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

Johan Sebastian Ohlendorff
University of Copenhagen

Anders Munch
University of Copenhagen

Thomas Alexander Gerds
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

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

2 Introduction

In medical research, the estimation of causal effects of treatments over time is often of interest. We consider a longitudinal continuous-time setting that is very similar to Rytgaard et al. (2022) in which patient characteristics can change at subject-specific times. This is the typical setting of registry data, which usually contains precise information about when events occur, e.g., information about drug purchase history, hospital visits, and laboratory measurements. This approach offers an advantage over discretized methods, as it eliminates the need to select a time grid mesh for discretization, which can affect both the bias and variance of the resulting estimator. A continuous-time approach would adapt to the events in the data. Furthermore, continuous-time data captures more precise information about when events occur, which may be valuable in a predictive sense. Let 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. 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.

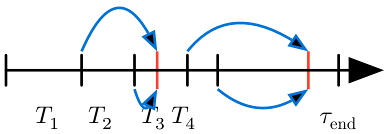

Figure 1: The "usual" approach where time is discretized. Each event time and its corresponding mark is rolled forward to the next time grid point, that is the values of the observations are updated based on the on the events occuring in the previous time interval.

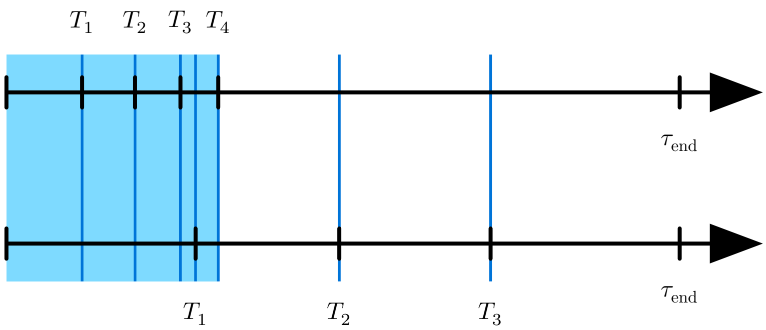

Figure 2: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let $t_1 < \ldots < t_m$ be all the event times in the sample. Then, given $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{t_r}\right]$, we regress back to $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{t_{r-1}}\right]$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

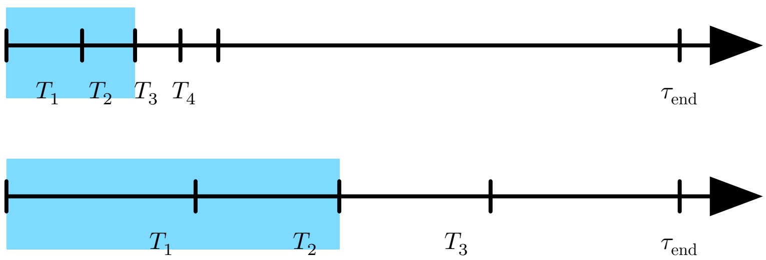

Figure 3: The figure illustrates the sequential regression approach proposed in this article. For each event k in the sample, we regress back on the history $\mathcal{F}_{T_{(k-1)}}$. That is, given $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{T_{(k)}}\right]$, we regress back to $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{T_{(k-1)}}\right]$. 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 may principally include covariates which do not change over time, but for simplicity of notation, we will include them among those that do change over time. Throughout, we assume that we have two treatment options, $A(t) = 0, 1$ (e.g., placebo and active treatment), where $A(t)$ denotes the treatment at time t. The time-varying confounders are assumed to take values in a finite subset $\mathcal{L} \subset \mathbb{R}^m$. These processes are assumed to be càdlàg, jump processes. Furthermore, the times at which the treatment and covariates may change are dictated entirely by the counting processes N^a and N^ℓ, respectively in the sense that $\Delta A(t) \neq 0$ only if $\Delta N^a(t) \neq 0$ and $\Delta L(t) \neq 0$ only if $\Delta N^\ell(t) \neq 0$.

Since our setup will be a competing risk setup, we also have counting processes representing the event of interest and the competing event. We let N^y and N^d be counting processes corresponding to the primary and competing event, respectively. 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 the jump process $(N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t))$, and we let $\mathcal{F}_t = \sigma((N^a(s), A(s), N^\ell(s), L(s), N^y(s), N^d(s)) \mid s \leq t)$ be the natural filtration generated by the processes up to time t. Furthermore, define σ-algebra $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right)\right) \vee \mathcal{F}_{T_{(k-1)}}$ and $\mathcal{F}_0 = \sigma(A_0, L_0)$, representing the information up to the k'th event. Here we tacitly allow the event times after the first event to be ∞. We observe $O = \mathcal{F}_{T_{(K)}} = \left(T_{(K)}, \Delta_{(K)}, T_{(K-1)}, \Delta_{(K-1)}, A\left(T_{(K-1)}\right), L\left(T_{(K-1)}\right), ..., A_0, L_0\right) \sim P \in \mathcal{M}$ where $\mathcal{M}$ is the set of all probability measures. We also impose the condition that the last event has to be a terminal event, i.e., $\Delta_{(K)} = y$ or d if $T_{(K-1)} < \infty$. Let $\pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ be the probability of being treated 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 and let $\pi_k\left(t, da, \mathcal{F}_{T_{(k-1)}}\right)$ be the corresponding probability measure. Let $\mu_k\left(t, \cdot, \mathcal{F}_{T_{(k-1)}}\right)$ be the probability measure for 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 $\lambda^x_{k-1}\left(t, \mathcal{F}_{T_{(k-1-1)}}\right)$ be the hazard of the k'th event at time t given $\mathcal{F}_{T_{(k-1)}}$ (Assumption 3). Our overall goal is to estimate the interventional cumulative incidence function at time τ,

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

where $\widetilde{N}^y_t$ is the potential outcome representing the counterfactual outcome $N^y_t = \sum_{k=1}^K \mathbb{1}\left\{T_{(k)} \leq t, \Delta_{(k)} = y\right\}$ had the treatment regime (starting and staying on treatment at every visitation time), possibly contrary to fact, been followed. The necessary conditions for the potential outcome framework are given in Theorem 1. We recognize at this point that visitation times may not actually be that irregular, but regularly scheduled as they are in a randomized control trial. This will likely *not* change our results for identification, estimation, and debiasing. For instance, even if $\Lambda^a_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ were fully known and deterministic, the efficient influence function given in Theorem 5 and the iterative conditional expectation formula (Equation 5 and Equation 4) are not likely to change based on the form given in these theorems.

Assumption 3 (Conditional distributions of jumps): We assume that the conditional distributions $P\left(T_{(k)} \in \cdot \mid \mathcal{F}_{T_{(k-1)}}\right) \ll m$ P-a.s., where m is the Lebesgue measure on $\mathbb{R}_+$.

From observational data, we will emulate a randomized controlled trial in continuous-time. In the continuous-time longitudinal setting, this can e.g., correspond to a trial in which there is perfect compliance with the treatment protocol. Our approach is inspired by the conditions of Ryalen (2024), but we do not apply martingale theory. Moreover, the identification conditions do not require the existence of an entire potential outcome process. Instead the conditions are stated for identification at the time horizon of interest. While our theory provides a potential outcome framework, it is unclear at this point how graphical models can be used to reason about the conditions (see Richardson & Robins (2013) for the discrete time approach). However, one may define a (stochastic)[1] intervention with respect to a local independence graph (Røysland et al. (2024)) but we do not further pursue this here. For simplicity, we assume that we are only interested in the effect of staying on treatment ($A(t) = 1$ for all $t > 0$) and starting on treatment ($A(0) = 1$).

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

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

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

[1]The reason we write stochastic is that the treatment consists of two components; the total compensator for the treatment and mark probabilities. Since these two components are inseparable, the intervention of interest is not a static intervention.

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

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

  - **Positivity**: The measure given by $dR = WdP$ where $W^*_t = \prod_{k=1}^{N_t} \left(\frac{\mathbb{1}\left\{A\left(T_{(k)}\right)=1\right\}}{\pi_k\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)}\right)^{\mathbb{1}\left\{\Delta_{(k)}=a\right\}} \frac{\mathbb{1}\{A_0=1\}}{\pi_0(L_0)}$ is a probability measure,

  where $N_t = \sum_{k=1}^{K-1} \mathbb{1}\left\{T_{(k)} \leq t\right\}$.

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

$$
\begin{aligned}
\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\}Y_\tau W_\tau\right] &= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\}\mathbb{1}\{T^a > \tau\}Y_\tau W_\tau\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \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)} \leq \tau < T_{(k)}\right\}\tilde{Y}_\tau W_\tau\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\}\tilde{Y}_\tau W_{T_{(k-1)}}\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \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), \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)} \leq \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), \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)} \leq \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), \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)} \leq \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), \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)} \leq \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)} \leq \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 $\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\}\tilde{Y}_\tau\right] = \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\}Y_\tau W_\tau\right]$ as needed. □

By the intersection property of conditional independence, we see that a sufficient condition for the first exchangeability condition in Equation 1 is that

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

An illustration of the consistency condition is given in Figure 4.

The second condition may in particular be too strict in practice as the future event times may be affected by prior treatment. 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, but the consistency condition is very similar. Specifically, let $\tilde{Y}_{\tau,k}$ be the potential outcome at event k corresponding to $\mathbb{1}\left\{T_{(k)} \leq \tau, \Delta_{(k)} = y\right\}$. 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 $0 \leq j < k$ and $k = 1, ..., K$. 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.

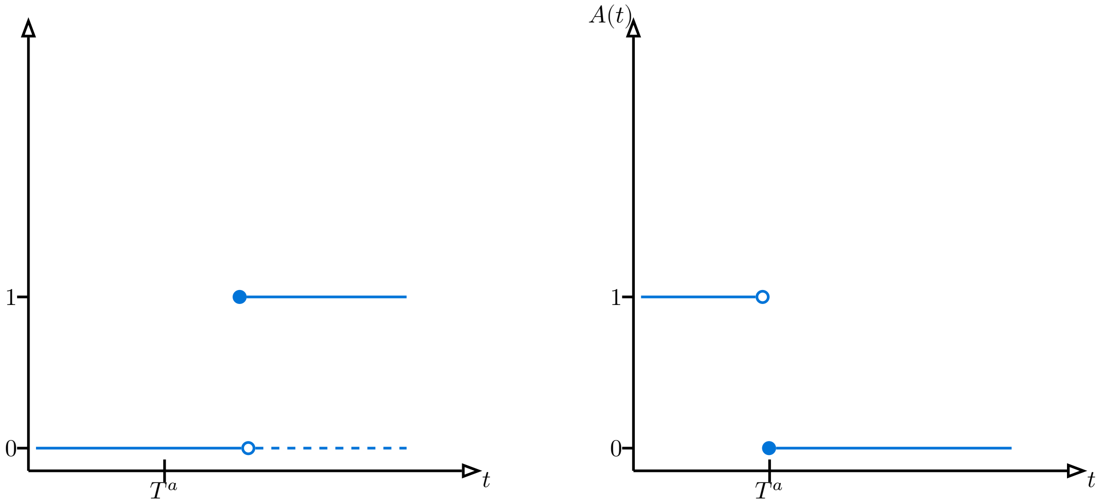

Figure 4: The figure illustrates the consistency condition for the potential outcome framework for single individual. The left panel shows the potential outcome process $\tilde{Y}_t$ (dashed) and the observed process Y_t (solid). The right panel shows the treatment process $A(t)$. At time T^a, the treatment is stopped and the processes may from then on diverge.

We are now ready to give an iterated conditional expectations formula for the target parameter in the case with no censoring. The algorithm can be foundin Theorem 2, but we will state more explicitly in the next section how to allow for censoring. Note that the iterative conditional expectations formula in terms of the event-specific cause-specific hazards and the density for the time-varying covariates (Theorem 3) actually shows that our target parameter is the same as the one given in Rytgaard et al. (2022) with no competing event (**TODO**: show this more explicitly) under the stated identifiability conditions.

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,

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

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, we find

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

by the law of iterated expectations and that

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

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

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

by the law of iterated expectations. □

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

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

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

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

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

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

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

4 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 $N^c(t) = \mathbb{1}\{C \le t\}$ 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 $\lambda^c_{k-1}\left(t, \mathcal{F}_{T_{(k-2)}}\right)$ 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 $S^c\left(t - \mid \mathcal{F}_{T_{(k-1)}}\right) = \prod_{s \in \left(T_{(k-1)}, t\right)} \left(1 - \lambda^c_{k-1}\left(s, \mathcal{F}_{T_{(k-1-1)}}\right) ds\right)$ 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 and predictable with respect 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 7 that the filtration generated by the random measures is necessarily the same as $\mathcal{F}_t^{\text{full}}$. We are now ready to state the main theorem which allows us to prove that the ICE IPCW estimator is valid. A simple implementation of the IPCW is provided in Section 4.1.

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

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

Consequently, we have that

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

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

$$
\begin{aligned}
\Lambda^{r,a}(\mathrm{d}t, \mathrm{d}a) &= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \pi_k\left(t, \mathrm{d}a, \mathcal{F}_{T_{(k-1)}}\right) \lambda_k^y\left(t, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}t \\
\Lambda^{r,\ell}(\mathrm{d}t, \mathrm{d}\ell) &= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \mu_k\left(t, \mathrm{d}\ell, \mathcal{F}_{T_{(k-1)}}\right) \lambda_k^\ell\left(t, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}t \\
\Lambda^{r,y}(\mathrm{d}t) &= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \lambda_k^y\left(t, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}t \\
\Lambda^{r,d}(\mathrm{d}t) &= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \lambda_k^d\left(t, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}t
\end{aligned}
$$

by Theorem 7 and by assumption is also the compensator for the filtration $\mathcal{F}_t^{\text{full}}$. By the innovation theorem (Section II.4.2 of Andersen et al. (1993)),

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

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 7 directly to the observed process to get that

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

where $\tilde{\Lambda}^x_k\left(dt \mid \mathcal{F}^{\text{obs}}_{\bar{T}_{(k-1)}}\right)$ is cause-specific cumulative hazard of the k'th event at time t given $\mathcal{F}^{\text{obs}}_{\bar{T}_{(k-1)}}$, and $\tilde{\pi}_k\left(a, t, \mathcal{F}^{\text{obs}}_{\bar{T}_{(k-1)}}\right)$ is the density of the treatment process at time t given $\mathcal{F}^{\text{obs}}_{\bar{T}_{(k-1)}}$, and $\tilde{\mu}_k\left(l, t, \mathcal{F}^{\text{obs}}_{\bar{T}_{(k-1)}}\right)$ is the density of the covariate process at time t given $\mathcal{F}^{\text{obs}}_{\bar{T}_{(k-1)}}$. Since the canonical compensator given in Equation 6 (Theorem 4.3.9 in Last & Brandt (1995)) determines uniquely the distribution of the marks and the event times, the theorem follows. □

In the next section, we will now write $T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right)$ instead of $\bar{T}_{(k)}, \bar{\Delta}_{(k)}, A(\bar{T}_k), L(\bar{T}_k)$ and $\mathcal{F}_{T_{(k)}}$ instead of $\mathcal{F}^{\text{obs}}_{T_{(k)}}$. **NOTE**: Change this.

4.1 Algorithm for IPCW Iterative Conditional Expectations Estimator

The following algorithm gives a simple implementation of the IPCW ICE 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:
  $$\hat{\eta}_k = \frac{\mathbb{1}\left\{T_{(k)} \leq \tau, \Delta_k \in \{a, \ell\}, k < K\right\} \hat{\nu}_k\left(\mathcal{F}^{-A}_{T_{(k)}}, \mathbf{1}\right)}{\hat{S}^c\left(T_{(k)} - \mid \mathcal{F}_{T_{(k-1)}}\right)}$$
  Then calculate the subject-specific pseudo-outcome
  $$\hat{R}_k = \frac{\mathbb{1}\left\{T_{(k)} \leq \tau, \Delta_k = y\right\}}{\hat{S}^c\left(T_{(k)} - \mid \mathcal{F}_{T_{(k-1)}}\right)} + \hat{\eta}_k$$
  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 $\hat{\Psi}_n = \frac{1}{n} \sum_{i=1}^n \hat{\nu}_0(L_i(0), 1)$.

It is recommended to use Equation 4 for estimating $\bar{Q}^g_{k,\tau}$ instead of direct computation Equation 2: The resulting integral representing the target parameter would, in realistic settings, be incredibly highly dimensional. Specialized approaches may yet exist (see the discussion).

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

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

Based on this definition, we can give a simple condition for the IPCW ICE estimator to be consistent in the $L^2(P)$ -norm.

Lemma 1: Define $P_k = P \mid_{\mathcal{F}_{T_{(k)}}}$ the restriction of P to the σ-algebra generated by the history of the first k events, and $P'_k = P \mid_{\mathcal{F}_{T_{(k)}}}$ the restriction of P to the $\sigma\left(T_{(k)}, \Delta_{(k)}\right) \vee \mathcal{F}_{T_{(k-1)}}$. Assume that $\| \hat{\nu}_{k+1} - \bar{Q}^g_{k+1,\tau} \|_{L^2(P_{k+1})} = o_P(1)$, $\| \hat{\Lambda}^c_{k+1} - \Lambda^c_{k+1} \|_{L^2(P'_{k+1})} = o_P(1)$. 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 of step 2 of the algorithm and assume that

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

where $\bar{Z}^a_{k,\tau}$ is defined as the integrand of Equation 4. Then,

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

Proof: By the triangle inequality,

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

The first term is $o_P(1)$ by assumption. The third term is $o_P(1)$ by Jensen's inequality and by assumption. That the second term is $o_P(1)$ follows from the fact again applying Jensen's inequality to see that

$$\begin{aligned}
&\| \mathbb{E}_P\left[\bar{Z}^a_{k,\tau}\left(\hat{S}^c, \hat{\nu}_{k+1}\right) \mid \mathcal{F}_{T_{(k)}} = \cdot\right] - \mathbb{E}_P\left[\bar{Z}^a_{k,\tau}\left(S^c, \hat{\nu}_{k+1}\right) \mid \mathcal{F}_{T_{(k)}} = \cdot\right] \|_{L^2(P_k)} \\
&\leq \sqrt{\int \left(\frac{1}{S^c(t_{k+1} - \mid f_k)} - \frac{1}{\hat{S}^c(t_{k+1} - \mid f_k)}\right)^2 \mathbb{1}\{t_{k+1} \leq \tau, d_{k+1} \in \{a, \ell\}\} \hat{\nu}^2_{k+1}(f_{k+1}) P_{k+1}(\mathrm{d} f_{k+1})} \\
&\leq \sqrt{\int \left(\frac{1}{S^c(t_{k+1} - \mid f_k)} - \frac{1}{\hat{S}^c(t_{k+1} - \mid f_k)}\right)^2 \mathbb{1}\{t_{k+1} \leq \tau\} P_{k+1}(\mathrm{d} f_{k+1})} \\
&\leq K \sqrt{\int \left(\hat{\Lambda}^c_{k+1}(t_{k+1} \mid f_k) - \Lambda^c_{k+1}(t_{k+1} \mid f_k)\right)^2 \mathbb{1}\{t_{k+1} \leq \tau\} P_{k+1}(\mathrm{d} f_{k+1})}
\end{aligned} \tag{7}$$

This shows that the second term is $o_P(1)$. In the last equality, we used that the function $x \mapsto \exp(-x)$ is Lipschitz continuous (since we assume that the hazards are uniformly bounded) on the set of possible values for the estimated cumulative hazard and the cumulative hazard. □

5 Efficient estimation

NOTE: Write introduction to efficiency theory.

We want to use machine learning estimators of the nuisance parameters, so to get inference in a non-parametric setting, 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. To do so, we introduce some additional notation and let

$$\bar{Q}^g_{k,\tau}\left(u \mid \mathcal{F}_{T_{(k)}}\right) = p_{ka}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k\ell}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{ky}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right), u < \tau \tag{8}$$

which, additionally can also be estimated with an ICE IPCW procedure (but we won't need to!).

One of the main features here is that the efficient influence function is given in terms of the martingale for the censoring process which may be simpler computationally to implement. In an appendix, we compare it with the efficient influence function derived in Rytgaard et al. (2022).

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

$$\varphi^*(P) = \frac{\mathbb{1}\{A_0 = 1\}}{\pi_0(L(0))} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\left\{A\left(T_{(j)}\right) = 1\right\}}{\pi_j\left(T_{(j)}, \mathcal{F}_{T_{(j-1)}}\right)} \right)^{\mathbb{1}\left\{\Delta_{(j)} = a\right\}} \frac{1}{\prod_{j=1}^{k-1} S^c\left(T_{(j)} - \mid \mathcal{F}_{T_{(j-1)}}\right)} \mathbb{1}\left\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\right\}$$
$$\times \left(\left(\bar{Z}^a_{k,\tau} - \bar{Q}^g_{k-1,\tau}\right) + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \left(\bar{Q}^g_{k-1,\tau}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) - \bar{Q}^g_{k-1,\tau}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\right) \frac{1}{S^c\left(u- \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} M^c_k(\mathrm{d}u) \right) \tag{9}$$
$$+ \bar{Q}^g_{0,\tau} - \Psi_\tau(P),$$

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

Proof: Define (update notation)

$$\begin{aligned}
\bar{Z}^a_{k,\tau}(s, t_k, d_k, l_k, a_k, f_{k-1}) &= \frac{I(t_k \leq s, d_k = \ell)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)} \bar{Q}^g_{k,\tau}(a_{k-1}, l_k, t_k, d_k, f_{k-1}) \\
&\quad + \frac{I(t_k \leq s, d_k = a)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)} \\
&\qquad \times \int \bar{Q}^g_{k,\tau}(\tilde{a}_k, l_{k-1}, t_k, d_k, f_{k-1}) \pi^*_{k-1}\left(t_k, \mathcal{F}_{T_{(k-1-1)}}\right) \nu_A(\mathrm{d}\tilde{a}_k) \\
&\quad + \frac{I(t_k \leq s, d_k = y)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)}, s \leq \tau
\end{aligned} \tag{10}$$

and let

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

We compute the efficient influence function by taking the Gateaux derivative of the above with respect to P, by discretizing the time. Note that this is not a rigorous argument showing that the efficient influence function is given by Equation 9. To formally prove that is the efficient influence function, we would have to compute the pathwise derivative of the target parameter along parametric submodels with a given score function. We will use two well-known "results" for the efficient influence function.

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

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,

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

$$= \frac{\delta_{\mathcal{F}_{T_{(k-1)}}(f_{k-1})}}{P\Big(\mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)} \Bigg(\bar{Z}^a_{k,\tau} - \bar{Q}^g_{k-1,\tau}\Big(\tau, \mathcal{F}_{T_{(k-1)}}\Big) +$$

$$+ \int_{T_{(k-1)}}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{T_{(k-1)}}^{t_k} \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda^x_{k-1}\Big(s, \mathcal{F}_{T_{(k-1-1)}}\Big) \mathrm{d}s\Big)} \Big(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\Big(s, \mathcal{F}_{T_{(k-1-1)}}\Big) \mathrm{d}s\Big)$$

$$P_{\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big) \Bigg)$$

$$+ \int_{T_{(k-1)}}^{\tau} \Bigg(\frac{I(t_k \leq \tau, d_k \in \{a, \ell\})}{\exp\Big(-\int_{T_{(k-1)}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\Big)} \cdot \left(\frac{\pi^*_{k-1}\Big(t_k, \mathcal{F}_{T_{(k-1-1)}}\Big)}{\pi_{k-1}\Big(t_k, \mathcal{F}_{T_{(k-1-1)}}\Big)} \right)^{I(d_k=a)} \frac{\partial}{\partial \varepsilon} \bar{Q}^{a,\varepsilon}_{k-1,\tau}(a_k, l_k, t_k, d_k, f_{k-1}) \Big((1-\varepsilon)P + \varepsilon \delta_{\mathcal{F}_{T_{(k)}}}\Big)\Big|_{\varepsilon=0}$$

$$P_{\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)$$

Now note for the second term, we can write

$$\int_{T_{(k-1)}}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{T_{(k-1)}}^{t_k} \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda^x_{k-1}\Big(s, \mathcal{F}_{T_{(k-1-1)}}\Big) \mathrm{d}s\Big)} \Big(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\Big(s, \mathcal{F}_{T_{(k-1-1)}}\Big) \mathrm{d}s\Big)$$

$$P_{\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big))$$

$$= \int_{T_{(k-1)}}^{\tau} \int_{s}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) P_{\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)$$

$$\frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda^x_{k-1}\Big(s, \mathcal{F}_{T_{(k-1-1)}}\Big) \mathrm{d}s\Big)} \Big(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\Big(s, \mathcal{F}_{T_{(k-1-1)}}\Big) \mathrm{d}s\Big))$$

$$= \int_{T_{(k-1)}}^{\tau} \big(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\big)$$

$$\frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda^x_{k-1}\Big(s, \mathcal{F}_{T_{(k-1-1)}}\Big) \mathrm{d}s\Big)} \Big(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\Big(s, \mathcal{F}_{T_{(k-1-1)}}\Big) \mathrm{d}s\Big))$$

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

5.1 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 9 includes a censoring martingale. To estimate this martingale, we would need to have estimators of $\bar{Q}^g_{k,\tau}(t)$ (defined in Equation 8) 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 $\bar{Q}^g_{k,\tau}(t)$ directly due to the aforementioned high dimensionality of integrals. The IPCW approach we have given in Section 4.1 also would be prohibitively computationally expensive (at the very least if we use flexible machine learning estimators). 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 4.1.
2. First obtain the estimates $\tilde{\nu}_{k,\tau}$ by regressing $\hat{R}_{k+1}$ on $\Big(T_{(k)}, \Delta_{(k)}, A\Big(T_{(k)}\Big), \mathcal{F}_{T_{(k-1)}}\Big)$ (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 $\bar{Q}^g_{k,\tau}\Big(t, \mathcal{F}_{T_{(k)}}\Big)$ by

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

on the interevent level as we explained in step 1 of Section 4.1.

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

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

The resulting one-step estimator is given by

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

Under regularity conditions and empirical process, the one-step estimator is asymptotically linear and locally efficient. Conditions for the remainder term are given in Theorem 6. Conditions for the empirical process term are not stated here *yet*.

We have the following rate result for $\hat{\nu}^*_{k,\tau}$ which may be used in conjunction with Theorem 6.

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, that is

$$\sqrt{\int\int_{t_{k-1}}^{\tau}\left(\lambda^x_{k+1}(t \mid f_k) - \hat{\lambda}^x_{k+1}(t \mid f_k)\right)^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)} = o_P(1)$$

for $x = a, \ell, d, y$. Then for the derivatives of $\hat{\nu}^*_{k,\tau}$ and $\bar{Q}^g_{k,\tau}$, we have

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

where $P^*_k = m \otimes P|_{\mathcal{F}_{T_{(k)}}}$ and m is the Lebesgue measure on the interval $[0, \tau]$.

Proof: **Somewhat incomplete**. By the triangle inequality,

$$\| \hat{\nu}^{*,\prime}_{k,\tau} - \bar{Q}'^g_{k,\tau} \|_{L^2(P)} \leq \sqrt{\int\int_{t_k}^{\tau}\left(\hat{S}_{k+1}(t|f_k)\hat{\lambda}^y_{k+1}(t \mid f_k) - S_{k+1}(t|f_k)\lambda^y_{k+1}(t \mid f_k)\right)^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)}$$
$$+\sqrt{\int\int_{t_k}^{\tau}\left(\hat{\lambda}^a_{k+1}(t \mid f_k)\hat{S}_{k+1}(t|f_k) - S_{k+1}(t|f_k)\lambda^a_{k+1}(t \mid f_k)\right)^2(\tilde{\nu}(1, t, a, f_k))^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)}$$
$$+\sqrt{\int\int_{t_k}^{\tau}\left(\hat{\lambda}^\ell_{k+1}(t \mid f_k)\hat{S}_{k+1}(t|f_k) - S_{k+1}(t|f_k)\lambda^\ell_{k+1}(t \mid f_k)\right)^2(\tilde{\nu}(a_{k-1}, t, a, f_k))^2 dt P_{\mathcal{F}_{T_{(k)}}}(df_k)}$$
$$+\sqrt{\int\int_{t_k}^{\tau}\left(\tilde{\nu}^*_{k+1,\tau}(t, \ldots) - \bar{Q}^{-L}_{k+1}(t, \ldots)\right)^2\left(\sum_{x=a,l,y} S_{k+1}(t|f_k)\lambda^x_{k+1}(t \mid f_k)\right) dt P_{\mathcal{F}_{T_{(k-1)}}}(df_k)}$$

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

5.2 Remainder term

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

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

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

where

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

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

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

$$\begin{aligned} \varphi_k^*(P) = & \frac{\mathbb{1}\{A_0 = 1\}}{\pi_0(L(0))} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\left\{A\left(T_{(j)}\right) = 1\right\}}{\pi_j\left(T_{(j)}, \mathcal{F}_{T_{(j-1)}}\right)}\right)^{\mathbb{1}\{\Delta_{(j)} = a\}} \frac{1}{\prod_{j=1}^{k-1} S^c\left(T_{(j)} - \mid \mathcal{F}_{T_{(j-1)}}\right)} \mathbb{1}\left\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\right\} \\ & \times \left(\left(\bar{Z}_{k,\tau}^a - \nu_{k-1}\right) + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} (\nu_{k-1}^*(\tau) - \nu_{k-1}^*(u)) \frac{1}{S^c\left(u - \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} M_k^c(\mathrm{d}u)\right) \end{aligned}$$

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

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

Also note that

$$\mathbb{E}_{P_0}[\varphi_0^*(P)] + \Psi_\tau(P) - \Psi_\tau(P_0) = \mathbb{E}_{P_0}\left[\nu_0(L_0) - \bar{Q}_{0,\tau}^g(L_0)\right]. \tag{11}$$

Apply the law of iterated expectation to the efficient influence function in Equation 9 to get

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

where

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

Now note that

$$
\begin{aligned}
&\mathbb{E}_{P_0}\Big[h_k\Big(\mathcal{F}_{T_{(k)}}\Big) \mid \mathcal{F}_{T_{(k-1)}}\Big] \\
&= \mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}(S^c,\nu_k) \mid \mathcal{F}_{T_{(k-1)}}\Big] - \nu_{k-1,\tau}\Big(\mathcal{F}_{T_{(k-1)}}\Big) \\
&+\mathbb{E}_{P_0}\left[\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\Big(\nu^*_{k-1}\Big(\tau \mid \mathcal{F}_{T_{(k-1)}}\Big) - \nu^*_{k-1,\tau}\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)\Big)\frac{1}{S^c\Big(u- \mid \mathcal{F}_{T_{(k-1)}}\Big)S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}M^c_k(\mathrm{d}u)\right) \mid \mathcal{F}_{T_{(k-1)}}] \\
&= \mathbb{E}_{P_0}\big[\bar{Z}^a_{k,\tau}\big(S^c_0,\bar{Q}^g_{k,\tau}\big)\big) \mid \mathcal{F}_{T_{(k-1)}}\big] - \nu_{k-1,\tau}\Big(\mathcal{F}_{T_{(k-1)}}\Big) \\
&+\mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}(S^c,\nu_k) \mid \mathcal{F}_{T_{(k-1)}}\Big] - \mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}\Big(S^c,\bar{Q}^g_{k,\tau}\Big) \mid \mathcal{F}_{T_{(k-1)}}\Big] \\
&+\mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}\Big(S^c,\bar{Q}^g_{k,\tau}\Big) \mid \mathcal{F}_{T_{(k-1)}}\Big] - \mathbb{E}_{P_0}\big[\bar{Z}^a_{k,\tau}\big(S^c_0,\bar{Q}^g_{k,\tau}\big)\big) \mid \mathcal{F}_{T_{(k-1)}}\big] \\
&+\int_{T_{(k-1)}}^{\tau}\Big(\nu^*_{k-1}\Big(\tau \mid \mathcal{F}_{T_{(k-1)}}\Big) - \nu^*_{k-1,\tau}\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)\Big)\frac{S^c_0\Big(u- \mid \mathcal{F}_{T_{(k-1)}}\Big)S_0\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}{S^c\Big(u- \mid \mathcal{F}_{T_{(k-1)}}\Big)S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}\Big(\Lambda^c_{k,0}\Big(du \mid \mathcal{F}_{T_{(k-1)}}\Big) - \Lambda^c\Big(du \mid \mathcal{F}_{T_{(k-1)}}\Big)\Big)
\end{aligned}
$$

by a martingale argument. Noting that,

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

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

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

Since also

$$
\begin{aligned}
&\int \mathbb{1}\{t_1 < \dots < t_k < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^k \left(\frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})}\right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^k S_0^c(t_j - \mid f_{j-1})}{\prod_{j=1}^k S^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
&\qquad \times \left(\mathbb{E}_{P_0}\left[\bar{Z}^a_{k+1,\tau}\left(S_0^c, \bar{Q}^g_{k+1,\tau}\right) \mid \mathcal{F}_{T_{(k)}} = f_k\right] - \nu_{k,\tau}(f_k)\right) P_{\mathcal{F}_{T_{(k)}}}(df_k) \\
&+ \int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left(\frac{\pi_{0,j}(t_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\}, \dots, d_{k-1} \in \{\ell, a\}\} \\
&\qquad \times \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}(S^c, \nu_k) \mid \mathcal{F}_{T_{(k-1)}} = f_{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)}} = f_{k-1}\right] P\left(\mathcal{F}_{T_{(k-1)}}\right)(df_{k-1}) \\
&= \int \mathbb{1}\{t_1 < \dots < t_k < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^k \left(\frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})}\right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^k S_0^c(t_j - \mid f_{j-1})}{\prod_{j=1}^k S^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
&\qquad \times \left(\bar{Q}^g_{k,\tau}(f_k) - \nu_{k,\tau}(f_k)\right) P_{\mathcal{F}_{T_{(k)}}}(df_k) \qquad (12) \\
&+ \int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left(\frac{\pi_{0,j}(t_k, f_{j-1})}{\pi_j(t_k, f_{j-1})}\right)^{\mathbb{1}\{d_j=a\}} \frac{\prod_{j=1}^{k-1} S_0^c(t_j - \mid f_{j-1})}{\prod_{j=1}^{k-1} S^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} \\
&\quad \int \mathbb{1}\{t_k < \tau\} \frac{S_0^c(t_k - \mid f_{k-1})}{S^c(t_k - \mid f_{k-1})} \\
&\qquad \times \sum_{d_k = a, \ell} \left(\nu_k(t_k, d_k, g_k(a_k, d_k, f_{k-1}), l_k, f_{k-1}) - \bar{Q}^g_{k,\tau}(t_k, d_k, g_k(a_k, d_k, f_{k-1}), l_k, f_{k-1})\right) P_{T_{(k)}, \Delta_{(k)}, L(T_{(k)}) \mid \mathcal{F}_{T_{(k-1)}}}(df_k \mid f_{k-1}) P\left(\mathcal{F}_{T_{(k-1)}}\right)(df_{k-1}) \\
&= \int \mathbb{1}\{t_1 < \dots < t_k < \tau\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left(\frac{\pi_{0,j}(t_j, f_{k-1})}{\pi_j(t_j, f_{j-1})}\right)^{\mathbb{1}\{d_k=a\}} \left(\left(\frac{\pi_{0,k}(k_j, f_{k-1})}{\pi_k(t_k, f_{k-1})}\right)^{\mathbb{1}\{d_k=a\}} - 1\right) \frac{\prod_{j=1}^k S_0^c(t_j - \mid f_{j-1})}{\prod_{j=1}^k S^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
&\qquad \times \left(\bar{Q}^g_{k,\tau}(f_k) - \nu_{k,\tau}(f_k)\right) P_{\mathcal{F}_{T_{(k)}}}(df_k)
\end{aligned}
$$

where we set $g_k(a_k, d_k, f_{k-1}) = 1$ for $k > 1$. By combining Equation 11 and Equation 12, we are done. □

Note that by the triangle inequality

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

6 Real data application

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

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

We also want to compare with other methods.

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

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

An implementation is given in `ic_calculate.R` and `continuous_time_functions.R` and a simple run with simulated data can be run in `test_against_rtmle.R`.

7 Simulation study

The data generating mechanism should be based on real data given in Section 6. Note that the simulation procedure follows the DAG in Figure 6. Depending on the results from the data application, we should consider:

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

8 Discussion

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

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

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

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

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

and that

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

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

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

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

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

We also note that an iterative pseudo-value regression-based approach (Andersen et al. (2003)) may also possible, but is not further pursued in this article due to the computation time of the resulting procedure. Our ICE IPCW estimator also allows us to handle the case where the censoring distribution depends on time-varying covariates.

A potential other issue with the estimation of the nuisance parameters are that the history is high dimensional. This may yield issues with regression-based methods. If we adopt a TMLE approach, we may be able to use collaborative TMLE (van der Laan & Gruber, 2010) to deal with the high dimensionality of the history.

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

Bibliography

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

Andersen, P. K., Klein, J. P., & Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, *90*(1), 15–27. https://doi.org/10.1093/biomet/90.1.15

Cai, W., & Laan, M. J. van der. (2019). One-Step Targeted Maximum Likelihood Estimation for Time-to-Event Outcomes. *Biometrics*, *76*(3), 722–733. https://doi.org/10.1111/biom.13172

Chamapiwa, E. (2018). *Application of Marginal Structural Models (MSMs) to Irregular Data Settings*. The University of Manchester (United Kingdom).

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

Didelez, V. (2008). Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 245–264. https://doi.org/https://doi.org/10.1111/j.1467-9868.2007.00634.x

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

Gill, R. D., & Robins, J. M. (2001). Causal Inference for Complex Longitudinal Data: The Continuous Case. *The Annals of Statistics*, *29*(6), 1785–1811. https://doi.org/10.1214/aos/1015345962

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

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

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

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

Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

Richardson, T. S., & Robins, J. M. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, *128*(30), 2013.

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

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

Westling, T., Luedtke, A., Gilbert, P. B., & Carone, M. (2024). Inference for treatment-specific survival curves using machine learning. *Journal of the American Statistical Association*, *119*(546), 1541–1553.

9 Appendix

9.1 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 $\mathbb{F}_k = \mathcal{W} \times (\mathbb{R}_+ \times \{1, ..., d\} \times \mathcal{X})^k$.

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

1. We have $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \Delta_{(k)}, X\left(T_{(k)}\right)\right) \vee \mathcal{F}_{T_{(k-1)}}$ 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
$$\begin{aligned} &F_n(dw, dt_1, d\delta_1, dx_{11}, ..., dx_{1d}, ..., dt_n, d\delta_n, dx_{n1}, ..., dx_{nd}) \\ &= \left(\prod_{i=1}^n \mathbb{1}\{t_{i-1} < t_i\} \prod_{u \in (t_{i-1}, t_i)}\left(1 - \sum_{j=1}^d \Lambda_{i,j}(du, f_{i-1})\right) \sum_{j=1}^d \delta_j(d\delta_i) \zeta_{i,j}(dx_{ij}, t_i, f_{i-1}) \Lambda_{i,j}(dt_i, f_{i-1})\right) \mu(dw), \end{aligned}$$
and $f_k = (t_k, d_k, x_k, ...t_1, d_1, x_1, w) \in \mathbb{F}_k$ 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 $\mathcal{F}^N_t = \sigma(N((0, s], \cdot) \mid s \leq t) \vee \sigma(W) = \mathcal{F}_t$ 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 $i = 1, ..., d$.

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. □

9.2 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).

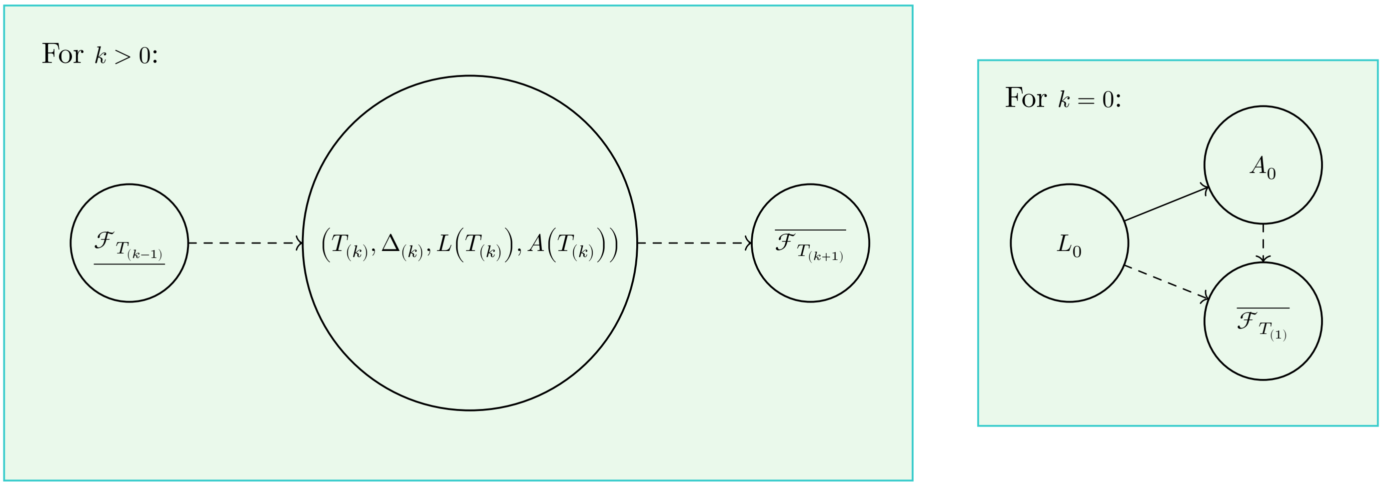

Figure 5: A DAG representing the relationships between the variables of O. The dashed lines indicate multiple edges from the dependencies in the past and into the future.

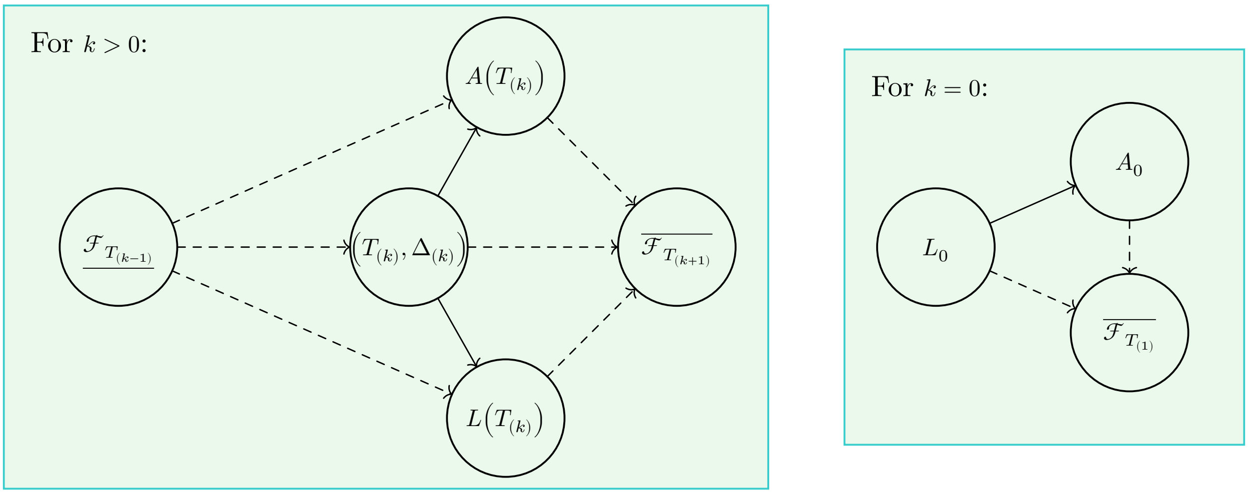

Figure 6: A DAG for simulating the data generating mechanism. The dashed lines indicate multiple edges from the dependencies in the past and into the future. Here $\underline{\mathcal{F}_{T_{(k)}}}$ is the history up to and including the k'th event and $\overline{\mathcal{F}_{T_{(k)}}}$ is the history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

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

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

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

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

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

$$+ \int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} (1 - B_{k-1}(u)) M_k^y(du) + \int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} (0 - B_{k-1}(u)) M_k^d(du)$$

$$+ \frac{1}{S^c\left(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\right)} I\left(T_{(k)} \leq \tau, \Delta_{(k)} = \ell, k < K\right) \Big(\bar{Q}_{k,\tau}^g\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, \ell, \mathcal{F}_{T_{(k-1)}}\right)$$

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

$$+ \int \bar{Q}_{1,\tau}^g(a, L_0) \pi_0^*\left(0, \mathcal{F}_{T_{(0-1)}}\right) \nu_A(\mathrm{d}a) - \Psi_\tau(P)$$

We find immediately that

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

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

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

$$- \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} (1) \Lambda_k^y(du) - \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} (0) \Lambda_k^d(du)$$

$$- \int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} B_{k-1}(u) M_k^\bullet(\mathrm{d}u)$$

$$+ \bar{Z}_{k,\tau}\left(T_{(k)}, \Delta_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right), \mathcal{F}_{T_{(k-1)}}\right) + \int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} B_{k-1}(u) M_k^c(du) \Bigg]$$

$$+ \int \bar{Q}_{1,\tau}^g(a, L_0) \pi_0^*\left(0, \mathcal{F}_{T_{(0-1)}}\right) \nu_A(\mathrm{d}a) - \Psi_\tau(P)$$

Now note that

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

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

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

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

Let us calculate the second integral

$$\bar{Q}^g_{k-1,\tau}(\tau)\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{1}{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}\lambda^\bullet_{k-1}\Big(s,\mathcal{F}_{T_{(k-1-1)}}\Big)\mathrm{d}s$$
$$=\bar{Q}^g_{k-1,\tau}(\tau)\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}{\Big(S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)\Big)^2}\lambda^\bullet_{k-1}\Big(s,\mathcal{F}_{T_{(k-1-1)}}\Big)\mathrm{d}s$$
$$=\bar{Q}^g_{k-1,\tau}(\tau)\left(\frac{1}{S^c\Big(T_{(k)}\wedge\tau\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(T_{(k)}\wedge\tau\mid\mathcal{F}_{T_{(k-1)}}\Big)}-1\right)$$

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

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

Now note that

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

Setting this into the previous integral, we get

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

Thus, we find

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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