_{T_{(k-1)}^i}) \right)^{\mathbb{1}\{k \neq K_i+1\}} \pi_{t \in (0, \tau_{\max}]} \left( 1 - d(\mathbb{1}\{t \in (T_{(k-1)}^i, T_{(k)}^i)\}) \Lambda_{\theta^*}^c(t \mid \mathcal{F}_{T_{(k-1)}^i}) \right) \right) \\ &= \operatorname{argmax}_{\theta^* \in \Theta^*} \left( \prod_{k \in S} \prod_{i=1}^n \left( d\Lambda_{\theta^*}^c(T_{(k)}^i \mid \mathcal{F}_{T_{(k-1)}^i}) \right)^{\mathbb{1}\{k < K_i+1\}} \pi_{t \in (0, \tau_{\max}]} \left( 1 - d(\mathbb{1}\{t \in (T_{(k-1)}^i, T_{(k)}^i)\}) \Lambda_{\theta^*}^c(t \mid \mathcal{F}_{T_{(k-1)}^i}) \right) \right) \end{aligned}$$

So we see that the maximization problem corresponds exactly to finding the maximum likelihood estimator on a pooled data set!

Other methods provide means of estimating the cumulative intensity  $\Lambda^x$  directly instead of splitting it up into  $K$  separate parameters. There exist only a few methods for estimating the cumulative intensity  $\Lambda^x$  directly (see [Liguori et al. \(2023\)](#) for neural network-based methods and [Weiss & Page \(2013\)](#) for a forest-based method).

Alternatively, we can use temporal difference learning to avoid iterative estimation of  $\bar{Q}_{k,\tau}^g$  altogether ([Shirakawa et al., 2024](#)).

One other direction is to use Bayesian methods. Bayesian methods may be particular useful for this problem since they do not have issues with finite sample size. They are also an excellent alternative to frequentist Monte Carlo methods for estimating the target parameter with [Equation 2](#) because they offer uncertainty quantification directly through simulating the posterior distribution whereas frequentist simulation methods do not.

A potential other issue with the estimation of the nuisance parameters are that the history is high dimensional. 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.

There is also the possibility for functional efficient estimation using the entire interventional cumulative incidence curve as our target parameter. There exist some methods for baseline interventions in survival analysis ([Cai & Laan \(2019\)](#); [Westling et al. \(2024\)](#)).

## Bibliography

- 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>
- Cai, W., & Laan, M. J. van der. (2019). One-Step Targeted Maximum Likelihood Estimation for Time-to-Event Outcomes. *Biometrics*, 76(3), 722–733. <https://doi.org/10.1111/biom.13172>
- Chamapiwa, E. (2018). *Application of Marginal Structural Models (MSMs) to Irregular Data Settings*. The University of Manchester (United Kingdom).
- 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>
- Didelez, V. (2008). Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1), 245–264. <https://doi.org/https://doi.org/10.1111/j.1467-9868.2007.00634.x>
- 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.
- Gill, R. D., & Robins, J. M. (2001). Causal Inference for Complex Longitudinal Data: The Continuous Case. *The Annals of Statistics*, 29(6), 1785–1811. <https://doi.org/10.1214/aos/1015345962>

- Hubbard, A. E., Kherad-Pajouh, S., & van der Laan, M. J. (2016). Statistical inference for data adaptive target parameters. *The International Journal of Biostatistics*, 12(1), 3–19.
- Laan, M. J. van der, & Gruber, S. (2012). *The International Journal of Biostatistics*, 8(1). <https://doi.org/doi:10.1515/1557-4679.1370>
- 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>
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.
- 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](https://doi.org/10.1007/978-1-4419-9782-1_4)
- Ryalen, P. (2024). *On the role of martingales in continuous-time causal inference*.
- 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., 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.
- 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>
- Weiss, J. C., & Page, D. (2013). Forest-Based Point Process for Event Prediction from Electronic Health Records. In H. Blockeel, K. Kersting, S. Nijssen, & F. Železný (Eds.), *Machine Learning and Knowledge Discovery in Databases: Machine Learning and Knowledge Discovery in Databases*. [https://doi.org/10.1007/978-3-642-40994-3\\_35](https://doi.org/10.1007/978-3-642-40994-3_35)
- Westling, T., Luedtke, A., Gilbert, P. B., & Carone, M. (2024). Inference for treatment-specific survival curves using machine learning. *Journal of the American Statistical Association*, 119(546), 1541–1553.

## 11 Appendix

### 11.1 Finite dimensional distributions and compensators

Let  $(\tilde{X}(t))_{t \geq 0}$  be a  $d$ -dimensional cadlag jump process, where each component  $i$  is two-dimensional such that  $\tilde{X}_i(t) = (N_i(t), X_i(t))$  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. Assume that the number of event times are bounded by a finite constant  $K$ . Furthermore, let  $\mathcal{F}_t = \sigma(\tilde{X}(s) \mid s \leq t) \vee \sigma(W)$  be the natural filtration. For each component  $\tilde{X}_i$ , let the corresponding random measure be given by

$$N_i(dt, dx) = \sum_{k: T_{(k)} < \infty} \delta_{(T_{(k)}, X(T_{(k)}))}(dt, dx).$$

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

**Theorem 7** (Finite-dimensional distributions): Under the stated conditions of this section, we have

1. We have  $\mathcal{F}_{T_{(k)}} = \sigma(T_{(k)}, \Delta_{(k)}, X(T_{(k)})) \vee \mathcal{F}_{T_{(k-1)}}$  and  $\mathcal{F}_0 = \sigma(W)$ .
2. There exist stochastic kernels  $\Lambda_{k,i}$  from  $\mathbb{F}_k$  to  $\mathbb{R}$  and  $\zeta_{k,i}$  from  $\mathbb{F}_k \times \mathbb{R}_+$  to  $\mathbb{R}_+$  such that the compensator for  $N_i$  is given by,

$$\Lambda_i(dt, dx) = \sum_{k: T_{(k)} < \infty} \mathbb{1}\{T_{(k-1)} < t \leq T_{(k)}\} \zeta_{k,i}(dx, t, \mathcal{F}_{T_{(k-1)}}) \Lambda_{k,i}(dt, \mathcal{F}_{T_{(k-1)}}).$$

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

3. The distribution of  $\mathcal{F}_{T_{(m)}}$  is given by

$$F_n(dw, dt_1, d\delta_1, dx_{11}, \dots, dx_{1d}, \dots, dt_n, d\delta_n, dx_{n1}, \dots, dx_{nd}) \\ = \left( \prod_{i=1}^n \mathbb{1}\{t_{i-1} < t_i\} \prod_{u \in (t_{i-1}, t_i)} \pi \left( 1 - \sum_{j=1}^d \Lambda_{i,j}(du, f_{i-1}) \right) \sum_{j=1}^d \delta_j(d\delta_i) \zeta_{i,j}(dx_{ij}, t_i, f_{i-1}) \Lambda_{i,j}(dt_i, f_{i-1}) \right) \mu(dw),$$

and  $f_k = (t_k, d_k, x_k, \dots, t_1, d_1, x_1, w) \in \mathbb{F}_k$  for  $n \in \mathbb{N}$ . Here  $\pi$  denotes the product integral (Gill & Johansen, 1990).

*Proof:* To prove 1, we first note that since the number of events are bounded, we the *minimality* condition of Theorem 2.5.10 of Last & Brandt (1995), the filtration  $\mathcal{F}_t^N = \sigma(N((0, s], \cdot) \mid s \leq t) \vee \sigma(W) = \mathcal{F}_t$  where

$$N(dt, dx) = \sum_{k: T_{(k)} < \infty} \delta_{(T_{(k)}, \tilde{X}(T_{(k)}))}(dt, dx)$$

Thus  $\mathcal{F}_{T_{(k)}} = \sigma(T_{(k)}, \tilde{X}(T_{(k)})) \vee \mathcal{F}_{T_{(k-1)}}$  and  $\mathcal{F}_0 = \sigma(W)$  in view of (2.2.44) of Last & Brandt (1995). To get 1, simply note that since the counting processes do not jump at the same time, there is a one-to-one corresponding between  $\Delta_{(k)}$  and  $N^i(T_{(k)})$  for  $i = 1, \dots, d$ .

To prove 2, simply use Theorem 4.1.11 of Last & Brandt (1995) which states that

$$\Lambda(dt, dx) = \sum_{k: T_{(k)} < \infty} \mathbb{1}\{T_{(k-1)} < t \leq T_{(k)}\} \frac{P((T_{(k)}, \tilde{X}(T_{(k)})) \in (dt, dx) \mid \mathcal{F}_{T_{(k-1)}})}{P(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}})}$$

is a  $P$ - $\mathcal{F}_t$  martingale. We can write that

$$\frac{P((T_{(k)}, \tilde{X}(T_{(k)})) \in (dt, dx) \mid \mathcal{F}_{T_{(k-1)}})}{P(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}})} = P(\tilde{X}(T_{(k)}) \in dx \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t) \frac{P(T_{(k)} \in dt \mid \mathcal{F}_{T_{(k-1)}})}{P(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}})}$$

Now write  $dx = (dm, dx_1, \dots, dx_d)$ , so we can write by the no simultaneous jumps condition,

$$P(\tilde{X}(T_{(k)}) \in dx \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t) \\ = \sum_{j=1}^d \delta_j(dm) P(\Delta_{(k)} = j \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t) P(X_j(T_{(k)}) \in dx_j \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t, \Delta_{(k)} = j) \prod_{l \neq j} \delta_{(X_l(T_{(k-1)}))}(dx_l)$$

Now note that

$$N_i(dt, dx) = N(dt, \mathcal{X}_1, \{0\}, \dots, \mathcal{X}_i, \{1\}, \dots, \mathcal{X}_d, \{0\})$$

so we find the compensator of  $N_i$  to be

$$\Lambda_i(dt, dx) = \sum_k \mathbb{1}\{T_{(k-1)} < t \leq T_{(k)}\} P\left(\Delta_{(k)} = i \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t\right) P\left(X_i(T_{(k)}) \in dx \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} = t, \Delta_{(k)} = i\right) \frac{P\left(T_{(k)} \in dt \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)}$$

Letting

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

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

completes the proof of 2.

3. is simply a straightforward extension of Proposition 1/Theorem 3 of [Ryalen \(2024\)](#)

or an application of Theorem 8.1.2 of [Last & Brandt \(1995\)](#). It also follows from iterative applications of 2.

□

## 11.2 Simulating the data

One classical causal inference perspective requires that we know how the data was generated up to unknown parameters (NPSEM) (Pearl, 2009). This approach has only to a small degree been discussed in continuous-time causal inference literature ([Røysland et al. \(2024\)](#)). This is initially considered in the uncensored case, but is later extended to the censored case. For a moment, we ponder how the data was generated given a DAG. The DAG given in the section provides a useful tool for simulating continuous-time data, but not for drawing causal inference conclusions. For the event times, we can draw a figure representing the data generating mechanism which is shown in [Figure 5](#). Some, such as [Chamapiwa \(2018\)](#), write down this DAG, but with an arrow from  $T_{(k)}$  to  $L(T_{(k)})$  and  $A(T_{(k)})$  instead of displaying a multivariate random variable which they deem the “time-as-confounder” approach to allow for irregularly measured data (see [Figure 6](#)). Fundamentally, this arrow would only be meaningful if the event time was known prior to the treatment and covariate value measurements, which they might not be. This can make sense if the event is scheduled ahead of time, but for, say, a stroke the time is not measured prior to the event. Because a cause must precede an effect, this makes the arrow invalid from a philosophical standpoint. On the other hand, DAGs such as the one in [Figure 5](#), are not informative about the causal relationships between the variables. This issue with simultaneous events is likely what has led to the introduction of local independence graphs ([Didelez \(2008\)](#)) but is also related to the notion that the treatment times are not predictable (that is knowable just prior to the event) as in [Ryalen \(2024\)](#).

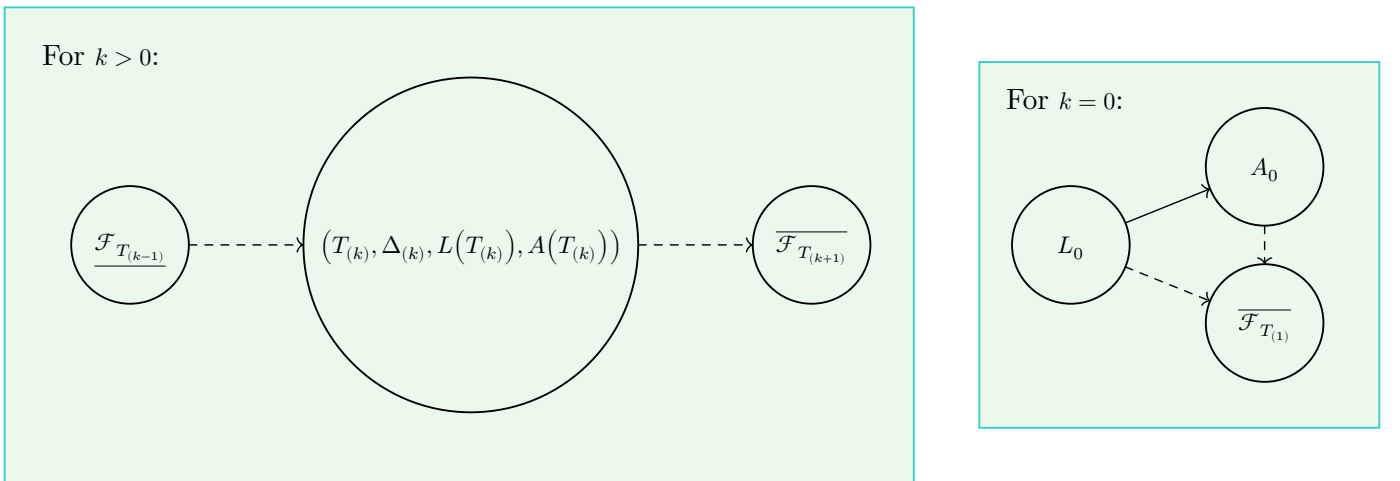


Figure 5: A DAG representing the relationships between the variables of  $O$ . The dashed lines indicate multiple edges from the dependencies in the past and into the future.



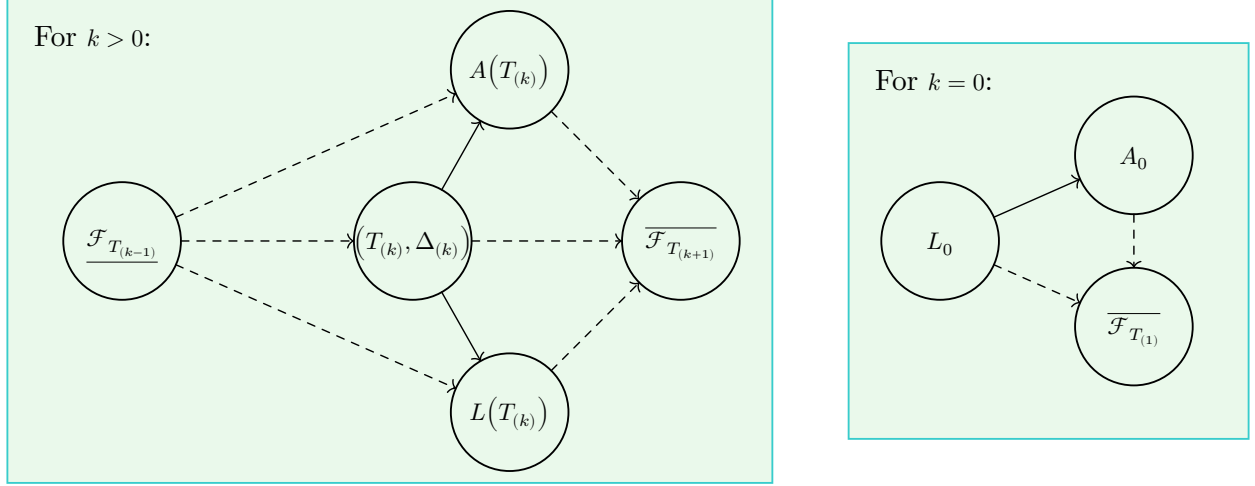


Figure 6: A DAG for simulating the data generating mechanism. The dashed lines indicate multiple edges from the dependencies in the past and into the future. Here  $\mathcal{F}_{T_{(k)}}$  is the history up to and including the  $k$ 'th event and  $\overline{\mathcal{F}_{T_{(k)}}}$  is the history after and including the  $k$ 'th event.

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

Let  $B_{k-1}(u) = (\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(u)) \frac{1}{\exp(-\sum_{x=a,\ell,d,y} \int_{T_{(k-1)}}^u \lambda_{k-1}^x(w, \mathcal{F}_{T_{(k-1)}}) dw)}$  and  $S(u | \mathcal{F}_{T_{(k-1)}}) = \exp(-\sum_{x=a,\ell,d,y} \int_{T_{(k-1)}}^u \lambda_{k-1}^x(w, \mathcal{F}_{T_{(k-1)}}) dw)$  and  $S^c(u | \mathcal{F}_{T_{(k-1)}}) = \exp(-\int_{T_{(k-1)}}^u \lambda_{k-1}^c(w, \mathcal{F}_{T_{(k-1)}}) dw)$ . We claim that the efficient influence function can also be written as:

$$\begin{aligned} \varphi_\tau^*(P) = & \sum_{k=1}^K \prod_{j=0}^{k-1} \left( \frac{\pi_{j-1}^*(T_{(j)}, A(T_{(j)}), \mathcal{F}_{T_{(j-1-1)}})}{\pi_{j-1}(T_{(j)}, \mathcal{F}_{T_{(j-1-1)}})} \right)^{I(\Delta_{(j)}=a)} \frac{I(\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} \leq \tau)}{\exp(-\sum_{1 \leq j < k} \int_{T_{(j-1)}}^{T_{(j)}} \lambda_{j-1}^c(s, \mathcal{F}_{T_{(j-1)}}) ds)} \left[ \right. \\ & \int_{T_{(k-1)}}^\tau \frac{1}{S^c(u | \mathcal{F}_{T_{(k-1)}})} \left( \int_{\mathcal{A}} \bar{Q}_{k,\tau}^g(L(T_{(k-1)}), a_k, s, a, \mathcal{F}_{T_{(k)}}) \pi_k^*(s, a_k, \mathcal{F}_{T_{(k-1)}}) \nu_A(da_k) - B_{k-1}(u) \right) M_k^a(du) \\ & + \int_{T_{(k-1)}}^\tau \frac{1}{S^c(u | \mathcal{F}_{T_{(k-1)}})} \left( \mathbb{E}_P \left[ \bar{Q}_{k,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] - B_{k-1}(u) \right) M_k^\ell(du) \\ & + \int_{T_{(k-1)}}^\tau \frac{1}{S^c(u | \mathcal{F}_{T_{(k-1)}})} (1 - B_{k-1}(u)) M_k^y(du) + \int_{T_{(k-1)}}^\tau \frac{1}{S^c(u | \mathcal{F}_{T_{(k-1)}})} (0 - B_{k-1}(u)) M_k^d(du) \\ & + \frac{1}{S^c(T_{(k)} | \mathcal{F}_{T_{(k-1)}})} I(T_{(k)} \leq \tau, \Delta_{(k)} = \ell, k < K) \left( \bar{Q}_{k,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), T_{(k)}, \ell, \mathcal{F}_{T_{(k-1)}}) \right. \\ & \left. \left. - \mathbb{E}_P \left[ \bar{Q}_{k-1,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), \widetilde{T}_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid \widetilde{T}_{(k)} = T_{(k)}, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \right] \\ & + \int \bar{Q}_{1,\tau}^g(a, L_0) \pi_0^*(0, a, \mathcal{F}_{T_{(0-1)}}) \nu_A(da) - \Psi_\tau(P) \end{aligned}$$

We find immediately that

$$\begin{aligned}
\varphi_\tau^*(P) = & \sum_{k=1}^K \prod_{j=0}^{k-1} \left( \frac{\pi_{j-1}^*(T_{(j)}, A(T_{(j)}), \mathcal{F}_{T_{(j-1-1)}})}{\pi_{j-1}(T_{(j)}, \mathcal{F}_{T_{(j-1-1)}})} \right)^{I(\Delta_{(j)}=a)} \frac{I(\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} \leq \tau)}{\exp\left(-\sum_{1 \leq j < k} \int_{T_{(j-1)}}^{T_{(j)}} \lambda_{j-1}^c(s, \mathcal{F}_{T_{(j-1)}}) ds\right)} \left[ \right. \\
& - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})} \left( \int_{\mathcal{A}} \bar{Q}_{k,\tau}^g(L(T_{(k-1)}), a_k, s, a, \mathcal{F}_{T_{(k)}}) \pi_k^*(s, a_k, \mathcal{F}_{T_{(k-1)}}) \nu_A(da_k) \right) \Lambda_k^a(du) \\
& - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})} \left( \mathbb{E}_P \left[ \bar{Q}_{k,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \Lambda_k^\ell(du) \\
& - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})} (1) \Lambda_k^y(du) - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})} (0) \Lambda_k^d(du) \\
& - \int_{T_{(k-1)}}^\tau \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})} B_{k-1}(u) M_k^\bullet(du) \\
& + \bar{Z}_{k,\tau}(T_{(k)}, \Delta_{(k)}, L(T_{(k)}), A(T_{(k)}), \mathcal{F}_{T_{(k-1)}}) + \int_{T_{(k-1)}}^\tau \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})} B_{k-1}(u) M_k^c(du) \left. \right] \\
& + \int \bar{Q}_{1,\tau}^g(a, L_0) \pi_0^*(0, a, \mathcal{F}_{T_{(0-1)}}) \nu_A(da) - \Psi_\tau(P)
\end{aligned}$$

Now note that

$$\begin{aligned}
& \int_{T_{(k-1)}}^\tau (\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(u)) \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}})} (N_k^\bullet(ds) - \tilde{Y}_{k-1}(s) \lambda_{k-1}^\bullet(s, \mathcal{F}_{T_{(k-1)}}) ds) \\
& = (\bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(T_{(k)})) \frac{1}{S^c(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}) S(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}})} \mathbb{1}\{T_{(k)} \leq \tau\} \\
& - \bar{Q}_{k-1,\tau}^g(\tau) \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(s, \mathcal{F}_{T_{(k-1)}}) ds \\
& + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{\bar{Q}_{k-1,\tau}^g(u)}{S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(s, \mathcal{F}_{T_{(k-1)}}) ds
\end{aligned}$$

Let us calculate the second integral

$$\begin{aligned}
& \bar{Q}_{k-1,\tau}^g(\tau) \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(s, \mathcal{F}_{T_{(k-1)}}) ds \\
& = \bar{Q}_{k-1,\tau}^g(\tau) \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}})}{(S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}}))^2} \lambda_{k-1}^\bullet(s, \mathcal{F}_{T_{(k-1)}}) ds \\
& = \bar{Q}_{k-1,\tau}^g(\tau) \left( \frac{1}{S^c(T_{(k)} \wedge \tau \mid \mathcal{F}_{T_{(k-1)}}) S(T_{(k)} \wedge \tau \mid \mathcal{F}_{T_{(k-1)}})} - 1 \right)
\end{aligned}$$

where the last line holds by the Duhamel equation (or using that the antiderivative of  $-\frac{f'}{f^2}$  is  $\frac{1}{f}$ ). The first of these integrals is equal to

$$\begin{aligned}
& \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{\bar{Q}_{k+1,\tau}^g(u)}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(u, \mathcal{F}_{T_{(k-1)}}) du \\
&= \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \left[ \int_0^u S(s \mid \mathcal{F}_{T_{(k-1)}}) \Lambda_{k-1}^a(ds, \mathcal{F}_{T_{(k-1)}}) \right. \\
&\quad \times \left( \int_{\mathcal{A}} \bar{Q}_{k+1,\tau}^g(L(T_{(k-1)}), a_k, s, a, \mathcal{F}_{T_{(k-1)}}) \pi_k^*(s, a_k, \mathcal{F}_{T_{(k-1)}}) \nu_A(da_k) \right) \\
&\quad + \int_0^u S(s \mid \mathcal{F}_{T_{(k-1)}}) \Lambda_{k-1}^\ell(ds, \mathcal{F}_{T_{(k-1)}}) \\
&\quad \times \left( \mathbb{E}_P \left[ \bar{Q}_{k+1,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \\
&\quad \left. + \int_0^u S(s \mid \mathcal{F}_{T_{(k-1)}}) \Lambda_{k-1}^y(ds, \mathcal{F}_{T_{(k-1)}}) \right] \\
&\times \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(u, \mathcal{F}_{T_{(k-1)}}) du \\
&= \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \int_s^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(u, \mathcal{F}_{T_{(k-1)}}) du \left[ S(s \mid \mathcal{F}_{T_{(k-1)}}) \Lambda_{k-1}^a(ds, \mathcal{F}_{T_{(k-1)}}) \right. \\
&\quad \times \left( \int_{\mathcal{A}} \bar{Q}_{k+1,\tau}^g(L(T_{(k-1)}), a_k, s, a, \mathcal{F}_{T_{(k-1)}}) \pi_k^*(s, a_k, \mathcal{F}_{T_{(k-1)}}) \nu_A(da_k) \right) \\
&\quad + S(s \mid \mathcal{F}_{T_{(k-1)}}) \Lambda_{k-1}^\ell(ds, \mathcal{F}_{T_{(k-1)}}) \\
&\quad \times \left( \mathbb{E}_P \left[ \bar{Q}_{k+1,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \\
&\quad \left. + S(s \mid \mathcal{F}_{T_{(k-1)}}) \Lambda_{k-1}^y(ds, \mathcal{F}_{T_{(k-1)}}) \right]
\end{aligned}$$

Now note that

$$\begin{aligned}
& \int_s^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}})S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(u, \mathcal{F}_{T_{(k-1)}}) du \\
&= \frac{1}{S^c(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}})S(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}})} - \frac{1}{S^c(s \mid \mathcal{F}_{T_{(k-1)}})S(s \mid \mathcal{F}_{T_{(k-1)}})}
\end{aligned}$$

Setting this into the previous integral, we get

$$\begin{aligned}
& - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda_{k-1}^\bullet(u, \mathcal{F}_{T_{(k-1)}}) du \left[ \Lambda_{k-1}^a(ds, \mathcal{F}_{T_{(k-1)}}) \right. \\
&\quad \times \left( \int_{\mathcal{A}} \bar{Q}_{k+1,\tau}^g(L(T_{(k-1)}), a_k, s, a, \mathcal{F}_{T_{(k-1)}}) \pi_k^*(s, a_k, \mathcal{F}_{T_{(k-1)}}) \nu_A(da_k) \right) \\
&\quad + \Lambda_{k-1}^\ell(ds, \mathcal{F}_{T_{(k-1)}}) \\
&\quad \times \left( \mathbb{E}_P \left[ \bar{Q}_{k+1,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \\
&\quad \left. + \Lambda_{k-1}^y(ds, \mathcal{F}_{T_{(k-1)}}) \right] \\
&+ \frac{1}{S^c(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}})S(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}})} \bar{Q}_{k-1,\tau}^g(\tau \wedge T_{(k)})
\end{aligned}$$

Thus, we find

$$\begin{aligned}
& \int_{T_{(k-1)}}^{\tau} \left( \bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(u) \right) \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}})} \left( N_k^\bullet(ds) - \tilde{Y}_{k-1}(s) \lambda_{k-1}^\bullet(s, \mathcal{F}_{T_{(k-1)}}) ds \right) \\
&= \left( \bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(T_{(k)}) \right) \frac{1}{S^c(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}) S(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}})} \mathbb{1}\{T_{(k)} \leq \tau\} \\
&\quad - \bar{Q}_{k-1,\tau}^g \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(s, \mathcal{F}_{T_{(k-1)}}) ds \\
&\quad + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{\bar{Q}_{k-1,\tau}^g(u)}{S^c(u \mid \mathcal{F}_{T_{(k-1)}}) S(u \mid \mathcal{F}_{T_{(k-1)}})} \lambda_{k-1}^\bullet(s, \mathcal{F}_{T_{(k-1)}}) ds \\
&= \left( \bar{Q}_{k-1,\tau}^g(\tau) - \bar{Q}_{k-1,\tau}^g(T_{(k)}) \right) \frac{1}{S^c(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}) S(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}})} \mathbb{1}\{T_{(k)} \leq \tau\} \\
&\quad - \left( \bar{Q}_{k-1,\tau}^g(\tau) \left( \frac{1}{S^c(T_{(k)} \wedge \tau \mid \mathcal{F}_{T_{(k-1)}}) S(T_{(k)} \wedge \tau \mid \mathcal{F}_{T_{(k-1)}})} \right) - 1 \right) \\
&\quad - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda_{k-1}^\bullet(u, \mathcal{F}_{T_{(k-1)}}) du \left[ \Lambda_{k-1}^a(ds, \mathcal{F}_{T_{(k-1)}}) \right. \\
&\quad \quad \times \left( \int_{\mathcal{A}} \bar{Q}_{k+1,\tau}^g(L(T_{(k-1)}), a_k, s, a, \mathcal{F}_{T_{(k-1)}}) \pi_k^*(s, a_k, \mathcal{F}_{T_{(k-1)}}) \nu_A(da_k) \right) \\
&\quad \quad + \Lambda_{k-1}^\ell(ds, \mathcal{F}_{T_{(k-1)}}) \\
&\quad \quad \times \left( \mathbb{E}_P \left[ \bar{Q}_{k+1,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \\
&\quad \quad \left. + \Lambda_{k-1}^y(ds, \mathcal{F}_{T_{(k-1)}}) \right] \\
&\quad + \frac{1}{S^c(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}) S(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}})} \bar{Q}_{k-1,\tau}^g(\tau \wedge T_{(k)}) \\
&= - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda_{k-1}^\bullet(u, \mathcal{F}_{T_{(k-1)}}) du \left[ \Lambda_{k-1}^a(ds, \mathcal{F}_{T_{(k-1)}}) \right. \\
&\quad \quad \times \left( \int_{\mathcal{A}} \bar{Q}_{k+1,\tau}^g(L(T_{(k-1)}), a_k, s, a, \mathcal{F}_{T_{(k-1)}}) \pi_k^*(s, a_k, \mathcal{F}_{T_{(k-1)}}) \nu_A(da_k) \right) \\
&\quad \quad + \Lambda_{k-1}^\ell(ds, \mathcal{F}_{T_{(k-1)}}) \\
&\quad \quad \times \left( \mathbb{E}_P \left[ \bar{Q}_{k+1,\tau}^g(L(T_{(k)}), A(T_{(k-1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \\
&\quad \quad \left. + \Lambda_{k-1}^y(ds, \mathcal{F}_{T_{(k-1)}}) \right] + \bar{Q}_{k-1,\tau}^g(\tau)
\end{aligned}$$