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 Introduction

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

Marginal structural models (MSMs), introduced by Robins (1986), are a widely used approach for estimating causal effects from observational data, particularly in the presence of time-varying confounding and treatment. MSMs typically require that data be recorded on a discrete time scale, capturing all relevant information available to the clinician at each treatment decision point and for the outcome.

However, many real-world datasets — such as health registries — are collected in continuous time, with patient characteristics updated at irregular, subject-specific times. These datasets often include detailed, timestamped information on events and biomarkers, such as drug purchases, hospital visits, and laboratory results. Analyzing data in its native continuous-time form avoids the need for discretization, which can introduce bias and increase variance depending on the choice of time grid (Ferreira Guerra et al. (2020); Ryalen et al. (2019)).

In this paper, we consider a longitudinal continuous-time framework similar to that of Rytgaard et al. (2022). We establish identification criteria for the causal effect of treatment on an outcome within this setting. Like Rytgaard et al. (2022), we adopt a nonparametric approach and focus on estimation and inference through the efficient influence function, yielding nonparametrically locally efficient estimators via a one-step procedure.

To this end, we propose an inverse probability of censoring iterative conditional expectation (ICE-IPCW) estimator, which, like that of Rytgaard et al. (2022), iteratively updates nuisance parameters. A key innovation in our method is that these updates are performed by indexing backwards through the number of events rather than through calendar time. Moreover, our estimator addresses challenges associated with the high dimensionality of the target parameter by employing inverse probability of censoring weighting (IPCW). The distinction between event-based and time-based updating is illustrated in Figure 1 and Figure 2. To the best of our knowledge, no general estimation procedure has yet been proposed for the components involved in the efficient influence function.

Continuous-time methods for causal inference in event history analysis have also been explored by Røysland (2011) and Lok (2008). Røysland (2011) developed identification criteria using a formal martingale framework based on local independence graphs, enabling causal effect estimation in continuous time via a change of measure. Lok (2008) similarly employed a martingale approach but focused on structural nested models to estimate a different type of causal parameter—specifically, a conditional causal effect. However, such estimands may be more challenging to interpret than marginal causal effects.

A key challenge shared by these approaches is the need to model intensity functions, which can be difficult to estimate accurately. While methods such as Cox proportional hazards (Cox (1972)) and Aalen additive hazards (Aalen (1980)) are commonly used for modeling intensities, they are often inadequate in the presence of time-varying confounding, as they do not naturally account for the full history of time-varying covariates. Consequently, summary statistics of covariate history are typically used to approximate the intensity functions.

In this paper, we propose a simple solution to this issue for settings with a limited number of events. Our approach enables the use of existing regression techniques from the survival analysis and point process literature to estimate the necessary intensities, providing a practical and flexible alternative.

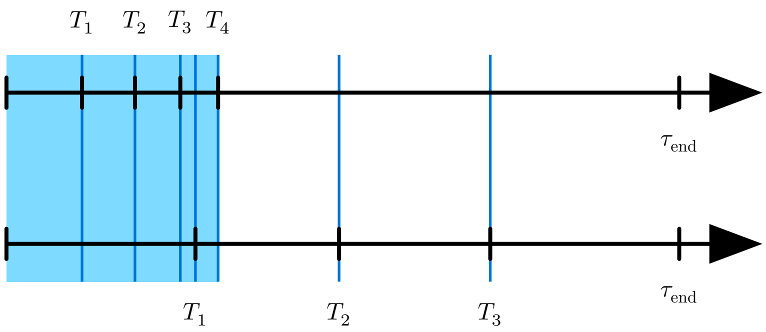

Figure 1: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let $t_1 < \ldots < t_m$ be all the event times in the sample. Then, given $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{t_r}\right]$, we regress back to $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{t_{r-1}}\right]$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

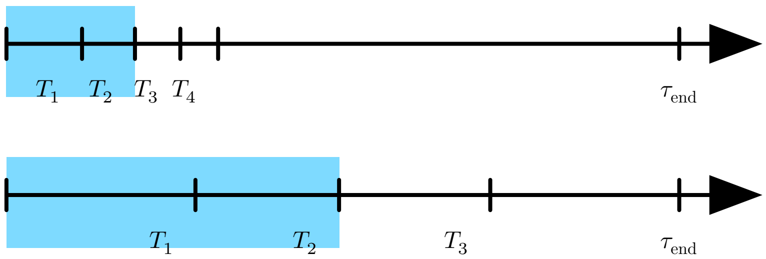

Figure 2: The figure illustrates the sequential regression approach proposed in this article. For each event k in the sample, we regress back on the history $\mathcal{F}_{T_{(k-1)}}$. That is, given $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{T_{(k)}}\right]$, we regress back to $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{T_{(k-1)}}\right]$. In the figure, $k = 3$.

2 Setting and Notation

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

At baseline, we record the values of the treatment $A(0)$ and the time-varying covariates $L(0)$ and let $\mathcal{F}_0 = \sigma(A(0), L(0))$ be the σ-algebra corresponding to the baseline information. We assume that we have two treatment options over time so that $A(t) \in \{0, 1\}$ (e.g., placebo and active treatment), where $A(t)$ denotes the treatment at time $t \geq 0$. The time-varying confounders $L(t)$ at time $t > 0$ are assumed to take values in a finite subset $\mathcal{L} \subset \mathbb{R}^m$, so that $L(t) \in \mathcal{L}$ for all $t \geq 0$. We assume that the stochastic processes $(L(t))_{t\geq 0}$ and $(A(t))_{t\geq 0}$ are càdlàg (right-continuous with left limits), jump processes. The fact that they are jump processes by Assumption 1 means that for P –almost all $\omega \in \Omega$, the processes $L(\omega, \cdot)$ and $A(\omega, \cdot)$ have at most $K - 1$ jumps. Furthermore, we require that the times at which the treatment and covariate values may change are dictated entirely by the counting processes $(N^a(t))_{t\geq 0}$ and $(N^\ell(t))_{t\geq 0}$, respectively in the sense that $\Delta A(t) \neq 0$ only if $\Delta N^a(t) \neq 0$ and $\Delta L(t) \neq 0$ only if $\Delta N^\ell(t) \neq 0$.

We also have counting processes representing the event of interest $(N^y(t))_{t\geq 0}$ and the competing event $(N^d(t))_{t\geq 0}$. Later, we will introduce the counting process N^c for the censoring process. For all counting processes involved, we assume for simplicity that the jump times differ with probability 1 (Assumption 2). Thus, we have observations from a the jump process $\alpha(t) = (N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t))$, and the natural filtration $(\mathcal{F}_t)_{t\geq 0}$ is given by $\mathcal{F}_t = \sigma(\alpha(s) \mid s \leq t) \vee \mathcal{F}_0$. Let $T_{(k)}$ be the k'th ordered jump time of α, that is $T_0 = 0$ and $T_{(k)} = \inf\left\{t > T_{(k-1)} \mid \alpha(t) \neq \alpha\left(T_{(k-1)}\right)\right\} \in [0, \infty]$ be the time of the k'th event and let $\Delta_{(k)} \in \{y, d, a, \ell\}$ be the status of the k'th event, i.e., $\Delta_{(k)} = x$ if $\Delta N^x\left(T_{(k)}\right) = 1$, so that

1. each $T_{(k)}$ is a $\mathcal{F}_t$ stopping time.
2. $T_{(k)} < T_{(k+1)}$ if $T_{(k)} < \infty$.
3. $T_{(k+1)} = \infty$ if $T_{(k)} = \infty$ or $\Delta_{(k-1)} \in \{y, d\}$.

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

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

where δ_x denotes the Dirac measure on $(\mathbb{R}_+ \times (\{y, d, a, \ell\} \times \{0,1\} \times \mathcal{L}))$. It follows that the filtration $(\mathcal{F}_t)_{t\geq 0}$ is the natural filtration of the random measure N^α (Theorem 8 (i)). Thus, the random measure N^α carries the same information as the stochastic process $(\alpha(t))_{t\geq 0}$. This will be critical for identification of the causal effect of interest and dealing with right-censoring.

Furthermore, it follows that the stopping time σ-algebra $\mathcal{F}_{T_{(k)}}$ associated with stopping time $T_{(k)}$ fulfills that $\mathcal{F}_{T_{(k)}} = \sigma(T_{(k)}, \Delta_{(k)}, A(T_{(k)}), L(T_{(k)})) \vee \mathcal{F}_{T_{(k-1)}}$. $\mathcal{F}_{T_{(k)}}$ represents the information up to the k'th event. We will interpret $\mathcal{F}_{T_{(k)}}$ as a random variable instead of a σ-algebra, whenever it is convenient to do so and also make the implicit assumption that whenever we condition on $\mathcal{F}_{T_{(k)}}$, we only consider the cases where $T_{(k)} < \infty$ and $\Delta_{(k)} \in \{a, \ell\}$.

We observe $O = \mathcal{F}_{T_{(K)}} = \left(T_{(K)}, \Delta_{(K)}, T_{(K-1)}, \Delta_{(K-1)}, A(T_{(K-1)}), L(T_{(K-1)}), \ldots, A(0), L(0)\right) \sim P \in \mathcal{M}$ where $\mathcal{M}$ is the statistical model, i.e., a set of probability measures. Let $\pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ be the probability of being treated at the k'th event given $\Delta_{(k)} = a, T_{(k)} = t$, and $\mathcal{F}_{T_{(k-1)}}$. Similarly, let $\mu_k\left(t, \cdot, \mathcal{F}_{T_{(k-1)}}\right)$ be the probability measure for the covariate value $\Delta_{(k)} = \ell, T_{(k)} = t$, and $\mathcal{F}_{T_{(k-1)}}$. Let also $\Lambda_k^x\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right)$ be the cumulative cause-specific hazard measure* for the k'th event of type x given $\mathcal{F}_{T_{(k-1)}}$.

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. The K'th event is terminal.

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.†

3 Causal framework

Our overall goal is to estimate the interventional cumulative incidence function at time τ,

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

where $\tilde{N}^y(t)$ is the potential outcome (a counting process with at most one jump) 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 g, possibly contrary to fact, been followed. For simplicity, we assume that the treatment regime specifies that $A(t) = 1$ for all $t \geq 0$. This means that treatment is administered at each visitation time. In terms of this data, this means that we must have $A(0) = 1$ and $A(T_{(k)}) = 1$ whenever $\Delta_{(k)} = a$ and $T_{(k)} < t$. We now define the càdlàg weight process $(W(t))_{t\geq 0}$ given by

*Let $T \in (0, \infty]$ and $X \in \mathcal{X}$ be random variables. Then the cause-specific cumulative hazard measure is given by $\Lambda_x(\mathrm{d}t) = \mathbb{1}\{P(T \geq t\} > 0\} \frac{P(T \in \mathrm{d}t, X = x)}{P(T \geq t)}$ (Appendix A5.3 of Last & Brandt (1995)).

†If the resulting martingales M^x, are of locally bounded variation, then the processes are orthogonal $\left[M^x, M^{x'}\right]_t = 0$ for $x \neq x'$ P-a.s, where $[\cdot, \cdot]$ is the quadratic covariation process.

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

where $N_t = \sum_k \mathbb{1}\left\{T_{(k)} \leq t\right\}$ is the number of events up to time t, and we consider the observed data target parameter $\Psi_\tau^{\text{obs}} : \mathcal{M} \to \mathbb{R}_+$[‡] given by

$$\Psi_\tau^{\text{obs}}(P) = \mathbb{E}_P[N^y(\tau) W(\tau)]. \tag{2}$$

We provide both martingale and non-martingale conditions for the identification $\left(\Psi_\tau^g(P) = \Psi_\tau^{\text{obs}}(P)\right)$ of the mean potential outcome in Theorem 1 and Theorem 2, respectively. One can also define a (stochastic) intervention with respect to a local independence graph (Røysland et al. (2024)) but we do not further pursue this here. 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.[§]

3.1 Identification of the causal effect (martingale approach)

Let $N_t^a(\cdot) = N^\alpha((0, t] \times \{a\} \times \cdot \times \mathcal{L})$ be the random measure on $(\mathbb{R}_+ \times \{0, 1\})$ for the treatment process and let $\Lambda_t^a(\cdot)$ be the corresponding P-$\mathcal{F}_t$ compensator. We consider a martingale approach for the identification of causal effects similar to the approach taken in Ryalen (2024).[¶]

To do this this, we define the stopping time T^a as the time of the first visitation event where the treatment plan is not followed, i.e.,

$$T^a = \inf_{t \geq 0} \{A(t) = 0\} = \begin{cases} \inf_{k > 1} \left\{T_{(k)} \mid \Delta_{(k)} = a, A\left(T_{(k)}\right) \neq 1\right\} \text{ if } A(0) = 1 \\ 0 \text{ otherwise} \end{cases}$$

Overall T^a acts as a coarsening variable, limiting the ability to observe the full potential outcome process. An illustration of the consistency condition in Theorem 1 is provided in Figure 3. Informally, the consistency condition states that the potential outcome process $\widetilde{N}^y(t)$ coincides $N^y(t)$ if the treatment plan is *followed* at time t.

To fully phrase the causal inference problem as a missing data problem, we also need an exchangeability condition. The intuition behind the exchangeability condition in Theorem 1 is that the outcome process $\widetilde{N}^y$ should be independent of both the timing of treatment visits and treatment assignment, conditional on observed history.

We also briefly discuss the positivity condition, which ensures that $(W(t))_{t \geq 0}$ is a uniformly integrable martingale with $\mathbb{E}_P[W(t)] = 1$ for all $t \in [0, \tau_{\text{end}}]$ by Equation 8. This guarantees that the observed data parameter $\Psi_\tau^{\text{obs}}(P)$ is well-defined.

Note that instead of conditioning on the entire potential outcome process in the exchangeability condition, we could have simply conditioned on a single potential outcome variable $\tilde{T}_y := \inf\left\{t > 0 \mid \widetilde{N}^y(t) = 1\right\} \in [0, \infty]$[‖] and added that information at baseline[**].

We can also state the time-varying exchangeability condition of Theorem 1 explicitly in terms of the observed data: Let $\mathcal{H}_{T_{(k)}}$ be the corresponding stopping time σ-algebra for the k'th event with respect to the filtration $\{\mathcal{H}_t\}$ given in Theorem 1. In light of the canonical compensator (Theorem 8 (ii)), we see immediately that the exchangeability condition is fulfilled if $A\left(T_{(k)}\right) \perp \tilde{T}_y \mid T_{(k)}, \Delta_{(k)} = a, \mathcal{F}_{T_{(k-1)}}$ and the cause-specific cumulative hazards for $T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}, \tilde{T}_y$ for treatment visits only depend on $\mathcal{F}_{T_{(k-1)}}$ and not on $\tilde{T}_y$.

[‡]Note that by fifth equality of Appendix S1.2 of Rytgaard et al. (2022), this is the same as the target parameter in Rytgaard et al. (2022) with no competing event.

[§]see Richardson & Robins (2013) for the discrete time variant, i.e., single world intervention graphs.

[¶]The overall difference between Ryalen (2024) and our exchangeability condition is that the P-$\mathcal{F}_t$ compensator $\Lambda_t^a(\{1\})$ is not required to be the P-$\left(\mathcal{F}_t \vee \sigma\left(\widetilde{N}^y\right)\right)$ compensator for N^a.

[‖]A competing event occuring corresponds to $\tilde{T}_y = \infty$

[**]Note that $\mathbb{1}\left\{\tilde{T}_y \leq t\right\} = \widetilde{N}^y(t)$ for all $t > 0$ because $\left(\widetilde{N}^y(t)\right)_{t \geq 0}$ jumps at most once.

Theorem 1 (Martingale identification of mean potential outcome): Define

$$\zeta(t,m,a,l) := \sum_k \mathbb{1}\left\{T_{(k-1)} < t \le T_{(k)}\right\} \left(\frac{\mathbb{1}\{a=1\}}{\pi_k\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)} \right)^{\mathbb{1}\{m=a\}}, \tag{3}$$

If *all* of the following conditions hold:

- **Consistency**: $\widetilde{N}^y(t)\mathbb{1}\{T^a > t\} = N^y(t)\mathbb{1}\{T^a > t\} \quad P-\text{a.s.}$
- **Exchangeability**: Define $\mathcal{H}_t := \mathcal{F}_t \vee \sigma\left(\widetilde{N}^y\right)$. The P-$\mathcal{F}_t$ compensator for N^a $\Lambda_t^a(\cdot)$ is also the P-$\mathcal{H}_t$ compensator and
$$\widetilde{N}^y(t) \perp A(0) \mid L(0), \forall t \in (0, \tau_{\text{end}}].$$
- **Positivity**: $\mathbb{E}_P[\int \mathbb{1}\{t \le \tau_{\text{end}}\} \, |\zeta(t,m,a,l) - 1| \, W(t-)N(d(t,m,a,l))] < \infty$ and $\mathbb{E}_P[W(0)] = 1$.

Then,

$$\Psi_t^g(P) = \Psi_t^{\text{obs}}(P)$$

for all $t \in (0, \tau_{\text{end}}]$.

Proof: We shall use that the likelihood ratio solves a specific stochastic differential equation (we use what is essentially Equation (2.7.8) of Andersen et al. (1993)), but present the argument using Theorem 10.2.2 of Last & Brandt (1995) as the explicit conditions are not stated in Andersen et al. (1993). First, let

$$\begin{aligned}\psi_{k,x}\left(t, \mathcal{F}_{T_{(k-1)}}, \mathrm{d}(m,a,l)\right) &= \mathbb{1}\{x=a\}\left(\delta_1(\mathrm{d}a)\pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right) + \delta_0(\mathrm{d}a)\left(1 - \pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)\right)\right)\delta_{L\left(T_{(k-1)}\right)}(\mathrm{d}l) \\ &+ \mathbb{1}\{x=\ell\}\mu_k\left(dl, t, \mathcal{F}_{T_{(k-1)}}\right)\delta_{A\left(T_{(k-1)}\right)}(\mathrm{d}a) \\ &+ \mathbb{1}\{x \in \{y,d\}\}\delta_{A\left(T_{(k-1)}\right)}(\mathrm{d}a)\delta_{L\left(T_{(k-1)}\right)}(\mathrm{d}l).\end{aligned} \tag{4}$$

We shall use that the P-$\mathcal{F}_t$ compensator of N^α is given by

$$\Lambda^\alpha(d(t,m,a,l)) = \sum_k \mathbb{1}\left\{T_{(k-1)} < t \le T_{(k)}\right\} \sum_{x=a,\ell,y,d} \delta_x(\mathrm{d}m)\psi_{k,x}(t, d(a,l))\Lambda_k^x\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right), \tag{5}$$

see e.g., Theorem 8 (ii). Second, let $\Phi(d(t,x)) = \mathbb{1}\{t \le \tau_{\text{end}}\}N^\alpha(d(t,x))$ and $\nu(d(t,x)) = \mathbb{1}\{t \le \tau_{\text{end}}\}\Lambda^\alpha(d(t,x))$ be the restricted random measure and its compensator. We define P-$\mathcal{F}_t$ predictable, $\mu(d(t,x)) := \zeta(t,x)\nu(d(t,x))$. Here, we use the shorthand notation $x = (m,a,l)$. The likelihood ratio process $L(t)$ given in (10.1.14) of Last & Brandt (1995) is defined by

$$\begin{aligned}L(t) = \mathbb{1}\{t < T_\infty \wedge T_\infty(\nu)\}L_0 &\prod_{n: T_{(n)} \le t} \zeta\left(T_{(n)}, \Delta_{(n)}, A\left(T_{(n)}\right), L\left(T_{(n)}\right)\right) \\ &\prod_{\substack{s \le t \\ N_{s-} = N_s}} \frac{1 - \bar{\nu}\{s\}}{1 - \bar{\mu}\{s\}} \exp\left(\int \mathbb{1}\{s \le t\}(1 - \zeta(s,x))\nu^c(d(s,x))\right) \\ &+ \mathbb{1}\{t \ge T_\infty \wedge T_\infty(\nu)\} \liminf_{s \to T_\infty \wedge T_\infty(\nu)} L(s).\end{aligned} \tag{6}$$

Here $T_\infty := \lim_n T_n$, $T_\infty(\nu) := \inf\{t \ge 0 \mid \nu((0,t] \times \{y,d,a,\ell\} \times \{0,1\} \times \mathcal{L}) = \infty\}$, $\bar{\mu}(\cdot) := \mu(\cdot \times \{y,d,a,\ell\} \times \{0,1\} \times \mathcal{L})$, $\bar{\nu}(\cdot) := \nu(\cdot \times \{y,d,a,\ell\} \times \{0,1\} \times \mathcal{L})$, $\nu^c(\mathrm{d}(s,x)) := \mathbb{1}\{\bar{\nu}\{s\} = 0\}\nu(d(s,x))$, and $L_0 := W(0) = \frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))}$.

First, we will show that $L(t) = W(t)$, where $W(t)$ is the weight process defined in Equation 1.

Note that by Assumption 1, $T_\infty = \infty \quad P$-a.s. and thus $T_\infty(\nu) = T_\infty = \infty$ in view Theorem 4.1.7 (ii) since $\bar{\nu}\{t\} < \infty$ for all $t > 0$.

Second, note that $\bar{\nu} = \bar{\mu}$. This follows since

$$
\begin{aligned}
\bar{\nu}(A) &= \int_{A\times\{y,d,a,\ell\}\times\{0,1\}\times\mathcal{L}} \zeta(t,m,a,l)\mu(d(t,m,a,l)) \\
&= \int_{A\times\{y,d,\ell\}\times\{0,1\}\times\mathcal{L}} \zeta(t,m,a,l)\mu(d(t,m,a,l)) + \int_{A\times\{a\}\times\{0,1\}\times\mathcal{L}} \zeta(t,m,a,l)\mu(d(t,m,a,l)) \\
&= \int_{A\times\{y,d,\ell\}\times\{0,1\}\times\mathcal{L}} 1\mu(d(t,m,a,l)) + \int_{A\times\{a\}\times\{0,1\}\times\mathcal{L}} \zeta(t,m,a,l)\mu(d(t,m,a,l)) \\
&= \mu(A\times\{y,d,\ell\}\times\{0,1\}\times\mathcal{L}) + \sum_k \mathbb{1}\left\{T_{(k-1)} < t \le T_{(k)}\right\}\Lambda^a\left(dt \mid \mathcal{F}_{T_{(k-1)}}\right) \\
&= \mu(A\times\{y,d,a,\ell\}\times\{0,1\}\times\mathcal{L}) = \bar{\mu}(A),
\end{aligned}
$$

for Borel measurable sets $A \subseteq \mathbb{R}_+$, where the last step follows from the form of the compensator (Equation 5). Thus

$$
\prod_{\substack{s\le t\\ N_{s-}=N_s}} \frac{1-\bar{\nu}\{s\}}{1-\bar{\mu}\{s\}} \exp\left(\int \mathbb{1}\{s\le t\}(1-\zeta(s,x))\nu^c(d(s,x))\right) = 1,
$$

and hence

$$
\begin{aligned}
L(t) &= L_0 \prod_{n:T_{(n)}\le t} \zeta\left(T_{(n)}, \Delta_{(n)}, A\left(T_{(n)}\right), L\left(T_{(n)}\right)\right) \\
&\overset{\text{def.}}{=} W(t).
\end{aligned}
$$

Let $V(s,x) = \zeta(s,x) - 1 + \frac{\bar{\nu}\{s\}-\bar{\mu}\{s\}}{1-\bar{\mu}\{s\}} = \zeta(s,x) - 1$. $L(t)$ will fulfill that

$$
L(t) = L_0 + \int \mathbb{1}\{s\le t\} V(s,x) L(s-)[\Phi(d(s,x)) - \nu(d(s,x))]
$$

if

$$
\begin{aligned}
\mathbb{E}_P[L_0] &= 1, \\
\bar{\mu}\{t\} &\le 1, \\
\bar{\mu}\{t\} = 1 &\text{ if } \bar{\nu}\{t\} = 1, \\
\bar{\mu}[T_\infty \wedge T_\infty(\mu)] = 0 &\text{ and } \bar{\nu}[T_\infty \wedge T_\infty(\nu)] = 0.
\end{aligned}
\tag{7}
$$

by Theorem 10.2.2 of Last & Brandt (1995).

The first condition holds by positivity. The second condition holds by the specific choice of compensator since $\sum_x \Lambda_k^x\left(\{t\}, \mathcal{F}_{T_{(k-1)}}\right) \le 1$ for all $k = 1, \ldots, K$ and $t \in (0, \tau_{\text{end}}]$ (Theorem A5.9 of Last & Brandt (1995)). The third holds since $\bar{\mu} = \bar{\nu}$ and the fourth holds since $T_\infty = T_\infty(\nu) = T_\infty(\mu) = \infty$.

Thus,

$$
W(t) = \frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))} + \int_0^t W(s-)V(s,x)(\Phi(d(s,x)) - \nu(d(s,x))).
\tag{8}
$$

Then we shall show that

$$
M_t^* := \int \widetilde{N}^y(t)\mathbb{1}\{s\le t\}V(s,x)L(s-)[N(d(s,x)) - \Lambda(d(s,x))]
\tag{9}
$$

is a zero mean uniformly integrable martingale. This follows if

$$
\mathbb{E}_P\left[\int \widetilde{N}^y(t)\ |V(s,x)|\ L(s-)\Phi(d(s,x))\right] < \infty.
$$

and if $(\omega,s,x) \mapsto \widetilde{N}^y(t)\ |V(s,x)|\ L(s-)$ is P-$\mathcal{H}_s$ predictable by Exercise 4.1.22 of Last & Brandt (1995). Since

$$
\mathbb{E}_P\left[\int \widetilde{N}^y(t)\ |V(s,x)|\ L(s-)\Phi(d(s,x))\right] \le \mathbb{E}_P\left[\int \mathbb{1}\{s\le \tau_{\text{end}}\}\ |V(s,x)|\ L(s-)N(d(s,x))\right] < \infty
$$

and $(\omega,s) \mapsto \widetilde{N}^y(t)$ is predictable with respect to $\mathcal{H}_s$, $(\omega,s) \mapsto L(s-)$ is P-$\mathcal{H}_s$ predictable (càglàd and adapted; Theorem 2.1.10 of Last & Brandt (1995)), $(\omega,s,x) \mapsto V(s,x)$ is P-$\mathcal{H}_s$ predictable (Theorem 2.2.22 of Last & Brandt (1995)), so that $(\omega,s) \mapsto \widetilde{N}^y(t)\ |V(s,x)|\ L(s-)$ is P-$\mathcal{H}_s$ predictable, and the desired martingale result for Equation 9 follows. This in turn implies by Equation 8:

$$
\begin{aligned}
\mathbb{E}_P\left[\widetilde{N}^y_t W(t)\right] &= \mathbb{E}_P\left[\widetilde{N}^y_t W(0)\right] + \mathbb{E}_P[M^*_t] \\
&= \mathbb{E}_P\left[\widetilde{N}^y_t W(0)\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\widetilde{N}^y_t \mid \mathcal{F}_0\right] W(0)\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\widetilde{N}^y_t \mid L(0)\right] W(0)\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\widetilde{N}^y_t \mid L(0)\right] \mathbb{E}_P[W(0) \mid L(0)]\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\widetilde{N}^y_t \mid L(0)\right] 1\right] \\
&= \mathbb{E}_P\left[\widetilde{N}^y_t\right],
\end{aligned}
$$

where we use the baseline exchangeability condition and the law of iterated expectation. □

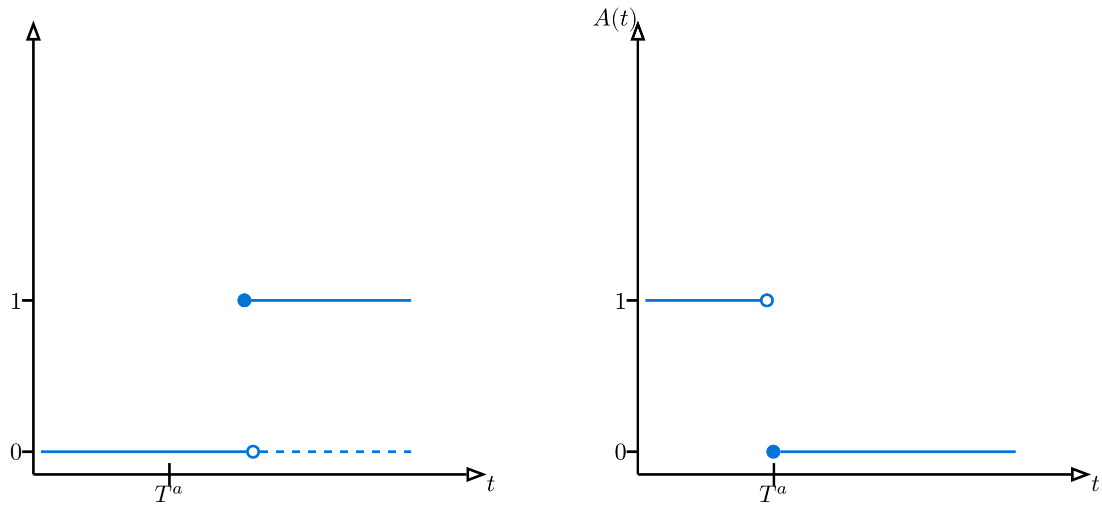

Figure 3: The figure illustrates the consistency condition for the potential outcome framework for single individual. The left panel shows the potential outcome process $\widetilde{N}^y(t)$ (dashed) and the observed process $N^y(t)$ (solid). The right panel shows the treatment process $A(t)$. At time T^a, the treatment is stopped and the processes may from some random future point diverge from each other.

3.2 Identification of the causal effect (non-martingale approach)

In this subsection, we present a non-martingale approach for the identification of causal effects, and the conditions are stated for identification at the time horizon of interest.

Theorem 2: Assume **Consistency** and **Positivity** as in Theorem 1 for a single timepoint τ (in the positivity condition replace τ_{end} with τ). Additionally, we assume that:

- **Exchangeability**: We have

$$
\begin{aligned}
&\widetilde{N}^y(\tau)\mathbb{1}\left\{T_{(j)} \leq \tau < T_{(j+1)}\right\} \perp A\left(T_{(k)}\right) \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \geq k > 0, \\
&\widetilde{N}^y(\tau)\mathbb{1}\left\{T_{(j)} \leq \tau < T_{(j+1)}\right\} \perp A(0) \mid L(0), \quad \forall j \geq 0.
\end{aligned}
\tag{10}
$$

Then the estimand of interest is identifiable, i.e.,

$$\Psi^g_\tau(P) = \Psi^{\text{obs}}_\tau(P).$$

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

$$
\begin{aligned}
\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} N^y_\tau W(\tau)\Big] &= \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\}\mathbb{1}\{T^a > \tau\} N^y_\tau W(\tau)\Big] \\
&= \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\}\mathbb{1}\{T^a > \tau\} \widetilde{N}^y(\tau) W(\tau)\Big] \\
&= \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau) W(\tau)\Big] \\
&= \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau) W\big(T_{(k-1)}\big)\Big] \\
&= \mathbb{E}_P\left[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau) \left(\frac{\mathbb{1}\Big\{A\big(T_{(k-1)}\big) = 1\Big\}}{\pi_{k-1}\Big(T_{(k-1)}, \mathcal{F}_{T_{(k-2)}}\Big)}\right)^{\mathbb{1}\{\Delta_{(k-1)} = a\}} W\big(T_{(k-2)}\big)\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau) \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}, A\big(T_{(k-1)}\big)\Big]\right. \\
&\qquad \left. \times \left(\frac{\mathbb{1}\Big\{A\big(T_{(k-1)}\big) = 1\Big\}}{\pi_{k-1}\Big(T_{(k-1)}, \mathcal{F}_{T_{(k-2)}}\Big)}\right)^{\mathbb{1}\{\Delta_{(k-1)} = a\}} W\big(T_{(k-2)}\big)\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau) \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}\Big]\right. \\
&\qquad \left. \times \left(\frac{\mathbb{1}\Big\{A\big(T_{(k-1)}\big) = 1\Big\}}{\pi_{k-1}\Big(T_{(k-1)}, \mathcal{F}_{T_{(k-2)}}\Big)}\right)^{\mathbb{1}\{\Delta_{(k-1)} = a\}} W\big(T_{(k-2)}\big)\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau) \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}\Big]\right. \\
&\qquad \left. \times \mathbb{E}_P\left[\left(\frac{\mathbb{1}\Big\{A\big(T_{(k-1)}\big) = 1\Big\}}{\pi_{k-1}\Big(A\big(T_{(k-1)}\big), \mathcal{F}_{T_{(k-2)}}\Big)}\right)^{\mathbb{1}\{\Delta_{(k-1)} = a\}} \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}\right] W\big(T_{(k-2)}\big)\right] \\
&= \mathbb{E}_P\Big[\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau) \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}\Big] W\big(T_{(k-2)}\big)\Big] \\
&= \mathbb{E}_P\Big[\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau) \mid \mathcal{F}_{T_{(k-3)}}, \Delta_{(k-2)}, T_{(k-2)}, A\big(T_{(k-2)}\big)\Big] W\big(T_{(k-2)}\big)\Big]
\end{aligned}
$$

Iteratively applying the same argument, we get that $\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} \widetilde{N}^y(\tau)\Big] = \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(k-1)} \leq \tau < T_{(k)}\Big\} N^y_\tau W(\tau)\Big]$ as needed. □

By the intersection property of conditional independence, we see that a sufficient condition for the first exchangeability condition in Equation 10 is that

$$
\begin{aligned}
&\widetilde{N}^y(\tau) \perp A\big(T_{(k)}\big) \mid T_{(j)} \leq \tau < T_{(j+1)}, \mathcal{F}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \geq k > 0, \\
&\mathbb{1}\Big\{T_{(j)} \leq \tau < T_{(j+1)}\Big\} \perp A\big(T_{(k)}\big) \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \geq k > 0.
\end{aligned}
$$

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 $\mathbb{1}\Big\{T_{(k)} \leq \tau, \Delta_{(k)} = y\Big\}$. Then the exchangeability condition is that $\tilde{Y}_{\tau,k} \perp A\big(T_{(j)}\big) \mid \mathcal{F}_{T_{(j-1)}}, T_{(j)}, \Delta_{(j)} = a)$ 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.

3.3 Iterated representation of the target parameter

In this section, we present a simple iterated representation of the observed data target parameter $\Psi_\tau^{\text{obs}}(P)$. We give an iterated conditional expectations formula for the target parameter in the case with no censoring. To do so, define

$$S_k\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = \prod_{s \in (0,t]} \left(1 - \sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right)\right), k = 1, ..., K$$

where $\prod_{s \in (0,t]}$ is the product integral over the interval $(0, t]$ (Gill & Johansen (1990)). We discuss more thoroughly the implications of this representation in the next section, where we deal with right-censoring.

Theorem 3: Let $\bar{Q}_{K,\tau}^g = \mathbb{1}\left\{T_{(K)} \le \tau, \Delta_{(K)} = y\right\}$ and

$$\begin{aligned}\bar{Q}_{k-1,\tau}^g\left(\mathcal{F}_{T_{(k-1)}}\right) = \mathbb{E}_P\Big[&\mathbb{1}\left\{T_{(k)} \le \tau, \Delta_{(k)} = \ell\right)\bar{Q}_{k,\tau}^g\left(A\left(T_{(k-1)}\right), L\left(T_{(k)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \\ &+ \mathbb{1}\left\{T_{(k)} < \tau, \Delta_{(k)} = a\right)\bar{Q}_{k,\tau}^g\left(1, L\left(T_{(k-1)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \\ &+ \mathbb{1}\left\{T_{(k)} < \tau, \Delta_{(k)} = y\right) \,\Big|\, \mathcal{F}_{T_{(k-1)}}\Big],\end{aligned} \tag{11}$$

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

$$\Psi_\tau^{\text{obs}}(P) = \mathbb{E}_P\left[\bar{Q}_{0,\tau}^g(1, L(0))\right]. \tag{12}$$

Furthermore,

$$\bar{Q}_{k-1,\tau}^g\left(\mathcal{F}_{T_{(k-1)}}\right) = p_{k-1,a}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k-1,\ell}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k-1,y}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) \tag{13}$$

where,

$$\begin{aligned}p_{k-1,a}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) &= \int_{\left(T_{(k-1)}, t\right)} S_k\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)\bar{Q}_{k,\tau}^g\left(1, L\left(T_{(k-1)}\right), s, a, \mathcal{F}_{T_{(k-1)}}\right)\Lambda_k^a\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right) \\ p_{k-1,\ell}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) &= \int_{\left(T_{(k-1)}, t\right)} S_k\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right) \\ &\qquad \left(\mathbb{E}_P\left[\bar{Q}_{k,\tau}^g\left(A\left(T_{(k-1)}\right), L\left(T_{(k)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\right)\Lambda_k^\ell\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right) \\ p_{k-1,y}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) &= \int_{\left(T_{(k-1)}, t\right]} S_k\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)\Lambda_k^y\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right), \quad t \le \tau.\end{aligned}$$

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

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

for $k = 0, ..., K$ satisfies the desired property of Equation 11. First, we find

$$
\begin{aligned}
\bar{Q}^g_{k,\tau} &= \mathbb{E}_P\left[\sum_{j=k+1}^K W_{k,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid \mathcal{F}_{T_{(k)}}\right] \\
&= \mathbb{E}_P\Bigg(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\}W_{k,k+1} \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} \in \{a,\ell\}\right\}\mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^K W_{k+1,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right]\Bigg) \\
&= \mathbb{E}_P\Bigg(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\}W_{k,k+1} \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\right\}\mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^K W_{k+1,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\right\}\mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^K W_{k+1,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\Bigg| \mathcal{F}_{T_{(k)}}\right] \\
&= \mathbb{E}_P\Bigg(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} \qquad (14) \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\right\}\mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^K W_{k+1,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\right\}\mathbb{E}_P\left[\mathbb{E}_P\left[\sum_{j=k+2}^K W_{k+1,j}\mathbb{1}\left\{T_{(j)} \le \tau, \Delta_{(j)} = y\right\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right]\Bigg| \mathcal{F}_{T_{(k)}}\Bigg) \\
&= \mathbb{E}_P\Big(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\right\}\mathbb{E}_P\left[W_{k,k+1}\bar{Q}^g_{k+1,\tau}\left(A\left(T_{(k+1)}\right), L\left(T_{(k)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k)}}\right) \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\right\}\mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}\left(A\left(T_{(k)}\right), L\left(T_{(k+1)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k)}}\right) \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \mid \mathcal{F}_{T_{(k)}}\Big) \\
&= \mathbb{E}_P\Big(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\right\}\mathbb{E}_P\left[W_{k,k+1}\bar{Q}^g_{k+1,\tau}\left(A\left(T_{(k+1)}\right), L\left(T_{(k)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k)}}\right) \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \\
&\qquad +\mathbb{1}\left\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\right\}\bar{Q}^g_{k+1,\tau}\left(A\left(T_{(k)}\right), L\left(T_{(k+1)}\right), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k)}}\right) \mid \mathcal{F}_{T_{(k)}}\Big)
\end{aligned}
$$

by the law of iterated expectations and that

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

for all $j = 1, ..., k-1$ and $k = 1, ..., K$. The first desired statement about $\bar{Q}^g_{k,\tau}$ simply follows from the fact that

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

by the law of iterated expectations in the second step from which Equation 11 follows. A similar calculation shows that $\Psi^{\text{obs}}_\tau(P) = \mathbb{E}_P[\bar{Q}^g_{0,\tau}(1, L(0))]$ and so Equation 12 follows. This shows the first statement.

We now show the second statement. Since $\Lambda^x_k\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right)$ is the cumulative cause-specific hazard given $\mathcal{F}_{T_{(k-1)}}$ and that the event was of type x, it follows that (A5.29 of Last & Brandt (1995))

$$P\Big(\big(T_{(k)},\Delta_{(k)}\big)\in d(t,m)\mid \mathcal{F}_{T_{(k-1)}}\Big)=\sum_{x=a,\ell,d,y}\delta_x(\mathrm{d}m)\prod_{s\in\left(T_{(k-1)},t\right)}\left(1-\sum_{x=\ell,a,d,y}\Lambda_k^x\Big(\mathrm{d}s,\mathcal{F}_{T_{(k-1)}}\Big)\right)\Lambda_k^z\Big(\mathrm{d}t,\mathcal{F}_{T_{(k-1)}}\Big),\tag{15}$$

whenever $T_{(k-1)}<\infty$ and $\Delta_{(k-1)}\in\{a,\ell\}$, so we get Equation 13 by plugging in Equation 15 to the second last equality of Equation 14. □

4 Censoring

In this section, we introduce a right-censoring time $C>0$ at which we stop observing the multivariate jump process $\alpha(t)$. We present conditions such that our ICE-IPCW estimator will be consistent for the target parameter. Specifically $N^c(t)=\mathbb{1}\{C\le t\}$ the counting process for the censoring process and let T^e further denote the (uncensored) terminal event time given by $T^e=\inf_{t>0}\{N^y(t)+N^d(t)=1\}$. Throughout the rest of the paper, we will assume that the process N^c does not jump at the same time as the processes N^a,N^ℓ,N^y,N^d.

Let $\beta(t)=(\alpha(t),N^c(t))$ be the fully observable multivariate jump process in $[0,\tau_{\text{end}}]$. Then, we observe the trajectories of the process given by $t\mapsto\beta(t\wedge C\wedge T^e):=\tilde\beta(t)$, and the observed filtration is given by $\mathcal{F}_t^{\tilde\beta}=\sigma(\beta(s\wedge C\wedge T^e)\mid s\le t)=\mathcal{F}_{t\wedge C}\vee\mathcal{G}_{t\wedge T^e}$[††], where $\mathcal{G}_t=\sigma(N^c(s)\mid s\le t)$ denotes the filtration generated by the censoring process (for a stopping time $T>0$, we use $\mathcal{F}_{t\wedge T}$ to denote stopping time σ-algebra given by the stopping time $t\wedge T$.). Let $\big(\bar T_{(k)},\bar\Delta_{(k)},A(\bar T_k),L(\bar T_k)\big)$ be the observed data given by

$$\bar T_{(k)}=C\wedge T_{(k)}$$

$$\bar\Delta_{(k)}=\begin{cases}\Delta_{(k)} \text{ if } C>T_{(k)}\\ \text{c otherwise}\end{cases}$$

$$A(\bar T_k)=\begin{cases}A\big(T_{(k)}\big)\text{if } C>T_{(k)}\\ A\big(T_{(k-1)}\big) \text{ otherwise}\end{cases}$$

$$L(\bar T_k)=\begin{cases}L\big(T_{(k)}\big) \text{ if } C>T_{(k)}\\ L\big(T_{(k-1)}\big) \text{ otherwise}\end{cases}$$

for $k=1,\dots,K$. From this, it follows that

$$\mathcal{F}_{T_{(k)}}=\mathcal{F}^{\tilde\beta}_{\bar T_{(k)}}\quad\text{if } \bar\Delta_{(k)}\neq c.$$

Also define the full data filtration for β given by $\mathcal{F}_t^\beta=\sigma(\beta(s)\mid s\le t)=\sigma(\alpha(s),N^c(s))\mid s\le t)$. Let $\tilde\Lambda_k^c\Big(\mathrm{d}t,\mathcal{F}^\beta_{\bar T_{(k-1)}}\Big)$ be the cause-specific cumulative hazard (measure) of the k'th event at time t given the observed history $\mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}$ and define the corresponding censoring survival function $\tilde S^c\Big(t\mid\mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\Big)=\prod_{s\in\left(T_{(k-1)},t\right]}\Big(1-\tilde\Lambda_k^c\Big(\mathrm{d}s,\mathcal{F}^\beta_{\bar T_{(k-1)}}\Big)\Big)$. This determines the observational probability that you will be censored after time t given the observed history up to $\mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}$.

We provide the conditions in terms of independent censoring. Our conditions are similar to those that may be found the literature based on independent censoring (Andersen et al. (1993); Definition III.2.1) or local independence conditions (Røysland et al. (2024); Definition 4). Heuristically, one may think of independent censoring in this setting as

$$\begin{aligned}&P\Big(T_{(k)}\in[t,t+\mathrm{d}t),\Delta_{(k)}=x,A\big(T_{(k)}\big)=m,L\big(T_{(k)}\big)=l\mid\mathcal{F}_{T_{(k-1)}},T_{(k)}\ge t\Big)\\&=P\Big(\bar T_{(k)}\in[t,t+\mathrm{d}t),\bar\Delta_{(k)}=x,A(\bar T_k)=m,L(\bar T_k)=l\mid\mathcal{F}_{T_{(k-1)}},\bar T_{(k)}\ge t\Big),\qquad x\neq c.\end{aligned}$$

for uncensored histories, i.e., when $\bar\Delta_{(k-1)}\neq c$. We are now ready to state the main theorem which proves that the ICE-IPCW estimator is valid. Later, we provide an implementation and algorithm for the ICE-IPCW estimator.

[††]The fact that the stopped filtration and the filtration generated by the stopped process are the same is not obvious but follows by Theorem 2.2.14 of Last & Brandt (1995). This can fail if $\mathcal{F}_0$ includes all null-sets.

Theorem 4: Assume that the compensator Λ^α of N^α with respect to the filtration $\mathcal{F}_t^\beta$ is also the compensator with respect to the filtration $\mathcal{F}_t$. Then for uncensored histories, we have

$$\begin{aligned}
&\mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\} P\left(\left(\bar{T}_n, \bar{\Delta}_n, A(\bar{T}_n), L(\bar{T}_n)\right) \in d(t,m,a,l) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) \\
&= \mathbb{1}\left\{\bar{T}_{n-1} < t, \bar{\Delta}_{(n-1)} \neq c\right\} \left(\tilde{S}\left(t- \mid \mathcal{F}_{T_{(n-1)}}\right) \sum_{x=a,\ell,d,y} \delta_x(\mathrm{d}m) \psi_{n,x}(t, d(a,l)) \Lambda_n^x\left(\mathrm{d}t, \mathcal{F}_{T_{(n-1)}}\right) \right. \\
&\left. + \delta_{(c, A(T_{(n-1)}), L(T_{(n-1)}))}(\mathrm{d}(m,a,l)) \tilde{\Lambda}_n^c\left(\mathrm{d}t, \mathcal{F}_{T_{(n-1)}}\right)\right),
\end{aligned} \tag{16}$$

where $\psi_{n,x}$ was defined in Equation 4 and $\mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\} \tilde{S}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = \mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\} \prod_{s \in \left(T_{(k-1)}, t\right]} \Big(1 - \sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right) - \tilde{\Lambda}_k^c\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right)\Big)$.

Further suppose that $\mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\} \prod_{s \in \left(T_{(k-1)}, t\right]} \left(1 - \sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right) - \tilde{\Lambda}_k^c\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right)\right) = \mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\} \prod_{s \in \left(T_{(k-1)}, t\right]} \left(1 - \sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right)\right) \prod_{s \in \left(T_{(k-1)}, t\right]} \left(1 - \tilde{\Lambda}_k^c\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right)\right)$ P-a.s. and that $\tilde{S}^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) > \eta$ for all $t \in (0, \tau]$ and $n \in \{1, \dots, K\}$ P-a.s. for some $\eta > 0$.

Then, the ICE-IPCW estimator is consistent for the target parameter, i.e.,

$$\begin{aligned}
\mathbb{1}\left\{\bar{\Delta}_{(k-1)} \neq c\right\} \bar{Q}^g_{k-1,\tau} = \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \neq c\right\} \mathbb{E}_P \Bigg[&\frac{\mathbb{1}\left\{\bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = \ell\right\}}{\tilde{S}^c\left(\bar{T}_{(k-1)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \bar{Q}^g_{k,\tau}\left(A(\bar{T}_{k-1}), L(\bar{T}_k), \bar{T}_{(k)}, \bar{\Delta}_{(k)}, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \\
&+ \frac{\mathbb{1}\left\{\bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = a\right\}}{\tilde{S}^c\left(\bar{T}_{(k-1)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \bar{Q}^g_{k,\tau}\left(1, L(\bar{T}_{k-1}), \bar{T}_{(k)}, \bar{\Delta}_{(k)}, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \\
&+ \frac{\mathbb{1}\left\{\bar{T}_{(k)} \leq \tau, \bar{\Delta}_{(k)} = y\right\}}{\tilde{S}^c\left(\bar{T}_{(k-1)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \Bigg| \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} \Bigg]
\end{aligned} \tag{17}$$

for $k = K, \dots, 1$ and

$$\Psi^{\mathrm{obs}}_\tau(P) = \mathbb{E}_P\left[\bar{Q}^g_{0,\tau}(1, L(0))\right]. \tag{18}$$

Proof: Under the local independence condition, the compensator of the random measure $N^\alpha(\mathrm{d}(t,m,a,l))$ with respect to the filtration $\mathcal{F}_t^\beta$, can be given by (Theorem 8 (ii))

$$\Lambda^\alpha(\mathrm{d}(t,m,a,l)) = \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \sum_{x=a,\ell,d,y} \delta_x(\mathrm{d}m) \psi_{k,x}\left(t, d(a,l), \mathcal{F}_{T_{(k-1)}}\right) \Lambda_k^m\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right).$$

In view of this, the stopped process $N^{\tilde{\beta}}$ has by the optional sampling theorem (the corollary after I Theorem 18 of Protter (2005) since $C \wedge T^e < \infty$) a compensator given by

$$\begin{aligned}
\Lambda^{\tilde{\beta}}(\mathrm{d}(t,m,a,l)) = \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\right\} &\sum_{x=a,\ell,d,y} \delta_x(\mathrm{d}m) \psi_{k,x}\left(t, d(a,l), \mathcal{F}_{T_{(k-1)}}\right) \Lambda_k^m\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right) \\
&+ \delta_{(c, A(T_{(k-1)}), L(T_{(k-1)}))}(\mathrm{d}(m,a,l)) \tilde{\Lambda}_k^c\left(\mathrm{d}t \mid \mathcal{F}_{T_{(k-1)}}\right)
\end{aligned} \tag{19}$$

with respect to the filtration $\mathcal{F}_t^{\tilde{\beta}}$ (to obtain the censoring term, we find a compensator for $N^{\tilde{\beta}}$ with respect to $\mathcal{F}^{\tilde{\beta}}$ again by Theorem 8 (ii) and obtain the component corresponding to censoring).

Let $N_{\boldsymbol{X}}$ denotes the space of all point process trajectories with mark space $\boldsymbol{X}$. Note that as $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n)}}$ is generated by finitely many random variables, there is a measurable function h_n with the property that $h_n\left((A(0), L(0)), N^{\tilde{\beta}}\left(\cdot \wedge T_{(n)}\right)\right) = \mathcal{F}_{T_{(n)}}$. We can define then stochastic kernel from $\{0,1\} \times \mathcal{L} \times N_{\boldsymbol{X}}$ to $\mathbb{R}_+ \times \boldsymbol{X}$ with $\boldsymbol{X} = \{a, y, \ell, d, c\} \times \{0,1\} \times \mathcal{L}$ by

$$\rho((l_0,a_0),\varphi,d(t,m,a,l)) = \sum_{k=1}^{K} \mathbb{1}\{\zeta_{n-1}(\varphi) < t \leq \zeta_n(\varphi)\} \sum_{x=a,\ell,d,y} \delta_x(\mathrm{d}m)\psi_{k,x}(t,d(a,l),h_n(\varphi)))\Lambda_k^x(dt,h_n(\varphi))$$
$$+\delta_{c,\eta^a_{n-1}(\varphi),\eta^\ell_{n-1}(\varphi)}(\mathrm{d}(m,a,l))\tilde{\Lambda}_k^c(dt \mid h_n(\varphi))$$

where ζ_n denotes the projection operator to the n'th event, where η_j^a is the projection operator to the j'th events treatment mark and η_j^ℓ is the projection operator to the j'th events covariate mark. It is the canonical compensator because $\rho\Big((L(0),A(0)),N^{\tilde{\beta}},d(t,m,a,l)\Big)$ is a compensator of the process $N^{\tilde{\beta}}$ with respect to the filtration $\mathcal{F}_t^{\tilde{\beta}}$ (Last & Brandt (1995); p. 130).

The conditional distribution of the event and marks then follows from this via Theorem 4.3.8 of Last & Brandt (1995). The theorem specifically states that

$$P\Big(\big(\bar{T}_k,\bar{\Delta}_k,A(\bar{T}_k),L(\bar{T}_k)\big) \in d(t,m,a,l) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\Big)$$
$$= \mathbb{1}\{\bar{T}_{k-1} < t\} \prod_{s\in(\bar{T}_{k-1},t)} \Big(1-\rho\big((l_0,a_0),\bar{N}_{\bar{T}_{k-1}},ds,\{a,y,l,d,y\}\times\{0,1\}\times\mathcal{L}\big)\Big)\rho\big((l_0,a_0),\bar{N}_{\bar{T}_{k-1}},d(t,m,a,l)\big).$$

where $N^{\tilde{\beta}}_{\bar{T}_n}$ is a shorthand for $N^{\tilde{\beta}}\big(\cdot \wedge \bar{T}_{(n)}\big)$. We find

$$\mathbb{1}\big\{\bar{T}_{k-1} < t, \bar{\Delta}_{(k-1)} \neq c\big\}\alpha\Big((l_0,a_0),N^{\tilde{\beta}}_{\bar{T}_{k-1}},d(t,m,a,l)\Big) = \mathbb{1}\big\{\bar{\Delta}_{(k-1)} \neq c\big\}\Bigg(\sum_{x=a,\ell,d,y} \delta_x(\mathrm{d}m)\psi_{k,x}\Big(t,d(a,l),\mathcal{F}_{T_{(k-1)}}\Big)\Lambda_k^x\Big(dt,\mathcal{F}_{T_{(k-1)}}\Big)$$
$$+\delta_{c,L(T_{(k-1)}),A(T_{(k-1)})}(\mathrm{d}(m,a,l))\tilde{\Lambda}_k^c\Big(dt \mid \mathcal{F}_{T_{(k-1)}}\Big)\Bigg)$$

and

$$\mathbb{1}\big\{\bar{T}_{k-1} < t, \bar{\Delta}_{(k-1)} \neq c\big\}\alpha\Big((l_0,a_0),N^{\tilde{\beta}}_{\bar{T}_n},ds,\{a,y,l,d,y\}\times\{0,1\}\times\mathcal{L}\Big) = \mathbb{1}\big\{\bar{\Delta}_{(k-1)} \neq c\big\}\Bigg(\sum_{x=a,\ell,d,y} \Lambda_k^x\Big(dt,\mathcal{F}_{T_{(k-1)}}\Big) + \tilde{\Lambda}_k^c\Big(dt \mid \mathcal{F}_{T_{(k-1)}}\Big)\Bigg).$$

From this, we get Equation 16. Applying this to the right hand side of Equation 17 shows that it is equal to Equation 11.

□

Note that Equation 16 also ensures that all hazards (other than censoring) and mark probabilities are identifiable from censored data if we can show that the censoring survival factorizes. We provide two criteria for this.

Theorem 5: Assume that for each $k = 1, ..., K$,

$$\Big(T_{(k)},\Delta_{(k)},A\big(T_{(k)}\big),L\big(T_{(k)}\big)\Big) \perp C \mid \mathcal{F}_{T_{(k-1)}}$$

Then the survival function factorizes

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

and the local independence statement given in Equation 16 holds.

Proof: By the stated independence condition, it follows immediately that

$$\tilde{S}\Big(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\Big) = P\Big(T_{(k)} > t, C > t \vert\ \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\Big) = S\Big(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\Big)S^c\Big(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\Big)$$

All we need to show for the first statement is,

$$S^c\Big(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\Big) = \prod_{s\in(0,t]} \Big(1-\tilde{\Lambda}_k^c\Big(\mathrm{d}s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\Big)\Big)$$

which follows "directly" by calculating $P\left(\bar{T}_{(k)} \in \mathrm{d}t, \Delta_{(k)} = c \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$ and from this we get the observed cause-specific hazards.

The local independence statement follows by letting $\tilde{\mathcal{F}}^{\beta} = \sigma(\alpha(s) \mid s \leq t) \vee \mathcal{Z}_0$, where $\mathcal{Z}_0 = \sigma(A(0), L(0), C)$. Evidently, $\mathcal{F}_t \subseteq \mathcal{F}^{\beta}_t \subseteq \tilde{\mathcal{F}}^{\beta}_t$. Under the independence assumption, by the use of the canonical compensator, the compensator for N^{α} for $\mathcal{F}_t$ is also the compensator for $\tilde{\mathcal{F}}^{\beta}_t$. Let M^{α} denotes the corresponding martingale decomposition. It follows that:

$$\begin{aligned}
&\mathbb{E}_P\left[M^{\alpha}((0,t] \times \{a,\ell,y,d,c\} \times \{0,1\} \times \mathcal{L}) \mid \mathcal{F}^{\beta}_s\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[M^{\alpha}((0,t] \times \{a,\ell,y,d,c\} \times \{0,1\} \times \mathcal{L}) \mid \tilde{\mathcal{F}}^{\beta}_s\right] \mid \mathcal{F}^{\beta}_s\right] \\
&\overset{\text{(i)}}{=} \mathbb{E}_P\left[M^{\alpha}((0,s] \times \{a,\ell,y,d,c\} \times \{0,1\} \times \mathcal{L}) \mid \mathcal{F}^{\beta}_s\right] \\
&\overset{\text{(ii)}}{=} M^{\alpha}((0,s] \times \{a,\ell,y,d,c\} \times \{0,1\} \times \mathcal{L})
\end{aligned}$$

which implies the desired statement. In part (i), we that the martingale is a martingale for $\tilde{\mathcal{F}}^{\beta}_t$. In part (ii), we use that the martingale is $\mathcal{F}^{\alpha}_t$-adapted. □

Not cleaned

Finally, we need to show that the survival function factorizes

$$\begin{aligned}
\tilde{S}\left(t- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\right) &= \prod_{s \in (0,t)} \left(1 - \left(\sum_{x=a,\ell,y,d} \Lambda^x_k\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right) + \tilde{\Lambda}^c_k\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\right) \\
&= \prod_{s \in (0,t)} \left(1 - \sum_{x=a,\ell,y,d} \Lambda^x_k\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right) \prod_{s \in (0,t]} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)
\end{aligned} \tag{20}$$

for in doing so, we can apply Equation 16 and Equation 20 show that the right hand side of Equation 17 is the same as the left hand side of Equation 17. But the factroization holds since

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

since the processes ζ and γ do not jump at the same time, the last factor factorizes, i.e.,

$$\prod_{v \in (s,t) \gamma(v) \neq \gamma(v-) \vee \zeta(v) \neq \zeta(v-)} (1 - \Delta(\zeta + \gamma)) = \prod_{\substack{v \in (s,t) \\ \gamma(v) \neq \gamma(v-)}} (1 - \Delta\gamma) \prod_{\substack{v \in (s,t) \\ \zeta(v) \neq \zeta(v-)}} (1 - \Delta\zeta)$$

Further suppose that $[M^c, M^x] = 0$ for $x \in \{a, \ell, d, y\}$. Then, $\tilde{S}\left(t \mid \mathcal{F}_{T_{(n-1)}}\right)$ is given by

$$\tilde{S}\left(t \mid \mathcal{F}_{T_{(n-1)}}\right) = S\left(t \mid \mathcal{F}_{T_{(n-1)}}\right) \tilde{S}^c\left(t \mid \mathcal{F}_{T_{(n-1)}}\right)$$

For this, we only that the censoring does not jump at the same time as any of the other counting processes. To show this, consider the quadratic covariation which by orthogonality implies

$$0 = \left[M^c, \sum_x M^x\right] = \int_0^{\cdot} \Delta\tilde{\Lambda}_c \sum_{x=a,\ell,y,d} d\Lambda_x$$

Using this, we have

$$\begin{aligned}
0 &= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \int_{\left(T_{(k-1)},t\right]} \Delta\tilde{\Lambda}_c \left(\sum_{x=a,\ell,y,d} d\Lambda_x\right) \\
&= \sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \int_{\left(T_{(k-1)},t\right]} \Delta\tilde{\Lambda}^c_k\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right) \left(\sum_{x=a,\ell,y,d} \Lambda^x_k\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right)),
\end{aligned}$$

so that $\mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \int_{\left(T_{(k-1)},t\right]} \Delta\left(\tilde{\Lambda}^c_k\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\left(\sum_{x=a,\ell,y,d} \Lambda^x_k\left(\mathrm{d}s, \mathcal{F}_{T_{(k-1)}}\right)\right)\right) = 0$. Taking the expectations on both sides, we have

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

It follows that for every t with $\tilde{S}\left(t- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\right) > 0$,

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

If we can further argue that $\Delta \tilde{\Lambda}^c_{k+1} + \sum_x \Delta \Lambda^x_{k+1}\left(t, \mathcal{F}_{T_{(k+1-1)}}\right) = 1 \Rightarrow \Delta \tilde{\Lambda}^c_{k+1} = 1 \vee \sum_x \Delta \Lambda^x_{k+1}\left(t, \mathcal{F}_{T_{(k+1-1)}}\right) = 1$ because these are the only points at which $\tilde{S}$ can be zero, but not the two other survival functions. Alternatively, it may be assumed that $\Delta \tilde{\Lambda}^c_{k+1} + \sum_x \Delta \Lambda^x_{k+1}\left(t, \mathcal{F}_{T_{(k+1-1)}}\right) < 1$ to get the desired statement.

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

A simple implementation of the IPCW is provided in Section 4.1.

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 $\hat{\Psi}_n = \frac{1}{n} \sum_{i=1}^n \hat{\nu}_0(L_i(0), 1)$.

The high dimensionality of the second representation given in Theorem 3 makes it likely that is not useful for computing the target parameter in practice. Though, specialized approaches may yet exist (see the discussion).

It is recommended to use Equation 17 for estimating $\bar{Q}^g_{k,\tau}$ instead of direct computation Equation 13: The resulting integral representing the target parameter would, in realistic settings, be incredibly highly dimensional.

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 17 for censoring to derive the efficient influence function. To do so, we introduce some additional notation and let

$$\bar{Q}^g_{k,\tau}\left(u \mid \mathcal{F}_{T_{(k)}}\right) = p_{ka}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k\ell}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{ky}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right), u < \tau \tag{21}$$

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). The efficient influence function may change slightly if the censoring distribution cannot be assumed to be continuous, but only slightly since we use the product integral instead of the exponential function, see e.g., Theorem 8 of Gill & Johansen (1990). This results in having to multiply by $\frac{1}{1-\Delta \Lambda^c_t}$.

Theorem 6 (Efficient influence function): The efficient influence function is given by

$$\begin{aligned}\varphi^*(P) = {} & \frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))} \sum_{k=1}^{K} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\left\{A\left(T_{(j)}\right)=1\right\}}{\pi_j\left(T_{(j)}, \mathcal{F}_{T_{(j-1)}}\right)} \right)^{\mathbb{1}\{\Delta_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} S^c\left(T_{(j)} - \mid \mathcal{F}_{T_{(j-1)}}\right)} \mathbb{1}\left\{\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} < \tau\right\} \\ & \times \left(\left(\bar{Z}^a_{k,\tau} - \bar{Q}^g_{k-1,\tau}\right) + \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \left(\bar{Q}^g_{k-1,\tau}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) - \bar{Q}^g_{k-1,\tau}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)\right) \frac{1}{S^c\left(u- \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} M^c_k(\mathrm{d}u) \right) \\ & + \bar{Q}^g_{0,\tau} - \Psi_\tau(P),\end{aligned} \tag{22}$$

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

Proof: Define (update notation)

$$\begin{aligned}\bar{Z}^a_{k,\tau}(s, t_k, d_k, l_k, a_k, f_{k-1}) = {} & \frac{I(t_k \le s, d_k = \ell)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)} \bar{Q}^g_{k,\tau}(a_{k-1}, l_k, t_k, d_k, f_{k-1}) \\ & + \frac{I(t_k \le s, d_k = a)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)} \\ & \quad \times \int \bar{Q}^g_{k,\tau}(\tilde{a}_k, l_{k-1}, t_k, d_k, f_{k-1}) \pi^*_{k-1}\left(t_k, \mathcal{F}_{T_{(k-1-1)}}\right) \nu_A(\mathrm{d}\tilde{a}_k) \\ & + \frac{I(t_k \le s, d_k = y)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)}, s \le \tau\end{aligned} \tag{23}$$

and let

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

We compute the efficient influence function by taking the Gateaux derivative of the above with respect to P, by discretizing the time. Note that this is not a rigorous argument showing that the efficient influence function is given by Equation 22. 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.

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

and

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

We will recursively calculate the derivative,

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

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

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

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

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

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

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

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

Now note for the second term, we can write

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

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

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

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

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

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

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 22 includes a censoring martingale. To estimate this martingale, we would need to have estimators of $\bar{Q}^g_{k,\tau}(t)$ (defined in Equation 21) 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 7. 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 7.

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

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

Lemma 1: 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 7 (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

$$z_k\left(\mathcal{F}_{T_{(k)}}\right) = \left(\left(\frac{\pi_{k-1,0}\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)}{\pi_{k-1}\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)}\right)^{\mathbb{1}\{\Delta_{(k)} = a\}} - 1\right) \left(\bar{Q}_{k-1,\tau}^{g}\left(\mathcal{F}_{T_{(k-1)}}\right) - \nu_{k-1,\tau}\left(\mathcal{F}_{T_{(k-1)}}\right)\right)$$

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

$$+ \left(\frac{\pi_{k-1,0}\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)}{\pi_{k-1}\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)}\right)^{\mathbb{1}\{\Delta_{(k)} = a\}} \int_{T_{(k-1)}}^{\tau} V_k\left(u, \mathcal{F}_{T_{(k-1)}}\right) \nu_{k-1,\tau}^*\left(du \mid \mathcal{F}_{T_{(k-1)}}\right),$$

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

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

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

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

$$\mathbb{E}_{P_0}[\varphi_0^*(P)] + \Psi_\tau(P) - \Psi_\tau(P_0) = \mathbb{E}_{P_0}\left[\nu_0(L(0)) - \bar{Q}_{0,\tau}^g(L(0))\right]. \tag{24}$$

Apply the law of iterated expectation to the efficient influence function in Equation 22 to get

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

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

where

$$h_k\Big(\mathcal{F}_{T_{(k)}}\Big) = \bar{Z}^a_{k,\tau} - \nu_{k-1} + \int_{T_{(k-1)}}^{\tau\wedge T_{(k)}} \Big(\nu^*_{k-1}\Big(\tau \mid \mathcal{F}_{T_{(k-1)}}\Big) - \nu^*_{k-1}\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)\Big) \frac{1}{S^c\Big(u- \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} M^c_k(\mathrm{d}u).$$

Now note that

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

by a martingale argument. Noting that,

$$\begin{aligned}
&\mathbb{E}_{P_0}\Bigg[\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)}}\Bigg] \\
&= \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

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

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^c_0(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\}\} \\
&\quad \times \left(\mathbb{E}_{P_0}\left[\bar{Z}^a_{k+1,\tau}\left(S^c_0, \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^c_0(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 \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^c_0(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\}\} \\
&\quad \times \left(\bar{Q}^g_{k,\tau}(f_k) - \nu_{k,\tau}(f_k)\right) P_{\mathcal{F}_{T_{(k)}}}(df_k) \qquad (25) \\
&+ \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^c_0(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^c_0(t_k - \mid f_{k-1})}{S^c(t_k - \mid f_{k-1})} \\
&\quad \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^c_0(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\}\} \\
&\quad \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 24 and Equation 25, we are done. □

Note that by the triangle inequality

$$
\begin{aligned}
&\left| \int_{T_{(k-1)}}^{\tau} \int_{T_{(k-1)}}^{u} \left(\frac{S_0\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}{S\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)} - 1\right)\frac{S^c_0\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\Lambda^c_{k,0}\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right) - \Lambda^c\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\left(\nu^*_{k-1,\tau}\left(du \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\right| \\
&\leq \nu^*_{k-1,\tau}(\tau) \sup_{u \in (T_{(k-1)}, \tau)} \left| \int_{T_{(k-1)}}^{u} \left(\frac{S_0\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)}{S\left(s \mid \mathcal{F}_{T_{(k-1)}}\right)} - 1\right)\frac{S^c_0\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}{S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)}\left(\Lambda^c_{k,0}\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right) - \Lambda^c\left(ds \mid \mathcal{F}_{T_{(k-1)}}\right)\right)\right|
\end{aligned}
$$

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 5. 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 $j \in S \subseteq \{1, ..., K\}$ 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

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

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

$$
\begin{aligned}
&\operatorname{argmax}_{\theta^* \in \Theta^*} \prod_{i=1}^n \prod_{t \in (0, \tau_{\max}]} \left(d\Lambda^c_\theta(t \mid \mathcal{F}^i_{t-}\right)^{\Delta N^c_i(t)} \left(1 - d\Lambda^c_\theta(t \mid \mathcal{F}^i_{t-}\right)^{1-\Delta N^c_i(t)} \\
&= \operatorname{argmax}_{\theta^* \in \Theta^*} \prod_{i=1}^n \left(\prod_{k \in S, k \leq K_i(\tau)+1} \left(d\Lambda^c_{\theta^*}\left(T^i_{(k)} \mid \mathcal{F}_{T^i_{(k-1)}}\right)\right)^{\mathbb{1}\{k \neq K_i+1\}} \prod_{t \in (0,\tau_{\max}]} \left(1 - d\left(\mathbb{1}\left\{t \in \left(T^i_{(k-1)}, T^i_{(k)}\right)\right\}\right) \Lambda^c_{\theta^*}\left(t \mid \mathcal{F}_{T^i_{(k-1)}}\right)\right)\right) \\
&\quad \times \prod_{i=1}^n \left(\prod_{k \notin S, k \leq K_i(\tau)+1} \left(d\Lambda^c_{\theta_k}\left(T^i_{(k)} \mid \mathcal{F}_{T^i_{(k-1)}}\right)\right)^{\mathbb{1}\{k \neq K_i+1\}} \prod_{t \in (0,\tau_{\max}]} \left(1 - d\left(\mathbb{1}\left\{t \in \left(T^i_{(k-1)}, T^i_{(k)}\right)\right\}\right) \Lambda^c_{\theta_k}\left(t \mid \mathcal{F}_{T^i_{(k-1)}}\right)\right)\right)
\end{aligned}
$$

and that

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

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 $\bar{Q}^g_{k,\tau}$ 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 13 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

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. *Mathematical Statistics and Probability Theory: Proceedings, Sixth International Conference, Wisła (Poland), 1978*, 1–25.

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

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

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

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

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

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

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

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

Protter, P. E. (2005). *Stochastic differential equations*. Springer.

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.

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

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

Ryalen, P. (2024). *On the role of martingales in continuous-time causal inference.*

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

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

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

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

Shirakawa, T., Li, Y., Wu, Y., Qiu, S., Li, Y., Zhao, M., Iso, H., & van der Laan, M. (2024, April). *Longitudinal Targeted Minimum Loss-based Estimation with Temporal-Difference Heterogeneous Transformer* (Issue arXiv:2404.04399). arXiv.

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

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

Let $\mathcal{F}_{T_{(k)}}$ be the stopping time 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 8 (Finite-dimensional distributions): Under the stated conditions of this section:

(i). 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 $\mathcal{F}_0 = \sigma(W)$. Furthermore, $\mathcal{F}_t^{\bar{N}} = \sigma(\bar{N}((0,s],\cdot) \mid s \leq t) \vee \sigma(W) = \mathcal{F}_t$, where

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

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

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

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

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

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

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

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

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

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

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

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

We then have,

$$\frac{P\left(\left(T_{(k)}, \Delta_{(k)}, X\left(T_{(k)}\right)\right) \in d(t,m,x) \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)}$$

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

Letting

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

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

completes the proof of (ii).

□

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 4. Some, such as Chamapiwa (2018), write down this DAG, but with an arrow from $T_{(k)}$ to $L(T_{(k)})$ and $A(T_{(k)})$ instead of displaying a multivariate random variable which they deem the "time-as-confounder" approach to allow for irregularly measured data (see Figure 5). 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 4, 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).

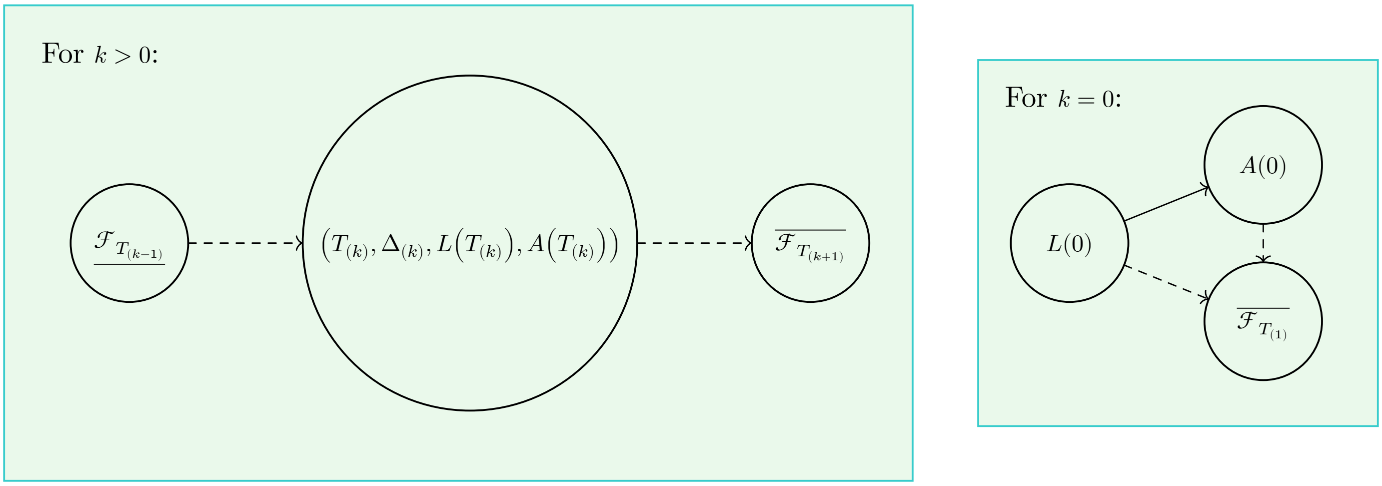

Figure 4: A DAG representing the relationships between the variables of O. The dashed lines indicate multiple edges from the dependencies in the past and into the future.

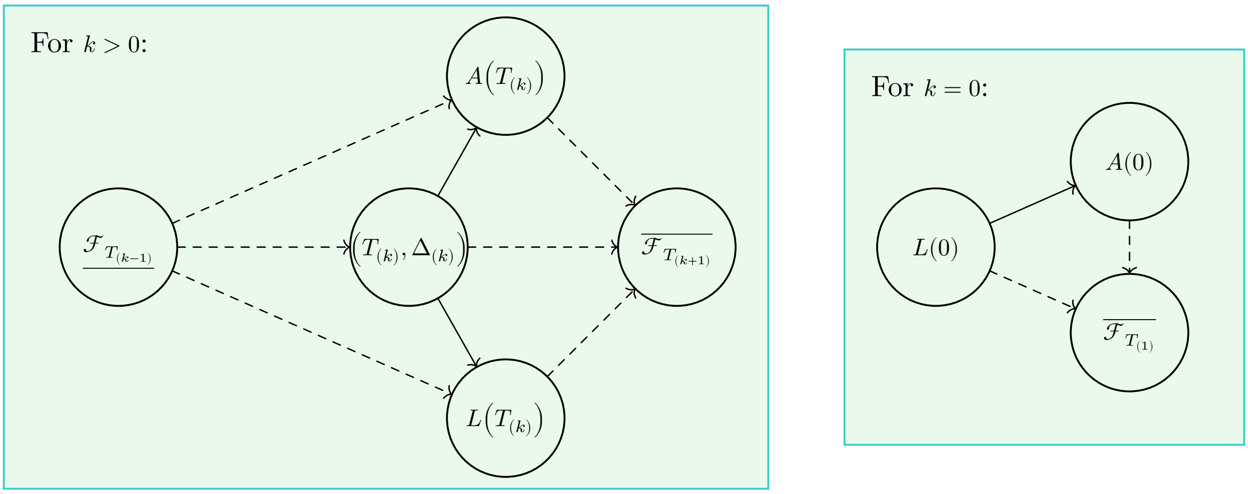

Figure 5: A DAG for simulating the data generating mechanism. The dashed lines indicate multiple edges from the dependencies in the past and into the future. Here $\underline{\mathcal{F}_{T_{(k)}}}$ is the history up to and including the k'th event and $\overline{\mathcal{F}_{T_{(k)}}}$ is the history 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:

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

We find immediately that

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

Now note that

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

Let us calculate the second integral

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

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

$$
\begin{aligned}
&\int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{\bar{Q}^g_{k+1,\tau}(u)}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \Lambda^{\bullet}_{k-1}\Big(u, \mathcal{F}_{T_{(k-1-1)}}\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) \\
&\qquad\qquad \times \Bigg(\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) \Bigg) \\
&+ \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) \\
&\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) \\
&+ \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) \\
&\qquad\qquad \times \Bigg(\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) \Bigg) \\
&+ S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^\ell_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1-1)}}\Big) \\
&\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) \\
&+ S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^y_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1-1)}}\Big) \Bigg]
\end{aligned}
$$

Now note that

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

Setting this into the previous integral, we get

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

Thus, we find

$$
\begin{aligned}
&\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\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)}\Big(N^{\bullet}_k(\mathrm{d}s)-\tilde{Y}_{k-1}(s)\Lambda^{\bullet}_{k-1}\Big(s,\mathcal{F}_{T_{(k-1-1)}}\Big)\mathrm{d}s\Big)\\
&=\left(\bar{Q}^g_{k-1,\tau}(\tau)-\bar{Q}^g_{k-1,\tau}\Big(T_{(k)}\Big)\right)\frac{1}{S^c\Big(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big)S\Big(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big)}\mathbb{1}\Big\{T_{(k)}\le\tau\Big\}\\
&-\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\\
&+\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(s,\mathcal{F}_{T_{(k-1-1)}}\Big)\mathrm{d}s\\
&=\left(\bar{Q}^g_{k-1,\tau}(\tau)-\bar{Q}^g_{k-1,\tau}\Big(T_{(k)}\Big)\right)\frac{1}{S^c\Big(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big)S\Big(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big)}\mathbb{1}\Big\{T_{(k)}\le\tau\Big\}\\
&-\left(\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)}\right)-1\right)\\
&-\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\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\times\left(\mathbb{E}_P\Big[\bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k)}\Big),A\Big(T_{(k-1)}\Big),T_{(k)},\Delta_{(k)},\mathcal{F}_{T_{(k-1)}}\Big)\mid T_{(k)}=s,\Delta_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\Big]\right)\\
&+\Lambda^{y}_{k-1}\Big(ds,\mathcal{F}_{T_{(k-1-1)}}\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)\\
&=-\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\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\times\left(\mathbb{E}_P\Big[\bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k)}\Big),A\Big(T_{(k-1)}\Big),T_{(k)},\Delta_{(k)},\mathcal{F}_{T_{(k-1)}}\Big)\mid T_{(k)}=s,\Delta_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\Big]\right)\\
&+\Lambda^{y}_{k-1}\Big(ds,\mathcal{F}_{T_{(k-1-1)}}\Big)\Bigg]+\bar{Q}^g_{k-1,\tau}(\tau)
\end{aligned}
$$