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 τ under prespecified treatment regimens. Most of the existing estimators based on observational data require a projection onto a discretized time scale Rose & van der Laan (2011). We consider a recently developed continuous-time approach to causal inference in this setting Rytgaard et al. (2022), which theoretically allows preservation of the precise event timing on a subject level. Working on a continuous-time scale may improve the predictive accuracy and reduce the loss of information. However, continuous-time extensions of the standard estimators comes at the cost of increased computational burden. We 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

- [x] Clean up figures.
- [x] Clean up existence of compensator + integral.
- [x] identifiability. My potential outcome approach. Add figure for potential outcome processes. Show full identification formula without reweighting
- [x] Censoring. Independent censoring IPCW rigorously.
- [x] 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 τ_{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. Without loss of register data applications, we assume that the maximum number of treatment and covariate changes of an individual is bounded by $K = 10,000$. Practically, we shall adapt K to our data and our target parameter. 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^ℓ, 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 τ. 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.

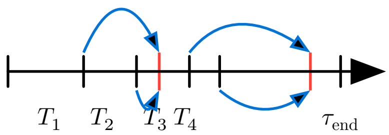

Figure 1: The "usual" approach where time is discretized. Each event time and its corresponding mark is rolled forward to the next time grid point, that is the values of the observations are updated based on the on the events occuring in the previous time interval.

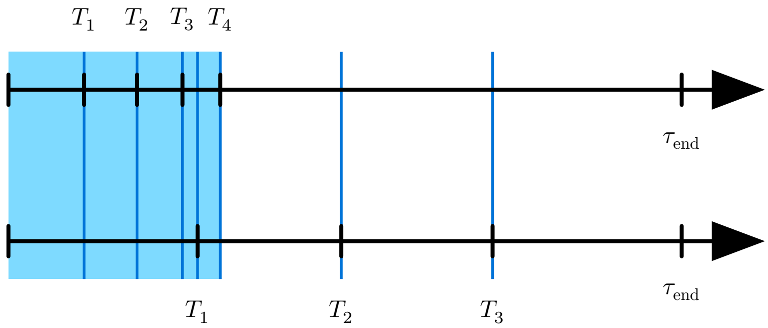

Figure 2: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let $t_1 < \ldots < t_m$ be all the event times in the sample. Then, given $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{t_r}\right]$, we regress back to $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{t_{r-1}}\right]$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

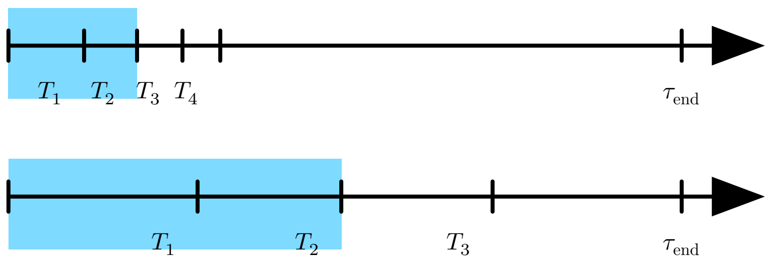

Figure 3: The figure illustrates the sequential regression approach proposed in this article. For each event k in the sample, we regress back on the history $\mathcal{F}_{T_{(k-1)}}$. That is, given $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{T_{(k)}}\right]$, we regress back to $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{T_{(k-1)}}\right]$. 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 $L(t) : \Omega \to \mathbb{R}^m$ and $A(t) : \Omega \to \mathbb{R}$ 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^ℓ, 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^y_t. 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 $\mathcal{F}_t = \sigma((N^a_s, A(s), N^\ell_s, L(s), N^y_s, N^d_s) \mid s \le t)$ be the natural filtration generated by the process up to time t. We observe $\mathcal{F}_{T_{(K)}} = \left(T_{(K)}, \Delta_{(K)}, T_{(K-1)}, \Delta_{(K-1)}, A\left(T_{(K-1)}\right), L\left(T_{(K-1)}\right), ..., A_0, L_0\right) \sim P \in \mathcal{M}$, where $T_{(k)}$ may be possibly ∞ 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 $\pi_k\left(da, \mathcal{F}_{T_{(k-1)}}\right)$ 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 $\mu_k\left(d\ell, t, \mathcal{F}_{T_{(k-1)}}\right)$ 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 $\lambda^x_{k-1}\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ 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 $P\left(T_{(k)} \in \cdot \mid \mathcal{F}_{T_{(k-1)}}\right) \ll m$ P-a.s., and $P\left(A\left(T_{(k)}\right) \in \cdot \mid T_{(k)} = t, \Delta_{(k)} = a, \mathcal{F}_{T_{(k-1)}}\right) \ll \nu_a$ P-a.s. and $P\left(L\left(T_{(k)}\right) \in \cdot \mid T_{(k)} = t, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right) \ll \nu_\ell$ P-a.s., where m is the Lebesgue measure on $\mathbb{R}_+$, ν_a is a measure on $\mathcal{A}$, and ν_ℓ is a measure on $\mathcal{L}$.

4 Target parameter and potential outcome framework

We now take an interventionalist stance to causal inference such as the one given in Ryalen (2024). In the interventionalist 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 it 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 $\Psi^g_\tau : \mathcal{M} \to \mathbb{R}$ is the mean interventional potential outcome at time τ,

$$\Psi^g_\tau(P) = \mathbb{E}_P\left[\tilde{N}^y_\tau\right],$$

where $\tilde{N}^y_t$ 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.,

$$T^a = \inf_{t \ge 0}\{A(t) = 0\} = \inf_{k > 1}\left\{T_{(k)} \mid \Delta_{(k)} = a, A\left(T_{(k)}\right) \ne 1\right\} \wedge \infty \mathbb{1}\{A(0) = 1\}$$

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

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

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

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

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

Then the estimand of interest is identifiable by

$$\Psi^g_\tau(P) = \mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P[Y_t W_t].$$

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

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

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

For the intersection property of conditional independence, a sufficient condition for Equation 1 is that

$$
\tilde{Y}_\tau \perp A\left(T_{(k)}\right) \mid T_{(j)} \le \tau < T_{(j+1)}, \mathcal{F}_{T_{(k)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \ge k > 0,
$$

$$
\mathbb{1}\left\{T_{(j)} \le \tau < T_{(j+1)}\right\} \perp A\left(T_{(k)}\right) \mid \tilde{Y}_\tau, \mathcal{F}_{T_{(k)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \ge k > 0.
$$

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 $\mathbb{1}\left\{T_{(k)} \le \tau, \Delta_{(k)} = y\right\}$. Then the exchangeability condition is that $\tilde{Y}_{\tau,k} \perp A\left(T_{(j)}\right) \mid \mathcal{F}_{T_{(j-1)}}, T_{(j)}, \Delta_{(j)} = a)$ for $0 \le j < k$ and $k = 1, \ldots, K$. 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.

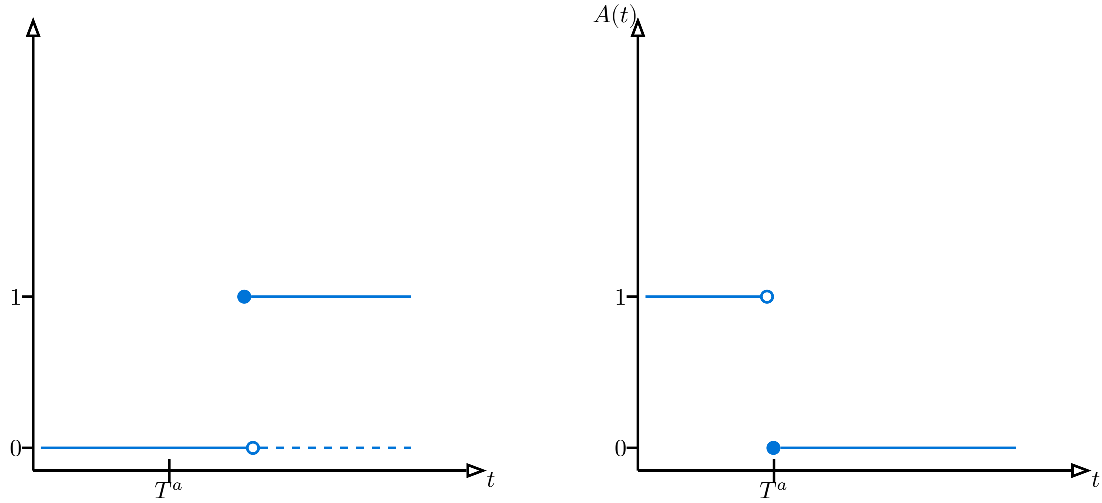

Figure 4: The figure illustrates the consistency condition for the potential outcome framework. The left panel shows the potential outcome process $\tilde{Y}_t$ (dashed) and the observed process Y_t (solid). The right panel shows the treatment process $A(t)$.

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 $\bar{Q}_K = \mathbb{1}\left\{T_{(K)} \leq \tau, \Delta_{(K)} = y\right\}$ and $\bar{Q}_k = \mathbb{E}_P\left[\sum_{j=k+1}^{K} W_{k,j}\mathbb{1}\left\{T_{(j)} \leq \tau, \Delta_{(j)} = y\right\} \mid \mathcal{F}_{T_{(k)}}\right]$. Then,

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

and

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

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

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

by the law of iterated expectations and that

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

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

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

by the law of iterated expectations. □

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

$$p_{ka}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = \int_{\mathcal{F}_{T_{(k-1)}}}^{t} S_k\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)\bar{Q}^g_{k+1,\tau}\left(1, L\left(T_{(k-1)}\right), s, a, \mathcal{F}_{T_{(k-1)}}\right)\lambda^a_k\left(s, \mathcal{F}_{T_{(k)}}\right)ds$$

$$p_{k\ell}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right)$$
$$= \int_{\mathcal{F}_{T_{(k-1)}}}^{t} S_k\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)\left(\mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}\left(A\left(T_{(k-1)}\right), L\left(T_{(k)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\right)\lambda^y_k\left(s, \mathcal{F}_{T_{(k)}}\right)ds$$

$$p_{ky}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = \int_{\mathcal{F}_{T_{(k-1)}}}^{t} S_k\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)\lambda^y_k\left(s, \mathcal{F}_{T_{(k)}}\right)ds$$

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

$$\bar{Q}^g_{k,\tau} = p_{ka}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k\ell}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{ky}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) \tag{2}$$

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

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

Since $\lambda^x_{k-1}\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ 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 $\left(T_{(k)}, \Delta_{(k)}\right)$ given $\mathcal{F}_{T_{(k-1)}}$ is given by

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

Putting this into the expectation of Equation 3, we get the claim. □

5 Censoring

In this section, we introduce a right-censoring time $C > 0$ at which we stop observing the multivariate jump process $Z(t) = \left(N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t)\right)$. Denote by $N^c(t)$ the counting process for the censoring process and its filtration $\mathcal{G}_t = \sigma(N^c(s) \mid s \le t)$. 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

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

Let $\lambda^c_{k-1}\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ 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 $S^c\left(t- \mid \mathcal{F}_{T_{(k-1)}}\right) = \prod_{s\in\left(T_{(k-1)},t\right)}\left(1 - \lambda^c_{k-1}\left(s, \mathcal{F}_{T_{(k-1)}}\right)ds\right)$ 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 $\mathcal{F}^{\text{full}}_{t\wedge C\wedge T^e} \subseteq \mathcal{F}^{\text{full}}_t$. From this, we get the observed data,

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

Denote by

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

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 6 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

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

Consequently, we have that

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

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

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

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

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

by Theorem 6 and by assumption also the compensator for the filtration $\mathcal{F}_t^{\text{full}}$. By the innovation theorem Andersen et al. (1993),

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

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 6 directly to the observed process to get that

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

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. □

Interestingly, Equation 2 corresponds exactly to the target parameter of Rytgaard et al. (2022) by plugging in the definitions of $\bar{Q}^g_{k,\tau}$ 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 τ.

- For each event point $k = K, K-1, ..., 1$ (starting with $k = K$):
  1. Obtain $\hat{S}^c\left(t \mid \mathcal{F}_{T_{(k-1)}}\right)$ 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:
  $$\hat{\eta}_k = \frac{\mathbb{1}\left\{T_{(k)} \leq \tau, \Delta_k \in \{a, \ell\}, k < K\right\} \hat{\nu}_k\left(\mathcal{F}^{-A}_{T_{(k)}}, \mathbf{1}\right)}{\hat{S}^c\left(T_{(k)} \mid \mathcal{F}^{-A}_{T_{(k-1)}}, \mathbf{1}\right)}$$
  Then calculate the subject-specific pseudo-outcome
  $$\hat{R}_k = \frac{\mathbb{1}\left\{T_{(k)} \leq \tau, \Delta_k = y\right\}}{\hat{S}^c\left(T_{(k)} \mid \mathcal{F}^{-A}_{T_{(k-1)}}, \mathbf{1}\right)} + \hat{\eta}_k$$
  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 $\hat{\nu}_{k-1} : \mathcal{H}_{k-1} \to \mathbb{R}_+$.
- 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 $\bar{Q}^g_{k,\tau}$ have the interpretation of the heterogenous causal effect after k events.

For now, we recommend Equation 4 for estimating $\bar{Q}^g_{k,\tau}$ 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 $\| \hat{\nu}_{k+1} - \bar{Q}^g_{k+1,\tau} \|_{L^2(P)} = o_P(1)$, $\| \hat{\Lambda}^c_k - \Lambda^c_k \|_{L^2(P)} = o_P(1)$. For the censoring, we need that $\hat{\Lambda}^c_k$ and Λ^c_k are uniformly bounded, that is $\hat{\Lambda}^c_k(t \mid f_{k-1}) \leq C$ and $\Lambda^c_k(t \mid (f_{k-1})) \leq C$ 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

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

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

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

Proof: By the triangle inequality,

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

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

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

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

5.2 Alternative nuisance parameter estimators

An alternative is to estimate the entire cumulative hazards Λ^x at once instead of having K separate parameters: There are very few methods for marked point process estimation but see Liguori et al. (2023) for methods mostly based on neural networks or Weiss & Page (2013) for a forest-based method. As a final alternative, we can use temporal difference learning to avoid iterative estimation of $\bar{Q}^a, \tilde{Q}$ Shirakawa et al. (2024). Most point process estimators are actually on the form given in terms of ref:intensity.

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

$$\varphi^*(P) = \frac{\mathbb{1}\{A_0 = 1\}}{\pi_0(L(0))} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\left\{A\left(T_{(j)}\right) = 1\right\}}{\pi_j\left(T_{(j)}, \mathcal{F}_{T_{(j-1)}}\right)} \right)^{\mathbb{1}\left\{\Delta_{(j)}=a\right\}} \frac{1}{\prod_{j=1}^{k-1} S^c\left(T_{(j)} - \mid \mathcal{F}_{T_{(j-1)}}\right)} \mathbb{1}\left\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\right\}$$
$$\times \left(\left(\bar{Z}^a_{k,\tau} - \bar{Q}^g_{k-1,\tau}\right) + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \left(\bar{Q}^g_{k-1,\tau}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) - \bar{Q}^g_{k-1,\tau}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\right) \frac{1}{S^c\left(u - \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} M^c_k(\mathrm{d}u) \right) \quad (8)$$
$$+\bar{Q}^g_{0,\tau} - \Psi_\tau(P),$$

where $M^c_k(t) = \mathbb{1}\left\{T_{(k-1)} < t \le T_{(k)}\right\}\left(N^c(t) - \Lambda^c\left(t \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$ is the martingale for the censoring process.

Proof: Define (sorry about the notation!)

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

and let

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

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.

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

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,

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

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

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon} \bar{Q}^{a,\varepsilon}_{k-1,\tau}(a_{k-1}, l_{k-1}, t_{t-1}, d_{k-1}, f_{k-2}) \Big((1-\varepsilon) P + \varepsilon \delta_{\mathcal{F}_{T_{(k-1)}}} \Big) \Big|_{\varepsilon=0} \\
&= \frac{\delta_{\mathcal{F}_{T_{(k-1)}}}(f_{k-1})}{P\Big(\mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)} \Bigg(\bar{Z}^a_{k,\tau} - \bar{Q}^g_{k-1,\tau}\Big(\tau, \mathcal{F}_{T_{(k-1)}}\Big) + \\
&+ \int_{T_{(k-1)}}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{T_{(k-1)}}^{t_k} \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda^x_{k-1}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s\Big)} \Big(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s \Big) \\
&\qquad P_{\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big) \Bigg) \\
&+ \int_{T_{(k-1)}}^{\tau} \Big(\frac{I(t_k \leq \tau, d_k \in \{a, \ell\})}{\exp\Big(-\int_{T_{(k-1)}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\Big)} \cdot \left(\frac{\pi^*_{k-1}\Big(t_k, a_k, \mathcal{F}_{T_{(k-1-1)}}\Big)}{\pi_{k-1}\Big(t_k, \mathcal{F}_{T_{(k-1-1)}}\Big)} \right)^{I(d_k = a)} \frac{\partial}{\partial \varepsilon} \bar{Q}^{a,\varepsilon}_{k-1,\tau}(a_k, l_k, t_k, d_k, f_{k-1}) \Big((1-\varepsilon) P + \varepsilon \delta_{\mathcal{F}_{T_{(k)}}} \Big) \Big|_{\varepsilon=0} \\
&\qquad P_{\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)
\end{aligned}
$$

Now note for the second term, we can write

$$
\begin{aligned}
&\int_{T_{(k-1)}}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{T_{(k-1)}}^{t_k} \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda^x_{k-1}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s\Big)} \Big(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s \Big) \\
&\qquad P_{\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)) \\
&= \int_{T_{(k-1)}}^{\tau} \int_{s}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) P_{\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big) \\
&\qquad \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda^x_{k-1}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s\Big)} \Big(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s \Big)) \\
&= \int_{T_{(k-1)}}^{\tau} \big(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s) \big) \\
&\qquad \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda^x_{k-1}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s\Big)} \Big(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s \Big))
\end{aligned}
$$

by an exchange of integerals. Combining the results iteratively gives the result. □

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 $\bar{Q}^g_{k,\tau}(t)$ at a sufficiently dense grid of time points t. Unfortunately, the event-specific cause-specific hazards $\hat{\Lambda}^x_k$ cannot readily be used to estimate $\bar{Q}^g_{Q,\tau_k}$ 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 $\bar{Q}^g_{k,\tau}$. Instead, we split up the estimation the estimation into two parts for $\bar{Q}^g_{k,\tau}$. For each k, the procedure constructs two new estimators of $\bar{Q}^g_{k,\tau}$:

1. $\hat{\nu}_{k,\tau}\Big(\mathcal{F}_{T_{(k)}}\Big)$ 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 $\left(A\left(T_{(k)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)$ (i.e., we do not include the latest covariate value). Given cause-specific estimators $\hat{\Lambda}^x_{k+1}$ for $x = a, l, d, y$, we estimate $\bar{Q}^g_{k,\tau}\left(t, \mathcal{F}_{T_{(k)}}\right)$ by

$$\hat{\nu}^*_{k,\tau}\left(t \mid \mathcal{F}_{T_{(k)}}\right) = \int_0^{t-T_{(k)}} \prod_{s \in \left(0, u - T_{(k)}\right)} \left(1 - \sum_{x=a,l,d,y} \hat{\Lambda}^x_{k+1}\left(ds \mid \mathbf{1}, \mathcal{F}^{-A}_{T_{(k)}}\right)\right) \left[\hat{\Lambda}^y_{k+1}\left(du \mid \mathbf{1}, \mathcal{F}^{-A}_{T_{(k)}}\right)\right.$$
$$+\tilde{\nu}_{k+1,\tau}\left(1, u + T_{(k)}, a, \mathbf{1}, \mathcal{F}^{-A}_{T_{(k-1)}}\right) \hat{\Lambda}^a_{k+1}\left(du \mid \mathbf{1}, \mathcal{F}^{-A}_{T_{(k)}}\right)$$
$$\left. +\tilde{\nu}_{k+1,\tau}\left(1, u + T_{(k)}, \ell, \mathbf{1}, \mathcal{F}^{-A}_{T_{(k-1)}}\right) \hat{\Lambda}^\ell_{k+1}\left(du \mid \mathbf{1}, \mathcal{F}^{-A}_{T_{(k)}}\right)\right]$$

on the interevent level.

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

$$\varphi^*\left(\hat{P}^*_n\right) = \frac{\mathbb{1}\{A_0 = 1\}}{\hat{\pi}_0(L(0))} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\left\{A\left(T_{(j)}\right) = 1\right\}}{\hat{\pi}\left(T_{(j)}, A\left(T_{(j)}\right), \mathcal{F}_{T_{(j-1)}}\right)}\right)^{\mathbb{1}\left\{\Delta_{(j)} = a\right\}} \frac{1}{\prod_{j=1}^{k-1} \hat{S}^c\left(T_{(j)} - \mid \mathcal{F}_{T_{(j-1)}}\right)} \mathbb{1}\left\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\right\}$$
$$\times \left(\bar{Z}^a_{k,\tau}\left(\hat{S}^c, \hat{\nu}_{k,\tau}\right) - \hat{\nu}_{k-1,\tau}\left(\mathcal{F}_{T_{(k-1)}}\right) + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \left(\hat{\nu}^*_{k-1}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) - \hat{\nu}^*_{k-1,\tau}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\right) \frac{1}{\hat{S}^c\left(u - \mid \mathcal{F}_{T_{(k-1)}}\right) \hat{S}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} M^c_k(\mathrm{d}u)\right))$$
$$+\hat{\nu}_{0,\tau}(1, \mathcal{F}_0) - \mathbb{P}_n \hat{\nu}_{0,\tau}(1, \cdot)$$

The resulting estimator is given by

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

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

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

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

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

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

Proof: By the triangle inequality,

$$\| \hat{\nu}^*_{k,\tau}(\cdot) - \bar{Q}^g_{k,\tau}(\cdot) \|_{L^2(P)} \leq \sqrt{\int \int_{t_k}^{\tau} \left(\hat{S}_{k+1}(t|f_k) \hat{\lambda}^y_{k+1}(t \mid f_k) - S_{k+1}(t|f_k) \lambda^y_{k+1}(t \mid f_k)\right)^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)}$$
$$+\sqrt{\int \int_{t_k}^{\tau} \left(\hat{\lambda}^a_{k+1}(t \mid f_k) \hat{S}_{k+1}(t|f_k) - S_{k+1}(t|f_k) \lambda^a_{k+1}(t \mid f_k)\right)^2 (\tilde{\nu}(1, t, a, f_k))^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)}$$
$$+\sqrt{\int \int_{t_k}^{\tau} \left(\hat{\lambda}^\ell_{k+1}(t \mid f_k) \hat{S}_{k+1}(t|f_k) - S_{k+1}(t|f_k) \lambda^\ell_{k+1}(t \mid f_k)\right)^2 (\tilde{\nu}(a_{k-1}, t, a, f_k))^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)}$$
$$+\sqrt{\int \int_{t_k}^{\tau} \left(\tilde{\nu}^*_{k+1,\tau}(t, \ldots) - \bar{Q}^{-L}_{k+1}(t, \ldots)\right)^2 \left(\sum_{x=a,l,y} S_{k+1}(t|f_k) \lambda^x_{k+1}(t \mid f_k)\right) dt P_{\mathcal{F}_{T_{(k-1)}}}(df_k)}$$

The last term is $o_P\left(n^{-\frac{1}{4}}\right)$ by assumption. By bounding $\tilde{\nu}$, the first three terms are then also $o_P\left(n^{-\frac{1}{4}}\right)$. 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 $o_P\left(n^{-\frac{1}{4}}\right)$. □

7.1 Remainder term

We now consider the efficient influence function, occuring in the remainder term. First define

$$\varphi_k^*(P) = \frac{\mathbb{1}\{A_0 = 1\}}{\pi_0(L(0))} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\left\{A\left(T_{(j)}\right) = 1\right\}}{\pi_j\left(T_{(j)}, \mathcal{F}_{T_{(j-1)}}\right)} \right)^{\mathbb{1}\left\{\Delta_{(j)}=a\right\}} \frac{1}{\prod_{j=1}^{k-1} S^c\left(T_{(j)} - \mid \mathcal{F}_{T_{(j-1)}}\right)} \mathbb{1}\left\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\right\}$$

$$\times \left(\left(\bar{Z}_{k,\tau}^a - \bar{Q}_{k-1,\tau}^g\right) + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \left(\bar{Q}_{k-1,\tau}^g\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) - \bar{Q}_{k-1,\tau}^g\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\right) \frac{1}{S^c\left(u - \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} M_k^c(\mathrm{d}u) \right)$$

for $k > 0$ and define $\varphi_0^*(P) = \bar{Q}_{0,\tau}^g - \Psi_\tau(P)$, so that

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

Apply the law of iterated expectation to the efficient influence function in Equation 8 to get

$$\mathbb{E}_{P_0}[\varphi_k^*(P)] = \int \mathbb{1}\{t_1 < \ldots < t_{k-1} < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left(\frac{\pi_{0,j}(t_k, f_{j-1})}{\pi_j(t_k, f_{j-1})} \right)^{\mathbb{1}\{d_j = a\}} \frac{\prod_{j=1}^{k-1} S_0^c(t_j - \mid f_{j-1})}{\prod_{j=1}^{k-1} S^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \ldots, d_{k-1} \in \{\ell, a\}\}$$

$$\times (\mathbb{E}_P\left[h_k\left(\mathcal{F}_{T_{(k)}}\right) \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\right] P_{\mathcal{F}^{\text{full}}_{T_{k-1}}}(df_{k-1})$$

where $h_k\left(\mathcal{F}_{T_{(k)}}\right) = \bar{Z}_{k,\tau}^a - \bar{Q}_{k-1,\tau}^g + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \left(\bar{Q}_{k-1,\tau}^g\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) - \bar{Q}_{k-1,\tau}^g\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\right) \frac{1}{S^c\left(u - \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} M_k^c(\mathrm{d}u) \mid \mathcal{F}_{T_{(k-1)}}]$.

Now note that

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

by a martingale argument. Noting that,

$$\mathbb{E}_{P_0}\left[\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\left(\nu^*_{k-1}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right)-\nu^*_{k-1,\tau}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\frac{1}{S^c\left(u- \mid \mathcal{F}_{T_{(k-1)}}\right)S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}M^c_k(\mathrm{d}u) \mid \mathcal{F}_{T_{(k-1)}}\right]$$
$$=\int_{T_{(k-1)}}^{\tau}\left(\nu^*_{k-1}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right)-\nu^*_{k-1,\tau}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\frac{S^c_0\left(u- \mid \mathcal{F}_{T_{(k-1)}}\right)S_0\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(u- \mid \mathcal{F}_{T_{(k-1)}}\right)S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\Lambda^c_{k,0}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)-\Lambda^c\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$
$$=\int_{T_{(k-1)}}^{\tau}\int_0^u\frac{S^c_0\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)S_0\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)S\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\Lambda^c_{k,0}\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)-\Lambda^c\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$
$$=\int_{T_{(k-1)}}^{\tau}\int_0^u\left(\frac{S_0\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}{S\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}-1\right)\frac{S^c_0\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\Lambda^c_{k,0}\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)-\Lambda^c\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$
$$+\int_{T_{(k-1)}}^{\tau}\int_0^u\left(\frac{S^c_0\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}\right)\left(\Lambda^c_{k,0}\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)-\Lambda^c\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$
$$=\int_{T_{(k-1)}}^{\tau}\int_0^u\left(\frac{S_0\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}{S\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}-1\right)\frac{S^c_0\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\Lambda^c_{k,0}\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)-\Lambda^c\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$
$$+\int_{T_{(k-1)}}^{\tau}\left(\frac{S^c_0\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}-1\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$

it follows that

$$\mathbb{E}_{P_0}\left[h_k\left(\mathcal{F}_{T_{(k)}}\right) \mid \mathcal{F}_{T_{(k-1)}}\right]$$
$$=\mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(S^c_0,\bar{Q}^g_{k,\tau}\right)\right) \mid \mathcal{F}_{T_{(k-1)}}]-\nu_{k-1,\tau}\left(\mathcal{F}_{T_{(k-1)}}\right)$$
$$+\mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}(S^c,\nu_k) \mid \mathcal{F}_{T_{(k-1)}}\right]-\mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(S^c,\bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}_{T_{(k-1)}}\right]$$
$$+\int_{T_{(k-1)}}^{\tau}\left(\frac{S^c_0\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}-1\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)-\bar{Q}^g_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$
$$+\int_{T_{(k-1)}}^{\tau}\int_0^u\left(\frac{S_0\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}{S\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}-1\right)\frac{S^c_0\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\Lambda^c_{k,0}\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)-\Lambda^c\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$

Since also,

$$\mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}(S^c,\nu_k) \mid \mathcal{F}_{T_{(k-1)}}\right]-\mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(S^c,\bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}_{T_{(k-1)}}\right]$$
$$=\int \mathbb{1}\{t_k<\tau\}\frac{S^c_0(t_k- \mid f_{k-1})}{S^c(t_k- \mid f_{k-1})}\sum_{d_k=a,\ell}\left(\nu_k(l_k,\mathbb{1}\mathbb{1}\{d_k=a\}+,d_k,t_k,f_{k-1})-\bar{Q}^g_{k,\tau}(l_k,\mathbb{1}\mathbb{1},d_k,t_k,f_{k-1})\right)P_{T_{(k)},\Delta_{(k)},A\left(T_{(k)}\right),L\left(T_{(k)}\right) \mid \mathcal{F}_{T_{(k-1)}}}(df_k \mid f_{k-1})$$

Letting

$$z_k\left(\mathcal{F}_{T_{(k-1)}}\right)=\left(\frac{\pi_{k-1,0}}{\pi_{k-1}}-1\right)\left(\mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}(S^c,\nu_k) \mid \mathcal{F}_{T_{(k-1)}}\right]-\mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(S^c,\bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}_{T_{(k-1)}}\right]\right) \text{ wrong?}$$
$$+\frac{\pi_{k-1,0}}{\pi_{k-1}}\int_{T_{(k-1)}}^{\tau}\left(\frac{S^c_0\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}-1\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)-\bar{Q}^g_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$
$$+\frac{\pi_{k-1,0}}{\pi_{k-1}}\int_{T_{(k-1)}}^{\tau}\int_0^u\left(\frac{S_0\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}{S\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}-1\right)\frac{S^c_0\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\Lambda^c_{k,0}\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)-\Lambda^c\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$$

, it follows that the remainder term can be written plainly as,

$$\sum_{k=1}^{K} \int \mathbb{1}\{t_1 < \ldots < t_{k-1} < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-2} \left(\frac{\pi_{0,j}(t_k, f_{j-1})}{\pi_j(t_k, f_{j-1})}\right)^{\mathbb{1}\{d_j = a\}} \frac{\prod_{j=1}^{k-1} S_0^c(t_j - \mid f_{j-1})}{\prod_{j=1}^{k-1} S^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \ldots, d_{k-1} \in \{\ell, a\}\}$$
$$\times z_k(f_{k-1}) P_{\mathcal{F}^{\text{full}}_{T_{k-1}}}(d f_{k-1})$$

8 Data-adaptive choice of K

In practice, we will want to use K_τ to be equal to 1 + maximum number of non-terminal events up to τ in the sample. It turns out, under the boundedness condition of the number of events, that an estimator that is asymptotically linear with efficient influence function $\varphi^*_\tau(P)(\max_i \kappa_i(\tau))$ is also asymptotically linear with efficient influence function $\varphi^*_\tau(P)(K_\tau)$ where K_τ is the last event point such that $P(\kappa_i(\tau) = K_\tau) > 0$.

Sketch: We want to use $K = K_n = \max_i \kappa_i(\tau)$. If we can do asymptotically and efficient inference for K_n, then we can also do it for a limiting $K_n \leq K$. Assume that the estimator is asymptotically linear with efficient influence function $\varphi^*_\tau(P)(K_n)$. Then by Assumption 1, there exists a K_{lim} which is the last point such that $P(K_n = K_{\text{lim}}) > 0$. Then, K_n converges to K_{lim} (by independence), and moreover, under standard regularity conditions such as strict positivity,

$$(\mathbb{P}_n - P)(\varphi^*_\tau(P)(K_n) - \varphi^*_\tau(P)(K))$$

is $o_P\left(n^{-\frac{1}{2}}\right)$, so if have asymptotic linearity in terms of $\varphi^*_\tau(P)(K_n)$, then we automatically have it for the original estimator for $\varphi^*_\tau(P)(K_{\text{lim}})$

9 Issues relating to rare patient histories

In the case of irregular data, we may have few people with many events. In that case, the iterative conditional expectations estimator will fail, because there are not enough people at each event point (see Table 1). We are then left with three options:

- Pooling
- A data-adaptive target parameter (Hubbard et al. (2016)), where the number of events considered (not all events) for the parameter is chosen based on the data, that is $\Psi_\tau(P) = \sum_{k=1}^{K_n} \mathbb{E}_P\left[\mathbb{1}\left\{\widetilde{T_{(k)}} \leq \tau, \widetilde{\Delta_{(k)}} = y\right\}\right]$, where K_n is selected based on the data. This is essentially done by sample splitting/cross-validation.
- Event-adaptive model selection, where the complexity of the models for each event point is based on how many data points are available at that event point (parametric models for very few data points, nonparametric models for many data points). By sample splitting, we should be able to take into account the data-adaptive model selection.

k	0	1	2	3	4	5
$\tilde{Y}_k(\tau)$	10000	8540	5560	2400	200	4
$\Delta A\left(T_{(k)}\right)$	6000	3560	1300	100	2	NA
$\Delta L\left(T_{(k)}\right)$	2540	2000	1100	100	2	NA

9.1 Pooling

Some people have complex histories. There may be very few of these people in the sample, so how do we do estimate the cause-specific hazard for the censoring in, say, the first step? In the artificial data example, there are only 4 people at the last time point.

We propose to pool the regressions across event points: Let us say that we want to estimate the cause-specific hazard for the censoring at event $k+1$ among people who are at risk of being censored at the $k+1$'th event, that is they either had a treatment change or a covariate change at their k event. If this population in the sample is very small, then we could do as follows. We delete the first event for these observations. Then the number of covariates is reduced by one, so we have the same number of covariates as we did for the people who are at risk of having an event at the k'th event. We combine these two data sets into one and regress the cause-specific hazard for the censoring at event "k". This provides a data set with correlated observations, which likely is not biased as we are not interested in variance estimation for parameters appearing in the regression.

To estimate the regression for the time-varying covariates, one could do:
- Not intervene on the last two or three time points, letting certain parts of the data generating mechanism be observational, that is $\pi_j^*\left(t, \cdot, \mathcal{F}_{T_{(j-1)}}\right) = \pi_j\left(t, \mathcal{F}_{T_{(j-1)}}\right)$ for $j = 4, 5$.
- Another is to make a Markov-like assumption in the interventional world, i.e.,
$$\mathbb{E}_Q\left[\sum_{j=1}^{3} I\left(T_{(j)} \leq \tau, \Delta_{(j)} = y\right) \mid \mathcal{F}_0\right] = \mathbb{E}_Q\left[\sum_{j=6}^{8} I\left(T_{(j)} \leq \tau, \Delta_{(j)} = y\right) \mid \mathcal{F}_{T_{(5)}}\right]$$
So we separately estimate the target parameter on the left hand side and use it to estimate the one on the right when we need to, pooling the data from the last three events with the data from the first three events.

Other possible methods are:
- Use an estimation procedure that is similar to Shirakawa et al. (2024) or use hazards which are estimated all at once.

- Bayesian methods may be useful since they do not have issues with finite sample size. They are also a natural way of dealing with the missing data problem. However, nonparametric Bayesian methods are not (yet) able to deal with a large number of covariates.

9.2 Other ideas

Some other issues are that the covariates are (fairly) high dimensional. This may yield issues with regression-based methods.
- Use Early-stopping cross-validation described as follows: First fit models with no covariates. Then we fit a model with the covariates from the last event. Determine if this is improves model fit via cross-validation and then we move on to the two latest changes and so on. Stop when the model fit does not improve. Theorem 2 of Schuler & van der Laan (2022) states that the convergence rates for an empirical risk minimizer are preserved. CTMLE also does something very similar (van der Laan & Gruber, 2010). This way, we may only select variables that are important in the specification of the treatment and outcome mechanism.

9.3 Topics for further research

Interestingly, $\int \bar{Q}^g_{0,\tau}(a, L_0)\nu_A(\mathrm{d}a)$ is a heterogenous causal effect. Can we estimate heterogenous causal effects in this way?

Time-fixed time-varying treatment could probably be interesting within a register-based study since it may be easier to define treatment in an interval rather than two define on, each time point, if the patient is on the treatment or not.

It may also sometimes be the case that some time-varying covariates are measured regularly instead of at subject-specific times. In this case, we may be able to do something similar to the above.

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

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.

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

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

Schuler, A., & van der Laan, M. (2022, May). *The Selectively Adaptive Lasso.*

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

10 Appendix

10.1 Finite dimensional distributions and compensators

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

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

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

Theorem 6 (Finite-dimensional distributions): Under the stated conditions of this section, we have
1. We have $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \Delta_{(k)}, X\left(T_{(k)}\right)\right) \vee \mathcal{F}_{T_{(k-1)}}$ 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,
$$\Lambda_i(dt, dx) = \sum_{k: T_{(k)} < \infty} \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \zeta_{k,i}\left(dx, t, \mathcal{F}_{T_{(k-1)}}\right) \Lambda_{k,i}\left(dt, \mathcal{F}_{T_{(k-1)}}\right).$$
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 $X_i\left(T_{(k)}\right)$ given $\mathcal{F}_{T_{(k-1)}}$ and $T_{(k)}$.
3. The distribution of $\mathcal{F}_{T_{(n)}}$ is given by
$$\begin{aligned} &F_n(dw, dt_1, d\delta_1, dx_{11}, \ldots, dx_{1d}, \ldots, dt_n, d\delta_n, dx_{n1}, \ldots, dx_{nd}) \\ &= \left(\prod_{i=1}^n \mathbb{1}\{t_{i-1} < t_i\} \prod_{u \in (t_{i-1}, t_i)} \left(1 - \sum_{j=1}^d \Lambda_{i,j}(du, f_{i-1})\right) \sum_{j=1}^d \delta_j(d\delta_i) \zeta_{i,j}(dx_{ij}, t_i, f_{i-1}) \Lambda_{i,j}(dt_i, f_{i-1})\right) \mu(dw), \end{aligned}$$
and $f_k = (t_k, d_k, x_k, \ldots t_1, d_1, x_1, w) \in \mathbb{F}_k$ for $n \in \mathbb{N}$. Here $\prod$ 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

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

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

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

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

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

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

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

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

Now note that

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

so we find the compensator of N_i to be

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

Letting

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

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

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.
□

10.2 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 continous-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 $L\left(T_{(k)}\right)$ and $A\left(T_{(k)}\right)$ 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 are. 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).

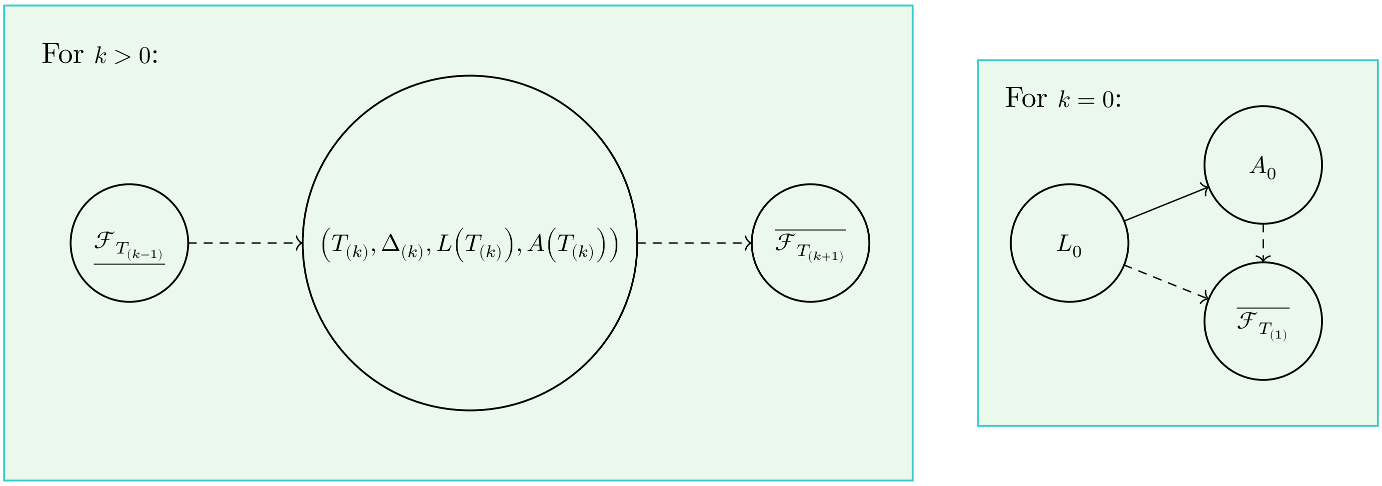

Figure 5: A DAG representing the relationships between the variables of O. The dashed lines indicate multiple edges from the dependencies in the past and into the future.

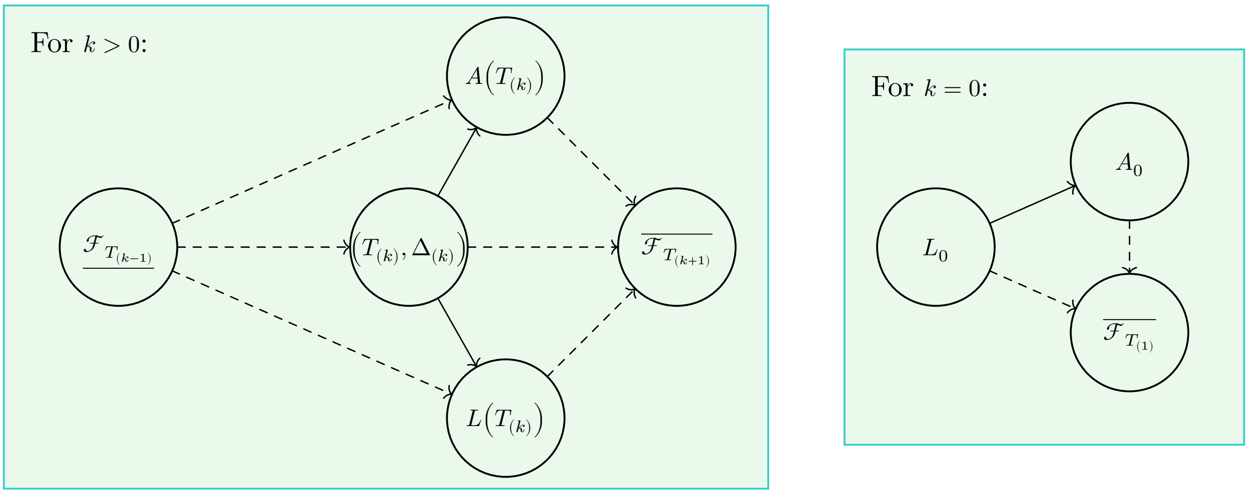

Figure 6: A DAG for simulating the data generating mechanism or such as those that may be found in Chamapiwa (2018). The dashed lines indicate multiple edges from the dependencies in the past and into the future. Here $\underline{\mathcal{F}_{T_{(k)}}}$ is the history up to and including the k'th event and $\overline{\mathcal{F}_{T_{(k)}}}$ is the history after and including the k'th event.

10.3 Remainder term for $K = 2$

Taking the mean of the EIF with respect to P_0 gives

$$\varphi_\tau^*(P)] = \mathbb{E}_{P_0}\Bigg[\sum_{k=1}^K \prod_{j=0}^{k-1} \left(\frac{\pi_{j-1}^*\Big(T_{(j)}, A\big(T_{(j)}\big), \mathcal{F}_{T_{(j-1-1)}}\Big)}{\pi_{j-1}\Big(T_{(j)}, \mathcal{F}_{T_{(j-1-1)}}\Big)}\right)^{I\left(\Delta_{(j)}=a\right)} \frac{I\Big(\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} \leq \tau\Big)}{\exp\Big(-\sum_{1\leq j<k}\int_{T_{(j-1)}}^{T_{(j)}} \lambda_{j-1}^c\Big(s, \mathcal{F}_{T_{(j-1)}}\Big) \mathrm{d}s\Big)}\Bigg[$$

$$\int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \left(\int_{\mathcal{A}} \bar{Q}_{k,\tau}^g\Big(L\big(T_{(k-1)}\big), a_k, s, a, \mathcal{F}_{T_{(k)}}\Big) \pi_k^*\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k) - B_{k-1}(u)\right) \Big(\Lambda_{k,0}^a(du) - \Lambda_k^a(du)\Big)$$

$$+ \int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_P\Big[\bar{Q}_{k,\tau}^g\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big] - B_{k-1}(u)\Big) \Big(\Lambda_{k,0}^\ell(du) - \Lambda_k^\ell$$

$$+ \int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} (1 - B_{k-1}(u)) \Big(\Lambda_{k,0}^y(du) - \Lambda_k^y(du)\Big) + \int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} - B_{k-1}(u) \Big(\Lambda_{k,0}^d(du) - \Lambda_k^d(du$$

$$+ \mathbb{1}\{k < K\} \int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_{P_0}\Big[\bar{Q}_{k-1,\tau}^g\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), \widetilde{T_{(k)}}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid \widetilde{T_{(k)}} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]$$

$$- \mathbb{E}_P\Big[\bar{Q}_{k-1,\tau}^g\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), \widetilde{T_{(k)}}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid \widetilde{T_{(k)}} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \Lambda_{k,0}^c(du)\Big]\Big]$$

$$+ \mathbb{E}_{P_0}\Big[\int \bar{Q}_{1,\tau}^g(a, L_0) \pi_0^*\Big(0, a, \mathcal{F}_{T_{(0-1)}}\Big) \nu_A(\mathrm{da})\Big] - \Psi_\tau(P)$$

We need to calculate

$$P, P_0) = \Psi_\tau(P) - \Psi_\tau(P_0) + P_0(\varphi_\tau^*(P))$$

$$= \mathbb{E}_{P_0}\Bigg[\sum_{k=1}^K \prod_{j=0}^{k-1} \left(\frac{\pi_{j-1}^*\Big(T_{(j)}, A\big(T_{(j)}\big), \mathcal{F}_{T_{(j-1-1)}}\Big)}{\pi_{j-1}\Big(T_{(j)}, \mathcal{F}_{T_{(j-1-1)}}\Big)}\right)^{I\left(\Delta_{(j)}=a\right)} \frac{I\Big(\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} \leq \tau\Big)}{\exp\Big(-\sum_{1\leq j<k}\int_{T_{(j-1)}}^{T_{(j)}} \lambda_{j-1}^c\Big(s, \mathcal{F}_{T_{(j-1)}}\Big) \mathrm{d}s\Big)}\Bigg[$$

$$\int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \left(\int_{\mathcal{A}} \bar{Q}_{k,\tau}^g\Big(L\big(T_{(k-1)}\big), a_k, s, a, \mathcal{F}_{T_{(k)}}\Big) \pi_k^*\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k) - B_{k-1}(u)\right) \Big(\Lambda_{k,0}^a(du) - \Lambda_k^a(du)\Big)$$

$$+ \int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_P\Big[\bar{Q}_{k,\tau}^g\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big] - B_{k-1}(u)\Big) \Big(\Lambda_{k,0}^\ell(du) - \Lambda_k^\ell($$

$$+ \int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} (1 - B_{k-1}(u)) \Big(\Lambda_{k,0}^y(du) - \Lambda_k^y(du)\Big) + \int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} - B_{k-1}(u) \Big(\Lambda_{k,0}^d(du) - \Lambda_k^d(du$$

$$+ \mathbb{1}\{k < K\} \int_{T_{(k-1)}}^\tau S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_{P_0}\Big[\bar{Q}_{k-1,\tau}^g\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), \widetilde{T_{(k)}}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid \widetilde{T_{(k)}} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]$$

$$- \mathbb{E}_P\Big[\bar{Q}_{k-1,\tau}^g\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), \widetilde{T_{(k)}}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid \widetilde{T_{(k)}} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \Lambda_{k,0}^\ell(du)\Big]\Big]$$

$$+ \mathbb{E}_{P_0}\Big[\int \bar{Q}_{1,\tau}^g(a, L_0) - \bar{Q}_{1,\tau_0}^g(a, L_0) \pi_0^*\Big(0, a, \mathcal{F}_{T_{(0-1)}}\Big) \nu_A(\mathrm{da})\Big]$$

By the Duhamel equation,

$$\begin{aligned}
&\bar{Q}_{1,\tau}^g(a, L_0) - \bar{Q}_{1,\tau_0}^g \\
&= S_0(s)(\text{bla} - \text{bla}_0) + (S(s) - S_0(s)) \text{ bla} \\
&= S_0(s)(\text{bla } Q - \text{bla}_0 Q + \text{bla}_0 Q - \text{bla}_0 Q_0) - \int S_0 B_{k-1} \sum_x (\Lambda_{0,x} - \Lambda_x)
\end{aligned}$$

The second term gives that we can ignore B_k:

$$\int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \left(\int_{\mathcal{A}} \bar{Q}_{k,\tau}^{g}\Big(L\big(T_{(k-1)}\big), a_k, s, a, \mathcal{F}_{T_{(k)}}\Big) \pi_k^*\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \big(\Lambda_{k,0}^a(du) - \Lambda_k^a(du)\big)$$

$$+ \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_P\Big[\bar{Q}_{k,\tau}^{g}\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \big(\Lambda_{k,0}^\ell(du) - \Lambda_k^\ell(du)\big)$$

$$+ \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \big(\Lambda_{k,0}^y(du) - \Lambda_k^y(du)\big)$$

$$+ \mathbb{1}\{k < K\} \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_{P_0}\Big[\bar{Q}_{k-1,\tau}^{g}\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), \widetilde{T_{(k)}}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid \widetilde{T_{(k)}} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]$$

$$- \mathbb{E}_P\Big[\bar{Q}_{k-1,\tau}^{g}\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), \widetilde{T_{(k)}}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid \widetilde{T_{(k)}} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \Lambda_{k,0}^\ell(du)$$

$$= \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \left(\int_{\mathcal{A}} \bar{Q}_{k,\tau}^{g}\Big(L\big(T_{(k-1)}\big), a_k, s, a, \mathcal{F}_{T_{(k)}}\Big) \pi_k^*\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \big(\Lambda_{k,0}^a(du) - \Lambda_k^a(du)\big)$$

$$+ \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_P\Big[\bar{Q}_{k,\tau}^{g}\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \big(-\Lambda_k^\ell(du)\big)$$

$$+ \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \big(\Lambda_{k,0}^y(du) - \Lambda_k^y(du)\big)$$

$$+ \mathbb{1}\{k < K\} \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_{P_0}\Big[\bar{Q}_{k-1,\tau}^{g}\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), \widetilde{T_{(k)}}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid \widetilde{T_{(k)}} = u, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \Lambda_{k,0}^\ell(du)$$

$$= \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \left(\int_{\mathcal{A}} \bar{Q}_{k,\tau_0}^{g}\Big(L\big(T_{(k-1)}\big), a_k, s, a, \mathcal{F}_{T_{(k)}}\Big) \pi_k^*\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \Lambda_{k,0}^a(du)$$

$$- \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \left(\int_{\mathcal{A}} \bar{Q}_{k,\tau}^{g}\Big(L\big(T_{(k-1)}\big), a_k, s, a, \mathcal{F}_{T_{(k)}}\Big) \pi_k^*\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \Lambda_k^a(du)$$

$$+ \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \left(\int_{\mathcal{A}} \bar{Q}_{k,\tau}^{g}\Big(L\big(T_{(k-1)}\big), a_k, s, a, \mathcal{F}_{T_{(k)}}\Big) - \bar{Q}_{k,\tau_0}^{g}\Big(L\big(T_{(k-1)}\big), a_k, s, a, \mathcal{F}_{T_{(k)}}\Big) \pi_k^*\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \Lambda_{k,0}^a(du)$$

$$+ \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_{P_0}\Big[\bar{Q}_{k,\tau}^{g} \cdot \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big] - \mathbb{E}_{P_0}\Big[\bar{Q}_{k,\tau_0}^{g} \cdot \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \Lambda_{k,0}^\ell(du)$$

$$+ \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_{P_0}\Big[\bar{Q}_{k,\tau_0}^{g}\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \Lambda_{k,0}^\ell(du)$$

$$- \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Big(\mathbb{E}_P\Big[\bar{Q}_{k,\tau}^{g}\Big(L\big(T_{(k)}\big), A\big(T_{(k-1)}\big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\Big) \Lambda_k^\ell(du)$$

$$+ \int_{T_{(k-1)}}^{\tau} S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) \frac{S_0^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \big(\Lambda_{k,0}^y(du) - \Lambda_k^y(du)\big)$$

Adding the first term together in the sum with the last term, we have

$$
\begin{aligned}
\varphi^*_\tau(P)] &= \mathbb{E}_{P_0^*}\Bigg[\int_0^\tau S_0(u \mid \mathcal{F}_0)\left(\frac{\pi_0(L(0))}{\pi(L(0))}\frac{S_0^c(u \mid \mathcal{F}_0)}{S^c(u \mid \mathcal{F}_0)} - 1\right)\left(\bar{Q}^g_{0,\tau_0}(ds, \mathcal{F}_0) - \bar{Q}^g_{0,\tau}(ds, \mathcal{F}_0)\right) \\
&+ \int_0^\tau S_0(u \mid \mathcal{F}_0)\frac{\pi_0(L(0))}{\pi(L(0))}\frac{S_0^c(u \mid \mathcal{F}_0)}{S^c(u \mid \mathcal{F}_0)}\left(\sum_{x=a,\ell}\mathbb{E}_{P_0^*}\left[\bar{Q}^g_{1,\tau} - \bar{Q}^g_{1,\tau_0} \mid T_{(1)} = s, \Delta_{(1)} = x, \mathcal{F}_0\right]\Lambda^x_{1,0}(du)\right)\Bigg] \\
&= \mathbb{E}_{P_0^*}\left[\int_0^\tau S_0(u \mid \mathcal{F}_0)\left(\frac{\pi_0(L(0))}{\pi(L(0))}\frac{S_0^c(u \mid \mathcal{F}_0)}{S^c(u \mid \mathcal{F}_0)} - 1\right)\left(\bar{Q}^g_{0,\tau_0}(ds, \mathcal{F}_0) - \bar{Q}^g_{0,\tau}(ds, \mathcal{F}_0)\right)\right] \\
&+ \mathbb{E}_{P_0^*}[\int_{T_{(k-1)}}^\tau S_0\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\frac{\pi_0(L(0))}{\pi(L(0))}\frac{S_0^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} \\
&\times \left(\sum_{x=a,\ell}\mathbb{E}_{P_0^*}\left[\int_0^\tau S_0\left(s \mid \mathcal{F}_{T_{(1)}}\right)\left(\frac{\pi_{1,0}\left(T_{(1)}, A\left(T_{(1)}\right), \mathcal{F}_0\right)}{\pi_1\left(T_{(1)}, A\left(T_{(1)}\right), \mathcal{F}_0\right)}\frac{S_0^c\left(s \mid \mathcal{F}_{T_{(1)}}\right)}{S^c\left(u \mid \mathcal{F}_{T_{(1)}}\right)} - 1\right)\left(\bar{Q}^g_{1,\tau_0}\left(ds, \mathcal{F}_{T_{(1)}}\right) - \bar{Q}^g_{1,\tau}\left(ds, \mathcal{F}_{T_{(1)}}\right)\right) \mid T_{(1)} = u, \Delta_{(1)} = x, \mathcal{F}_0\right]\Lambda^x_{1,0}(
\end{aligned}
$$

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

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

$$
\begin{aligned}
\varphi^*_\tau(P) = \sum_{k=1}^K \prod_{j=0}^{k-1}&\left(\frac{\pi^*_{j-1}\left(T_{(j)}, A\left(T_{(j)}\right), \mathcal{F}_{T_{(j-1-1)}}\right)}{\pi_{j-1}\left(T_{(j)}, \mathcal{F}_{T_{(j-1-1)}}\right)}\right)^{I\left(\Delta_{(j)}=a\right)}\frac{I\left(\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} \leq \tau\right)}{\exp\left(-\sum_{1 \leq j < k}\int_{T_{(j-1)}}^{T_{(j)}} \lambda^c_{j-1}\left(s, \mathcal{F}_{T_{(j-1)}}\right)\mathrm{d}s\right)}\Bigg[\\
&\int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\int_A \bar{Q}^g_{k,\tau}\left(L\left(T_{(k-1)}\right), a_k, s, a, \mathcal{F}_{T_{(k)}}\right)\pi^*_k\left(s, a_k, \mathcal{F}_{T_{(k-1)}}\right)\nu_A(da_k) - B_{k-1}(u)\right)M^a_k(du) \\
&+ \int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\mathbb{E}_P\left[\bar{Q}^g_{k,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right] - B_{k-1}(u)\right)M^\ell_k(du) \\
&+ \int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}(1 - B_{k-1}(u))M^y_k(du) + \int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)}(0 - B_{k-1}(u))M^d_k(du) \\
&+ \frac{1}{S^c\left(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\right)}I\left(T_{(k)} \leq \tau, \Delta_{(k)} = \ell, k < K\right)\left(\bar{Q}^g_{k,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, \ell, \mathcal{F}_{T_{(k-1)}}\right)\right. \\
&\qquad \left.- \mathbb{E}_P\left[\bar{Q}^g_{k-1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), \widetilde{T_{(k)}}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid \widetilde{T_{(k)}} = T_{(k)}, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\right)\Bigg] \\
&+ \int \bar{Q}^g_{1,\tau}(a, L_0)\pi^*_0\left(0, a, \mathcal{F}_{T_{(0-1)}}\right)\nu_A(\mathrm{d}a) - \Psi_\tau(P)
\end{aligned}
$$

We find immediately that

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

Now note that

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

Let us calculate the second integral

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

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}^g_{k+1,\tau}(u)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \lambda^{\bullet}_{k-1}\Big(u, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}u \\
&= \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \Bigg[\int_0^u S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^a_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \\
&\qquad \times \left(\int_{\mathcal{A}} \bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k-1)}\Big), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\Big) \pi^*_k\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \\
&+ \int_0^u S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^{\ell}_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \\
&\qquad \times \left(\mathbb{E}_P\Big[\bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k)}\Big), A\Big(T_{(k-1)}\Big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\right) \\
&+ \int_0^u S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^{y}_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big)\Bigg] \\
&\times \frac{1}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \lambda^{\bullet}_{k-1}\Big(u, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}u \\
&= \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \int_s^{\tau \wedge T_{(k)}} \frac{1}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \lambda^{\bullet}_{k-1}\Big(u, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}u \Bigg[S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^a_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \\
&\qquad \times \left(\int_{\mathcal{A}} \bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k-1)}\Big), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\Big) \pi^*_k\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \\
&+ S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^{\ell}_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \\
&\qquad \times \left(\mathbb{E}_P\Big[\bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k)}\Big), A\Big(T_{(k-1)}\Big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\right) \\
&+ S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^{y}_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big)\Bigg]
\end{aligned}
$$

Now note that

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

Setting this into the previous integral, we get

$$
\begin{aligned}
&-\int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda^{\bullet}_{k-1}\Big(u, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}u \Bigg[\Lambda^a_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \\
&\qquad \times \left(\int_{\mathcal{A}} \bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k-1)}\Big), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\Big) \pi^*_k\Big(s, a_k, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \\
&+ \Lambda^{\ell}_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \\
&\qquad \times \left(\mathbb{E}_P\Big[\bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k)}\Big), A\Big(T_{(k-1)}\Big), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\Big]\right) \\
&+ \Lambda^{y}_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big)\Bigg] \\
&+ \frac{1}{S^c\Big(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big)} \bar{Q}^g_{k-1,\tau}\Big(\tau \wedge T_{(k)}\Big)
\end{aligned}
$$

Thus, we find

$$\int_{T_{(k-1)}}^{\tau} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(u)\right) \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} \left(N^{\bullet}_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^{\bullet}_{k-1}\left(s, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}s\right)$$

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

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

$$+ \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{\bar{Q}^g_{k-1,\tau}(u)}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} \lambda^{\bullet}_{k-1}\left(s, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}s$$

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

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

$$- \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda^{\bullet}_{k-1}\left(u, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}u \Bigg[\Lambda^a_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right)$$

$$\times \left(\int_{\mathcal{A}} \bar{Q}^g_{k+1,\tau}\left(L\left(T_{(k-1)}\right), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\right) \pi^*_k\left(s, a_k, \mathcal{F}_{T_{(k-1)}}\right) \nu_A(\mathrm{d}a_k)\right)$$

$$+\Lambda^{\ell}_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right)$$

$$\times \left(\mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\right)$$

$$+\Lambda^y_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right)\Bigg]$$

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

$$= - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda^{\bullet}_{k-1}\left(u, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}u \Bigg[\Lambda^a_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right)$$

$$\times \left(\int_{\mathcal{A}} \bar{Q}^g_{k+1,\tau}\left(L\left(T_{(k-1)}\right), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\right) \pi^*_k\left(s, a_k, \mathcal{F}_{T_{(k-1)}}\right) \nu_A(\mathrm{d}a_k)\right)$$

$$+\Lambda^{\ell}_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right)$$

$$\times \left(\mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\right)$$

$$+\Lambda^y_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right)\Bigg] + \bar{Q}^g_{k-1,\tau}(\tau)$$